Nonlinear Estimation of Material Abundances in Hyperspectral Images With ℓ_1-Norm Spatial Regularization

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

***Abstract*—Integrating spatial information into hyperspectral unmixing procedures has been shown to have a positive effect on the estimation of fractional abundances due to the inherent spatial–spectral duality in hyperspectral scenes. However, current research works that take spatial information into account are mainly focused on the linear mixing model. In this paper, we investigate how to incorporate spatial correlation into a nonlinear abundance estimation process. A nonlinear unmixing algorithm operating in reproducing kernel Hilbert spaces, coupled with a ℓ_1-type spatial regularization, is derived. Experiment results, with both synthetic and real hyperspectral images, illustrate the effectiveness of the proposed scheme.**

***Index Terms*—Hyperspectral imaging, ℓ_1-norm regularization, nonlinear spectral unmixing, spatial regularization.**

I. Introduction

HYPERSPECTRAL imaging provides 2-D spatial images over many contiguous bands. The high spectral resolution allows to identify and quantify distinct materials from remotely observed data. This area has received considerable attention in the last decade. Applications include land use analysis, mineral detection, environment monitoring, field surveillance, etc. Typically, the observed reflectance at each pixel is a spectral mixture of several material signatures, termed endmembers, due to the limited spatial resolution of certain observation devices and diversity of materials. For this reason, spectral unmixing is an important issue for hyperspectral data processing to investigate information in each pixel [1].

There have been significant efforts during the past decade to solve linear unmixing problems, and several methods have been successfully derived within the context of hyperspectral imaging. Nevertheless, linear models can only accurately capture simple interactions between elements, e.g., in situations where the mixing of materials is not intimate and multiple scattering is negligible [1]. Recently, several researchers have started aggressively exploring nonlinear unmixing techniques. In [2] and [3], the authors extended the collection of endmembers by explicitly adding nonlinear artificial cross-terms of pure spectral signatures and used a conventional linear unmixing algorithm over this new dictionary. In [4], a Bayesian inference algorithm dedicated to the generalized bilinear mixture model was proposed. In [5], the same strategy was applied to estimate the fractional abundances of a polynomial post nonlinear mixing model. Manifold learning techniques were investigated in [6] and [7], and unmixing algorithms using geodesic distances or manifold-related regularization terms were proposed. Moreover, algorithms operating in reproducing kernel Hilbert spaces were considered to process hyperspectral data. In [8], a fuzzy-input fuzzy-output support vector machine was designed for subpixel image classification. Kernel-based nonlinear unmixing approaches were investigated in [9] and [10]. These algorithms were mainly obtained by replacing each inner product between endmember spectra, in the cost functions to be optimized, by a kernel function. This can be viewed as a nonlinear distortion map applied to the spectral signature of each material, which is of little physical interest in solving the unmixing problem because the nonlinear nature of mixing may also be characterized by nonlinear interactions of the materials. Physically inspired kernel-based models were introduced in [11] to circumvent this drawback, where we modeled each mixed pixel by a linear mixture of endmember spectra coupled with an additive nonlinear interaction term. The latter was used to model nonlinear effects of photon interactions and was defined in a reproducing kernel Hilbert space. In [12]–[14], a more complete and sophisticated theory and new methods were derived to automatically adjust the balance between linear and nonlinear components. A postnonlinear mixing model was also introduced in [15].

Although several linear and nonlinear state-of-the-art unmixing techniques have shown interesting performance for this problem, they have been mostly focused on exploiting spectral information available in the hyperspectral space. These approaches consist of considering pixel vectors as if they were independent from their neighboring pixels. However, one of the distinguishing properties of remotely sensed data is that they convey multivariate information into a 2-D pictorial representation [16]. Subsequently, instead of disregarding spatial contextual information, hyperspectral analysis techniques should benefit from the inherent spatial–spectral duality in

Manuscript received June 10, 2012; revised December 6, 2012 and March 11, 2013; accepted May 10, 2013. This work was supported by the Agence Nationale pour la Recherche, France, (Hypanema project, ANR-12-BS03-003).

J. Chen and C. Richard are with the Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur, 06108 Nice, France (e-mail: jie.chen@unice.fr; cedric.richard@unice.fr).

P. Honeine is with the Institut Charles Delaunay, Centre National de la Recherche Scientifique (CNRS), Université de Technologie de Troyes, 10010 Troyes Cedex, France (e-mail: paul.honeine@utt.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2264392

hyperspectral scenes. Following this idea, researchers have attempted to exploit spatial information in hyperspectral image analysis. Spatial preprocessing techniques were investigated for endmember determination in [17]–[19]. Spatial correlation was incorporated into hyperspectral image classification algorithms in [20] and [21]. Concerning the unmixing problem, a nonnegative matrix factorization-type problem regularized by the ℓ_1-norm of differences between neighboring pixels was proposed in [22]. A projected subgradient method was used to solve this problem. In [23], a Markov random field was proposed to model the spatial dependence of the pixels within classes. Bayesian inference was then used to estimate the model parameters. In [24], total variation was employed for spatial regularization to enhance the unmixing performance. The alternating direction method of multipliers was used to solve the regularized problem. Some other works also showed that incorporating spatial information can have a positive impact on the hyperspectral unmixing process [25], [26].

To the best of our knowledge, spatial regularization has not yet been extensively studied in a nonlinear unmixing process. As nonlinear unmixing is itself an important but challenging issue, it appears difficult to address these two problems simultaneously. A new nonlinear model was proposed in our work [12], where we assumed that a nonlinear mixture can be decomposed into a linear trend and an additive nonlinear fluctuation term in a reproducing kernel Hilbert space to model nonlinear effects. Based on this promising advance within the area of nonlinear unmixing, in this paper, we take spatial information into account in the unmixing process using ℓ_1-norm spatial regularization. An optimization method based on split-Bregman iterations is proposed to deal with this problem that suffers from the nonlinearity of the model and the nonsmoothness of the regularization term. Experiments with both synthetic and real data are conducted to validate the proposed method.

II. Formulation of the Problem

A. Notations

Suppose that the hyperspectral image under study has 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 facilitate the presentation, we transform this 3-D image into a $L \times N$ matrix, with $N = w \times h$ the total number of pixels. Then, let

- $n \in \{1, \ldots, N\}$ be the sequential index of pixels in the image.
- $\boldsymbol{r}_n = [r_{n1}, r_{n2}, \ldots, r_{nL}]^\top$ be the $(L \times 1)$ observed reflectance vector for the pixel n, which consists of a mixture of R endmember spectra.
- $\boldsymbol{M} = [\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_R]$ be the $(L \times R)$ endmember target matrix, where each column $\boldsymbol{m}_i$ is an endmember spectral signature. For expository convenience, we denote by $\boldsymbol{m}_{\lambda_\ell}^\top$ the ℓ-th row of $\boldsymbol{M}$, namely, the vector of the R endmember signatures at the ℓ-th wavelength band.
- $\boldsymbol{\alpha}_n = [\alpha_{n1}, \alpha_{n2}, \ldots, \alpha_{nR}]^\top$ be the $(R \times 1)$ abundance vector associated to the pixel n.
- $\boldsymbol{A} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N]$ be the matrix composed of all the abundance vectors.
- $\mathbf{1}_N$ be the $(N \times 1)$ all-one vector and $\mathbf{I}_N$ be the $(N \times N)$ identity matrix.

Other symbols will be defined in the context where they are used.

