# Spatial regularization for nonlinear unmixing of hyperspectral data with vector-valued kernel functions

Rita Ammanouil, André Ferrari, Cédric Richard, Jean-Yves Tourneret

# SPATIAL REGULARIZATION FOR NONLINEAR UNMIXING OF HYPERSPECTRAL DATA WITH VECTOR-VALUED KERNEL FUNCTIONS

*Rita Ammanouil\*, André Ferrari\*, Cédric Richard\*, Jean-Yves Tourneret[†]*

\* Lagrange laboratory, University of the Côte d'Azur, France
[†] IRIT laboratory, University of Toulouse, France
{rita.ammanouil, andre.ferrari, cedric.richard}@unice.fr, jean-yves.tourneret@enseeiht.fr

## ABSTRACT

This communication introduces a new framework for incorporating spatial regularization into a nonlinear unmixing procedure dedicated to hyperspectral data. The proposed model promotes smooth spatial variations of the nonlinear component in the mixing model. The spatial regularizer and the nonlinear contributions are jointly modeled by a vector-valued function that lies in a reproducing kernel Hilbert space (RKHS). The unmixing problem is strictly convex and reduces to a quadratic programming (QP) problem. Simulations on synthetic data illustrate the effectiveness of the proposed approach.

***Index Terms—*** Hyperspectral images, nonlinear unmixing, spatial regularization, vector-valued RKHS.

## 1. INTRODUCTION

Spectral unmixing (SU) is one of the most prominent tools for analyzing hyperspectral data [1]. This is partly due to the fact that hyperspectral sensors have a low spatial resolution. Hence, the area covered by a pixel usually encloses more than one material. The goal of SU is to identify the pure spectra (endmembers) in the scene, and to estimate their proportions (abundances) in each pixel.

Most of the SU algorithms proposed in the literature consider the linear mixing model (LMM), which represents the spectrum of a pixel as a linear combination of the endmembers weighted by their abundances [2]. The LMM is a simple yet very representative model, which was extensively studied in the literature [3]. However, some scenes such as vegetated and urban areas exhibit strong nonlinear effects [4, 5], which can be detected with statistical hypothesis testing [6]. This has led to the refinement of the LMM through the development of nonlinear mixing models [7]. Most of these models incorporate a new term in addition to the prevailing linear term. For example, bilinear models [8–12] incorporate the contributions of the pairwise products of the endmembers. The authors of [13] present a multilinear mixing model as an extension of bilinear models, that incorporates higher order products between the endmembers. In contrast with the previously cited models, the authors of [14] propose a non-parametric nonlinear model that lies in an RKHS [15]. This model does not require a specific expression for the nonlinear term but is defined in a rather general manner through a kernel function. Another strategy for nonlinear unmixing is proposed in [16], where the authors use a Gaussian process latent variable model (GPLVM).

Similarly to [14], the GPLVM framework is able to incorporate many forms of nonlinearities.

In this paper, we introduce a new spatial regularizer acting on the nonlinear contributions. We adopt the same strategy as in [14], i.e., we resort to a kernel-based nonlinear mixing model. Compared to [14], we go one step further by using an RKHS of vector-valued functions rather than scalar-valued functions [17, 18]. More precisely, each output of the vector-valued function represents the nonlinear contribution of the mixing model at a given pixel. We consider a special class of kernels, known as separable kernels [18]. In particular, these kernels are defined as the product of two terms, a scalar-valued kernel acting on the input, and a matrix-valued kernel encoding the closeness between the outputs. The first and the second term play a central role in defining the nonlinearity and the spatial regularization, respectively. The closeness between the outputs of the function, i.e. the nonlinear contributions at different pixels, is defined using a graph, see [19–21] and references therein for application of graph-based regularizations in image processing. In particular, the proposed spatial regularization consists of penalizing the $\ell_2$-norm of the difference between the outputs that appear as connected in the graph and hence promotes smoothness. Being solely defined by an appropriate design of the kernel, the spatial regularization is relatively transparent from the optimization problem point of view, which is then shown to reduce to a quadratic problem. To the best of our knowledge, there is no nonlinear model in the literature that promotes smooth nonlinear contributions. Nevertheless, several works considered other types of prior information for nonlinear unmixing. The authors of [22] incorporated a total variation over the abundances, which promotes piecewise smooth abundances in the scene. The authors of [23] introduced a robust nonlinear matrix factorization unmixing algorithm that promotes sparse nonlinear contributions. In contrast with the previously cited models, we promote smooth nonlinear contributions over the scene. This prior is justified by the spatial smoothness inherently present in natural scenes.

