NONLINEAR UNMIXING OF HYPERSPECTRAL IMAGES USING A SEMIPARAMETRIC MODEL AND SPATIAL REGULARIZATION

Jie Chen[*], Cédric Richard[*], Alfred O. Hero III[†]

[*] Université de Nice Sophia-Antipolis, CNRS, France
[†]University of Michigan, Ann Arbor, USA
E-mail: {jie.chen, cedric.richard}@unice.fr, hero@eecs.umich.edu

ABSTRACT

Incorporating spatial information into hyperspectral unmixing procedures has been shown to have positive effects, due to the inherent spatial-spectral duality in hyperspectral scenes. Current research works that consider spatial information are mainly focused on the linear mixing model. In this paper, we investigate a variational approach to incorporating spatial correlation into a nonlinear unmixing procedure. A nonlinear algorithm operating in reproducing kernel Hilbert spaces, associated with an ℓ_1 local variation norm as the spatial regularizer, is derived. Experimental results, with both synthetic and real data, illustrate the effectiveness of the proposed scheme.

Index Terms— Nonlinear unmixing, ℓ_1-norm regularization, spatial regularization, split Bregman iteration, hyperspectral data.

1. INTRODUCTION

Hyperspectral imaging provides two dimensional spatial images over many contiguous spectral bands. The high spectral resolution allows a comprehensive and quantitive analysis of materials in remotely observed data. This area has received considerable attention in the last decade, see [1] for a survey.

Usually, observed reflectance at each pixel is a spectral mixture of several material signatures, called endmembers, due to limited spatial resolution of observation devices and diversity of materials. Consequently, spectral unmixing has become an important issue for hyperspectral data processing [2]. There have been significant efforts during the past decade to address the linear unmixing problem for hyperspectral data [3–5]. Nevertheless, the linear model can only capture simple interactions between elements, e.g., in situations where the mixing of materials is not intimate and multiple scattering effects are negligible [1, 2]. Recently, several researchers have begun exploring nonlinear unmixing techniques. In [6], nonlinear unmixing was proposed based on the bilinear model and Bayesian inference. Post-nonlinear mixing models were discussed in [7, 8]. Unmixing algorithms using geodesic distances and other manifold learning based techniques were investigated in [9–12]. In addition, algorithms operating in reproducing kernel Hilbert spaces (RKHS) have been proposed for hyperspectral unmixing. Nonlinear unmixing with intuitive kernels was investigated in [13]. Physically-inspired kernel-based models were introduced in [14], where each mixed pixel is modeled by a linear mixture of endmember spectra, coupled with an additive nonlinear interaction term to model nonlinear effects of photon interactions. In [15–17], a more complete and sophisticated theory related to this strategy was presented. See [18] for an overview of recent advances in nonlinear unmixing modeling.

Beyond simply providing rich spectral information, remotely sensed data convey information about the spatial variability of spectral content in the 2D terrain [19]. Subsequently, hyperspectral analysis techniques should benefit from the inherent spatial-spectral duality in hyperspectral scenes. Following this idea, researchers have attempted to exploit spatial information for hyperspectral image unmixing. An NMF problem regularized with the ℓ_1-norm of differences between neighboring pixels was introduced in [20]. In [21], a Markov random field was proposed to model the spatial structure underlying pixels within classes. In [22], unmixing by a multi-resolution sticky Hierarchical Dirichlet Process model was used to account for spatial correlations. In [23], total variation was used for spatial regularization in order to enhance unmixing performance. Some other works also showed that incorporating spatial information can have a positive effect on unmixing processes [24–26]. Nevertheless, all these works were conducted with a linear mixing model. Rarely, if ever, have nonlinear mixing models incorporating spatial information been considered in the literature. As nonlinear unmixing is already an important but challenging issue, it appears complicated to address these two problems simultaneously. Based on the promising results of nonlinear unmixing in RKHS [15], in this paper, we propose a novel way to incorporate spatial information into the nonlinear unmixing process using ℓ_1-norm spatial regularization, i.e., a local version of the total variation smoothness regularizer in image reconstruction. An optimization method based on split-Bregman iterations is proposed to deal with the nonlinearity of the model and the non-smoothness of the regularizer.