B. Formulation of the Nonlinear Unmixing Problem With Spatial Regularization

Suppose that the materials in a scene have been determined by some endmember extraction algorithm. The unmixing problem boils down to estimating the abundance vector associated to each pixel. One way to exploit spatial information is to define an appropriate criterion to be optimized, e.g., by considering extra penalty terms to promote the similarity of fractional abundances between neighboring pixels. The rationale is that homogeneous regions within which correlation among neighboring pixels is high potentially exist in real images. This suggests that increasing spatial homogeneity should tend to increase the accuracy of the representation of spectral objects and to suppress high-spatial-frequency artifacts [27], although there are risks to remove small features.

Taking the spatial relationship among pixels into consideration, the unmixing problem can then be solved by minimizing a general cost function with respect to the matrix $\boldsymbol{A}$

$$J(\boldsymbol{A}) = J_{\mathrm{err}}(\boldsymbol{A}) + \eta J_{\mathrm{sp}}(\boldsymbol{A}) \tag{1}$$

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

$$\boldsymbol{A} \succeq \mathbf{0}$$
$$\boldsymbol{A}^\top \mathbf{1}_R = \mathbf{1}_N.$$

Recent works have raised the question of relaxing the sum-to-one constraint. Indeed, poor estimates of the endmember signatures may affect the performance of the unmixing process. This constraint is maintained in the following as a baseline for comparison with existing approaches. However, in Appendix, we also briefly consider the case where the sum-to-one constraint is removed from the model. In the general expression (1), the function J_{err} represents the modeling error, and J_{sp} is a regularization term to promote the similarity of the fractional abundances within neighboring pixels. The nonnegative parameter η controls the tradeoff between data fidelity and pixel similarity. In [22] for instance, the ℓ_1-norm of the differences of abundance vectors in the neighborhood of each pixel was used as the spatial regularizer. In [24], anisotropic total variation norm was considered for linear sparse unmixing. In [26], a so-called fuzzy local information proportion was used for incorporating spatial information. In [21] and [23], a Markov random field was proposed to model spatial dependence. All these approaches have shown a substantial advantage of using spatial information for hyperspectral data unmixing, although there are risks to remove small and significant features. Preprocessing can be conducted to alleviate such hazards by separating these features in advance [28].

Rarely, if ever, have nonlinear mixing models incorporating spatial information been considered in the literature. In this paper, we intend to formulate J_{err} as in our recent work [12] because this approach has shown excellent performance and low computational cost. For self-containedness, let us briefly review part of this work. Consider the general unmixing process, acting between the entries $r_{n,\ell}$ of the 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 $\psi_{\boldsymbol{\alpha}_n}$ being an unknown nonlinear function to be estimated that defines the interaction between the endmember spectra, in the proportion $\boldsymbol{\alpha}_n$, and with $\boldsymbol{e}_n$ being the estimation error. This requires us to consider a general problem

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

with μ being a positive parameter that controls the tradeoff 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 [12], we defined them by a linear trend parameterized by the abundance vector $\boldsymbol{\alpha}_n$, combined with a nonlinear fluctuation function ψ_n, namely

$$\psi_{\boldsymbol{\alpha}_n}(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{\alpha}_n^\top \boldsymbol{m}_{\lambda_\ell} + \psi_n(\boldsymbol{m}_{\lambda_\ell}) \tag{3}$$

where ψ_n can be any real-valued function of 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 the efficient and accurate resolution of inverse problems of form (2) by exploiting the central idea of this research area, known as the *kernel trick* [29]. This trick has been widely used for solving nonlinear regression and classification problems [30], [31]. We proposed in [12] to conduct data unmixing (2) and (3) by solving the following least square support vector regression (LS-SVR) problem:

$$\begin{aligned} \boldsymbol{\alpha}^*_n, \psi^*_n = \arg\min_{\boldsymbol{\alpha}_n, \psi_n} \frac{1}{2}\left(\|\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|^2_{\mathcal{H}} + \frac{1}{\mu}\|\boldsymbol{e}_n\|^2\right) \\ \text{subject to} \quad \boldsymbol{\alpha}_n \succeq \mathbf{0} \quad \text{and} \quad \mathbf{1}^\top \boldsymbol{\alpha}_n = 1 \end{aligned} \tag{4}$$

where $\boldsymbol{e}_n$ is the $(L \times 1)$ misadjustment error vector with the 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 (2). It can be shown that problem (4) is convex so that it can be solved exactly by the duality theory. This so-called K-Hype method was introduced in [12]. A very efficient extension, called SK-Hype, consisting of relaxing the sum-to-one constraint and automatically adjusting the balance between $\boldsymbol{\alpha}_n^\top \boldsymbol{m}_{\lambda_\ell}$ and $\psi_n(\boldsymbol{m}_{\lambda_\ell})$ via an appropriate parameterization was also presented in this paper. In the following, we restrict our attention to K-Hype, and we maintain the sum-to-one constraint as a baseline for comparison with existing approaches. However, in Appendix, we also briefly address the case where the sum-to-one constraint is removed from the model. Finally, considering all the pixels of the image to process, the modeling error to be optimized writes

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

subject to the nonnegativity and sum-to-one constraints over the abundance vectors. 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 into account the spatial correlation between pixels, we shall use ℓ_1-type regularizers of the form [22], [24] to promote piecewise constant transitions in the fractional abundance of each endmember among neighboring pixels. The regularization function is expressed as

$$J_{\text{sp}}(\boldsymbol{A}) = \sum_{n=1}^{N}\sum_{m \in \mathcal{N}(n)} \|\boldsymbol{\alpha}_n - \boldsymbol{\alpha}_m\|_1 \tag{5}$$

where $\|\ \|_1$ denotes the vector ℓ_1-norm and $\mathcal{N}(n)$ denotes the set of neighbors of the pixel n. Without any loss of generality, in this paper, we define the neighborhood of the pixel n by taking the four nearest pixels $n-1$ and $n+1$ (row adjacency) and $n-w$ 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- and right-hand neighbors, 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 (5) can be rewritten in matrix form as

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

with $\boldsymbol{H}$ being the $(N \times 4N)$ matrix $(\boldsymbol{H}_\leftarrow\ \boldsymbol{H}_\rightarrow\ \boldsymbol{H}_\uparrow\ \boldsymbol{H}_\downarrow)$ and $\|\ \|_{1,1}$ being the sum of the ℓ_1-norms of the columns of a matrix. Obviously, other boundaries for the neighborhood $\mathcal{N}(n)$ may be used by simply defining the appropriate matrix $\boldsymbol{H}$. Finally, note that this regularization function is convex but nonsmooth.

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

$$\begin{aligned} \boldsymbol{A}^*, \boldsymbol{\psi}^* = \arg\min_{\boldsymbol{A}, \boldsymbol{\psi}} \sum_{n=1}^{N}\frac{1}{2}\left(\|\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|^2_{\mathcal{H}} + \frac{1}{\mu}\|\boldsymbol{e}_n\|^2\right) \\ + \eta\|\boldsymbol{A}\boldsymbol{H}\|_{1,1} \\ \text{subject to} \quad \boldsymbol{A} \succeq 0 \quad \text{and} \quad \boldsymbol{A}^\top \mathbf{1}_R = \mathbf{1}_N \end{aligned} \tag{6}$$

where η controls the tradeoff between model fitting in each pixel and similarity among neighboring pixels. The constraints over $\boldsymbol{A}$ define a convex set $\mathcal{S}_A$. For ease of exposition, in the formulation of optimization problems, we will write $\boldsymbol{A} \in \mathcal{S}_A$.

