Band selection for nonlinear unmixing of hyperspectral images as a maximal clique problem

Tales Imbiriba[(1)], *Student Member, IEEE*, José Carlos Moreira Bermudez[(1)], *Senior Member, IEEE*, Cédric Richard[(2)], *Senior Member, IEEE*

***Abstract*—Kernel-based nonlinear mixing models have been applied to unmix spectral information of hyperspectral images when the type of mixing occurring in the scene is too complex or unknown. Such methods, however, usually require the inversion of matrices of sizes equal to the number of spectral bands. Reducing the computational load of these methods remains a challenge in large scale applications. This paper proposes a centralized band selection (BS) method for supervised unmixing in the reproducing kernel Hilbert space (RKHS). It is based upon the coherence criterion, which sets the largest value allowed for correlations between the basis kernel functions characterizing the selected bands in the unmixing model. We show that the proposed BS approach is equivalent to solving a maximum clique problem (MCP), i.e., searching for the biggest complete subgraph in a graph. Furthermore, we devise a strategy for selecting the coherence threshold and the Gaussian kernel bandwidth using coherence bounds for linearly independent bases. Simulation results illustrate the efficiency of the proposed method.**

***Index Terms*—Hyperspectral data, nonlinear unmixing, band selection, kernel methods, maximum clique problem.**

I. Introduction

The unmixing of spectral information acquired by hyperspectral sensors is at the core of many remote sensing applications such as land use analysis, mineral detection, environment monitoring and field surveillance [1], [2]. Such information is typically mixed at the pixel level due to the low resolution of hyperspectral devices or because distinct materials are combined into a homogeneous mixture [3]. The observed reflectances then result from mixtures of several pure material signatures present in the scene, called endmembers. Considering that the endmembers have been identified, hyperspectral unmixing (HU) refers to estimating the proportional contribution of each endmember to each pixel in a scene.

The linear mixture model is widely used to identify and quantify pure components in remotely sensed images due to its simple physical interpretation. Though the linear model leads to simple unmixing algorithms and facilitates implementation, there are many situations to which it is not applicable. These include scenes where there is complex radiation scattering among several endmembers, as may happen in some vegetation areas [4]. In such situations, nonlinear mixing models must be considered [5], [6]. Several model-based nonlinear HU methods assume prior knowledge of the mixing that actually occurs in the scene, e.g., see [5]–[9]. However, the complexity of real mixing mechanisms, and the lack of prior knowledge about them in many practical applications has led to the consideration of flexible nonlinear mixing models that can model generic nonlinear functions. Many model-free nonlinear HU strategies have been proposed, including classification of abundances coefficients [10], graph-based approximate geodesic distances [11], [12], and kernel-based algorithms [13]–[17]. Kernel methods provide a non-parametric representation of functional spaces, and can model nonlinear mixings of arbitrary characteristics [5], [6], [13]–[16], [18].

Kernel-based methods are efficient machine learning techniques [19]–[21] that consist mainly of linear algorithms operating in high dimensional reproducing kernel Hilbert spaces (RKHS), into which the data have been mapped using kernel functions [19]. Working in such high dimensional feature spaces is possible due to the so-called *kernel trick*, which allows the computation of inner products in the feature space through a kernel function in the input space [22]. A limitation of kernel methods for HU is that they usually require the inversion of matrices whose dimensions equal the number of spectral bands. Thus, reducing their computational cost remains a challenge for their use in large-scale applications.

A possible way of reducing computational cost is to perform band selection (BS) prior to unmixing [23]. The literature proposes many BS strategies including information-based methods [24], [25], band prioritization techniques [26], [27], sparsity promoting methods [28], [29], spectral distance strategies [26], [30], and a combination of manifold ranking and saliency detection [31]. However, most efforts were focused on discriminating endmembers since these methods deal with the classification of spectral patterns, see e.g. [24], [29], [32] and references therein. When the objective is to unmix spectral information, subspace projection techniques [33] tend to be preferred over BS [26], [28], [34], [35] for reducing the complexity of linear unmixing processes. The reason is that high dimensional data in linearly-mixed images are confined to a low-dimensional simplex defined by the endmembers [3]. However, the simplex property is not preserved in the presence of nonlinearly-mixed pixels [5], rendering projection techniques less attractive.

In [25], [29], [32], the authors propose different unsupervised BS strategies for classification of spectral patterns. A

T. Imbiriba and J.-C. M. Bermudez are with the Department of Electrical Engineering, Federal University of Santa Catarina at Florianópolis, SC, 88040-900, Brazil (e-mail: talesim@gmail.com; j.bermudez@ieee.org). C. Richard is with the University of Nice Sophia-Antipolis, Nice 06108, France (e-mail: cedric.richard@unice.fr), Lagrange Laboratory (CNRS, OCA), in collaboration with Morpheme team (INRIA Sophia-Antipolis). The work of J.-C. M. Bermudez was partly supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grants 305377/2009-4, 400566/2013-3 and 141094/2012-5. The work of C. Richard was partly supported by ANR grants ANR-12- BS03-003 (Hypanema), by the CNRS Imagin project under grant 2015OPTIMISME, by the BNPSI ANR Project no ANR-13- BS-03-0006-01.

multitask sparsity pursuit (MTSP) criterion is used in [29], while clustering is used in [32]. In [25], the authors combine a graph-based formulation with structure-aware measures for band informativeness and independence. They provide an efficient global search strategy using a dominant set extraction technique. Concerning BS techniques applied to HU, the authors in [34] select bands by applying the unmixing algorithm to a cross validation set. A discrimination metric is then applied, which is based on the root mean square error of the abundance estimates. In [28], simultaneous band selection and endmember detection is proposed based on the ICE algorithm, which performs a least squares minimization of the residual sum of squares (RSS). The *progressive band selection* (PBS) algorithm [26], sequentially selects bands that were previously ranked according to a prioritization criterion such as entropy or variance[1]. This method also applies a decorrelation procedure to eliminate highly correlated adjacent bands, and exploits the virtual dimension (VD) [36] to determine upper and lower bounds for the number of bands to select. The PBS method offers good performance for linear unmixing problems.

All the band selection approaches discussed above were proposed for unmixing of linearly mixed hyperspectral images. Their application to HU of nonlinearly mixed images is not trivial, especially considering the interplay of the choice of number of bands and the actual BS in determining unmixing quality. Recently, we presented initial results on a BS method to reduce the computational complexity of kernel-based nonlinear HU [37]. The method in [37] employs the kernel k-means algorithm to identify clusters of spectral bands in the corresponding RKHS. The cluster prototypes are then the selected bands. This method reduces significantly the computation time required for nonlinear unmixing without compromising the accuracy of abundance estimation. In this approach, however, bands are selected based solely on the distances between the RKHS images of the input vectors, neither as a function of the resulting accuracy of the unmixing procedure, nor as a function of the reconstruction error of any band from the bands in the dictionary. In addition, it requires an *a priori* definition of the final number of bands. Hence, some cluster prototypes can be close to others and degrade problem conditioning if this parameter is overestimated. To the best of our knowledge, [37] is the only method available in the literature for kernel-based nonlinear unmixing of hyperspectral data with band selection in the RKHS.

In this paper we propose new coherence-based approaches for band selection in the RKHS for application in kernel-based supervised nonlinear unmixing of hyperspectral images. The main advantage of a coherence-based method with respect to the approach in [37] is that it seeks a reduced number of bands that are able to represent well all the other bands in the RKHS (in the sense of the reconstruction error of any band from the bands in the dictionary) without requiring extensive calculations for selecting the basis elements. The effectiveness of the coherence criterion for online prediction of time series has been verified, for instance, in [38].

[1]Although PBS select bands in the original data space, we consider this method in the experiment section for benchmarking.

Initially, we introduce an automatic parameter selection strategy to be employed in the band selection and unmixing algorithms. Using this parameter set, we then propose a new greedy band selection algorithm. This greedy approach is appropriate for online settings due to its reduced complexity. Though it leads to significant complexity reductions with similar performance when compared to the full band unmixing, the online constraint requires a sequential dictionary construction, what may preclude the examination of a large amount of possibilities. We then address offline band selection problem by formulating it as a maximum clique problem. This new formulation allows a systematic search for the largest number of bands required to form a linearly independent RKHS basis for a given coherence level. The new method results in dictionaries of kernel functions, and thus spectral bands, that are less coherent than those obtained using kernel k-means initialized with dictionaries of the same size. The application of this method is verified to lead to better average performance than those obtained using either kernel k-means or the proposed greedy algorithm, with nonlinear unmixing performance equivalent to the full band solution.

This paper is organized as follows. Section II reviews the main concepts of hyperspectral image unmixing and two popular nonlinear mixing models that will be used to generate synthetic data for the simulations presented in Section VII. Section III reviews least-square support vector regression applied to hyperspectral unmixing and the SK-Hype algorithm for nonlinear unmixing. This algorithm will be used to compare the nonlinear unmixing results obtained using all the spectral bands and performing band selection prior to unmixing. Section IV introduces an original parameter selection strategy to be employed in the band selection algorithms. The new methodology aims to select a reduced number of bands that are able to represent well all the other bands in the RKHS (in the sense of the reconstruction error of any band from the bands in the dictionary) and that does not require extensive calculations for selecting the elements of this basis. Using this dictionary construction methodology, Section V proposes a new greedy band selection algorithm. Then, in Section VI we show that the band selection problem can be formulated as a maximum clique problem, which leads to a novel batch algorithm for band selection. Section VII presents simulation results with synthetic and real data to illustrate the performances of the proposed band selection methods.

II. Hyperspectral Images and unmixing

Observed pixels in HIs are usually modeled as a function, possibly nonlinear, of the endmembers and an additive noise that accounts for the measurement noise plus a modeling error, namely,

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

where $\boldsymbol{r} = [r_1, \ldots, r_L]^\top$ is a vector of observed reflectances in L spectral bands, $\boldsymbol{M} = [\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R]$ is the $L \times R$ matrix of R endmembers, whose i-th column $\boldsymbol{m}_i$ corresponds to an endmember, $\boldsymbol{e}$ is a white Gaussian noise (WGN) vector, and function $\boldsymbol{\psi}$ represents an unknown mixing mechanism. Several models of the form (1) were proposed in the literature,

depending on the linearity or nonlinearity of $\boldsymbol{\psi}$, the nature of mixture, and other properties [5].