2. FORMULATION OF THE PROBLEM

Consider an hyperspectral image with w pixels in each row, and h pixels in each column. Each pixel consists of a reflectance vector in L contiguous spectral bands. In order to keep the presentation simple, we transform this three dimensional image into an $L \times N$ matrix, with $N = w \times h$ the total number of pixels. Let $n \in \{1, \ldots, N\}$ be the sequential index of pixels. Suppose that the scene consists of R significant endmembers, each with a spectral signature $\boldsymbol{m}_i \in \mathbb{R}^L$. Let $\boldsymbol{r}_n \in \mathbb{R}^L$ be an observed hyperspectral pixel, and let $\boldsymbol{\alpha}_n \in \mathbb{R}^R$ be the vector of endmember abundances in the pixel $\boldsymbol{r}_n$. The matrix composed of all the abundance vectors is denoted by $\boldsymbol{A} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N]$. Let $\boldsymbol{M} = [\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R] \in \mathbb{R}^{L \times R}$ be the matrix of the endmember spectra. For the sake of convenience, the ℓ-th row of $\boldsymbol{M}$ is denoted by $\boldsymbol{m}_{\lambda_\ell}^\top \in \mathbb{R}^L$, that is, $\boldsymbol{m}_{\lambda_\ell}$ is the vector of the endmember signatures at the ℓ-th wavelength band. Finally, let $\boldsymbol{1}$ and $\boldsymbol{I}$ be the all-one vector and the identity matrix, respectively, with compatible sizes according to the context.

Similarly to many nonlinear unmixing approaches, we suppose that the material signatures in the scene have been determined by some endmember extraction algorithm. The unmixing problem boils

down to estimating the abundance vectors. To take the spatial relationship between pixels into consideration, the unmixing problem can be solved by minimizing a general cost function, with respect to $\boldsymbol{A}$, of the form

$$J(\boldsymbol{A}) = J_{\text{err}}(\boldsymbol{A}) + \eta\, J_{\text{sp}}(\boldsymbol{A}) \qquad (1)$$

subject to a non-negativity constraint on each entry of $\boldsymbol{A}$, and a sum-to-one constraint on each column of $\boldsymbol{A}$, namely, on each $\boldsymbol{\alpha}_n$. For ease of notation, these two physical constraints will be expressed by

$$\boldsymbol{A} \succeq \boldsymbol{0} \quad \text{and} \quad \boldsymbol{A}^\top \boldsymbol{1}_R = \boldsymbol{1}_N \qquad (2)$$

Recent work has raised the question of relaxing the sum-to-one constraint. The proposed algorithm can be easily adapted if this constraint is removed. In the experimental section, results subject to the non-negativity constraint will only be presented. In the general expression (1), the function J_{err} represents the modeling error and J_{sp} is a regularization term to promote similarity of the fractional abundances of neighboring pixels. Various regularizers have been defined in the literature [20, 23, 25]. The non-negative parameter η controls the trade-off between local data fidelity and pixel similarity.

Let us now present J_{err} and J_{sp} investigated in this paper. Consider the general unmixing process, acting between the entries $r_{n,\ell}$ of the observed reflectance vector, and the spectral signatures $\boldsymbol{m}_{\lambda_\ell}$ of the endmembers at each wavelength band λ_ℓ, defined as

$$r_{n,\ell} = \psi_{\boldsymbol{\alpha}_n}(\boldsymbol{m}_{\lambda_\ell}) + e_{n,\ell}$$

with $\psi_{\boldsymbol{\alpha}_n}$ an unknown nonlinear function to be estimated that defines the interaction between the endmember spectra, in the proportion $\boldsymbol{\alpha}_n$, and $\boldsymbol{e}_n$ the estimation error. This leads us to consider the general problem

$$\psi_{\boldsymbol{\alpha}_n}^* = \arg\min_{\psi_{\boldsymbol{\alpha}_n}} \frac{1}{2}\|\psi_{\boldsymbol{\alpha}_n}\|_{\mathcal{H}}^2 + \frac{1}{2\mu}\sum_{\ell=1}^{L}(r_{n,\ell} - \psi_{\boldsymbol{\alpha}_n}(\boldsymbol{m}_{\lambda_\ell}))^2 \quad (3)$$

with μ a positive parameter that controls the trade-off between structural error and misadjustment error. Clearly, this basic strategy may fail if the functionals $\psi_{\boldsymbol{\alpha}_n}$ cannot be adequately and finitely parameterized. In [15], we defined them by a linear trend parameterized by the abundance vector $\boldsymbol{\alpha}_n$, combined with a nonlinear fluctuation function ψ_n, namely,

$$\psi_{\boldsymbol{\alpha}_n}(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{\alpha}_n^\top \boldsymbol{m}_{\lambda_\ell} + \psi_n(\boldsymbol{m}_{\lambda_\ell}) \qquad (4)$$

where ψ_n can be any real-valued function in a reproducing kernel Hilbert space $\mathcal{H}$, endowed with the reproducing kernel κ such that $\psi_n(\boldsymbol{m}_{\lambda_\ell}) = \langle \psi_n, \kappa(\cdot, \boldsymbol{m}_{\lambda_\ell})\rangle$. Indeed, kernel-based methods lead to efficient and accurate resolution for inverse problems of the form (3) by exploiting the central idea of this research area, known as the *kernel trick*. We proposed in [15] to conduct data unmixing (3)–(4) by solving the following least-square support vector regression (LS-SVR) problem