III. Solving the Problem

Although the optimization problem (6) is convex, it cannot be solved easily because it combines an LS-SVR problem with a huge-dimensional nonsmooth regularization term. In order

to overcome this drawback, we rewrite (6) in the following equivalent form:

$$\min_{\boldsymbol{A}\in\mathcal{S}_A,\boldsymbol{\psi}} \quad \sum_{n=1}^{N}\frac{1}{2}\left(\|\boldsymbol{\alpha}_n\|^2+\|\psi_n\|_{\mathcal{H}}^2+\frac{1}{\mu}\|\boldsymbol{e}_n\|^2\right)+\eta\|\boldsymbol{U}\|_{1,1}$$
$$\text{subject to} \quad \boldsymbol{V}=\boldsymbol{A} \quad \text{and} \quad \boldsymbol{U}=\boldsymbol{V}\boldsymbol{H} \tag{7}$$

where two new matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, and two additional constraints, have been introduced. This variable-splitting approach was initially introduced in [32]. Matrix $\boldsymbol{U}$ will allow us to decouple the nonsmooth ℓ_1-norm regularization functional from the constrained LS-SVR problem. Matrix $\boldsymbol{V}$ will make the LS-SVR problem tractable by relaxing connections between pixels.

As studied in [32], 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 (6), the following formulation is obtained:

$$\begin{aligned}&\boldsymbol{A}^{(k+1)},\boldsymbol{\psi}^{(k+1)},\boldsymbol{V}^{(k+1)},\boldsymbol{U}^{(k+1)}\\&=\underset{\boldsymbol{A}\in\mathcal{S}_A,\boldsymbol{\psi},\boldsymbol{V},\boldsymbol{U}}{\arg\min}\sum_{n=1}^{N}\frac{1}{2}\left(\|\boldsymbol{\alpha}_n\|^2+\|\psi_n\|_{\mathcal{H}}^2+\frac{1}{\mu}\|\boldsymbol{e}_n\|^2\right)\\&\quad+\eta\|\boldsymbol{U}\|_{1,1}+\frac{\zeta}{2}\left\|\boldsymbol{A}-\boldsymbol{V}-\boldsymbol{D}_1^{(k)}\right\|_F^2\\&\quad+\frac{\zeta}{2}\left\|\boldsymbol{U}-\boldsymbol{V}\boldsymbol{H}-\boldsymbol{D}_2^{(k)}\right\|_F^2\end{aligned} \tag{8}$$

with

$$\begin{aligned}\boldsymbol{D}_1^{(k+1)}&=\boldsymbol{D}_1^{(k)}+\left(\boldsymbol{V}^{(k+1)}-\boldsymbol{A}^{(k+1)}\right)\\\boldsymbol{D}_2^{(k+1)}&=\boldsymbol{D}_2^{(k)}+\left(\boldsymbol{V}^{(k+1)}\boldsymbol{H}-\boldsymbol{U}^{(k+1)}\right)\end{aligned} \tag{9}$$

where $\|\ \|_F^2$ denotes the matrix Frobenius norm and ζ is a positive parameter. Because of the way that we have split the components 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 that we have to perform are as follows.

1) Step 1—Optimization With Respect to $\boldsymbol{A}$ and $\boldsymbol{\psi}$: The optimization problem (8) reduces to

$$\begin{aligned}\boldsymbol{A}^{(k+1)},\boldsymbol{\psi}^{(k+1)}&=\underset{\boldsymbol{A}\in\mathcal{S}_A,\boldsymbol{\psi}}{\arg\min}\sum_{n=1}^{N}\frac{1}{2}\\&\quad\times\left(\|\boldsymbol{\alpha}_n\|^2+\|\psi_n\|_{\mathcal{H}}^2+\frac{1}{\mu}\|\boldsymbol{e}_n\|^2+\zeta\left\|\boldsymbol{\alpha}_n-\boldsymbol{\xi}_n^{(k)}\right\|^2\right)\end{aligned}$$

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 decomposed into subproblems, each one involving an abundance vector $\boldsymbol{\alpha}_n$. This results from the use of the matrix $\boldsymbol{V}$ in the split iteration algorithm (8).

Let us now solve the local optimization problem

$$\begin{aligned}\boldsymbol{\alpha}_n^{(k+1)},\psi_n^{(k+1)}=\underset{\boldsymbol{\alpha}_n,\psi_n,\boldsymbol{e}_n}{\arg\min}\ &\frac{1}{2}\Bigg(\|\boldsymbol{\alpha}_n\|^2+\|\psi_n\|_{\mathcal{H}}^2+\frac{1}{\mu}\sum_{\ell=1}^{L}e_{n,\ell}^2\\&\qquad+\zeta\left\|\boldsymbol{\alpha}_n-\boldsymbol{\xi}_n^{(k)}\right\|^2\Bigg)\\\text{subject to}\quad & e_{n,\ell}=r_{n,\ell}-\left(\boldsymbol{\alpha}_n^\top\boldsymbol{m}_{\lambda_\ell}+\psi_n(\boldsymbol{m}_{\lambda_\ell})\right)\\&\boldsymbol{\alpha}_n\succeq 0\\&\boldsymbol{\alpha}_n^\top\mathbf{1}_R=1.\end{aligned} \tag{10}$$

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 to (10) is written as

$$\begin{aligned}\mathcal{L}_n=&\frac{1}{2}\left(\|\boldsymbol{\alpha}_n\|^2+\|\psi_n\|_{\mathcal{H}}^2+\frac{1}{\mu}\sum_{\ell=1}^{L}e_{n,\ell}^2+\zeta\left\|\boldsymbol{\alpha}_n-\boldsymbol{\xi}_n^{(k)}\right\|^2\right)\\&-\sum_{\ell=1}^{L}\beta_\ell\left(e_{n,\ell}-r_{n,\ell}+\boldsymbol{\alpha}_n^\top\boldsymbol{m}_{\lambda_\ell}+\psi_n(\boldsymbol{m}_{\lambda_\ell})\right)\\&-\sum_{r=1}^{R}\gamma_r\alpha_{n,r}+\lambda_n\left(\boldsymbol{\alpha}_n^\top\mathbf{1}_R-1\right)\end{aligned} \tag{11}$$

with $\gamma_{n,r}\geq 0$. The conditions for the optimality of $\mathcal{L}_n$ with respect to the primal variables lead us to

$$\begin{cases}\boldsymbol{\alpha}_n^*=\frac{1}{\zeta+1}\left(\sum_{\ell=1}^{L}\beta_{n,\ell}^*\boldsymbol{m}_{\lambda_\ell}+\boldsymbol{\gamma}_n^*-\lambda_n^*\mathbf{1}+\zeta\boldsymbol{\xi}_n^{(k)}\right)\\\psi_n^*=\sum_{\ell=1}^{L}\beta_{n,\ell}^*\kappa(\cdot,\boldsymbol{m}_{\lambda_\ell})\\e_{n,\ell}^*=\mu\beta_{n,\ell}^*\end{cases} \tag{12}$$

where κ is the reproducing kernel of $\mathcal{H}$. By substituting (12) into (11), we get the dual problem

$$\begin{aligned}&\max_{\boldsymbol{\beta}_n,\boldsymbol{\gamma}_n,\lambda_n}\mathcal{L}_n'(\boldsymbol{\beta}_n,\boldsymbol{\gamma}_n,\lambda_n)\\&=-\frac{\rho}{2\zeta}\begin{pmatrix}\boldsymbol{\beta}_n\\\hline\boldsymbol{\gamma}_n\\\hline\lambda_n\end{pmatrix}^\top\left(\begin{array}{c|c|c}\boldsymbol{K}_\psi&\boldsymbol{M}&-\boldsymbol{M}\mathbf{1}_R\\\hline\boldsymbol{M}^\top&\mathbf{I}&-\mathbf{1}_R\\\hline-\mathbf{1}_R^\top\boldsymbol{M}^\top&-\mathbf{1}_R^\top&R\end{array}\right)\\&\quad\times\begin{pmatrix}\boldsymbol{\beta}_n\\\hline\boldsymbol{\gamma}_n\\\hline\lambda_n\end{pmatrix}+\begin{pmatrix}\boldsymbol{r}_n-\rho\boldsymbol{M}\boldsymbol{\xi}_n^{(k)}\\\hline-\rho\boldsymbol{\xi}_n^{(k)}\\\hline\rho\boldsymbol{\xi}_n^{(k)\top}\mathbf{1}_R-1\end{pmatrix}^\top\begin{pmatrix}\boldsymbol{\beta}_n\\\hline\boldsymbol{\gamma}_n\\\hline\lambda_n\end{pmatrix}\\&\text{subject to}\quad\boldsymbol{\gamma}_n\succeq\mathbf{0}\\&\text{with}\quad\boldsymbol{K}_\psi=\frac{1}{\zeta}(\boldsymbol{K}+\mu\mathbf{I})+\boldsymbol{M}\boldsymbol{M}^\top\quad\text{and}\quad\rho=\frac{\zeta}{1+\zeta}\end{aligned} \tag{13}$$