A. The linear mixing model

The linear mixing model (LMM) considers only interactions of light rays with a single material, neglecting interactions between light and several materials [3]. The LMM assumes that $\boldsymbol{r}$ is a convex combination of the endmembers, namely,

$$\begin{aligned} &\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \boldsymbol{e} \\ &\text{subject to } \mathbf{1}^\top \boldsymbol{\alpha} = 1 \text{ and } \boldsymbol{\alpha} \succeq \mathbf{0} \end{aligned} \tag{2}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_R]^\top$ denotes the vector of abundances of each endmember in $\boldsymbol{M}$, and $\succeq$ is the entrywise $\geq$ operator. Being proportions, 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

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

where $\boldsymbol{m}_{\lambda_\ell}$ denotes the ℓ-th row of $\boldsymbol{M}$ written as a column vector. In the noiseless case ($e_\ell = 0$), the sum-to-one and positivity constraints over $\boldsymbol{\alpha}$ in (2) restrict the data to a simplex whose vertices are the endmembers.

B. Nonlinear mixing models

Most unmixing techniques require the definition of a mixing model that analytically describes how the endmembers combine to form the mixed spectrum $\boldsymbol{r}$. Ideally, mixing models should be physically motivated, linking its parameters to physical phenomena. Physically motivated models exist. For instance, [39] provides a description of the mixing mechanism that is based on radiative transfer, and that can accurately describe the light scattering happening in the observed scene. Unfortunately, however, such physically motivated models tend to lead to very complex unmixing problems. Hence, most HI unmixing methods rely on models that include reasonable simplifying assumptions and that lead to manageable unmixing problems.

Several nonlinear models have been proposed to describe complex mixing mechanisms [5], [6]. We now review two popular models that will be used later. These models permit the generation of synthetic images with a controlled level of nonlinearity that can be used to assess the performance of the proposed methods. The level of performance assessment using real images is compromised by the uncertainties of the actual mixing details (lack of ground truth).

The generalized bilinear model (GBM) [40], [41] is defined as follows:

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

where each parameter $\delta_{ij} \in [0, 1]$ characterizes the interaction of endmembers $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$, and $\odot$ denotes the Hadamard product. For simplicity, we shall consider a simplified version of this model where all the bilinear terms in (4) are weighted by a single parameter $\delta = \delta_{ij}$ for all (i, j).

The post nonlinear mixing model (PNMM) [41] is defined as follows:

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

where $\boldsymbol{g}$ is a nonlinear function applied to the noiseless LMM. Thanks to function $\boldsymbol{g}$, the PNMM specifies a large family of nonlinear mixing models via a single expression. For instance, the PNMM considered in [16] is given by

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

where $(\boldsymbol{v})^\xi$ denotes the exponentiation applied to each entry of $\boldsymbol{v}$. For $\xi = 2$, (6) is a bilinear model closely related to the GBM but without a linear term. The PNMM has been explored with different forms for $\boldsymbol{g}$ [42], [43].

The GBM and the PNMM models essentially describe situations where the light interacts first with an endmember, and then with a second one, before being captured by the hyperspectral sensor. Other nonlinear models can be considered depending on the characteristics of the scene [40], [41], [44]–[50]. More importantly, information about these characteristics is usually missing, and it makes sense to consider nonparametric models that do not rely on strong assumptions.

III. LS-SVR for hyperspectral unmixing

Kernel-based methods consist of mapping observations from the original input space into a feature space by means of a nonlinear function. Nonlinear regression problems can be addressed in an efficient way in this new space as they are converted to a linear problem. We shall now review the main definitions related to RKHS [20], [51]–[53].

A. Mercer kernels and RKHS

The theory of positive definite kernels emerged from the study of positive definite integral operators [54], and was further generalized for the study of positive definite matrices [55]. It was established that, to every positive definite function

$$\kappa : \mathcal{M} \times \mathcal{M} \to \mathbb{R} \tag{7}$$

defined over a non-empty compact $\mathcal{M} \subset \mathbb{R}^d$, there corresponds one and only one family of real-valued functions on $\mathcal{M}$ that defines a Hilbert space $\mathcal{H}$ endowed with an unique inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and the associated norm $\|\cdot\|_{\mathcal{H}}$, and admitting κ as a reproducing kernel [22]. This means that $\kappa(\cdot, \boldsymbol{m}) \in \mathcal{H}$ for all $\boldsymbol{m} \in \mathcal{M}$, and has the reproducing property defined as:

$$\psi(\boldsymbol{m}) = \langle \psi, \kappa(\cdot, \boldsymbol{m}) \rangle_{\mathcal{H}} \tag{8}$$

for all $\psi \in \mathcal{H}$ and $\boldsymbol{m} \in \mathcal{M}$. Replacing ψ by $\kappa(\cdot, \boldsymbol{m}')$ in (8) leads to:

$$\kappa(\boldsymbol{m}, \boldsymbol{m}') = \langle \kappa(\cdot, \boldsymbol{m}), \kappa(\cdot, \boldsymbol{m}') \rangle_{\mathcal{H}} \tag{9}$$

for all $\boldsymbol{m}, \boldsymbol{m}' \in \mathcal{M}$. Equation (9) is the origin of the now generic denomination reproducing kernel to refer to κ. Note that $\mathcal{H}$ can be restricted to the span of $\{\kappa(\cdot, \boldsymbol{m}) : \boldsymbol{m} \in \mathcal{M}\}$ because, according to the reproducing property (8), nothing outside this set affects ψ evaluated at any point of $\mathcal{M}$. Let us denote by φ the map from $\mathcal{M}$ to $\mathcal{H}$ that assigns $\kappa(\cdot, \boldsymbol{m})$ to $\boldsymbol{m}$. Relation (9) implies that $\kappa(\boldsymbol{m}, \boldsymbol{m}') = \langle \varphi(\boldsymbol{m}), \varphi(\boldsymbol{m}') \rangle_{\mathcal{H}}$. This

means that the kernel κ evaluates the inner product of any pair of elements of $\mathcal{M}$ mapped into $\mathcal{H}$ without any explicit knowledge of φ or $\mathcal{H}$. This principle is called the kernel trick.

Several kernel functions have been considered in a variety of applications during the past two decades [56]. Among the most frequently used kernels, we highlight the Gaussian kernel:

$$\kappa(\boldsymbol{m}, \boldsymbol{m}') = \exp\left(-\frac{\|\boldsymbol{m} - \boldsymbol{m}'\|^2}{2\sigma^2}\right) \tag{10}$$

where σ is the kernel bandwidth.

B. The Representer Theorem

The BS method proposed in this paper is based on the approximation of functions in a RKHS. The following is a well known result for nonlinear regressions using reproducing kernels.

Let $\kappa : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ be a kernel, and let $\mathcal{H}$ be the RKHS associated with it. Consider the least-squares approach to solve the problem of determining a function $\psi(\cdot)$ of $\mathcal{H}$ that minimizes the sum of L squared errors e_ℓ between samples r_ℓ of the desired response and the corresponding model output samples $\psi(\boldsymbol{m}_\ell) = \langle \psi(\cdot), \kappa(\cdot, \boldsymbol{m}_\ell)\rangle_{\mathcal{H}}$, namely,

$$\min_{\psi\in\mathcal{H}} \sum_{\ell=1}^{L} |r_\ell - \psi(\boldsymbol{m}_\ell)|^2. \tag{11}$$

By virtue of the representer theorem [53], the function $\psi(\cdot)$ of $\mathcal{H}$ minimizing (11) can be written as a kernel expansion in terms of available data:

$$\psi(\cdot) = \sum_{\ell=1}^{L} \alpha_\ell \, \kappa(\cdot, \boldsymbol{m}_\ell). \tag{12}$$

Regarding the hyperspectral unmixing problem, consider an observation r_ℓ at the ℓ-th wavelength, that is, the ℓ-th entry of $\boldsymbol{r}$, and the column vector $\boldsymbol{m}_{\lambda_\ell}$ of the R endmember signatures at the ℓ-th wavelength, that is, the (transposed) ℓ-th row of $\boldsymbol{M}$. By analogy with the LMM (3), we write:

$$r_\ell = \psi(\boldsymbol{m}_{\lambda_\ell}) + e_\ell \tag{13}$$

with ψ a real-valued function in a RKHS $\mathcal{H}$ that characterizes the nonlinear interactions between the endmembers, and e_ℓ an additive noise at the ℓ-th band. The next section describes the classical application of a kernel method to nonlinear unmixing of hyperspectral data. Different kernel-based unmixing methods could be considered. As the focus of this work is on the improvement that can be obtained using BS prior to unmixing, we consider a single state-of-the-art nonlinear unmixing algorithm (the SK-Hype), which been shown to provide good unmixing performance [5], [15], [16].