$$\boldsymbol{\alpha}_n^*, \psi_n^* = \arg\min_{\boldsymbol{\alpha}_n, \psi_n} \frac{1}{2}\Big(\|\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|_{\mathcal{H}}^2 + \frac{1}{\mu}\|\boldsymbol{e}_n\|^2\Big)$$

$$\text{subject to} \quad \boldsymbol{\alpha}_n \succeq \boldsymbol{0} \quad \text{and} \quad \boldsymbol{1}^\top \boldsymbol{\alpha}_n = 1 \qquad (5)$$

where $\boldsymbol{e}_n$ is the $(L \times 1)$ misadjustment error vector with ℓ-th entry $e_{n,\ell} = r_{n,\ell} - (\boldsymbol{\alpha}_n^\top \boldsymbol{m}_{\lambda_\ell} + \psi_n(\boldsymbol{m}_{\lambda_\ell}))$ as defined in (3). It can be shown that problem (5) is convex so that it can be solved exactly by the duality theory. This so-called K-Hype method was introduced in [15]. Finally, considering all the pixels of the image to process, the modeling error to be minimized is expressed as

$$J_{\text{err}}(\boldsymbol{A}, \boldsymbol{\psi}) = \frac{1}{2}\sum_{n=1}^{N}\Big(\|\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|_{\mathcal{H}}^2 + \frac{1}{\mu}\|\boldsymbol{e}_n\|^2\Big)$$

subject to the contraints in (2). In this expression, $\boldsymbol{A} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N]$ and $\boldsymbol{\psi} = \{\psi_n \in \mathcal{H}: n = 1, \ldots, N\}$.

In order to take spatial correlation between pixels into account, we shall use ℓ_1-type regularizers of the form [20, 23] to promote piecewise constant transitions in the fractional abundance of each endmember among neighboring pixels. The regularization function is expressed as

$$J_{\text{sp}}(\boldsymbol{A}) = \sum_{n=1}^{N}\sum_{m\in\mathcal{N}(n)}\|\boldsymbol{\alpha}_n - \boldsymbol{\alpha}_m\|_1 \qquad (6)$$

where $\|\ \|_1$ denotes the ℓ_1 norm, and $\mathcal{N}(n)$ the set of neighbors of the pixel n. Without loss of generality, in this paper, we define the neighborhood of a pixel n by taking the 4 nearest pixels $n - 1$ and $n + 1$ (row adjacency), $n - w$ and $n + w$ (column adjacency). In this case, let us define the $(N \times N)$ matrices $\boldsymbol{H}_\leftarrow$ and $\boldsymbol{H}_\rightarrow$ as the two linear operators that compute the difference between any abundance vector and its left-hand neighbor, and right-hand neighbor, respectively. Similarly, let $\boldsymbol{H}_\uparrow$ and $\boldsymbol{H}_\downarrow$ be the linear operators that compute that difference with the top neighbor and the down neighbor, respectively. With these notations, the regularization function (6) can be rewritten in matrix form as

$$J_{\text{sp}}(\boldsymbol{A}) = \|\boldsymbol{A}\boldsymbol{H}\|_{1,1}$$

with $\boldsymbol{H}$ the $(N \times 4N)$ matrix $\big(\boldsymbol{H}_\leftarrow\, \boldsymbol{H}_\rightarrow\, \boldsymbol{H}_\uparrow\, \boldsymbol{H}_\downarrow\big)$ and $\|\ \|_{1,1}$ the sum of the ℓ_1-norms of the columns of a matrix. Unfortunately, while this regularization function is convex, it is non-smooth.

Now considering both the mismodeling error J_{err} and the regularization term J_{sp}, the optimization problem becomes

$$\boldsymbol{A}^*, \boldsymbol{\psi}^* = \arg\min_{\boldsymbol{A}, \boldsymbol{\psi}}\sum_{n=1}^{N}\frac{1}{2}\Big(\|\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|_{\mathcal{H}}^2 + \frac{1}{\mu}\|\boldsymbol{e}_n\|^2\Big) + \eta\,\|\boldsymbol{A}\boldsymbol{H}\|_{1,1}$$

$$\text{subject to} \quad \boldsymbol{A} \succeq \boldsymbol{0} \quad \text{and} \quad \boldsymbol{A}^\top \boldsymbol{1}_R = \boldsymbol{1}_N \qquad (7)$$

The constraints over $\boldsymbol{A}$ define a convex set $\mathcal{S}_A$. For ease of exposition, we will denote the constraints by $\boldsymbol{A} \in \mathcal{S}_A$.