where $\boldsymbol{K}$ is the Gram matrix defined as $[\boldsymbol{K}]_{\ell p}=\kappa(\boldsymbol{m}_{\lambda_\ell},\boldsymbol{m}_{\lambda_p})$. The problem (13) is a convex quadratic programming problem with respect to the dual variables.

TABLE I
NONLINEAR UNMIXING WITH SPATIAL REGULARIZATION ALGORITHM

Initialization
Choose the kernel κ_{nlin} and calculate the kernel matrix $\boldsymbol{K}_{\text{nlin}}$.
Choose the regularization parameters μ and η in problem (6).
Initialize $\boldsymbol{U}^{(0)}$, $\boldsymbol{V}^{(0)}$, $\boldsymbol{D}_1^{(0)}$ and $\boldsymbol{D}_2^{(0)}$.
Repeat
Calculate primal variables:
- Calculate $\boldsymbol{A}^{(k+1)}$ by solving the QP problem (13) for each pixel.
- Calculate $\boldsymbol{V}^{(k+1)}$ using (15), and $\boldsymbol{U}^{(k+1)}$ using (17).

Update $\boldsymbol{D}_1^{(k+1)}$ and $\boldsymbol{D}_2^{(k+1)}$ using (9).
until stopping criterion is satisfied
Estimate Use the final $\boldsymbol{A}^{(k+1)}$ as the fractional abundances.

Finally, provided that the optimal dual variables $\boldsymbol{\beta}_n^*$, $\boldsymbol{\gamma}_n^*$, and λ_n^* have been determined, the vector of fractional abundances $\boldsymbol{\alpha}_n^{(k+1)}$ is estimated by

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

This process has to be repeated for $n = 1, \ldots, N$ in order to get $\boldsymbol{A}^{(k+1)}$. Matrices $\boldsymbol{A}^{(k+1)}$ and $\boldsymbol{V}^{(k+1)}$, whose calculation is presented hereafter, allow to evaluate $\boldsymbol{D}_1^{(k+1)}$ using (9).

2) Step 2—Optimization With Respect to $\boldsymbol{V}$: The optimization problem (8) now reduces to

$$\boldsymbol{V}^{(k+1)} = \arg\min_{\boldsymbol{V}} \left\|\boldsymbol{A}^{(k+1)} - \boldsymbol{V} - \boldsymbol{D}_1^{(k)}\right\|_F^2 + \left\|\boldsymbol{U}^{(k)} - \boldsymbol{V}\boldsymbol{H} - \boldsymbol{D}_2^{(k)}\right\|_F^2. \quad (14)$$

Equating to zero the derivative of the above expression with respect to $\boldsymbol{V}$ leads to

$$\left(\boldsymbol{A}^{(k+1)} - \boldsymbol{V} - \boldsymbol{D}_1^{(k)}\right) - \left(\boldsymbol{U}^{(k)} - \boldsymbol{V}\boldsymbol{H} - \boldsymbol{D}_2^{(k)}\right)\boldsymbol{H}^\top = 0$$

whose solution is then given by

$$\boldsymbol{V}^{(k+1)} = \left(\boldsymbol{A}^{(k+1)} - \boldsymbol{D}_1^{(k)} + \left(\boldsymbol{U}^{(k)} - \boldsymbol{D}_2^{(k)}\right)\boldsymbol{H}^\top\right)(\mathbf{I} + \boldsymbol{H}\boldsymbol{H}^\top)^{-1}. \quad (15)$$

3) Step 3—Optimization With Respect to $\boldsymbol{U}$: Finally, the optimization problem that we have to consider is as follows:

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

Its solution can be expressed via the well-known soft threshold function

$$\boldsymbol{U}^{(k+1)} = \text{Thresh}\left(\boldsymbol{V}^{(k+1)}\boldsymbol{H} + \boldsymbol{D}_2^{(k)}, \frac{\eta}{\zeta}\right) \quad (17)$$

where $\text{Thresh}(\cdot, \tau)$ denotes the componentwise application of the soft threshold function defined as [33]

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

4) Discussion: To conclude, problem (7) is solved by iteratively applying (8) and (9) until some stopping criterion is satisfied. It can be shown that, if the problem (8) has a solution $\boldsymbol{A}^*$ given any $\zeta > 0$, then the generated sequence $\boldsymbol{A}^{(k)}$ converges to the optimum $\boldsymbol{A}^*$ [34].

Problem (8) is addressed by solving three subproblems with respect to $\boldsymbol{A}$ [see (10)], with respect to $\boldsymbol{V}$ [see (14)], and with respect to $\boldsymbol{U}$ [see (16)], repeatedly. The first subproblem with respect to $\boldsymbol{A}$ is solved by considering the standard quadratic programming problem defined by (13), of size $L + R + 1$, subject to R inequality constraints and one equality constraint, for each pixel. It is equivalent to the K-Hype algorithm described in [12]. Its computational complexity does not depend on the nature of the nonlinearity in the mixture model as it is formulated in a reproducing kernel Hilbert space. The second subproblem with respect to $\boldsymbol{V}$ has an explicit solution (15) that involves the inverse of the matrix $(\boldsymbol{I} + \boldsymbol{H}\boldsymbol{H}^\top)$. The latter can be calculated once the neighborhood relationship is defined. The third subproblem with respect to $\boldsymbol{U}$ has also an explicit solution (17). The computational time required by the two latter stages is almost negligible compared to the resources required by the first stage. Finally, the overall running time of the algorithm is mainly dictated by the number of iterations on $\boldsymbol{A}$, $\boldsymbol{U}$, and $\boldsymbol{V}$. It can be adjusted to compromise between computational time and convergence accuracy. The algorithm is provided in pseudocode form in Table I.

IV. EXPERIMENT RESULTS

In this section, experiments on spatially correlated images are reported to validate the proposed algorithm.

A. Experiments With Synthetic Images

1) Simulation Scenario Settings: Two spatially correlated abundance maps were generated for the following experiments. The endmembers were randomly selected from the spectral library advanced spaceborne thermal emission and reflection radiometer (ASTER) [35]. Each signature of this library has reflectance values measured over 224 spectral bands, uniformly distributed in the interval of 3–12 μm. Two synthetic images identical to that in [24] were used.