C. LS-SVR: least squares support vector regression

In order to estimate ψ in the least squares sense, we can formulate the following convex optimization problem, also called LS-SVR [21]:

$$\begin{aligned} &\min_{\psi\in\mathcal{H}} \ \frac{1}{2}\|\psi\|_{\mathcal{H}}^2 + \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 \\ &\text{such that} \quad e_\ell = r_\ell - \psi(\boldsymbol{m}_{\lambda_\ell}), \quad \ell = 1, \ldots, L. \end{aligned} \tag{14}$$

Consider the Lagrangian function

$$\mathcal{L}(\psi, \boldsymbol{e}, \boldsymbol{\beta}) = \frac{1}{2}\|\psi\|_{\mathcal{H}}^2 + \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 - \sum_{\ell=1}^{L} \beta_\ell \left(e_\ell - r_\ell + \psi(\boldsymbol{m}_{\lambda_\ell})\right). \tag{15}$$

where $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_L]^\top$ is the vector of Lagrange multipliers. Using the directional derivative with respect to ψ [57], the conditions for optimality with respect to the primal variables ψ and e_ℓ are given by

$$\psi^* = \sum_{\ell=1}^{L} \beta_\ell \kappa(\cdot, \boldsymbol{m}_{\lambda_\ell}) \tag{16}$$

$$e_\ell^* = \mu\beta_\ell \tag{17}$$

Substituting (16) and (17) in (15), we obtain the following function to be maximized with respect to $\boldsymbol{\beta}$:

$$\mathcal{L}(\psi^*, \boldsymbol{e}^*, \boldsymbol{\beta}) = -\frac{1}{2}\,\boldsymbol{\beta}^\top(\boldsymbol{K} + \mu\boldsymbol{I})\,\boldsymbol{\beta} + \boldsymbol{\beta}^\top\boldsymbol{r}, \tag{18}$$

where $\boldsymbol{K}$ is the Gram matrix whose (i, j)-th entry is defined by $\kappa(\boldsymbol{m}_{\lambda_i}, \boldsymbol{m}_{\lambda_j})$. Now we can state the following dual problem:

$$\boldsymbol{\beta}^* = \arg\max_{\boldsymbol{\beta}} -\frac{1}{2}\,\boldsymbol{\beta}^\top(\boldsymbol{K} + \mu\boldsymbol{I})\,\boldsymbol{\beta} + \boldsymbol{\beta}^\top\boldsymbol{r}. \tag{19}$$

Its solution is obtained by solving the linear system:

$$\left(\begin{array}{c|c} -\boldsymbol{I} & \boldsymbol{K} + \mu\boldsymbol{I} \end{array}\right)\left(\frac{\boldsymbol{r}}{\boldsymbol{\beta}}\right) = \boldsymbol{0}. \tag{20}$$

Although the formulation (14)–(19) allows one to address an estimation problem in $\mathcal{H}$ by solving the linear system (20), this approach is computationally demanding since it involves the inversion of $L \times L$ matrices. This issue is critical, as modern hyperspectral image sensors employ hundreds of contiguous bands with an ever increasing spatial resolution. Hence, it is of major interest to consider band selection techniques that lead to significant computational cost reduction without noticeable quality loss. Considering (16), a possible strategy is to focus on a reduced-order model of the form:

$$\psi = \sum_{j\in\mathcal{I}_D} \beta_j \kappa(\cdot, \boldsymbol{m}_{\lambda_j}) \tag{21}$$

where $\mathcal{I}_D \subset \{1, \ldots, L\}$ is an M-element ($M < L$) subset of indexes. We shall call $\mathcal{D} = \{\kappa(\cdot, \boldsymbol{m}_{\lambda_j})\}_{j\in\mathcal{I}_D}$ the dictionary.

This reasoning is well established in the study of representation of nonlinear functions in a RKHS, and directly applies to the solution of the LS-SVR. The observed hyperspectral image is a nonlinear function of the L rows $\boldsymbol{m}_{\lambda_\ell}$, $\ell = 1, \ldots, L$, of $\boldsymbol{M}$, and kernel-based nonlinear unmixing algorithms rely on a good representation of this nonlinear function in the RKHS. Also, LS-SVR performs unmixing by maximizing convex criteria in this RKHS. Finally, the representer theorem establishes that the optimal solution can be written as a linear combination of kernelized inputs, in our case kernelized bands as shown in (16). Hence, a good strategy to construct $\mathcal{I}_D$ in (21) is to seek a reduced set of linearly independent functions $\kappa(\cdot, \boldsymbol{m}_{\lambda_j})$, $j \in \mathcal{I}_D$ that could be linearly combined to reconstruct any $\kappa(\cdot, \boldsymbol{m}_{\lambda_j})$, $j \notin \mathcal{I}_D$ with a reduced error.

One low complexity criterion that can be used to this end is the coherence [38], as will be shown latter.

Now, to solve the hyperspectral nonlinear unmixing problem frequently requires a formulation that includes the possibility of estimating the abundances associated with each endmember. This is not possible using the cost function (14), which is not parameterized by the endmember abundances. The next section briefly reviews the SK-Hype algorithm [15], which addresses this limitation by adding an extra linear term to the cost function and by introducing a new parameter u to adequately distribute the solution of the optimization problem between the linear and the nonlinear contributions.

D. The SK-Hype Algorithm

This section reviews the SK-Hype algorithm[2] for nonlinear unmixing of HIs [15]. It generalizes the LS-SVR method by considering a mixing model consisting of a linear trend parametrized by the abundance vector $\boldsymbol{\alpha}$ and a nonlinear residual component ψ. This model is given by:

$$r_\ell = u\,\boldsymbol{\alpha}^\top \boldsymbol{m}_{\lambda_\ell} + (1-u)\,\psi(\boldsymbol{m}_{\lambda_\ell}) + e_\ell \tag{22}$$

where $u \in [0,1)$ controls the amount of linear contribution to the model and $\psi(\cdot)$ is an unknown function in an RKHS $\mathcal{H}$. SK-Hype solves the optimization problem

$$\begin{aligned} \min_{\boldsymbol{\alpha},\psi,u} \; & \frac{1}{2}\left(\frac{1}{u}\|\boldsymbol{\alpha}\|^2 + \frac{1}{1-u}\|\psi\|^2_{\mathcal{H}}\right) + \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 \\ \text{subject to} \quad & \boldsymbol{\alpha} \succeq \mathbf{0},\ \mathbf{1}^\top\boldsymbol{\alpha} = \mathbf{1},\ \text{and} \\ & e_\ell = r_\ell - u\,\boldsymbol{\alpha}^\top\boldsymbol{m}_{\lambda_\ell} - (1-u)\,\psi(\boldsymbol{m}_{\lambda_\ell}). \end{aligned} \tag{23}$$

which is convex under mild continuity conditions [15]. Problem (23) is solved using a two stage alternating iterative procedure with respect to $(\boldsymbol{\alpha},\psi)$ and u. For fixed u and Lagrange multipliers $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, the dual problem of (23) is given by [15]

$$\begin{aligned} \max_{\boldsymbol{\beta},\boldsymbol{\gamma}} \; & G(u,\boldsymbol{\beta},\boldsymbol{\gamma}) = \\ & -\frac{1}{2}\begin{pmatrix}\boldsymbol{\beta}\\ \hline \boldsymbol{\gamma}\end{pmatrix}^\top \left(\begin{array}{c|c} \boldsymbol{K}_u + \mu\boldsymbol{I} & u\boldsymbol{M} \\ \hline u\boldsymbol{M}^\top & u\boldsymbol{I}\end{array}\right)\begin{pmatrix}\boldsymbol{\beta}\\ \hline \boldsymbol{\gamma}\end{pmatrix} \\ & + \begin{pmatrix}\boldsymbol{r}\\ \hline \mathbf{0}\end{pmatrix}^\top \begin{pmatrix}\boldsymbol{\beta}\\ \hline \boldsymbol{\gamma}\end{pmatrix} \\ & \text{subject to} \quad \boldsymbol{\gamma} \succeq \mathbf{0} \end{aligned} \tag{24}$$

with $\boldsymbol{K}_u = u\boldsymbol{M}\boldsymbol{M}^\top + (1-u)\boldsymbol{K}$. Solving (24) is equivalent to solving the linear system

$$\left(\begin{array}{c|c|c} -\boldsymbol{I} & \boldsymbol{K}_u + \mu\boldsymbol{I} & u\boldsymbol{M} \\ \hline \mathbf{0} & u\boldsymbol{M}^\top & u\boldsymbol{I}\end{array}\right)\begin{pmatrix}\boldsymbol{r}\\ \hline \boldsymbol{\beta}\\ \hline \boldsymbol{\gamma}\end{pmatrix} = \mathbf{0}. \tag{25}$$

Denoting $\boldsymbol{\beta}^*$ and $\boldsymbol{\gamma}^*$ the solutions of (24), the solution of the primal problem (23) for u fixed is [15]

$$\begin{cases} \boldsymbol{\alpha}^* = \frac{\boldsymbol{M}^\top\boldsymbol{\beta}^* + \boldsymbol{\gamma}^*}{\mathbf{1}^\top(\boldsymbol{M}^\top\boldsymbol{\beta}^* + \boldsymbol{\gamma}^*)} \\ \psi^* = (1-u)\sum_{\ell=1}^{L}\beta_\ell^*\,\kappa(\cdot,\boldsymbol{m}_{\lambda_\ell}) \\ e_\ell^* = \mu\,\beta_\ell^* \end{cases} \tag{26}$$

[2]Matlab code available at www.cedric-richard.fr

The alternating optimization is completed by using (26) in [15], defining the resulting cost function $J(u)$, solving

$$\min_u J(u) \quad \text{subject to} \quad 0 < u < 1 \tag{27}$$

and continue by iteratively solving (25) and (27) to find the global solution [15].

IV. Setting the Parameters for Band Selection

BS has been an active topic of research for classification of spectral patterns, see [24], [58]–[61] and references therein. Subspace projection techniques [33], [62], [63] tend, however, to be preferred over BS [26], [35] for reducing the complexity of linear unmixing processes. They use the property that high-dimensional hyperspectral data are confined to a low-dimensional simplex in linearly-mixed images with only a few endmembers [3]. This assumption becomes invalid when nonlinear mixing phenomena are involved. In a recent work [37], we introduced a BS strategy method that employs the kernel k-means algorithm to identify clusters of spectral bands in the RKHS where nonlinear unmixing is performed. The HU results obtained were encouraging. One drawback of the approach in [37] is the need for an arbitrary choice of the order of the nonlinear model, i.e., the number of selected bands (or, equivalently, the dimension of the dictionary). Another limitation of the method in [37] is that BS is performed based on the distances among different bands in the RKHS. Hence, the optimality of the solution is not driven by any direct measure of modeling accuracy. In this section, we introduce a new BS framework based on the coherence criterion [38] and two algorithms to address this problem. The first is a greedy approach, which is attractive for online applications which require reduced computational effort. The second is a centralized approach, in which the solution is found through a maximum clique search in a graph. Although these two approaches are connected, they differ in their formulation and in the characteristics of the sets of bands they select. The use of a coherence-based method allows the search for a reduced number of bands that are able to represent well all the other bands in the RKHS (in the sense of the reconstruction error of any band from the bands in the dictionary) without the need for extensive calculations to select the elements of this basis.