3. SOLVING THE PROBLEM

Although the optimization problem (7) is convex, it cannot be solved easily because it combines a functional regression problem with a large-dimensional non-smooth regularization term. In order to overcome this, we rewrite (7) in the following equivalent form

$$\min_{\boldsymbol{A}\in\mathcal{S}_A, \boldsymbol{\psi}}\sum_{n=1}^{N}\frac{1}{2}\Big(\|\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|_{\mathcal{H}}^2 + \frac{1}{\mu}\|\boldsymbol{e}_n\|^2\Big) + \eta\,\|\boldsymbol{U}\|_{1,1}$$

$$\text{subject to} \quad \boldsymbol{V} = \boldsymbol{A} \quad \text{and} \quad \boldsymbol{U} = \boldsymbol{V}\boldsymbol{H} \qquad (8)$$

where two new matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, and two additional constraints, have been introduced. This variable-splitting approach was initially proposed in [27]. The matrix $\boldsymbol{U}$ will allow us to decouple the non-smooth ℓ_1-norm regularizer from the constrained LS-SVR problem. The matrix $\boldsymbol{V}$ will make the LS-SVR problem tractable by relaxing connections between pixels.

As studied in [27], the split-Bregman iteration algorithm is an efficient method to deal with a broad class of ℓ_1-regularized problems. By applying this framework to (7), the following iterative formulation is obtained

$$\boldsymbol{A}^{(k+1)}, \boldsymbol{\psi}^{(k+1)}, \boldsymbol{V}^{(k+1)}, \boldsymbol{U}^{(k+1)}$$

$$= \arg\min_{\boldsymbol{A}\in\mathcal{S}_A, \boldsymbol{\psi}, \boldsymbol{V}, \boldsymbol{U}}\sum_{n=1}^{N}\frac{1}{2}\Big(\|\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|_{\mathcal{H}}^2 + \frac{1}{\mu}\|\boldsymbol{e}_n\|^2\Big) + \eta\|\boldsymbol{U}\|_{1,1}$$

$$+ \frac{\zeta}{2}\|\boldsymbol{A} - \boldsymbol{V} - \boldsymbol{D}_1^{(k)}\|_F^2 + \frac{\zeta}{2}\|\boldsymbol{U} - \boldsymbol{V}\boldsymbol{H} - \boldsymbol{D}_2^{(k)}\|_F^2 \qquad (9)$$

$$\max_{\boldsymbol{\beta}_n, \boldsymbol{\gamma}_n, \lambda_n} \mathcal{L}'_n(\boldsymbol{\beta}_n, \boldsymbol{\gamma}_n, \lambda_n) = -\frac{\rho}{2\zeta} \left(\begin{array}{c} \boldsymbol{\beta}_n \\ \hline \boldsymbol{\gamma}_n \\ \hline \lambda_n \end{array} \right)^\top \left(\begin{array}{c|c|c} \boldsymbol{K}_\psi & \boldsymbol{M} & -\boldsymbol{M}\mathbf{1}_R \\ \hline \boldsymbol{M}^\top & \mathbf{I} & -\mathbf{1}_R \\ \hline -\mathbf{1}_R^\top \boldsymbol{M}^\top & -\mathbf{1}_R^\top & R \end{array} \right) \left(\begin{array}{c} \boldsymbol{\beta}_n \\ \hline \boldsymbol{\gamma}_n \\ \hline \lambda_n \end{array} \right) + \left(\begin{array}{c} \boldsymbol{r}_n - \rho \boldsymbol{M} \boldsymbol{\xi}_n^{(k)} \\ -\rho \boldsymbol{\xi}_n^{(k)} \\ \rho \boldsymbol{\xi}_n^{(k)\top} \mathbf{1}_R - 1 \end{array} \right)^\top \left(\begin{array}{c} \boldsymbol{\beta}_n \\ \hline \boldsymbol{\gamma}_n \\ \hline \lambda_n \end{array} \right)$$
$$\text{(14)}$$

$$\text{subject to} \quad \boldsymbol{\gamma}_n \succeq \mathbf{0}$$

$$\text{with} \quad \boldsymbol{K}_\psi = \frac{1}{\zeta} (\boldsymbol{K} + \mu \mathbf{I}) + \boldsymbol{M}\boldsymbol{M}^\top \text{ and } \rho = \frac{\zeta}{1+\zeta}$$

---

with
$$\boldsymbol{D}_1^{(k+1)} = \boldsymbol{D}_1^{(k)} + (\boldsymbol{V}^{(k+1)} - \boldsymbol{A}^{(k+1)})$$
$$\boldsymbol{D}_2^{(k+1)} = \boldsymbol{D}_2^{(k)} + (\boldsymbol{V}^{(k+1)}\boldsymbol{H} - \boldsymbol{U}^{(k+1)}) \tag{10}$$

where $\| \|_F^2$ denotes the matrix Frobenius norm, and ζ is a positive parameter. Because of how we have split the terms of the cost function, we can now perform the above minimization efficiently by iteratively minimizing with respect to $(\boldsymbol{A}, \boldsymbol{\psi})$, $\boldsymbol{V}$ and $\boldsymbol{U}$ separately. The three steps we have to perform are:

Step 1 - Optimization with respect to $\boldsymbol{A}$ and $\boldsymbol{\psi}$: The optimization problem (9) reduces to

$$\boldsymbol{A}^{(k+1)}, \boldsymbol{\psi}^{(k+1)} = \arg\min_{\boldsymbol{A} \in \mathcal{S}_A, \boldsymbol{\psi}} \sum_{n=1}^{N} \frac{1}{2} \Big(\|\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|_{\mathcal{H}}^2 + \frac{1}{\mu} \|\boldsymbol{e}_n\|^2 + \zeta \|\boldsymbol{\alpha}_n - \boldsymbol{\xi}_n^{(k)}\|^2 \Big)$$

where $\boldsymbol{\xi}_n^{(k)} = \boldsymbol{V}_n^{(k)} + \boldsymbol{D}_{1,n}^{(k)}$. Here, $\boldsymbol{V}_n$ and $\boldsymbol{D}_{1,n}$ denote the n-th column of $\boldsymbol{V}$ and $\boldsymbol{D}_1$, respectively. It can be observed that this problem can be solved, independently, for each vector $\boldsymbol{\alpha}_n$. This results from the use of the matrix $\boldsymbol{V}$. Let us now solve the local optimization problem

$$\boldsymbol{\alpha}_n^{(k+1)}, \psi_n^{(k+1)}$$
$$= \arg\min_{\boldsymbol{\alpha}_n, \psi_n, \boldsymbol{e}_n} \frac{1}{2} \Big(\|\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|_{\mathcal{H}}^2 + \frac{1}{\mu} \sum_{\ell=1}^{L} e_{n,\ell}^2 + \zeta \|\boldsymbol{\alpha}_n - \boldsymbol{\xi}_n^{(k)}\|^2 \Big)$$
$$\text{subject to} \quad e_{n,\ell} = r_{n,\ell} - (\boldsymbol{\alpha}_n^\top \boldsymbol{m}_{\lambda_\ell} + \psi_n(\boldsymbol{m}_{\lambda_\ell}))$$
$$\boldsymbol{\alpha}_n \succeq 0 \quad \text{and} \quad \boldsymbol{\alpha}_n^\top \mathbf{1}_R = 1 \tag{11}$$

By introducing the Lagrange multipliers $\beta_{n,\ell}$, $\gamma_{n,\ell}$ and λ_n, where the superscript (k) of these variables has been omitted for simplicity of notation, the Lagrange function associated with (11) is equal to

$$\mathcal{L}_n = \frac{1}{2} \Big(\|\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|_{\mathcal{H}}^2 + \frac{1}{\mu} \sum_{\ell=1}^{L} e_{n,\ell}^2 + \zeta \|\boldsymbol{\alpha}_n - \boldsymbol{\xi}_n^{(k)}\|^2 \Big)$$
$$- \sum_{\ell=1}^{L} \beta_\ell (e_{n,\ell} - r_{n,\ell} + \boldsymbol{\alpha}_n^\top \boldsymbol{m}_{\lambda_\ell} + \psi_n(\boldsymbol{m}_{\lambda_\ell}))$$
$$- \sum_{r=1}^{R} \gamma_r \alpha_{n,r} + \lambda_n (\boldsymbol{\alpha}_n^\top \mathbf{1}_R - 1) \tag{12}$$

with $\gamma_{n,r} \geq 0$. The conditions for optimality of $\mathcal{L}_n$ are

$$\begin{cases} \boldsymbol{\alpha}_n^* = \frac{1}{\zeta+1} \Big(\sum_{\ell=1}^{L} \beta_{n,\ell}^* \boldsymbol{m}_{\lambda_\ell} + \boldsymbol{\gamma}_n^* - \lambda_n^* \mathbf{1} + \zeta \boldsymbol{\xi}_n^{(k)} \Big) \\ \psi_n^* = \sum_{\ell=1}^{L} \beta_{n,\ell}^* \kappa(\cdot, \boldsymbol{m}_{\lambda_\ell}) \\ e_{n,\ell}^* = \mu \beta_{n,\ell}^* \end{cases} \tag{13}$$

where κ denotes the reproducing kernel of $\mathcal{H}$. By substituting (13) into (12), we get the dual problem (14) (see above), where $\boldsymbol{K}$ is the Gram matrix defined as $[\boldsymbol{K}]_{\ell p} = \kappa(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p})$. The problem (14) is a convex quadratic programming problem with respect to the dual variables. Finally, provided that the optimal dual variables $\boldsymbol{\beta}_n^*$, $\boldsymbol{\gamma}_n^*$ and λ_n^* have been determined, the vector of fractional abundances is estimated by

$$\boldsymbol{\alpha}_n^* = \frac{1}{\zeta+1} \big(\boldsymbol{M}^\top \boldsymbol{\beta}_n^* + \boldsymbol{\gamma}_n^* - \lambda_n^* \mathbf{1} + \zeta \boldsymbol{\xi}_n^{(k)} \big)$$

This process has to be repeated for $n = 1, \ldots, N$ to get $\boldsymbol{A}^{(k+1)}$.