This paper is organized as follows. Section 2 introduces the nonlinear mixing model. Section 3 details the kernel design and the underlying spatial regularization. Section 4 presents the proposed unmixing algorithm. Finally, experimental results investigated in section 5 show the effectiveness of the proposed approach.

## 2. VECTOR-VALUED FORMULATION

Consider an hyperspectral image with $N$ pixels, estimated over $L$ spectral bands. According to the LMM [2], the spectrum of the $n$-th pixel is modeled as:

$$s_n = \sum_{i=1}^{M} a_{i,n} r_i + e_n, \ \forall n = 1, \dots N, \quad (1)$$

where $\boldsymbol{s}_n = [s_{1,n}, \ldots, s_{L,n}]^\top$ is the $L$-dimensional spectrum of the $n$-th pixel, $M$ is the number of endmembers, $a_{i,n}$ is the abundance of the $i$-th endmember in the $n$-th pixel, $\boldsymbol{r}_i$ is the $L$-dimensional spectrum of the $i$-th endmember, and $\boldsymbol{e}_n$ is a vector of white Gaussian noise. All vectors are column vectors. The abundances, being the relative contributions of the endmembers, are positive and usually sum to one [2], namely: $a_{i,n} \geq 0$ and $\sum_{i=1}^{M} a_{i,n} = 1$. As mentioned previously, most nonlinear mixing models incorporate an additional term within the LMM (1). In this work, we consider the nonlinear mixing model, known as khype, that was proposed in [14]:

$$s_{\ell,n} = \boldsymbol{r}_{\lambda_\ell}^\top \boldsymbol{a}_n + f_n(\boldsymbol{r}_{\lambda_\ell}) + e_{\ell,n}, \tag{2}$$

where $\boldsymbol{R} = [\boldsymbol{r}_1, \ldots, \boldsymbol{r}_M]$ is the $L \times M$ matrix of endmembers, $\boldsymbol{r}_{\lambda_\ell}$ is an $M \times 1$ vector formed with the elements of the $\ell$-th row of $\boldsymbol{R}$, $\boldsymbol{a}_n = [a_{1,n}, \ldots, a_{M,n}]^\top$ is the abundance vector of the $n$-th pixel, and $f_n$ is a scalar-valued function in an RKHS modelling the nonlinearity at any band. Let $\boldsymbol{f} = [f_1, \ldots, f_N]^\top$ be a vector-valued function. Equation (2) can be rewritten as follows:

$$s_{\lambda_\ell} = \boldsymbol{A}^\top \boldsymbol{r}_{\lambda_\ell} + \boldsymbol{f}(\boldsymbol{r}_{\lambda_\ell}) + \boldsymbol{e}_{\lambda_\ell}, \tag{3}$$

where $\boldsymbol{s}_{\lambda_\ell}$ and $\boldsymbol{e}_{\lambda_\ell}$ denote the $\ell$-th rows of $\boldsymbol{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N]$ and $\boldsymbol{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N]$ respectively. The aim of the next section is to show the relevance of the vector-valued formulation (3). In particular, we demonstrate the ability of vector-valued functions to incorporate prior information about the similarities between $f_1, \ldots, f_N$, the outputs of $\boldsymbol{f}$, through an appropriate kernel design.

## 3. KERNEL DESIGN AND REGULARIZATION

We will assume that the nonlinear function $\boldsymbol{f}$ in (3) lies in an RKHS of vector-valued functions, denoted by $\bar{\mathcal{H}}_{\boldsymbol{k}}$, associated with the following separable kernel function [17, 18]:

$$\begin{aligned} \bar{\boldsymbol{k}} : \quad \mathbb{R}^M \times \mathbb{R}^M \quad &\rightarrow \quad \mathbb{R}^{N \times N} \\ (\boldsymbol{r}_{\lambda_\ell}, \boldsymbol{r}_{\lambda_{\ell'}}) \quad &\rightarrow \quad \bar{\boldsymbol{k}}(\boldsymbol{r}_{\lambda_\ell}, \boldsymbol{r}_{\lambda_{\ell'}}), \end{aligned} \tag{4}$$

with

$$\bar{\boldsymbol{k}}(\boldsymbol{r}_{\lambda_\ell}, \boldsymbol{r}_{\lambda_{\ell'}}) = k(\boldsymbol{r}_{\lambda_\ell}, \boldsymbol{r}_{\lambda_{\ell'}}) \, \boldsymbol{\mathcal{E}}. \tag{5}$$