A. Coherence criterion for dictionary selection

Coherence is a parameter of fundamental interest for characterizing dictionaries of atoms in linear sparse approximation problems [64]. It was first introduced as an heuristic quantity for Matching Pursuit in [65]. Formal studies followed in [66], and were enriched for Basis Pursuit in [67], [68].

Consider a set of kernel functions $\{\kappa(\cdot,\boldsymbol{m}_{\lambda_\ell})\}_{\ell=1,\ldots,M}$ in $\mathcal{H}$. The definition of coherence was extended to RKHS as [38]:

$$\begin{aligned} \mu &= \max_{i\neq j} \left|\langle \kappa(\cdot,\boldsymbol{m}_{\lambda_i}), \kappa(\cdot,\boldsymbol{m}_{\lambda_j})\rangle_{\mathcal{H}}\right| \\ &= \max_{i\neq j} \left|\kappa(\boldsymbol{m}_{\lambda_i},\boldsymbol{m}_{\lambda_j})\right| \end{aligned} \tag{28}$$

where κ is a unit-norm kernel. Otherwise, replace $\kappa(\cdot,\boldsymbol{m}_{\lambda_i})$ with $\kappa(\cdot,\boldsymbol{m}_{\lambda_i})/\sqrt{\kappa(\boldsymbol{m}_{\lambda_i},\boldsymbol{m}_{\lambda_i})}$ in (28). Parameter μ is the largest absolute value of the off-diagonal entries in the Gram

matrix. It reflects the largest cross correlation in the dictionary $\{\kappa(\cdot, \boldsymbol{m}_{\lambda_\ell})\}_\ell$, and is equal to zero for every orthonormal basis. A dictionary is said to be incoherent when its coherence μ is small. Although its definition is rather simple, coherence possesses important properties [38]. In particular, it can be shown that the kernel functions in the dictionary $\mathcal{D} = \{\kappa(\cdot, \boldsymbol{m}_{\lambda_\ell})\}_{\ell=1,\ldots,M}$ are linearly independent if $(M-1)\mu < 1$. This sufficient condition illustrates that the coherence (28) provides valuable information on a dictionary at low computational cost. Other properties are discussed in [38].

Kernel-based dictionary learning methods usually consider approximate linear dependence conditions to evaluate whether a candidate kernel function $\kappa(\cdot, \boldsymbol{m}_{\lambda_i})$ can be reasonably well represented by a combination of the kernel functions that are already in the dictionary $\mathcal{D}$. To avoid excessive computational complexity, a greedy dictionary learning method has been introduced in [38]. It consists of inserting the candidate $\kappa(\cdot, \boldsymbol{m}_{\lambda_i})$ into the dictionary $\mathcal{D}$ provided its coherence is still below a given threshold μ_0, namely,

$$\max_{j \in \mathcal{I}_\mathcal{D}} |\kappa(\boldsymbol{m}_{\lambda_i}, \boldsymbol{m}_{\lambda_j})| \leq \mu_0 \tag{29}$$

where μ_0 is a parameter in $[0, 1)$ determining both the maximum coherence in $\mathcal{D}$ and its cardinality $|\mathcal{D}|$.

Using coherence criterion for BS allows to explicitly limit the correlation of kernel functions in the dictionary. This contrasts with the kernel k-means strategy, which starts from a number of dictionary elements prescribed by the user without taking the coherence of kernel functions into consideration.

B. Automatic parameter settings

A coherence-based BS in the RKHS includes the choice of three design parameters, namely the coherence threshold μ_0, the dictionary cardinality $|\mathcal{D}|$ and the kernel bandwidth σ. This section proposes a procedure for automatic parameter setting.

Let $\boldsymbol{K}_\sigma$ be the $L \times L$ Gram matrix whose (i, j)-th entry is defined by $\kappa_\sigma(\boldsymbol{m}_{\lambda_i}, \boldsymbol{m}_{\lambda_j})$, where κ_σ denotes the Gaussian kernel (10) parametrized by the bandwidth σ. Let $\mathcal{D}$ be an M-element dictionary with coherence μ and index set $\mathcal{I}_\mathcal{D}$. Then, as shown in [38], a sufficient condition for linear independence of the M elements of $\mathcal{D}$ is given by $(M-1)\mu < 1$. We write:

$$\mu < \frac{1}{(M-1)}. \tag{30}$$

We assume an initial objective to build a dictionary with (approximately) M linearly independent elements. We thus propose to set the coherence threshold μ_0 as:

$$\mu_0 = \frac{1}{(M-1)} \tag{31}$$

and adjust the bandwidth σ to obtain a Gram matrix $\boldsymbol{K}_\sigma$ whose entries are close to μ_0 in some sense. To allow some flexibility in the construction of the dictionary, we propose to adjust σ such that the average value of the off-diagonal entries $\boldsymbol{K}_{\sigma_{ij}}|_{(i \neq j)}$ is equal to μ_0. Hence, we set

$$\frac{2}{L^2 - L} \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} \boldsymbol{K}_{\sigma_{ij}} = \mu_0 \tag{32}$$

and determine σ^2 as the solution of the following optimization problem:

$$\begin{aligned} \sigma^2 = \arg\min_{\sigma^2} & \left(\frac{2}{L^2 - L} \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} [\boldsymbol{K}_{1_{ij}}]^{1/\sigma^2} - \mu_0 \right)^2 \\ \text{s. t.} \quad & \sigma^2 \in \mathbb{R}^+ \end{aligned} \tag{33}$$

where $\boldsymbol{K}_1 = \boldsymbol{K}_\sigma$ is the Gram matrix for $\sigma = 1$. We emphasize that since $\boldsymbol{K}_{\sigma_{ij}} \leq 1$, (32) is a decreasing function of σ^{-2}, and thus (33) has a unique solution. Finally, we determine $\boldsymbol{K}_\mathcal{D}$ as the largest sub-matrix of $\boldsymbol{K}_\sigma$ whose all off-diagonal entries satisfy (29), which will be addressed in the sequel.

V. Greedy Band Selection Algorithm

Inspired by the good results obtained in [38] for online data prediction, we propose in this section a greedy BS criterion using the parameters determined by the procedure described in Section IV-B.

The greedy coherence-based band selection (GCBS) algorithm starts with an initial guess M for the number of bands to be selected. The dictionary is then initialized by arbitrarily selecting the first band. The procedure then continues for the other bands, and is summarized by the following steps:

1) Given a maximum desired number of bands M, the coherence threshold μ_0 and the kernel bandwidth σ are determined using (31) and solving (33) respectively.
2) The Gram matrix $\boldsymbol{K}_\sigma$ is built. The first band is randomly selected and included in the dictionary index $\mathcal{I}_\mathcal{D}$.
3) The coherences between the bands not in the dictionary and bands in $\mathcal{I}_\mathcal{D}$ are sequentially compared with μ_0 and added to the dictionary if condition (29) is satisfied.

The GCBS is detailed in Algorithm 1. The inputs to Algorithm 1 are the maximum desired number M of bands and the $L \times L$ Gaussian kernel Gram matrix for $\sigma = 1$ and entries $\boldsymbol{K}_{1_{ij}} = \kappa(\boldsymbol{m}_{\lambda_i}, \boldsymbol{m}_{\lambda_j}) = \exp\left(-0.5\|\boldsymbol{m}_{\lambda_i} - \boldsymbol{m}_{\lambda_j}\|^2\right)$. It returns the index of selected bands and the Gaussian kernel bandwidth σ. Initialization occurs in line 1, where the index set $\mathcal{I}_\mathcal{D}$ is initialized with the first spectral band index, the number N_b of bands in the dictionary is set to one, and the coherence threshold μ_0 is adjusted according to (31). Next, σ^2 is determined in line 2 by solving problem (33), and the Gram matrix $\boldsymbol{K}_\sigma$ is computed in line 3 for the optimum σ^2. From line 4 to line 13 the algorithm sequentially tests all the $L-1$ remaining bands using condition (29). Breaking the parts down, in line 5 a zero vector $\boldsymbol{c}$ of length N_b is created, and the off diagonal terms $(\ell, \mathcal{I}_{\mathcal{D}_j})$ of the Gram matrix $\boldsymbol{K}_\sigma$ are stored in $\boldsymbol{c}$. If the maximum absolute value of the entries of $\boldsymbol{c}$ is less than the coherence threshold (line 9), then the ℓ-th band index is added to $\mathcal{I}_\mathcal{D}$, and N_b is incremented by one (lines 10 and 11). Finally, the algorithm returns the complete set of selected bands and the kernel bandwidth in line 14.

Although GCBS has a reduced computational cost, its dependence on the sequence of bands tested (initialization) may result in suboptimal solutions. The next section presents an alternative algorithm that leads to more effective solutions for batch mode applications.

Algorithm 1: Greedy Coherence-based BS (GCBS)

Input : The $L \times L$ Gram matrix $\boldsymbol{K}_1 = (\boldsymbol{K}_\sigma)_{\sigma=1}$, and the desired number M of atoms.

Output: The indices $\mathcal{I}_\mathcal{D}$ of selected atoms, and the Gaussian kernel bandwidth σ^2.

1 Initialization: $\mathcal{I}_\mathcal{D} = \{1\}$, $N_b = 1$, $\mu_0 = 1/(M-1)$;
2 Find σ^2 solving (33);
3 Compute $\boldsymbol{K}_\sigma$ using σ^2 obtained in line 2;
4 **for** $\ell := 2$ **to** L **do**
5 $\boldsymbol{c} := \boldsymbol{0}_{N_b \times 1}$;
6 **for** $j := 1$ **to** N_b **do**
7 $\boldsymbol{c}_j := \boldsymbol{K}_{\sigma_{\ell,\mathcal{I}_{\mathcal{D}_j}}}$;
8 **end**
9 **if** $\max(|\boldsymbol{c}_j|) \leq \mu_0$ **then**
10 Insert ℓ into $\mathcal{I}_\mathcal{D}$;
11 $N_b := N_b + 1$;
12 **end**
13 **end**
14 **return** $\mathcal{I}_\mathcal{D}$, σ^2;

VI. Batch band selection algorithm

This section proposes a new batch band selection algorithm in the RKHS. First, we show that the BS in RKHS can be formulated as the solution of a maximum clique problem. Then, we briefly review the major solutions available to solve this problem and justify the choice of the algorithm to be employed in our proposed BS method. Finally, we present the details of the proposed batch BS algorithm.