Step 2 - Optimization with respect to $\boldsymbol{V}$: The optimization problem (9) now reduces to

$$\boldsymbol{V}^{(k+1)} = \arg\min_{\boldsymbol{V}} \|\boldsymbol{A}^{(k+1)} - \boldsymbol{V} - \boldsymbol{D}_1^{(k)}\|_F^2$$
$$+ \|\boldsymbol{U}^{(k)} - \boldsymbol{V}\boldsymbol{H} - \boldsymbol{D}_2^{(k)}\|_F^2 \tag{15}$$

Equating to zero the gradient of this expression with respect to $\boldsymbol{V}$ directly gives us the solution

$$\boldsymbol{V}^{(k+1)} = \Big(\boldsymbol{A}^{(k+1)} - \boldsymbol{D}_1^{(k)} + (\boldsymbol{U}^{(k)} - \boldsymbol{D}_2^{(k)}) \boldsymbol{H}^\top \Big) (\mathbf{I} + \boldsymbol{H}\boldsymbol{H}^\top)^{-1}$$

Step 3 - Optimization with respect to $\boldsymbol{U}$: The optimization problem (9) reduces to

$$\boldsymbol{U}^{(k+1)} = \arg\min_{\boldsymbol{U}} \eta \|\boldsymbol{U}\|_{1,1} + \frac{\zeta}{2} \|\boldsymbol{U} - \boldsymbol{V}^{(k+1)}\boldsymbol{H} - \boldsymbol{D}_2^{(k)}\|_F^2 \tag{16}$$

Its solution is expressed via the well-known soft threshold function

$$\boldsymbol{U}^{(k+1)} = \text{Thresh}\big(\boldsymbol{V}^{(k+1)}\boldsymbol{H} + \boldsymbol{D}_2^{(k)}, \eta/\zeta \big) \tag{17}$$

where $\text{Thresh}(\cdot, \tau)$ denotes the component-wise application of the soft threshold function defined as [28]

$$\text{Thresh}(x, \tau) = \text{sign}(x) \max(|x| - \tau, 0)$$

Note that, as they are spatially invariant, the multiplications by $\boldsymbol{H}$ in the above expressions can be efficiently performed with an FFT.

4. EXPERIMENTAL RESULTS

4.1. Experiments with synthetic images

Two spatially correlated hyperspectral images were generated for the following experiments. The endmembers were randomly selected from the spectral library ASTER [29], where signatures have reflectance values measured over 224 spectral bands. Following [23], two spatially correlated abundance distributions, with $R = 5$ and $R = 9$ were used. See [23] for the data description. The reflectance vectors were generated with two nonlinear mixture models described hereafter, and corrupted by a zero-mean white Gaussian noise $\boldsymbol{v}_n$ with a SNR of 20 dB. The first mixture model was the bilinear model defined as $\boldsymbol{r}_n = \boldsymbol{M}\boldsymbol{\alpha}_n + \sum_{i=1}^{R} \sum_{j=i+1}^{R} \alpha_{n,i} \alpha_{n,j} \boldsymbol{m}_i \otimes \boldsymbol{m}_j + \boldsymbol{v}_n$, with $\otimes$ the Hadamard product. The second one was a post-nonlinear model (PNMM) given by $\boldsymbol{r}_n = (\boldsymbol{M}\boldsymbol{\alpha}_n)^{0.7} + \boldsymbol{v}_n$. Several algorithms were tested in order to compare their unmixing performance on these two images. Their tuning parameters were set by preliminary experiments: 1) The linear unmixing methods [3]: The fully constrained least-square method (FCLS) was tested. By relaxing the sum-to-one constraint, one obtains the nonnegative constrained least-square method (NCLS), which was also considered. 2) The

Table 1. RMSE comparison with the synthetic data.