The first data cube, denoted by IM1 and containing 75×75 pixels, was generated by using five signatures randomly selected from the ASTER library. Pure regions and mixed regions involving between two and five endmembers, distributed spatially in the form of square regions, were generated. The background pixels were defined as mixtures of the same five endmembers with the abundance vector

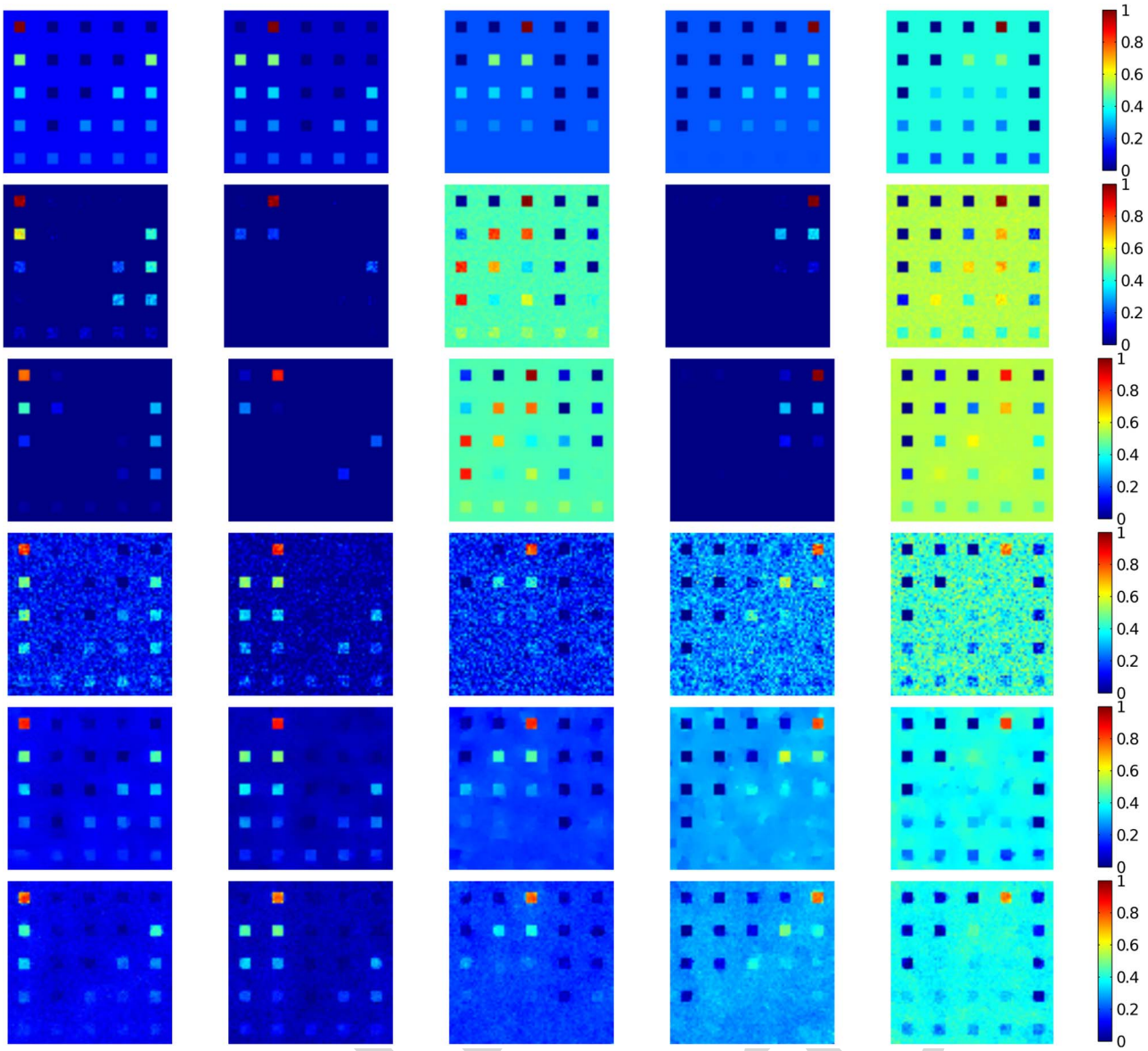

Fig. 1. Estimated abundance maps for IM1 image. For each row, from top to bottom: True abundance maps, FCLS, spatially regularized FCLS, K-Hype, proposed algorithm with four neighbors, and proposed algorithm with eight neighbors.

$[0.1149, 0.0741, 0.2003, 0.2055, 0.4051]^\top$. The first row in Fig. 1 shows the true fractional abundances for each endmember. The reflectance samples were generated with the two nonlinear mixture models defined hereafter, based on the five endmembers, and corrupted by a zero-mean white Gaussian noise $\boldsymbol{v}_n$ with an SNR of 20 dB. The first nonlinear mixture model is the bilinear one defined as

$$\boldsymbol{r}_n = \boldsymbol{M}\boldsymbol{\alpha}_n + \sum_{i=1}^{R}\sum_{j=i+1}^{R} \alpha_{n,i}\alpha_{n,j}\boldsymbol{m}_i \otimes \boldsymbol{m}_j + \boldsymbol{v}_n \qquad (18)$$

with $\otimes$ being the Hadamard product. The second one is a postnonlinear model (PNMM) given by

$$\boldsymbol{r}_n = (\boldsymbol{M}\boldsymbol{\alpha}_n)^\gamma + \boldsymbol{v}_n \qquad (19)$$

with an exponential value $\gamma = 0.7$ applied to the linear mixing model. At the end of this series of experiments with synthetic images, note that we will also consider a signal-dependent noise $\boldsymbol{v}_n$ to conform with conditions that may be experienced with new-generation sensors.

The second data cube, denoted by IM2 and containing 100×100 mixed pixels, was generated using nine endmember signatures. The abundance maps of the endmembers are the same as that for the image DC2 in [24]. Among these nine materials, only the 1st, 3rd, 5th, 8th, and 9th abundances are considered for pictorial illustration in Fig. 2. The first row of this figure depicts the true distribution of these five materials. Spatially homogeneous areas with sharp transitions can be clearly observed. Based on these abundance maps, a hyperspectral data cube was generated with the bilinear model (18) and with the postnonlinear model (19) applied to the nine endmember spectral signatures. The scene was also corrupted by a zero-mean white Gaussian noise $\boldsymbol{v}_n$ with an SNR of 20 dB.