A. Band selection as a maximum clique problem

To circumvent the limitations of GCBS in identifying all bands satisfying the prescribed coherence threshold, a possible approach is to seek the largest subset of spectral bands for which the coherences between all pairs of elements are below the threshold μ_0.

Consider a set of kernel functions $\{\kappa(\cdot, \boldsymbol{m}_{\lambda_\ell})\}_{\ell=1,\ldots,L}$. Determining a subset $\mathcal{D}$ with a prescribed coherence level can be viewed as a two-step procedure. The first step aims at listing all the pairs of functions that satisfy the coherence rule (29). This can be performed by constructing a $L \times L$ binary matrix $\boldsymbol{B}$ with entries defined as:

$$\boldsymbol{B}_{ij} = \begin{cases} 1 & \text{if } |\kappa(\boldsymbol{m}_{\lambda_i}, \boldsymbol{m}_{\lambda_j})| \leq \mu_0 \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

The second step consists of finding in $\boldsymbol{B}$, up to a simultaneous reordering of its rows and columns, the largest submatrix of only ones. This problem can be recast as determining a maximum clique in an undirected graph $\mathcal{G} = \{V, E\}$, where each vertex ℓ of $V = \{1, \ldots, L\}$ corresponds to a candidate function $\kappa(\cdot, \boldsymbol{m}_{\lambda_\ell})$, and edges in $E \subseteq V \times V$ connecting the vertices are defined by the adjacency matrix $\boldsymbol{B}$. Two vertices are said to be adjacent if they are connected by an edge. A complete subgraph of $\mathcal{G}$ is one whose vertices are pairwise adjacent. The maximal clique problem (MCP) consists of finding the maximal complete subgraph of $\mathcal{G}$ [69]. This problem is NP-Complete [70]. Figure 1 illustrates this problem in the context of BS. This figure shows, for instance, that the coherence of $\kappa(\cdot, \boldsymbol{m}_{\lambda_1})$ and $\kappa(\cdot, \boldsymbol{m}_{\lambda_4})$ is lower than the preset threshold μ_0, and the coherence of $\kappa(\cdot, \boldsymbol{m}_{\lambda_1})$ and $\kappa(\cdot, \boldsymbol{m}_{\lambda_2})$ is larger than μ_0. This graph has one maximum clique defined by the set of vertices $\mathcal{I}_\mathcal{D} = \{1, 3, 4, 5\}$, which means that the coherence of the dictionary $\mathcal{D} = \{\kappa(., \boldsymbol{m}_{\lambda_j})\}_{j \in \mathcal{I}_D}$ is lower than μ_0 and it has maximum cardinality. A vast literature exists on maximum clique problems (MCP), see [71] and references therein. The next section reviews the main algorithms for MCP.

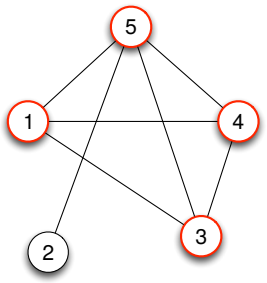

Fig. 1: The maximum clique problem (MCP)

B. The maximum clique problem

MCP has a wide range of practical applications arising in a number of domains such as bioinformatics, coding theory, economics, social network analysis, etc. Given its theoretical importance and practical interests, considerable efforts have been devoted for deriving exact and heuristic algorithms. Efficient exact methods have been designed mainly based on the branch-and-bound (B&B) framework. Dynamic bounds on the clique size are used to prune (or discard) branches during search, and then dramatically reducing the search space [72]. Although algorithms are now much faster and efficient than their past counterparts [73], the inherent complexity of exact methods can still lead to a prohibitive computation time when large problems are addressed [71]. To handle problems whose optimal solutions cannot be reached within a reasonable time, various heuristic and metaheuristic algorithms have been derived with the purpose of providing sub-optimal solutions in an acceptable time. In this paper, however, we shall focus on exact algorithms since our application concerns small graphs with a number of vertices equal to the number of bands.

Since the introduction of the Carraghan and Pardalos (CP) exact algorithm [72], many refinements have been proposed to improve its performance with a focus on two main issues. The first one is to tighten the upper bound on the maximum clique during search for the purpose of more efficient subtree pruning. The second one is to improve the branching rule, and then select the most promising vertices to expand candidate cliques. In [71], the authors classify the exact MCP algorithms into four groups, depending on their strategies for pruning and branching. The first group solves sub-clique problems for each vertex with iterative deepening and pruning strategies. Examples are the CP algorithm [72] and its improved version [74]. Both algorithms are sensitive to the order of vertices,

which can result in drastically different execution times for a given graph [74]. A second group is based on vertex coloring techniques [75]. The most prominent algorithms in this group use B&B strategies based on subgraph coloring. Examples of algorithms are BT and the recent MCQ, MCR, MaxCliqueDyn, BB-MaxClique, among others [71]. The third group improves the basic CP by tightening candidate sets via the removal of vertices that cannot be used to extend the current clique to a maximum clique. Along this line, three B&B algorithms, denoted DF, χ and χ+DF were proposed in [76]. The fourth group consists of the exact methods based on MaxSAT [77], which improve the techniques based on vertex coloring. The MaxCLQ algorithm proposed in [77] is considered to be very effective and solved the DIMACS problem (p_hat1000–3) for the first time [71]. A complex approach (ILS&MaxCLQ) that combines different algorithms such as the MaxCLQ, MCS and the ILS, was recently proposed [78]. A comparative discussion on exact methods is presented in [71]. The MaxCLQ and ILS&MaxCLQ were the only methods to solve all the presented problems, with the smallest CPU times for the former.

C. The Batch BS Algorithm

Similarly to the GCBS algorithm, this approach uses the methodology presented in Section IV-B to set the Gaussian kernel bandwidth and the coherence threshold. Thus, an initial guess for the maximum desired number M of bands in the dictionary is also needed. Once μ_0 and σ^2 are found, the kernel matrix $\boldsymbol{K}_\sigma$ can be computed. As discussed in Section VI-A solving this problem is equivalent to solve an MCP for a graph with adjacency matrix $\boldsymbol{B}$ built following (34). Then, an MCP algorithm (maxCQL) is used to find the indices of the maximum clique in $\boldsymbol{B}$, that is, the indices of the selected bands. This approach can be summarized as:

1) Initialize the algorithm by choosing M.
2) Find the parameters μ_0 and σ^2 using (31) and (33) respectively.
3) Build the Gram matrix $\boldsymbol{K}_\sigma$.
4) Compare all elements of $\boldsymbol{K}_\sigma$ with μ_0 and build the adjacency matrix $\boldsymbol{B}$ as in (34).
5) Use the MaxCLQ algorithm to solve the MCP and find the selected spectral bands.

The clique coherence-based band selection method is described in Algorithm 2. Similarly to Algorithm 1, the inputs are $\boldsymbol{K}_1$ and M. The adjacency matrix $\boldsymbol{B}$ in initialized with zeros (line 1), the vertices vector V with the indices of all available wavelengths, μ_0 following (31), and $\mathcal{I}_\mathcal{D}$ as an empty set. The kernel bandwidth is computed in line 2, and the Gram matrix is computed in line 3 for the optimum σ^2. In lines 4 through 10 every entry of the upper diagonal part of $\boldsymbol{B}$ is set according to (34). In line 11 the *MaxCLQ* algorithm is used to find the indices of the maximum clique in the graph. These indices are assigned to the dictionary index set $\mathcal{I}_\mathcal{D}$, which is returned in line 10 together with the kernel bandwidth.

Note that M is used in Algorithm 1 and Algorithm 2 as a design parameter, which is required to obtain the coherence threshold and the Gaussian kernel bandwidth. The number N_b of bands in the final dictionary can differ from M.

Algorithm 2: Clique Coherence-based BS (CCBS)

Input : The $L \times L$ Gram matrix $\boldsymbol{K}_1 = (\boldsymbol{K}_\sigma)_{\sigma=1}$, and the desired number M of atoms.
Output: The indices $\mathcal{I}_\mathcal{D}$ of selected atoms, and the Gaussian kernel bandwidth σ^2.

1 Initialization: $\boldsymbol{B} := \boldsymbol{0}_{L\times L}$, $V = \{1, \ldots, L\}$, $\mu_0 = 1/(M-1)$, $\mathcal{I}_{\mathcal{D}_c} = \{\emptyset\}$;
2 Find σ^2 solving (33);
3 $\boldsymbol{K}_\sigma$ using σ^2 obtained in line 2;
4 **for** $i := 1$ **to** $L-1$ **do**
5 **for** $j := i+1$ **to** L **do**
6 **if** $[\boldsymbol{K}_{\sigma_{ij}}] \leq \mu_0$ **then**
7 $\boldsymbol{B}_{ij} := 1$;
8 **end**
9 **end**
10 **end**
11 $\mathcal{I}_\mathcal{D} := \mathit{MaxCLQ}(V, \boldsymbol{B})$;
12 **return** $\mathcal{I}_\mathcal{D}$, σ^2;