The function $k(\cdot, \cdot)$ is a scalar-valued kernel such as the polynomial or Gaussian kernel, and $\boldsymbol{\mathcal{E}}$ is an $N \times N$ symmetric nonnegative matrix. Let $\boldsymbol{K}$ be the $L \times L$ Gram matrix associated with the scalar-valued kernel $k$, namely, $k_{\ell,\ell'} = k(\boldsymbol{r}_{\lambda_\ell}, \boldsymbol{r}_{\lambda_{\ell'}})$, and let $\bar{\boldsymbol{K}}$ be the $NL \times NL$ Gram matrix associated with the matrix-valued kernel $\bar{\boldsymbol{k}}$, namely, $\bar{\boldsymbol{k}}_{\ell,\ell'} = \bar{\boldsymbol{k}}(\boldsymbol{r}_{\lambda_\ell}, \boldsymbol{r}_{\lambda_{\ell'}})$. Given equation (5), we have:

$$\bar{\boldsymbol{K}} = \boldsymbol{K} \otimes \boldsymbol{\mathcal{E}}. \tag{6}$$

Moreover, the norm of $\boldsymbol{f}$ in $\bar{\mathcal{H}}_{\boldsymbol{k}}$ [17] is given by:

$$\|\boldsymbol{f}\|_{\bar{\mathcal{H}}_{\boldsymbol{k}}}^2 = \sum_{n,n'=1}^{N} \mathcal{E}_{n,n'}^\dagger \langle f_n, f_{n'} \rangle_{\mathcal{H}_k}, \tag{7}$$

where $\boldsymbol{\mathcal{E}}^\dagger$ is the pseudo inverse of $\boldsymbol{\mathcal{E}}$. The above expression shows that the norm of $\boldsymbol{f}$ is equal to the weighted sum of the pairwise inner products between the individual functions. From a regularization point of view, equation (7) can be used to promote structured similarities between the different functions through the design of $\boldsymbol{\mathcal{E}}$. Hereafter, we investigate the so-called "graph regularizer" [17] and provide the corresponding structure for the matrix $\boldsymbol{\mathcal{E}}$. Note that the

authors of [17] provided other examples of regularizers with the corresponding design of $\boldsymbol{\mathcal{E}}$.

Due to the inherent spatial correlation present in real images, spatially neighboring pixels usually have similar spectra. As a consequence, we assume that they are characterized by similar nonlinear contributions. This prior about the closeness between adjacent pixels can be modeled by a graph. We denote by $\boldsymbol{W} \in \mathbb{R}^{N \times N}$ the adjacency matrix of this graph [24]. When two pixels are adjacent, the corresponding nodes are connected by an edge and associated with a positive similarity weight $w_{n,n'} > 0$, otherwise $w_{n,n'}$ is set to zero. In accordance with the prior, the graph regularizer promotes similarity between the estimated nonlinearities at adjacent pixels in the image, hence connected nodes in the graph. It is defined as:

$$\|\boldsymbol{f}\|_{\mathcal{H}_k}^2 = \sum_{n=1}^{N} \|f_n\|_{\mathcal{H}_k}^2 w_{n,n} + \frac{1}{2} \sum_{n=1}^{N} \sum_{n'=1}^{N} \|f_n - f_{n'}\|_{\mathcal{H}_k}^2 w_{n,n'}. \tag{8}$$

Note that, (8) penalizes the norms of the individual functions in addition to the differences between each pair of functions, hence forcing them to be similar. Moreover, the strength of the similarity between each pair of functions is determined by the corresponding weight. More precisely, a high value of $w_{n,n'}$ promotes a strong similarity between $f_n$ and $f_{n'}$, and conversely, a low value of $w_{n,n'}$ promotes a weak similarity between the two functions. Using (7) and (8), some calculations show that $\boldsymbol{\mathcal{E}}^\dagger$ is related to $\boldsymbol{W}$ as follows:

$$\begin{cases} \mathcal{E}_{n,n'}^\dagger &= -w_{n,n'}, \text{ if } n \neq n', \\ \mathcal{E}_{n,n}^\dagger &= \sum_{n'=1}^{N} w_{n,n'}, \text{ otherwise.} \end{cases} \tag{9}$$

Finally, note that when $\boldsymbol{\mathcal{E}} = \boldsymbol{I}_N$, the norm of $\boldsymbol{f}$ reduces to the sum of the individual norms of its components $f_n$. This corresponds to processing all the functions independently without exploiting any regularization between them as in the khype model [14].