	DC1		DC2		Comp. time (ms/pixel)	
	Bilinear	PNMM	Bilinear	PNMM	IM1	IM2
FCLS	0.1730±0.0092	0.1316±0.0052	0.1680±0.0265	0.1444±0.0098	0.07	0.08
NCLS	0.1351±0.0131	0.1468±0.0071	0.0784±0.0076	0.1378±0.0135	0.06	0.07
spatial.-reg. FCLS	0.1729±0.0091	0.1311±0.0052	0.1676±0.0263	0.1381±0.0074	0.91	1.00
spatial.-reg. NCLS	0.1159±0.0044	0.1472±0.0069	0.0685±0.0053	0.1304±0.0097	0.85	0.90
K-Hype	0.0781±0.0050	0.0895±0.0072	0.0755±0.0080	0.1107±0.0104	5.7	6.0
NK-Hype	0.0771±0.0054	0.0873±0.0066	0.0919±0.0082	0.1059±0.0096	5.7	6.0
spatial.-reg. K-Hype (proposed)	0.0444±0.0016	0.0480±0.0480	0.0521±0.0033	0.0849±0.0042	56.5	68.8
spatial.-reg. NK-Hype (proposed)	0.0493±0.0026	0.0458±0.0042	0.0647±0.0032	0.0773±0.0044	55.1	69.8
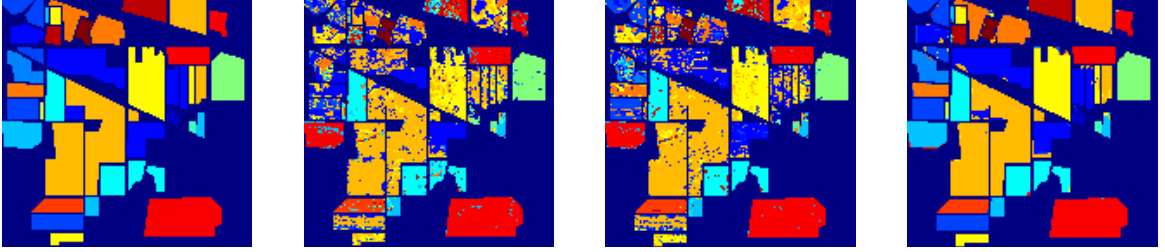
Fig. 1. Indian Pines classification map. From left to right: ground-truth, FCLS (61.36%), K-HYPE (71.39%), Proposed (96.80%).

spatially-regularized FCLS/NCLS: For comparison purposes, regularizer (6) was considered with FCLS/NCLS algorithms, solved by split-Bregman iterations. 3) The nonlinear unmixing algorithm K-Hype [15]: Unmixing was performed in this case by solving problem (5). Its nonnegative counterpart obtained by relaxing the sum-to-one constraint (NK-Hype) was also tested. The polynomial kernel defined by $\kappa(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_\ell}) = [1 + (\boldsymbol{m}_{\lambda_\ell} - 1/2)^\top (\boldsymbol{m}_{\lambda_\ell} - 1/2)/R^2]^2$ was used, as in [15]. 4) The proposed nonlinear algorithms incorporating spatial regularization: K-Hype and its nonnegative counterpart NK-Hype were both considered with spatial regularization. The parameter ζ was adjusted in an adaptive way based on primal and dual residual norms at each iteration, see [30]. Finally, the optimization algorithm was stopped when the number of iterations exceeded 10, or both $\frac{\|\boldsymbol{V} - \boldsymbol{A}\|_F}{N \times R}$ and $\frac{\|\boldsymbol{U} - \boldsymbol{V}\boldsymbol{H}\|_F}{4N \times R}$ became smaller than 10^{-5}. The RMSE

$$\text{RMSE} = \sqrt{\frac{1}{NR} \sum_{n=1}^N \|\boldsymbol{\alpha}_n - \boldsymbol{\alpha}_n^*\|^2} \qquad (18)$$

was used for comparing these algorithms, as reported in Table 1. Clearly, it can be observed that FCLS had large estimation errors. Relaxing the sum-to-one constraint with NCLS algorithm allowed to improve the performance in some cases, especially for DC2 with the bilinear model. The spatially-regularized FCLS and NCLS algorithms offered limited performance improvement. Nonlinear methods notably reduced this error in the mean sense, except for DC2 with the bilinear model. In this case, because most of the areas in the image are characterized by a dominant element with fractional abundance almost equal to one (see [23] for visual illustration), mixing phenomena associated with the bilinear model are significantly weaker. Finally, the proposed spatially-regularized methods showed lower errors than all other tested algorithms.

4.2. Experiments with AVIRIS data

In order to circumvent the difficulty that, in the literature, there is no available ground-truth for unmixing problems with real data, we adopted an indirect strategy to evaluate the proposed algorithm, via abundance-based classification. The estimated abundances were

Table 2. Classification performance with abundance-based features.