VII. Application

A. Simulation with synthetic data

This section presents simulation results using synthetic data to illustrate the performance of the proposed unmixing method under controlled conditions for which the abundance values are known. We constructed synthetic images using a set of 8 endmembers extracted from the spectral library of the ENVI software, which correspond to the spectral signatures of minerals present in the Cuprite mining field in Nevada. The minerals are alunite, calcite, epidote, kaolinite, buddingtonite, almandine, jarosite and lepidolite, and their spectra consisted of 420 contiguous bands, covering wavelengths from 0.3951μm to 2.56μm. We constructed two 2000-pixel hyperspectral images ($N = 2000$), each using 8 endmembers ($R = 8$) from the Cuprite data, and the simplified GBM or PNMM mixing models (see Section II) with $\delta = 1$ and $\xi = 0.7$, respectively. The abundances were obtained by uniformly sampling from the simplex, i.e., obeying the positivity and sum-to-one constraints. WGN was added to all images with power adjusted to produce a 21dB SNR. We consider the root mean square error (RMSE) in abundance estimation

$$\text{RMSE} = \sqrt{\frac{1}{NR}\sum_{n=1}^{N}\|\boldsymbol{\alpha}_n - \boldsymbol{\alpha}_n^*\|^2} \tag{35}$$

and the CPU time required for both BS (when applicable) and unmixing (averaged over 100 unmixings of the same HIs) to compare the different BS strategies. All unmixings were performed using a Gaussian kernel and considering either the full set of bands or smaller sets selected using the BS strategies presented in Section IV. SK-Hype was implemented for the full set of bands. The kernel bandwidth for SK-Hype was selected among the values $\sigma_{\text{skp}} \in \{0.5\sigma, \sigma, 2\sigma, 10\sigma, 20\sigma\}$ to obtain the minimum RMSE, where σ is the solution of (33), for $M = 30$. The global kernel k-means (GKKM) algorithm [37] implementation requires the number of bands to be fixed *a*

priori. To keep the computational complexity under control, we considered a selection approach based on the Akaike Information Criterion (AIC) and given by [79]

$$M = \arg\min_{M} [E(\nu_1, \ldots, \nu_M) + \lambda M] \tag{36}$$

where $\nu_1, \ldots, \nu_M$ are the centroids chosen by the method [37], and

$$E(\nu_1, \ldots, \nu_K) = \sum_{k=1}^{M} \sum_{\ell \in \mathcal{C}_k} \|\kappa(\cdot, \boldsymbol{m}_{\lambda_\ell}) - \nu_k\|^2_{\mathcal{H}}. \tag{37}$$

Each cluster $\mathcal{C}_k$ is then represented by the band ℓ_k corresponding to the closest point to its centroid ν_k:

$$\ell_k = \arg\min_{\ell \in \mathcal{C}_k} \|\kappa(\cdot, \boldsymbol{m}_{\lambda_\ell}) - \nu_k\|^2_{\mathcal{H}}. \tag{38}$$

The parameter λ controls the complexity of the model, and needs to be found empirically. The kernel bandwidth σ_{kkm} also needs to be selected for GKKM. A grid search was performed using a small part (200 pixels) of the synthetic image to find λ and σ_{kkm} that would lead to a good RMSE performance. The parameters were chosen among the values $\lambda \in \{2, 4, 6\}$ and $\sigma_{\text{kkm}} \in \{0.5\sigma,\ \sigma,\ 2\sigma,\ 10\sigma,\ 20\sigma\}$, again with σ being the solution of (33), for $M = 30$. The parameter set leading to the best performance in terms of RMSE for the abundances was then selected. This way we seek a good RMSE performance using the simplest possible model. To complete benchmarking, the Progressive Band Selection (PBS) method in [26] was considered using entropy for band prioritization, and spectral information divergence (SID) for band decorrelation. The algorithm was slightly modified to accept the specification of the number of bands[3]. To select the kernel bandwidth of the unmixing method, we adopted the same procedure used for the GKKM algorithm. It is important to notice that, in general, the abundance ground truth is not available from real data. Thus, the RMSE in abundance estimation could not be used in design as a measure to select model parameters. Hence, the SK-Hype, GKKM, and PBS designs used in this comparison are based on a quasi-optimal choice of parameters for these methods, which could not be determined in practice. The proposed design for the BS methods, however, can be employed in practical applications.

BS with the CCBS and GCBS algorithms was performed using $M \in \{5, 10, 20, 30\}$, with parameters μ_0 and σ adjusted using the methodology presented in Section IV-B. We emphasize that this parameter setting strategy assumes no *prior* knowledge about the abundance ground truth.

The simulation results are summarized in Tables I and II. In these tables, the first column shows the BS strategy considered prior to unmixing. SK-Hype in this column indicates the solution without BS. The symbol "(r)" besides CCBS or GCBS means that we have randomized the order of the bands prior to applying the BS strategy. The second column shows the obtained RMSE and the standard deviation (STD) in abundance estimation. The third column lists the average CPU time elapsed in the (BS + unmixing) process. Column four shows the number of selected bands N_b, and last column shows the coherence of the final dictionary. For both tables, the selected parameters for the SK-Hype and GKKM are $\sigma_{\text{skp}} = 0.5\sigma = 0.02515$, and $\lambda = 2$ and $\sigma_{\text{kkm}} = 2\sigma = 0.1006$.

Tables I and II show the results for HIs built with Cuprite endmembers and using, respectively, the GBM and the PNMM mixing models. Note that the RMSE obtained using the BS algorithms are very close to those obtained using all bands. Nevertheless the reduction in number of bands obtained through BS is at least tenfold. The computational complexity advantage of the BS methods is evidenced by the required average CPU time, which show reductions by factors ranging from 50 to 110, depending on the algorithm and parameter settings. Note also that the number of bands in the final dictionary tends to be larger than the value M used to initialize the algorithms. This increase in the anticipated number of bands is obtained to optimize the dictionary coherence, what is not possible in the GKKM algorithm. As expected, the number of bands remained the same for the clique algorithm (CCBS) for each value of M, and the slight changes in the RMSE results indicate that the maximum clique is not unique. For the greedy approach (GCBS), however, different numbers of bands are obtained at each execution due to initial randomization, and the results in terms of RMSE and CPU time vary slightly. In general, randomization did not have any significant impact on the results. Comparing the coherence levels of the different dictionaries is not meaningful, as coherence levels are directly comparable only for dictionaries built using kernels with the same bandwidth. Nevertheless, one notes from these tables that the level of coherence of the dictionary generated using GKKM is more than 20 times larger than the limit in (30). Even considering that (30) is a sufficient condition, such discrepancy is a strong indication of the possibility of a linearly dependent basis. The PBS method provided the highest RMSE in all cases. This result was expected since the unmixing method operates in a different space. Results similar to those described above were obtained using the HIs created with data extracted from the Pavia database. These results are described in detail in an extended report available at [80].

B. Simulation with real data

When working with real data, ground truth for the fractional abundances are rarely available. Thus, we compare the performances of the different algorithms using the reconstruction RMSE (RRMSE). The image used is shown in Figure 2. It is a scene from the Cuprite mining field site in Nevada, acquired by the AVIRIS instrument. It has originally 224 spectral bands, from which we have removed the water absorption bands, resulting in 188 bands. This scene has 7371 pixels and previous analysis identified five minerals (Sphene, Montmorillonite, Kaolinite, Dumortierite, and Pyrope) to have strong components in this particular region [81]. The endmember matrix was extracted using the VCA algorithm [63]. Table III shows the RRMSEs for this image using the different methods, where we present results for values of M up to the point

[3]The PBS algorithm as proposed in [26] starts with an initial number of bands, and then adjusts this number using a time-consuming strategy that requires performing HU at each iteration. Hence, we slightly modified the PBS algorithm by fixing the number of bands *a priori* to reduce complexity and to produce a fair comparison among the different strategies.

TABLE I: RMSE. 100 runs, 2000 pxl., 8 endmembers (Cuprite), SNR=21dB, GBM. μ_0 computed using Equation (31) for a given M, and σ is found solving problem (33).

Strategy	RMSE ± STD	Av. Time	N_b	μ
SK-Hype	0.0680 ± 0.0028	301.08 ± 17.93	420	-
GKKM	0.0664 ± 0.0026	25.40 ± 0.22	36	0.5893
$M = 5$, $\mu_0 = 0.25$, $\sigma = 0.2548$				
CCBS	0.0687 ± 0.0028	3.10 ± 0.14	10	0.2482
CCBS (r)	0.0687 ± 0.0028	3.13 ± 0.12	10	0.2482
GCBS	0.0724 ± 0.0031	2.91 ± 0.02	8	0.2482
GCBS (r)	0.0721 ± 0.0030	3.15 ± 0.15	7.13	0.2331
PBS	0.0766 ± 0.0032	2.68 ± 0.20	10	0.4367
$M = 10$, $\mu_0 = 0.1111$, $\sigma = 0.1320$				
CCBS	0.0678 ± 0.0027	2.85 ± 0.13	16	0.1108
CCBS (r)	0.0679 ± 0.0027	2.89 ± 0.17	16	0.1108
GCBS	0.0685 ± 0.0028	2.57 ± 0.02	16	0.1104
GCBS (r)	0.0688 ± 0.0028	2.65 ± 0.06	13.09	0.0996
PBS	0.0765 ± 0.0033	2.54 ± 0.08	16	0.8213
$M = 20$, $\mu_0 = 0.0526$, $\sigma = 0.0965$				
CCBS	0.0659 ± 0.0026	2.96 ± 0.15	21	0.0520
CCBS (r)	0.0660 ± 0.0026	3.01 ± 0.17	21	0.0520
GCBS	0.0670 ± 0.0027	2.59 ± 0.02	20	0.0525
GCBS (r)	0.0678 ± 0.0027	2.67 ± 0.08	15.95	0.0467
PBS	0.0762 ± 0.0033	2.84 ± 0.10	21	0.8545
$M = 30$, $\mu_0 = 0.0345$, $\sigma = 0.0503$				
CCBS	0.0637 ± 0.0024	5.54 ± 0.22	42	0.0339
CCBS (r)	0.0637 ± 0.0024	5.74 ± 0.18	42	0.0339
GCBS	0.0637 ± 0.0024	3.32 ± 0.04	41	0.0344
GCBS (r)	0.0644 ± 0.0025	2.83 ± 0.07	33.39	0.0326
PBS	0.0735 ± 0.0031	4.80 ± 0.07	42	0.8999

TABLE II: RMSE. 100 runs, 2000 pxl., 8 endmembers (Cuprite), SNR=21dB, PNMM. μ_0 computed using Equation (31) for a given M, and σ is found solving problem (33).