## 4. ESTIMATION ALGORITHM

In order to estimate the abundances and the nonlinear function, we propose to consider the following optimization problem:

$$\begin{aligned} \min_{\{\boldsymbol{e}_{\lambda_\ell}\}_{\ell=1}^{L}, \boldsymbol{f} \in \mathcal{H}_k, \boldsymbol{A}} \quad & \tfrac{1}{2} \sum_{\ell=1}^{L} \|\boldsymbol{e}_{\lambda_\ell}\|^2 + \tfrac{\lambda}{2} \|\boldsymbol{f}\|_{\bar{\mathcal{H}}_{\boldsymbol{k}}}^2 + \tfrac{\mu}{2} \|\boldsymbol{A}\|_{\mathrm{F}}^2 \\ \text{subject to} \quad & \boldsymbol{e}_{\lambda_\ell} = \boldsymbol{s}_{\lambda_\ell} - \boldsymbol{A}^\top \boldsymbol{r}_{\lambda_\ell} - \boldsymbol{f}(\boldsymbol{r}_{\lambda_\ell}) \\ & a_{i,n} \succeq 0 \quad \forall i = 1, \cdots M, n = 1, \cdots N, \\ & \sum_{i=1}^{M} a_{i,n} = 1 \quad \forall n = 1, \cdots, N, \end{aligned} \tag{10}$$

where $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$, and $\lambda$, $\mu$ are tuning parameters. The first term in the objective function (10) measures the square error between the observations and the estimated model. The second term in the objective function (10) is the $\ell_2$-norm of $\boldsymbol{f}$ in $\bar{\mathcal{H}}_{\boldsymbol{k}}$. This term incorporates the norms of its individual outputs in addition to their weighted differences (8). As a result, it constrains the regularity of the estimated functions and their pairwise differences depending on the kernel design. The third term in (10) is the Frobenius norm of $\boldsymbol{A}$ which constrains the norm of the estimated abundances. The relevance of having simultaneously two strictly convex regularizers is that it ensures the strict convexity of the objective function. The Lagrangian associated with problem (10) is:

$$\begin{aligned} \boldsymbol{L}(\boldsymbol{E}, \boldsymbol{f}, \boldsymbol{A}, \boldsymbol{V}, \boldsymbol{\Lambda}, \boldsymbol{u}) = \tfrac{1}{2} \sum_{\ell=1}^{L} \|\boldsymbol{e}_{\lambda_\ell}\|^2 + \tfrac{\lambda}{2} \|\boldsymbol{f}\|_{\bar{\mathcal{H}}_{\boldsymbol{k}}}^2 + \\ \tfrac{\mu}{2} \|\boldsymbol{A}\|_{\mathrm{F}}^2 + \sum_{\ell=1}^{L} \boldsymbol{v}_{\lambda_\ell}^\top (\boldsymbol{s}_{\lambda_\ell} - \boldsymbol{A}^\top \boldsymbol{r}_{\lambda_\ell} - \boldsymbol{f}(\boldsymbol{r}_{\lambda_\ell}) - \boldsymbol{e}_{\lambda_\ell}) \\ - \operatorname{tr}(\boldsymbol{\Lambda}^\top \boldsymbol{A}) - \boldsymbol{u}^\top (\boldsymbol{A}^\top \boldsymbol{1}_M - \boldsymbol{1}_N) \end{aligned} \tag{11}$$
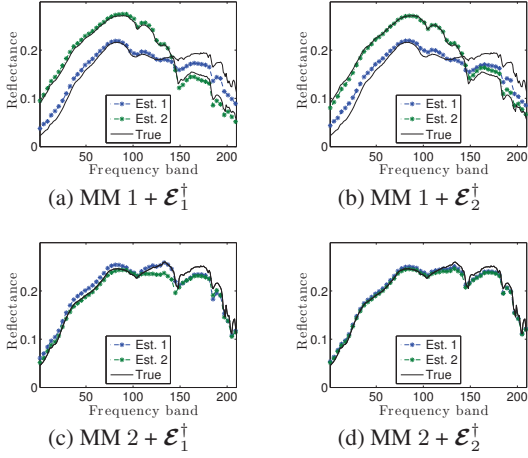
**Fig. 1**. True and estimated nonlinear parts in $s_1$ and $s_2$ obtained using the various settings for the bilinear coefficients and for $\mathcal{E}$.