	5%	10%	15%
FCLS	56.41	61.36	62.32
K-Hype	67.67	71.39	74.68
Proposed	93.82	96.80	97.02

used as features to feed a classifier, and classification results were compared with labeled classification ground-truth. The scene used in our experiment is the well-known data set captured on the Indian Pines region by AVIRIS. The scene comprises 145×145 samples, consisting of 220 contiguous spectral bands. The ground-truth data contains 16 mutually exclusive classes. This widely used benchmark data set is known to be dominated by mixed pixels, even if ground-truth information assigns each pixel to a unique class. In this experiment, the so-called unmixing based classification chain #4 in [31] was used. We tested FCLS, K-Hype, and the proposed algorithm for extracting abundance-based features. A one-against-all multi-class SVM with Gaussian kernel was applied to these data. We constructed five training sets by randomly selecting 5%, 10%, and 15% of the samples available per class. All the required parameters were optimized by preliminary experiments. Table 2 summarizes the classification accuracies of SVM operating on features extracted with the unmixing algorithms. Fig. 1 presents these results in the case of an SVM trained with 10% of the samples available per class. It appears that our nonlinear unmixing algorithms are more efficient than the linear one for feature extraction. Finally, we observe that spatial regularization greatly improved the classification accuracy.

5. CONCLUSION

We considered the problem of nonlinear unmixing of hyperspectral images. A nonlinear algorithm operating in reproducing kernel Hilbert spaces was proposed. Spatial information was incorporated using an ℓ_1-norm local variation regularizer. Split-Bregman iterations were used to solve this convex non-smooth optimization problem. Experiments illustrated the effectiveness of this scheme.

6. 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 Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[2] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Sig. Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

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

[4] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2010.

[5] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P Gader, and J. Chanussot, "Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approches," *IEEE J. Sel. Topic Appl. Earth Observ.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[6] A. Halimi, Y. Altman, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

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

[8] J. Chen, C. Richard, and P. Honeine, "Estimating abundance fractions of materials in hyperspectral images by fitting a post-nonlinear mixing model," in *Proc. IEEE WHISPERS*, Gainesville, FL. USA, Jun. 2013.

[9] R. Heylen, D. Burazerovic, and P. Scheunders, "Non-linear spectral unmixing by geodesic simplex volume maximization," *IEEE J. Sel. Topic Signal Process.*, vol. 5, no. 3, pp. 534–542, Jun. 2011.

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

[11] N. H. Nguyen, J. Chen, C. Richard, P. Honeine, and C. Theys, *Supervised nonlinear unmixing of hyperspectral images using a pre-image method*, vol. 59, pp. 417–437, EDP Sciences, 2013.

[12] P. Honeine and C. Richard, "Solving the pre-image problem in kernel machines: A direct method," in *Proc. IEEE MLSP*, Grenoble, France, Sept. 2009, pp. 1–6.

[13] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Proc. IEEE IGARSS*, Barcelona, Spain, Jul. 2007, pp. 4041–4044.

[14] J. Chen, C. Richard, and P. Honeine, "A novel kernel-based nonlinear unmixing scheme of hyperspectral images," in *Proc. ASILOMAR*, Pacific Grove, CA. USA, Nov. 2011, pp. 1898–1902.

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

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

[17] J. Chen, C. Richard, A. Ferrari, and P. Honeine, "Nonlinear unmixing of hyperspectral data with partially linear least-squares support vector regression," in *Proc. IEEE ICASSP*, Vancouver, Canada, May. 2013, pp. 2174–2178.

[18] 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 Sig. Process. Mag. Process.*, Jan. 2013 (to appear).

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

[20] A. Zymnis, S. J. Kim, J. Skaf, M. Parente, and S. Boyd, "Hyperspectral image unmixing via alternating projected subgradients," in *Proc. ASILOMAR*, Pacific Grove, CA. USA, Nov. 2007, pp. 1164–1168.

[21] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4239–4247, Nov. 2011.

[22] R. Mittelman, N. Dobigeon, and A. O. Hero, "Hyperspectral image unmixing using multiresolution sticky hierarchical Dirichlet process," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1556–1671, Apr. 2012.

[23] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.

[24] S. Jia and Y. Qian, "Spectral and spatial complexity-based hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3867–3879, Dec. 2007.

[25] A. Zare, "Spatial-spectral unmixing using fuzzy local information," in *Proc. IEEE IGARSS*, Vancouver, Canada, Jul. 2011, pp. 1139–1142.

[26] G. Martin and A. Plaza, "Region-based spatial preprocessing for endmember extraction and spectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 745–749, Jul. 2011.

[27] T. Goldstein and S. Osher, "The split Bregman method for L1 regularized problems," *SIAM J. Imaging Sci.,*, vol. 2, no. 2, pp. 323–343, Apr. 2009.

[28] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[29] A. M. Baldridge, S. J. Hook, C. I. Grove, and G. Rivera, "The ASTER spectral library version 2.0," *Remote Sens. of Environ.*, vol. 113, no. 4, pp. 711–715, Apr. 2009.

[30] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

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