Strategy	RMSE ± STD	Av. Time	N_b	μ
SK-Hype	0.0728 ± 0.0030	277.03 ± 4.30	420	-
GKKM	0.0729 ± 0.0030	25.52 ± 0.18	36	0.7760
$M = 5$, $\mu_0 = 0.25$, $\sigma = 0.2548$				
CCBS	0.0748 ± 0.0031	2.99 ± 0.10	10	0.2482
CCBS (r)	0.0749 ± 0.0031	3.12 ± 0.18	10	0.2482
GCBS	0.0764 ± 0.0032	2.85 ± 0.06	8	0.2482
GCBS (r)	0.0776 ± 0.0033	2.99 ± 0.15	7.13	0.2331
PBS	0.1027 ± 0.0049	3.30 ± 0.18	10	0.9081
$M = 10$, $\mu_0 = 0.1111$, $\sigma = 0.1320$				
CCBS	0.0746 ± 0.0031	2.85 ± 0.19	16	0.1108
CCBS (r)	0.0745 ± 0.0031	2.84 ± 0.14	16	0.1108
GCBS	0.0757 ± 0.0032	2.57 ± 0.04	16	0.1104
GCBS (r)	0.0757 ± 0.0031	2.64 ± 0.10	13.09	0.0996
PBS	0.0828 ± 0.0035	3.32 ± 0.09	16	0.9184
$M = 20$, $\mu_0 = 0.0526$, $\sigma = 0.0965$				
CCBS	0.0735 ± 0.0029	2.87 ± 0.12	21	0.0520
CCBS (r)	0.0737 ± 0.0029	2.96 ± 0.17	21	0.0520
GCBS	0.0753 ± 0.0031	2.55 ± 0.03	20	0.0525
GCBS (r)	0.0753 ± 0.0031	2.56 ± 0.04	15.95	0.0467
PBS	0.0836 ± 0.0035	3.77 ± 0.09	21	0.9184
$M = 30$, $\mu_0 = 0.0345$, $\sigma = 0.0503$				
CCBS	0.0740 ± 0.0029	5.41 ± 0.18	42	0.0339
CCBS (r)	0.0740 ± 0.0029	5.62 ± 0.19	42	0.0339
GCBS	0.0737 ± 0.0029	3.24 ± 0.04	41	0.0344
GCBS (r)	0.0742 ± 0.0030	2.74 ± 0.07	33.39	0.0326
PBS	0.0776 ± 0.0033	5.45 ± 0.08	42	0.9456

of diminishing improvements ($M = 150$). The RRMSE performance is compatible to that obtained using synthetic images, and the savings in computational complexity can be inferred from the CPU time reduction by a factor of at least 4 (for $M = 150$). The parameters used for the SK-Hype and GKMM were $\sigma_{\text{skp}} = 0.5\sigma = 0.0088545$, $\lambda = 2$ and $\sigma_{\text{kkm}} = 0.1006$.

TABLE III: Reconstruction RMSE for the Cuprite Scene.

Strategy	RRMSE ± STD	Time	N_b	μ
SK-Hype	0.0006 ± 0.0000	184.2852	188	-
GKKM	0.0064 ± 0.0000	17.0144	13	0.7982
$M = 5$, $\mu_0 = 0.2500$, $\sigma = 0.0916$				
CCBS	0.0155 ± 0.0002	13.5109	9	0.2454
GCBS	0.0129 ± 0.0001	15.9691	9	0.2454
$M = 30$, $\mu_0 = 0.0345$, $\sigma = 0.0174$				
CCBS	0.0101 ± 0.0001	26.1530	36	0.0336
GCBS	0.0102 ± 0.0001	26.1337	35	0.0341
$M = 50$, $\mu_0 = 0.0204$, $\sigma = 0.0113$				
CCBS	0.0099 ± 0.0001	38.1907	49	0.0199
GCBS	0.0092 ± 0.0001	19.4064	49	0.0202
$M = 70$, $\mu_0 = 0.0145$, $\sigma = 0.0087$				
CCBS	0.0089 ± 0.0001	24.5304	61	0.0141
GCBS	0.0089 ± 0.0001	26.5687	61	0.0130
$M = 120$, $\mu_0 = 0.0084$, $\sigma = 0.0059$				
CCBS	0.0079 ± 0.0000	42.4282	84	0.0084
GCBS	0.0077 ± 0.0000	62.1598	84	0.0084
$M = 150$, $\mu_0 = 0.0067$, $\sigma = 0.0051$				
CCBS	0.0074 ± 0.0000	55.4628	93	0.0067
GCBS	0.0076 ± 0.0000	56.0754	92	0.0067

A second example using real data was performed using the RELAB data from Brown University. These data include laboratory measured reflectances, and thus provide ground truth information. It consists of intimate mixtures of minerals (Anorthite, Olivine, Enstatite, and Magnetite) that were crushed and mixed together. The data is composed by the reflectances of the 4 pure minerals (endmembers) and of binary (Olivine/Enstatite, Olivine/Magnetite, and Olivine/Anorthite) and ternary (Olivine/Anorthite/Enstatite) mixtures. Each binary combination of minerals has 5 mixtures with different abundances for each endmembers (ranging form 0.1 to 0.95). The ternary mineral combinations have 7 spectra, considering also different abundances. Detailed information about this dataset can be found in [7], [82].

We performed simulations following the same procedure used for the Cuprite data. The results, however, are measured using the RMSE for the abundance estimations (not reconstruction error) since ground truth is available for this data. Table IV (Olivine/Enstatite mixture) and Table V (Olivine/Arnothite/Enstatite mixture) are representative of the results obtained, and show a good performance of the proposed BS methods. Although the full band SK-Hype presents the smallest RMSEs, the RMSEs obtained using the proposed BS methods (CCBS and GCBS) are comparable, specially for $M = 30$, indicating the possibility of a significant reduction in computational complexity. One should consider the statistical significance of these RMSE results with caution due to the small number of samples available. One also notes the considerable effect of dealing with a reduced amount of data on the CPU time, as it includes BS and unmixing. One

note, for instance, that the greedy approach GCBS needs CPU times that are 12 to 15 times smaller than the full band solution for $M = 30$. One notices, however, a large variation in CPU times for the MCP based algorithm (CCBS), especially for $M = 30$. The noticeable differences for different data sets are due to the solution of the maximal clique problem. Since MCP are NP-hard problems, the required CPU time for its solution can significantly change for different data sets. The fact that binary RELAB mixtures considered here are composed of only small numbers of mixtures (5 pixels) makes the processing time of solving a MCP more evident. This required MCP time is greatly diluted when larger data sets are considered, as could be verified in the other results presented. This indicates that the CCBS method is more advantageous for larger datasets, a common situation in hyperspectral image processing.

TABLE IV: Olivine/Enstatite

Strategy	RMSE ± STD	Time	N_b	μ
SK-Hype	0.0442 ± 0.0011	0.2499	211	-
GKKM	0.1883 ± 0.0317	1.8460	14	0.7969
$M = 5$, $\mu_0 = 0.2500$, $\sigma = 0.1034$				
CCBS	0.2045 ± 0.0375	0.1001	7	0.2490
GCBS	0.2441 ± 0.0500	0.0374	5	0.2097
$M = 10$, $\mu_0 = 0.1111$, $\sigma = 0.0524$				
CCBS	0.1430 ± 0.0192	0.0702	14	0.1079
GCBS	0.1505 ± 0.0209	0.0191	13	0.1078
$M = 20$, $\mu_0 = 0.0526$, $\sigma = 0.0273$				
CCBS	0.0907 ± 0.0069	0.2874	24	0.0517
GCBS	0.0941 ± 0.0081	0.0245	21	0.0492
$M = 30$, $\mu_0 = 0.0345$, $\sigma = 0.0182$				
CCBS	0.0705 ± 0.0035	0.6403	34	0.0339
GCBS	0.0674 ± 0.0034	0.0162	32	0.0331

TABLE V: Olivine/Arnothite/Enstatite

Strategy	RMSE ± STD	Time	N_b	μ
SK-Hype	0.1320 ± 0.0161	0.3034	211	-
GKKM	0.1325 ± 0.0150	1.8930	15	0.7694
$M = 5$, $\mu_0 = 0.2500$, $\sigma = 0.1084$				
CCBS	0.1358 ± 0.0154	0.0550	7	0.2441
GCBS	0.1585 ± 0.0229	0.0227	5	0.2212
$M = 10$, $\mu_0 = 0.1111$, $\sigma = 0.0589$				
CCBS	0.1224 ± 0.0125	0.0740	14	0.1110
GCBS	0.1323 ± 0.0158	0.0172	12	0.1076
$M = 20$, $\mu_0 = 0.0526$, $\sigma = 0.0341$				
CCBS	0.1132 ± 0.0118	0.0731	22	0.0506
GCBS	0.1179 ± 0.0131	0.0204	20	0.0491
$M = 30$, $\mu_0 = 0.0345$, $\sigma = 0.0242$				
CCBS	0.1123 ± 0.0131	0.1679	29	0.0339
GCBS	0.1166 ± 0.0143	0.0248	28	0.0323

More simulation results were obtained using hyperspectral images created with data extracted from the Pavia database and using real data from the RELAB database at Brown University. These results are described in detail in an extended report available at [80].

VIII. Conclusions

In this paper we proposed a centralized method for nonlinear unmixing of hyperspectral images, which employs band selection in in the reproducing kernel Hilbert space (RKHS).

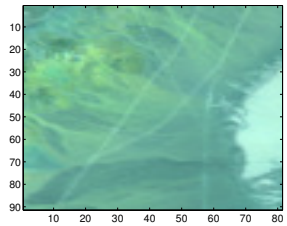

Fig. 2: Cuprite scene used in [81].

The proposed method is based on the coherence criterion, which incorporates a measure of the quality of the dictionary in the RKHS for the nonlinear unmixing. We showed that the proposed BS approach is equivalent to solving a maximum clique problem (MCP). Contrary to competing methods that do not include an efficient choice of the model parameters, the proposed method requires only an initial guess on the number of selected bands. Simulation results employing both synthetic and real data illustrate the quality of the unmixing results obtained with the proposed method, which leads to abundance estimations as accurate as those obtained using the full-band SK-Hype method, at a small fraction of the computational cost.