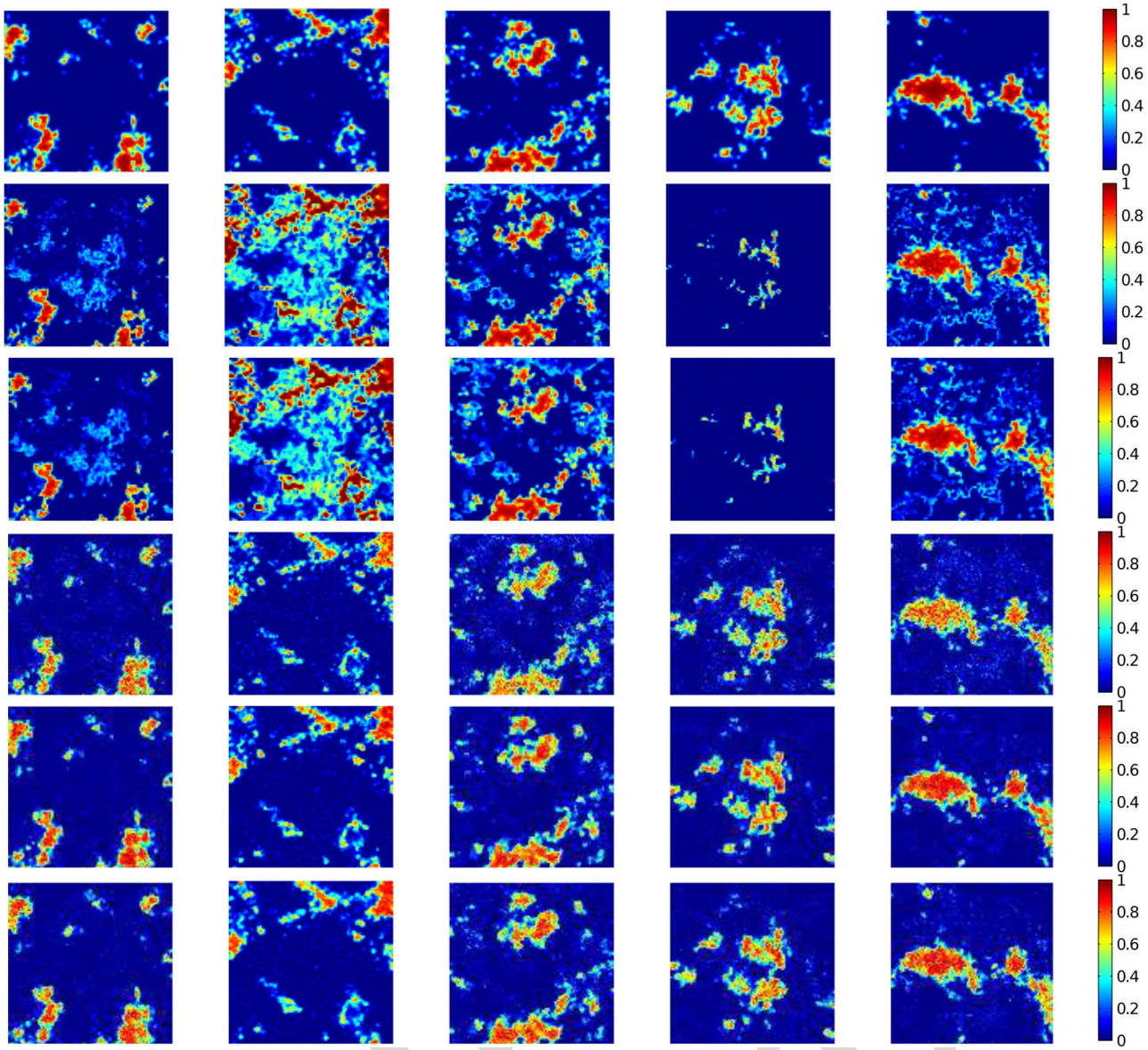

Fig. 2. Estimated abundance maps for IM2 image. For each row, from top to bottom: True abundance maps, FCLS, spatially regularized with four neighbors, FCLS, K-Hype, proposed algorithm with four neighbors, and proposed algorithm with eight neighbors.

2) Comparative Simulations: Several algorithms were tested in order to compare their unmixing performance on these two images. Their tuning parameters were set during preliminary experiments on independent data, via a simple search over the grids defined hereafter.

1) The linear unmixing methods [36]: The fully constrained least square method (FCLS) was run with the sum-to-one constraint strictly satisfied for comparability among algorithms. By relaxing the sum-to-one constraint, the nonnegative constrained least square method (NCLS) was also considered.
2) The spatially regularized FCLS/NCLS: In order to compare linear and nonlinear algorithms, we added the spatial regularization term (5) to the FCLS/NCLS algorithms. We conducted the split-Bregman iterations to solve these problems. We varied the spatial regularization parameter η from 0.0025 to 0.01 with an increment of 0.0025.
3) The nonlinear unmixing algorithm K-Hype [12]: Unmixing was performed in this case by solving problem (4). Its nonnegative counterpart obtained by relaxing the sum-to-one constraint (NK-Hype) was also tested. As in [12], the polynomial kernel defined by

$$\kappa(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_\ell}) = \left(1 + \frac{1}{R^2}(\boldsymbol{m}_{\lambda_\ell} - 1/2)^\top (\boldsymbol{m}_{\lambda_\ell} - 1/2)\right)^2 \tag{20}$$

was used. The parameter μ that controls the tradeoff between the misadjustment error and the regularization error was varied in the set $\{0.001, 0.005, 0.01, 0.05, 0.1\}$.
4) The proposed nonlinear algorithms incorporating spatial regularization: K-Hype and its nonnegative counterpart NK-Hype were both considered with spatial regularization. The polynomial kernel (20) was used, and the regularization parameter μ was varied in the same set

TABLE II
PARAMETER SETTINGS FOR THE COMPARATIVE SIMULATIONS ON IM1 AND IM2

	IM1		IM2	
	Bilinear	PNMM	Bilinear	PNMM
FCLS / spatial.-reg. FCLS	$\eta = 0.005$	$\eta = 0.005$	$\eta = 0.005$	$\eta = 0.01$
NCLS / spatial.-reg. NCLS	$\eta = 0.005$	$\eta = 0.0025$	$\eta = 0.005$	$\eta = 0.01$
K-Hype	$\mu = 0.1$	$\mu = 0.1$	$\mu = 0.01$	$\mu = 0.01$
NK-Hype	$\mu = 0.1$	$\mu = 0.1$	$\mu = 0.01$	$\mu = 0.05$
spatial.-reg. K-Hype (proposed)	$\mu = 0.005$, $\eta = 0.5$	$\mu = 0.005$, $\eta = 0.5$	$\mu = 0.005$, $\eta = 0.5$	$\mu = 0.005$, $\eta = 0.5$
spatial.-reg. NK-Hype (proposed)	$\mu = 0.005$, $\eta = 0.5$	$\mu = 0.005$, $\eta = 0.5$	$\mu = 0.005$, $\eta = 0.5$	$\mu = 0.005$, $\eta = 0.5$

TABLE III
RMSE COMPARISON (SIGNAL-INDEPENDENT NOISE): SEE EXPERIMENTS IN SECTION IV-A2

	IM1		IM2		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±0.048	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

TABLE IV
RMSE OF THE PROPOSED METHOD AS FUNCTION OF η: SEE EXPERIMENTS IN SECTION IV-A3

η	0.25	0.50	0.75	1.00
IM1	0.0548±0.0027	0.0456±0.0015	0.0479±0.0024	0.0508±0.0037
IM2	0.0576±0.0040	0.0521±0.0033	0.0530±0.0032	0.0567±0.0037

as above. The parameter ζ was adjusted in an adaptive way based on primal and dual residual norms at each iteration (see [37]). We varied the spatial regularization parameter η from 0.25 to 1 with increments of 0.25. Finally, the optimization algorithm was stopped when both $(\|\boldsymbol{V} - \boldsymbol{A}\|_F/N \times R)$ and $(\|\boldsymbol{U} - \boldsymbol{V}\boldsymbol{H}\|_F/4N \times R)$ became smaller than 10^{-5} or the number of iterations exceeded 10.

The above tests were performed on training images IM1 and IM2 to estimate the best parameter values in the sense that they minimize the estimation error (rmse) defined as

$$E = \sqrt{\frac{1}{NR}\sum_{n=1}^{N} \|\boldsymbol{\alpha}_n - \boldsymbol{\alpha}_n^*\|^2}.$$

These preliminary experiments led to the parameter settings shown in Table II. The performance, with standard deviation, on independent test images IM1 and IM2 is reported in Table III, and the estimated fractional abundances are represented in Figs. 1 and 2. For both images and both nonlinear mixture models, it can be observed that, when applied to nonlinearly mixed data, the linear method FCLS has large estimation errors. The abundance maps appear quite correct visually, but they are severely biased due to the nonlinearity of the mixing model. Relaxing the sum-to-one constraint with the NCLS algorithm allowed to improved the performance in some cases, particularly for IM2 with the bilinear model. The spatially regularized FCLS and NCLS algorithms offer limited performance improvement. Nonlinear methods notably reduce this error in the mean sense, except for IM2 with the bilinear model. In this case, note that most of the areas in the image are characterized by a dominant element with fractional abundance almost equal to one. Mixing phenomena with the bilinear model are thus limited, and the nonlinearity of the unmixing model supported by K-Hype-based algorithms suffers from this situation. They, however, provide high-resolution maps elsewhere. Finally, the proposed spatially regularized methods have lower reconstruction errors than the other proposed algorithms and clearer abundance maps.