where $V = [v_{\lambda_1}, \ldots, v_{\lambda_L}]^\top$, $\Lambda$, and $u$ are the Lagrange multipliers associated with the constraints in (10). Given that problem (10) is strictly convex, its solution can be found by solving the Lagrange dual problem [25]. Setting the derivatives of the Lagrangian w.r.t. the primal variables to zero yields:

$$\begin{cases} E^\star & = V \\ f^\star(\cdot) & = \sum_{\ell=1}^{L} k(\cdot, r_{\lambda_\ell}) \frac{v_{\lambda_\ell}}{\lambda} \\ A^\star & = \frac{1}{\mu}(R^\top V + \Lambda + \mathbf{1}_M u^\top) \end{cases} \quad (12)$$

Replacing the optimal variables in (11) by their expressions in (12), gives the Lagrangian dual function. Some calculations show that the Lagrange dual function can be written as a quadratic form (see equation (13)). The vectors $\widehat{v}$, $\widehat{\Lambda}$ in (13) are shorthand notations for $\text{vec}(V^\top)$ and $\text{vec}(\Lambda^\top)$, where $\text{vec}(\cdot)$ is an operator that stacks the columns of its input matrix on top of each other. The Lagrange dual problem consists of maximizing the Lagrange dual function (13), with the additional constraint $\Lambda \geq 0$ where the inequality is applied element wise. In other words, the dual problem reduces to solving a positively constrained quadratic problem. In the experiments, problem (13) is solved using a quadratic solver. When $\mathcal{E} = I_N$, i.e. no proximity between the functions is assumed, problem (13) separable and reduces to $N$ quadratic subproblems.

## 5. EXPERIMENTS ON SYNTHETIC DATA

### 5.1. Toy example

The proposed approach is first illustrated using a toy example. Eight endmembers were randomly selected from the ENVI software library. The endmembers spectra consist of $L = 210$ spectral bands uniformly sampled in the range from 395 to 2560 nm. Two nonlinearly mixed spectra, namely, $s_1$ and $s_2$, were generated such that:

$$\begin{cases} s_1 = & Ra_1 + u\left(\alpha_{1,1} Ra_1 \odot Ra_1 + \alpha_{1,2} Ra_2 \odot Ra_2\right) + e_1, \\ s_2 = & Ra_2 + u\left(\alpha_{2,1} Ra_1 \odot Ra_1 + \alpha_{2,2} Ra_2 \odot Ra_2\right) + e_2, \end{cases} \quad (15)$$

where "$\odot$" is the element wise product between two vectors, $u$ is an attenuation coefficient set to 0.5 in the experiments, and $\alpha_{i,j} \in [0\ 1]$ is the contribution of the bilinear term depending on $a_j$ in $s_i$. Note that the second term on the right hand side of the first equation in (15)

corresponds to the nonlinear contribution $[f_1(r_{\lambda_1}) \ldots f_1(r_{\lambda_L})]^\top$, and similarly for the second equation. Two cases are considered:

- MM 1: $\alpha_{1,1} = \alpha_{2,2} = 1$ and $\alpha_{1,2} = \alpha_{2,1} = 0$
- MM 2: $\alpha_{1,1} = \alpha_{2,2} = 0.5$ and $\alpha_{1,2} = \alpha_{2,1} = 0.5$

MM1 corresponds to the well-known polynomial post nonlinear mixing model (PPNM) [11]. MM2 corresponds to a bilinear model where the bilinear contributions simultaneously depend on both abundance vectors. In particular, setting all bilinear coefficients to 0.5 yields the same nonlinear contribution for $s_1$ and $s_2$. Finally, the two following cases were considered for the matrix $\mathcal{E}^\dagger$:

- $\mathcal{E}_1^\dagger$: $w_{1,1} = w_{2,2} = 1, w_{1,2} = w_{2,1} = 0$,
- $\mathcal{E}_2^\dagger$: $w_{1,1} = w_{2,2} = 1, w_{1,2} = w_{2,1} = 10$.