Acknowledgment

The authors would like to acknowledge the help of Professor John F. Mustard, from Brown University, for kindly providing us with the data from the RELAB database, and the anonymous reviewer for suggesting the use of this data.

References

[1] 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.

[2] D. Manolakis, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29–43, 2002.

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

[4] 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.

[5] 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.

[6] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 1844–1868, June 2014.

[7] J. F. Mustard and C. M. Pieters, "Photometric phase functions of common geologic minerals and applications to quantitative analysis of mineral mixture reflectance spectra," *Journal of Geophysical Research: Solid Earth*, vol. 94, no. B10, pp. 13619–13634, 1989.

[8] 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.

[9] 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, June 2012.

[10] P.-X. Li, B. Wu, and L. Zhang, "Abundance estimation from hyperspectral image based on probabilistic outputs of multi-class support vector machines," in *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05.*, July 2005, vol. 6, pp. 4315–4318.

[11] R. Heylen, D. Burazerovic, and P. Scheunders, "Non-linear spectral unmixing by geodesic simplex volume maximization," *Selected Topics in Signal Processing, IEEE Journal of*, , no. 99, pp. 1–1, 2011.

[12] R. Heylen and P. Scheunders, "A distance geometric framework for nonlinear hyperspectral unmixing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 1879–1888, June 2014.

[13] 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.

[14] X. Li, J. Cui, and L. Zhao, "Blind nonlinear hyperspectral unmixing based on constrained kernel nonnegative matrix factorization," *Signal, Image and Video Processing*, vol. 8, no. 8, pp. 1555–1567, 2012.

[15] 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, pp. 480–492, Jan 2013.

[16] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundances in hyperspectral images with ℓ_1-norm spatial regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2654–2665, May 2014.

[17] 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.

[18] R. Ammanouil, A. Ferrari, C. Richard, and S. Mathieu, "Nonlinear unmixing of hyperspectral data with vector-valued kernel functions," *Image Processing, IEEE Transactions on (to appear)*, 2017.

[19] V. N. Vapnik, *The nature of statistical learning theory*, Springer, New York, NY, 1995.

[20] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[21] 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.

[22] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, 1950.

[23] C.-I. Chang, *Hyperspectral data processing - Algorithm design and analysis*, Wiley, 2014.

[24] J. Feng, L. Jiao, F. Liu, T. Sun, and X. Zhang, "Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, high information, and low redundancy," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, pp. 2956–2969, 2015.

[25] G. Zhu, Y. Huang, J. Lei, Z. Bi, and F. Xu, "Unsupervised hyperspectral band selection by dominant set extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 227–239, Jan 2016.

[26] C.-I. Chang and K.-H. Liu, "Progressive band selection of spectral unmixing for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 4, pp. 2002–2017, 2014.

[27] K. Sun, X. Geng, and L. Ji, "Exemplar component analysis: A fast band selection method for hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 5, pp. 998–1002, May 2015.

[28] A. Zare and P. Gader, "Hyperspectral band selection and endmember detection using sparsity promoting priors," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 256–260, April 2008.

[29] Y. Yuan, G. Zhu, and Q. Wang, "Hyperspectral band selection by multitask sparsity pursuit," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 631–644, Feb 2015.

[30] M. Tian, J. Feng, B. Rivard, and C. Zhao, "A method to compute the n-dimensional solid spectral angle between vectors and its use for band selection in hyperspectral data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 50, pp. 141 – 149, 2016.

[31] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1279–1289, June 2016.

[32] Y. Yuan, J. Lin, and Q. Wang, "Dual-clustering-based hyperspectral band selection by contextual analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1431–1445, March 2016.

[33] J. M. Bioucas-Dias and A. Plaza, "An overview on hyperspectral unmixing: Geometrical, statistical, and sparse regression based approaches," *2011 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1135–1138, 2011.

[34] J. E. Ball, L. M. Bruce, and N. H. Younan, "Hyperspectral pixel unmixing via spectral band selection and dc-insensitive singular value decomposition," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 3, pp. 382–386, July 2007.

[35] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1575–1585, 2006.

[36] 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, March 2004.

[37] T. Imbiriba, J. C. M. Bermudez, C. Richard, and J.-Y. Tourneret, "Band selection in RKHS for fast nonlinear unmixing of hyperspectral images," in *2015 23rd European Signal Processing Conference (EUSIPCO)*, Aug 2015, pp. 1651–1655.

[38] 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.

[39] B. Hapke, "Bidirectional reflectance spectroscopy, 1, Theory," *Journal of Geophysical Research*, vol. 86, no. B4, pp. 3039–3054, 1981.

[40] 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, Nov. 2011.

[41] 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.

[42] 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.

[43] Y. Altmann, A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Supervised nonlinear spectral unmixing using a polynomial post nonlinear model for hyperspectral imagery," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 1009–1012.

[44] 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.

[45] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," in *SPIE Europe Remote Sensing*. International Society for Optics and Photonics, 2009, pp. 74770I–74770I.

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

[47] 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, Jan 1994.

[48] 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.

[49] 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.

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

[51] G. Kimeldorf and G. Wahba, "Some results on Tchebycheffian spline functions," *Journal of Mathematical Analysis and Applications*, vol. 33, pp. 82–95, 1971.

[52] E. Kreyszig, *Introductory functional analysis with applications*, vol. 81, wiley New York, 1989.

[53] Bernhard Schölkopf, Ralf Herbrich, and Alex J. Smola, *A Generalized Representer Theorem*, pp. 416–426, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[54] 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.

[55] E. H. Moore, "On properly positive hermitian matrices," *Bull. American Mathematical Society*, vol. 23, pp. 59, 1916.

[56] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, The MIT Press, 2001.

[57] H. Kadri, E. Duflos, M. Davy, P. Preux, and S. Canu, "General framework for nonlinear functional regression with reproducing kernel hilbert spaces," Research Report RR-6908, INRIA, 2009.

[58] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 564–568, 2008.

[59] P. Estévez, M. Tesmer, C. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Transactions on Neural Networks*, vol. 20, no. 2, pp. 189–201, 2009.

[60] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4158–4171, 2007.

[61] J. Feng, L. C. Jiao, X. Zhang, and T. Sun, "Hyperspectral band selection based on trivariate mutual information and clonal selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 7, pp. 4092–4105, 2014.

[62] 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.

[63] 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.

[64] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[65] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[66] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[67] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2558–2567, 2002.

[68] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[69] P. M. Pardalos and J. Xue, "The maximum clique problem," *Journal of global Optimization*, vol. 4, no. 3, pp. 301–328, 1994.

[70] R. M. Karp, *Reducibility among combinatorial problems*, Springer, 1972.

[71] Q. Wu and J.-K. Hao, "A review on algorithms for maximum clique problems," *European Journal of Operational Research*, vol. 242, no. 3, pp. 693–709, 2015.

[72] R. Carraghan and P. M. Pardalos, "An exact algorithm for the maximum clique problem," *Operations Research Letters*, vol. 9, no. 6, pp. 375–382, 1990.

[73] C.-M. Li, Z. Fang, and K. Xu, "Combining maxsat reasoning and incremental upper bound for the maximum clique problem," in *2013 IEEE 25th International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2013, pp. 939–946.

[74] P. R. J. Östergård, "A fast algorithm for the maximum clique problem," *Discrete Applied Mathematics*, vol. 120, no. 1, pp. 197–207, 2002.

[75] N. Biggs, "Some heuristics for graph coloring," *Graph Colourings, Longman, New York*, pp. 87–96, 1990.

[76] T. Fahle, "Simple and fast: Improving a branch-and-bound algorithm for maximum clique," in *AlgorithmsESA 2002*, pp. 485–498. Springer, 2002.

[77] C.-M. Li and Z. Quan, "An efficient branch-and-bound algorithm based on maxsat for the maximum clique problem.," in *AAAI*, 2010, vol. 10, pp. 128–133.

[78] E. Maslov, M. Batsyn, and P. M. Pardalos, "Speeding up branch and bound algorithms for solving the maximum clique problem," *Journal of Global Optimization*, vol. 59, no. 1, pp. 1–21, 2014.

[79] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*, vol. 1, Cambridge university press, Cambridge, 2008.

[80] T. Imbiriba, J. C. M. Bermudez, and C. Richard, "Technical report: Band selection for nonlinear unmixing of hyperspectral images as a maximal clique problem," *ArXiv e-prints, arXiv:1603.00437v2*, Sept. 2016.

[81] T. Imbiriba, J. C. M. Bermudez, C. Richard, and J.-Y. Tourneret, "Nonparametric detection of nonlinearly mixed pixels and endmember estimation in hyperspectral images," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1136–1151, March 2016.

[82] J. F. Mustard and C. M. Pieters, "Quantitative abundance estimates from bidirectional reflectance measurements," *Journal of Geophysical Research: Solid Earth*, vol. 92, no. B4, pp. E617–E626, 1987.

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, and his Doctorade degree from the Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, in 2016. He is currently serving as a postdoctoral researcher at the Digital Signal Processing Laboratory (LPDS) at UFSC. His research interests include audio and image processing, pattern recognition, 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. In the winter of 1992, he was a Visiting Researcher with the Department of Electrical Engineering, Concordia University. In 1994, he was a Visiting Researcher with the Department of Electrical Engineering and Computer Science, University of California, Irvine (UCI). He was a Visiting Researcher with Institut National Polytechnique de Toulouse, France, in the winter 2007, spring 2009 and spring 2011. He was a Visiting Researcher with Universit Nice Sophia-Antipolis, France, in spring 2010 and winter 2014. In 2012 he spent a sabbatical year at Institut National Polytechnique de Toulouse, France. 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 in the area of adaptive filtering from 1994 to 1996 and from 1999 to 2001. He also served as an Associate Editor of the EURASIP Journal of Advances on Signal Processing from 2006 to 2010. He is a member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society. He is an Associated Editor of the GRETSI journal Traitement du Signal.

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.