3) Influence of the Parameter η: The penalty term η controls the tradeoff between data fitting and similarity among neighboring pixels. In the case of $\eta = 0$, the algorithm reduces to the original K-Hype that only considers spectral information at each pixel. The larger the η, the flatter the image is. In order to illustrate this intuition, we varied parameter η from 0.25 to 1 for IM1 and IM2 generated via the bilinear model, with $\mu = 0.005$. Note that, according to Table IV, the optimal value of η is 0.5 in both cases. To illustrate this experiment, the results are represented for IM1 in Fig. 3.

4) Influence of the Neighborhood: In the above experiments, we have used the four nearest neighbors to construct the difference matrix $\boldsymbol{H}_4$, with the subscript (4) to specify the size of the neighborhood. Any other neighborhood could be considered, provided that the matrix $\boldsymbol{H}$ is properly defined. For illustration purpose, we also considered the larger matrix of the eight nearest neighbors of each pixel defined as $\boldsymbol{H}_8 = (\boldsymbol{H}_\nwarrow\, \boldsymbol{H}_\nearrow\, \boldsymbol{H}_4\, \boldsymbol{H}_\swarrow\, \boldsymbol{H}_\searrow)$, with $\boldsymbol{H}_\nwarrow$, $\boldsymbol{H}_\nearrow$, $\boldsymbol{H}_\swarrow$, $\boldsymbol{H}_\searrow$ being the four diagonal adjacency matrices. The estimation errors of the abundance fractions are reported in Table V. The spatial regularization parameters were set to half of the values in Table II, as twice the number of neighboring pixels is used. Abundance maps for the bilinear scenario provided by the proposed algorithm are shown in the last row of Figs. 1 and 2.

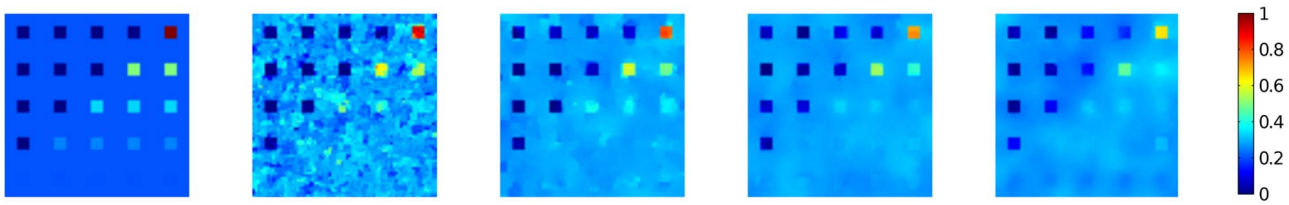

Fig. 3. Influence of spatial regularization parameter η. From left to right: True, $\eta = 0.25$, $\eta = 0.5$ (optimum), $\eta = 0.75$, and $\eta = 1$.

TABLE V
RMSE WITH EIGHT NEIGHBORING PIXELS (SIGNAL-INDEPENDENT NOISE): SEE EXPERIMENTS IN SECTION IV-A4

	IM1		IM2	
	Bilinear	PNMM	Bilinear	PNMM
spatial.-reg. K-Hype (proposed)	0.0509±0.0032	0.0570±0.0073	0.0557±0.0036	0.0916±0.0049
spatial.-reg. NK-Hype (proposed)	0.0568±0.0054	0.0564±0.0067	0.0701±0.0034	0.0858±0.0046

TABLE VI
RMSE COMPARISON (SIGNAL-DEPENDENT NOISE): SEE EXPERIMENTS IN SECTION IV-A5

	IM1		IM2	
	Bilinear	PNMM	Bilinear	PNMM
FCLS	0.1729±0.0091	0.1319±0.0053	0.1679±0.0265	0.1444±0.0096
NCLS	0.1346±0.0134	0.1473±0.0070	0.0790±0.0079	0.1373±0.0134
spatial.-reg. FCLS	0.1729±0.0091	0.1310±0.0053	0.1678±0.0265	0.1382±0.0072
spatial.-reg. NCLS	0.1154±0.0033	0.1474±0.0069	0.0690±0.0057	0.1229±0.0096
K-Hype	0.0784±0.0051	0.0887±0.0070	0.0928±0.0097	0.1194±0.0085
NK-Hype	0.0775±0.0056	0.0878±0.0067	0.0911±0.0080	0.1105±0.0092
spatial.-reg. K-Hype (proposed)	0.0445±0.0016	0.0485±0.0051	0.0517±0.0032	0.0842±0.0043
spatial.-reg. NK-Hype (proposed)	0.0492±0.0023	0.0476±0.0044	0.0640±0.0032	0.0762±0.0042

No significant improvement can be observed in these two cases, but obviously, the proper definition of a neighborhood is closely related to the structure of the images and must be driven by application needs. Fine spatial resolution can greatly improve scene understanding by magnifying subtle details in some cases but may lead to misleading interpretation in some other situations [38].

5) Test With a Signal-Dependent Noise: Unmixing algorithms proposed in the literature have usually been tested on images corrupted by an independent and identical distributed (i.i.d.) additive Gaussian noise. Due to the improved sensitivity of electronic components, this assumption may not be appropriate for data collected with new-generation sensors. Noise modeling and estimation in hyperspectral images has recently become an active subject of research. It is admitted that cameras provide images corrupted by two independent sources of noise, a signal-dependent noise and a signal-independent one [39]–[42]. The former results from the stochastic nature of the photon arrival/detection process. The latter results from sensor electronics and quantization process. We compared the unmixing algorithms on images IM1 and IM2 corrupted by the signal-dependent noise defined as

$$v_{n,\ell} = \tilde{r}_{n,\ell}^{\gamma} v_{n,\ell}^{(1)} + v_{n,\ell}^{(2)}$$

where $\tilde{r}_{n,\ell}$ is the ℓ-th wavelength band of the noise-free reflectance, $v_{n,\ell}^{(1)}$ and $v_{n,\ell}^{(2)}$ are two i.i.d. zero-mean Gaussian noises, and $0 \leq \gamma \leq 1$. Parameter γ was set to $(1/2)$, and the noise variances were set so that $\sigma^2_{v^{(1)}} = \sigma^2_{v^{(2)}}$ and the resulting SNR is 20 dB. Table VI gives the performance of the algorithms. Note that they were not notably affected by this noise setting. The proposed algorithm still exhibits the best performance.

TABLE VII
CLASSIFICATION ACCURACIES AFTER APPLYING SVM TO THREE DIFFERENT TYPES OF FEATURES (FCLS, K-HYPE, AND PROPOSED ALGORITHM)

	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

B. Experiments With AVIRIS Data

This part provides unmixing results for the proposed algorithm when applied on real hyperspectral data. The major difficulty in evaluating the performance of unmixing algorithms is that there are few existing ground-truth references for this purpose. However, in this section, we shall adopt an indirect strategy to circumvent this problem via the unmixing-based classification.