The first case ($\mathcal{E}^\dagger = \mathcal{E}_1^\dagger$) does not promote any a priori similarity. The resulting norm of $f$ given by (8) reduces in this case to the sum of the norms of the individual functions. The second case ($\mathcal{E}^\dagger = \mathcal{E}_2^\dagger$) promotes similarity between $f_1$ and $f_2$. In addition to the sum of norms of the individual functions $f_1$ and $f_2$, the norm of $f$ incorporates the difference between $f_1$ and $f_2$ weighted by $w_{2,1}$. As for the scalar kernel, a second order polynomial kernel is used:

$$k(r_{\lambda_\ell}, r_{\lambda_{\ell'}}) = (r_{\lambda_\ell}^\top r_{\lambda_{\ell'}})^2. \quad (16)$$

The feature map of the kernel as defined by (16) incorporates the pairwise products between the endmembers, which motivates its use with the bilinear model. Gaussian noise was added in order to reach the desired signal to noise ratio (SNR). For each case, 100 Monte Carlo runs were performed. The performance of the proposed approach is evaluated using the root mean square error (RMSE):

$$\text{RMSE}(A, A^\star) = \sqrt{\frac{\|A - A^\star\|_F^2}{MN}}. \quad (17)$$

The parameters $\lambda$ and $\mu$ were tested among the values $[10^{-3}, 5 \times 10^{-3}, 10^{-2}, 10^{-1}, 1, 10]$. The endmembers are assumed to be known in the experiments. Table 1 reports the RMSEs of the estimated abundances and the nonlinear parts obtained in each scenario with several values of the SNR and M. Table 2 reports the optimal tuning parameters for each case. As expected, penalizing the discrepancy between $f_1$ and $f_2$, that is, using $\mathcal{E}_2^\dagger$, gives better results when the nonlinear parts are equal. On the other hand, using $\mathcal{E}_2^\dagger$ for different functions $f_1$ and $f_2$ deteriorates the results. More importantly, improving the estimation of the nonlinear part also yields an improved estimation of the abundances. This shows the importance of the estimation of $f$, and the use of a correct prior. Figure 1 shows the true and estimated functions $f_1$ and $f_2$ for each case. Figures 1 (a) and (b) correspond to the case where $f_1$ and $f_2$ are different, they are represented by two solid lines. Whereas Figures 1 (c) and (d) correspond to the case where $f_1$ and $f_2$ are equal, hence they are both represented by one solid line. The estimated nonlinear contributions are represented using blue and green dashed lines respectively. Figures 1 (b) and (d) show how penalizing the difference between $f_1$ and $f_2$ yields closer estimations compared to Figures 1 (a) and (c).

### 5.2. Spatial data set

The proposed approach was tested on a synthetic image known as the spatial image [22]. The abundance maps for this image are the same as those used for the image IM2 in [22]. The image has $100 \times 100$ pixels, and is composed of 8 endmembers. As in the previous experiment, a bilinear mixing model was used where the bilinear coefficients depend on neighboring abundances. The attenuation parameter was set to $u = 0.5$. The endmembers spectra used in the

$$G(\widehat{\boldsymbol{v}}, \widehat{\boldsymbol{\Lambda}}, \boldsymbol{u}) = -\frac{1}{2} \left( \frac{\widehat{\boldsymbol{v}}}{\widehat{\boldsymbol{\Lambda}}} \right)^{\top} \left( \begin{array}{c|c|c} \boldsymbol{K}_{\widehat{\boldsymbol{v}}} & \frac{1}{\mu}\boldsymbol{R} \otimes \boldsymbol{I}_N & \frac{1}{\mu}(\boldsymbol{R}\boldsymbol{1}_M) \otimes \boldsymbol{I}_N \\ \hline \frac{1}{\mu}\boldsymbol{R}^{\top} \otimes \boldsymbol{I}_N & \frac{1}{\mu}\boldsymbol{I}_{MN} & \frac{1}{\mu}\boldsymbol{1}_M \otimes \boldsymbol{I}_N \\ \hline \frac{1}{\mu}(\boldsymbol{R}\boldsymbol{1}_M)^{\top} \otimes \boldsymbol{I}_N & \frac{1}{\mu}\boldsymbol{1}_M^{\top} \otimes \boldsymbol{I}_N & \frac{M}{\mu}\boldsymbol{I}_N \end{array} \right) \left( \frac{\widehat{\boldsymbol{v}}}{\widehat{\boldsymbol{\Lambda}}} \right)^{\top} + \left( \frac{\widehat{\boldsymbol{v}}}{\widehat{\boldsymbol{\Lambda}}} \right)^{\top} \left( \frac{\widehat{\boldsymbol{s}}}{\boldsymbol{0}_{MN}} \right), \quad (13)$$

where

$$\boldsymbol{K}_{\widehat{\boldsymbol{v}}} = (\boldsymbol{I}_{LN} + \frac{1}{\lambda}\bar{\boldsymbol{K}} + \frac{1}{\mu}(\boldsymbol{R}\boldsymbol{R}^{\top}) \otimes \boldsymbol{I}_N). \quad (14)$$

**Table 1**. RMSEs $(\times 10^{-2})$ for the abundances and the nonlinear part (left and right term in brackets respectively) obtained with the toy example.

|  |  | $\mathcal{E}_1^{\dagger}$ | | |
|---|---|---|---|---|
|  |  | $SNR = 40$ | $SNR = 30$ | $SNR = 20$ |
| M=3 | MM 1 | (1.28, 0.78) | (2.07, 1.05) | (4.55, 1.97) |
|  | MM 2 | (2.16, 0.78) | (2.54, 1.09) | (5.33, 2.17) |
| M=5 | MM 1 | (3.07, 1.46) | (2.60, 1.74) | (5.30, 3.34) |
|  | MM 2 | (2.32, 0.91) | (2.84, 1.25) | (4.72, 1.93) |
| M=8 | MM 1 | (2.12, 1.51) | (2.73, 1.81) | (5.59, 3.34) |
|  | MM 2 | (2.09, 0.80) | (3.21, 1.56) | (6.82, 3.43) |
|  |  | $\mathcal{E}_2^{\dagger}$ | | |
|  |  | $SNR = 40$ | $SNR = 30$ | $SNR = 20$ |
| M=3 | MM 1 | (2.36, 0.85) | (4.24, 1.97) | (5.42, 2.40) |
|  | MM 2 | (1.12, 0.63) | (1.68, 0.81) | (4.27, 1.65) |
| M=5 | MM 1 | (3.07, 1.46) | (2.60, 1.74) | (5.30, 3.34) |
|  | MM 2 | (1.53, 0.67) | (2.10, 1.00) | (4.48, 1.73) |
| M=8 | MM 1 | (4.74, 2.33) | (3.67, 2.11) | (5.97, 3.10) |
|  | MM 2 | (1.37, 0.77) | (2.86, 1.21) | (5.99, 2.85) |

**Table 2**. Optimal parameters $(\lambda, \mu)$ used with each algorithm in the toy example.

|  |  | $\mathcal{E}_1^{\dagger}$ | | |
|---|---|---|---|---|
|  |  | $SNR = 40$ | $SNR = 30$ | $SNR = 20$ |
| M=3 | MM 1 | (1, 0.1) | (1, 0.1) | (1, 0.1) |
|  | MM 2 | (0.1, 0.01) | (1, 0.1) | (1, 0.1) |
| M=5 | MM 1 | (1, 0.1) | (1, 0.1) | (10, 1) |
|  | MM 2 | (0.1, 0.01) | (1, 0.1) | (10, 1) |
| M=8 | MM 1 | (1, 0.1) | (1, 0.1) | (1, 0.1) |
|  | MM 2 | (1, 0.01) | (1, 0.01) | (1, 0.1) |
|  |  | $\mathcal{E}_2^{\dagger}$ | | |
|  |  | $SNR = 40$ | $SNR = 30$ | $SNR = 20$ |
| M=3 | MM 1 | (0.005, 0.005) | (1, 0.1) | (1, 0.1) |
|  | MM 2 | (1, 0.1) | (1, 0.1) | (1, 0.1) |
| M=5 | MM 1 | (0.01, 0.001) | (1, 0.1) | (10, 1) |
|  | MM 2 | (0.1, 0.01) | (1, 0.1) | (10, 1) |
| M=8 | MM 1 | (0.1, 0.01) | (1, 0.1) | (1, 0.1) |
|  | MM 2 | (0.1, 0.01) | (1, 0.01) | (1, 0.1) |

previous experiment were used in this experiment. A white Gaussian noise was added to the observations in order to get an SNR of 30 dB. The image was unmixed using three methods. The first method is the extended endmember matrix method (ExtM) [26]. It consists of extending the endmember matrix artificially with cross-spectra of pure materials. The second method is khype [14]. It was obtained by simply setting $\mathcal{E}$ to the identity matrix in our algorithm. The third method is the proposed approach used with $\mathcal{E} \neq \boldsymbol{I}$, i.e., with prior information regarding the similarity between the nonlinearities. For the latter method, the image was decomposed into $3 \times 3$ patches in order to reduce the computational complexity. In each patch, nonlinear parts at adjacent pixels are assumed to be similar. The similarity weights were tested among the values $[10, 50, 100]$. After preliminary tests, they were set to 50 in all experiments. Table 3 reports the RMSEs of the abundances and the nonlinear contributions for the three methods. The best scores in terms of the RMSEs are obtained with the proposed approach. Figure 2 shows the true and estimated nonlinear contributions at band #100 obtained with each method. Figure 2 (d) shows that incorporating the spatial prior resulted in visually smoother variations.

## 6. CONCLUSION

This communication proposed a new framework for incorporating spatial regularization in nonlinear unmixing. The proposed model promotes smooth spatial variations of the nonlinear components in the mixing model. Vector-valued RKHS make this algorithm computationally efficient as it reduces to a QP problem. The performance of the proposed approach was validated on synthetic data.

**Table 3**. Unmixing performance and optimal tuning parameters obtained with the spatial data set.

|  | Ext | khype | Prop. |
|---|---|---|---|
| RMSE$(A, A^{\star})$ | 0.0507 | 0.0380 | 0.0276 |
| RMSE$(F, F^{\star})$ | 0.0507 | 0.0213 | 0.0138 |
| $(\lambda, \mu)$ | — | (1, 0.01) | (1, 0.01) |



(a) True

(b) Ext
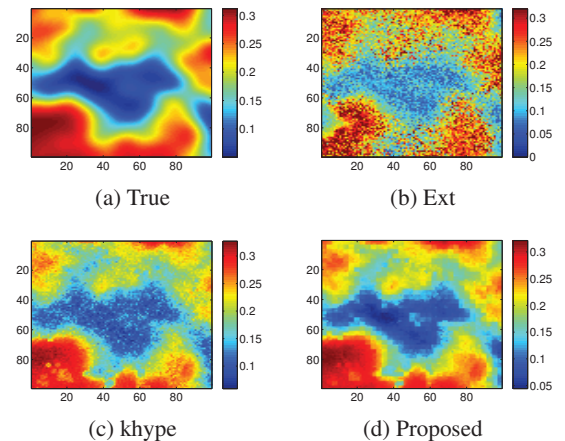
(c) khype

(d) Proposed

**Fig. 2**. True and estimated nonlinear contributions at band #100 obtained with the spatial data set using the extended endmember method (Ext), khype , and the proposed approach.

## 7. REFERENCES

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

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

[3] J. 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.

[4] N. Dobigeon, L. Tits, B. Somers, Y. Altmann, and P. Coppin, "A comparison of nonlinear mixing models for vegetated areas using simulated and real hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 1869–1878, 2014.

[5] I. Meganem, P. Deliot, X. Briottet, Y. Deville, and S. Hosseini, "Linear–quadratic mixing model for reflectances in urban environments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 544–558, 2014.

[6] Tales Imbiriba, Jose-Carlos M. Bermudez, Cédric Richard, and Jean-Yves Tourneret, "Nonparametric detection of nonlinearly mixed pixels and endmember estimation in hyperspectral images," *Image Processing, IEEE Transactions on*, vol. 25, no. 3, pp. 1136–1151, 2016.

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

[8] J. Nascimento and J. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," in *SPIE*, 2009.

[9] 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.

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

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

[12] O. Eches and M. Guillaume, "A bilinear–bilinear nonnegative matrix factorization method for hyperspectral unmixing," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 4, pp. 778–782, 2014.

[13] R. Heylen and P. Scheunders, "A multilinear mixing model for nonlinear spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 240–251, 2016.

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

[15] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*, Cambridge university press, 2004.

[16] 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, no. 10, pp. 2442–2453, 2013.

[17] T. Evgeniou, C. Micchelli, and M. Pontil, "Learning multiple tasks with kernel methods," in *Journal of Machine Learning Research*, 2005, pp. 615–637.

[18] M. Alvarez, L. Rosasco, and N. Lawrence, "Kernels for vector-valued functions: A review," *Foundations and Trends in Machine Learning*, vol. 4, no. 3, pp. 195–266, 2012.

[19] C. Couprie, L. Grady, L. Najman, J.-C. Pesquet, and H. Talbot, "Dual constrained tv-based regularization on graphs," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1246–1273, 2013.

[20] G. Chierchia N., Pustelnik, B. Pesquet-Popescu, and J.-C. Pesquet, "A nonlocal structure tensor-based approach for multi-component image recovery problems," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5531–5544, 2014.

[21] R. Ammanouil, A. Ferrari, and C. Richard, "A graph laplacian regularization for hyperspectral data unmixing," *arXiv preprint arXiv:1410.3699*, 2014.

[22] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundances in hyperspectral images with l1-norm spatial regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2654 – 2665, 2014.

[23] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4810–4819, 2015.

[24] L. J. Grady and J. R. Polimeni, *Discrete calculus*, Springer, 2010.

[25] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2008.

[26] 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.