Supervised classification of hyperspectral images is a very challenging but important goal because it generally involves a limited number of training data with unusually high-dimensional patterns. Several feature extraction techniques have been recommended throughout the literature, including principal component analysis (PCA) and independent component analysis (ICA). In [43], the authors explored an alternative strategy consisting of using spectral unmixing for feature extraction, prior to classification. They considered different unmixing-based processes to evaluate the feasibility of this strategy and to perceive the necessity of extracting pure spectral endmembers for classification purposes. The so-called unmixing chain #4 in [43] was found to be the most efficient one. It simply consists of averaging the training samples in each labeled class and uses these spectral signatures to unmix

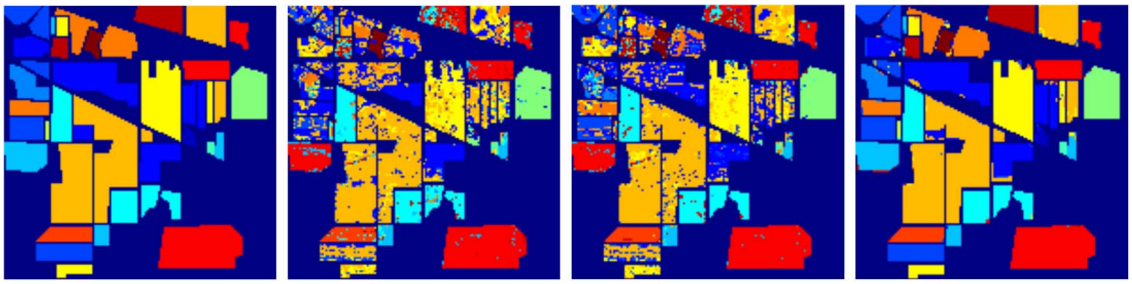

Fig. 4. Indian Pines classification map. From left to right: Ground truth, FCLS (61.36%), K-Hype (71.39%), and proposed algorithm (96.80%).

the original image. The features resulting from the unmixing of training samples are used to train an support vector machine (SVM) classifier. The latter is tested using the remaining labeled samples.

The scene used in our experiment is the well-known data set captured on the Indian Pines region by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). The scene comprises 145 $\times$ 145 samples, consisting of 220 contiguous spectral bands that cover wavelengths ranging from 0.40 to 2.5 μm, with spectral resolution approximately equal to 0.01 μm. Prior to analysis, noisy and water absorption bands were removed, yielding a total of 202 available spectral bands. The ground-truth data contains 16 mutually exclusive classes. The number of pixels in the smallest class is 20, while it is equal to 2468 pixels in the largest class. This widely used benchmark data set is known to be dominated by mixed pixels, even if ground-truth information assigns each pixel a unique class. In this experiment, we used FCLS, K-Hype, and the proposed algorithm with four nearest neighbors for unmixing-based feature extraction. A one-against-all multiclass SVM with Gaussian kernel was applied to these data. We constructed five training sets by randomly selecting 5%, 10%, and 15% of the ground-truth pixels. All the required parameters were optimized by a grid search procedure and fivefold cross-validation. The regularization parameter μ was set to 0.05 for K-Hype and for the proposed algorithm. In addition, for the latter, the spatial regularization parameter η was set to 0.5.

Table VII summarizes the classification accuracies of SVM operating on features extracted with FCLS, K-Hype, and the proposed algorithm. Fig. 4 presents these results in the case of SVM trained with 10% of the available samples per class. It appears that the two nonlinear unmixing algorithms are more effective than the linear one for feature extraction. This clearly means that our nonlinear unmixing model provides less confusing features between the heavily mixed-pixel classes that characterize the Indian Pines benchmark. Finally, we observe that spatial regularization allows to greatly improve the classification accuracy. Spatial homogeneity is a significant prior information for this problem, which allows to substantially improve the quality of the unmixing process.

V. Conclusion

Hyperspectral image unmixing can benefit from both spectral information and spatial information. In this paper, we presented a nonlinear abundance estimation algorithm for the hyperspectral unmixing task. The proposed algorithm integrates the spatial information using ℓ_1-norm regularization into a constrained LS-SVR problem. Split-Bregman iterations were used to solve this optimization problem. Experiments showed the advantage of introducing this spatial regularization into the nonlinear unmixing problem. Spatial correlation within the context of our nonlinear unmixing framework may take various forms. One can promote the similarity of abundance vectors between neighboring pixels, as considered in this paper, or the similarity of the nonlinear fluctuation functions in future works. We may also promote the similarity of the nonlinearity degrees as defined in [12] and derive more localized or adaptive solution strategies to reduce the computational complexity of unmixing algorithms that jointly consider spatial and spectral information.

Appendix

Recent works have raised the question of relaxing the sum-to-one constraint because poor estimates of the endmember signatures or misadjustment of the model may affect the performance of the unmixing process. The interested reader is addressed to [12] for a more detailed discussion on the subject. In this paper, we maintained this constraint for comparison purpose with existing approaches. We shall now provide the main result in the case where this constraint over $\boldsymbol{\alpha}_n$ is relaxed in problem (7). For clarity, let us denote by $\boldsymbol{h}_n$ the nonnormalized vector of abundances. The Lagrange function (11) becomes

$$\mathcal{L}_n = \frac{1}{2}\left(\|\boldsymbol{h}_n\|^2 + \|\psi_n\|_{\mathcal{H}}^2 + \frac{1}{\mu}\sum_{\ell=1}^{L} e_{n,\ell}^2 + \zeta\left\|\boldsymbol{h}_n - \boldsymbol{\xi}_n^{(k)}\right\|^2\right) - \sum_{\ell=1}^{L}\beta_\ell\left(e_{n,\ell} - r_{n,\ell} + \boldsymbol{h}_n^\top \boldsymbol{m}_{\lambda_\ell} + \psi_n(\boldsymbol{m}_{\lambda_\ell})\right) - \sum_{r=1}^{R}\gamma_r h_{n,r}$$

with $\gamma_{n,r} \geq 0$. The conditions for the optimality of $\mathcal{L}_n$ with respect to the primal variables lead us to

$$\begin{cases} \boldsymbol{h}_n^* = \frac{1}{\zeta+1}\left(\sum_{\ell=1}^{L}\beta_{n,\ell}^*\boldsymbol{m}_{\lambda_\ell} + \boldsymbol{\gamma}_n^* + \zeta\boldsymbol{\xi}_n^{(k)}\right) \\ \psi_n^* = \sum_{\ell=1}^{L}\beta_{n,\ell}^*\kappa(\cdot, \boldsymbol{m}_{\lambda_\ell}) \\ e_{n,\ell}^* = \mu\beta_{n,\ell}^*. \end{cases}$$

By substituting these conditions into the primal problem, we get the dual problem (21)

$$\begin{aligned}
&\max_{\boldsymbol{\beta}_n,\boldsymbol{\gamma}_n,\lambda_n} \mathcal{L}'_n(\boldsymbol{\beta}_n,\boldsymbol{\gamma}_n,\lambda_n) \\
&\quad = -\frac{\rho}{2\zeta}\begin{pmatrix}\boldsymbol{\beta}_n\\ \boldsymbol{\gamma}_n\end{pmatrix}^\top \left(\begin{array}{c|c}\boldsymbol{K}_\psi & \boldsymbol{M}\\ \hline \boldsymbol{M}^\top & \mathbf{I}\end{array}\right)\begin{pmatrix}\boldsymbol{\beta}_n\\ \boldsymbol{\gamma}_n\end{pmatrix} \\
&\qquad + \begin{pmatrix}\boldsymbol{r}_n - \rho\boldsymbol{M}\boldsymbol{\xi}_n^{(k)}\\ -\rho\boldsymbol{\xi}_n^{(k)}\end{pmatrix}^\top \begin{pmatrix}\boldsymbol{\beta}_n\\ \boldsymbol{\gamma}_n\end{pmatrix} \\
&\text{subject to} \quad \boldsymbol{\gamma}_n \succeq \mathbf{0} \\
&\qquad \text{with} \quad \boldsymbol{K}_\psi = \frac{1}{\zeta}(\boldsymbol{K}+\mu\mathbf{I}) + \boldsymbol{M}\boldsymbol{M}^\top \quad \text{and} \quad \rho = \frac{\zeta}{1+\zeta}.
\end{aligned} \tag{21}$$

Provided that the optimal dual variables $\boldsymbol{\beta}_n^*$ and $\boldsymbol{\gamma}_n^*$ have been determined, the solution $\boldsymbol{h}_n^*$ is given by

$$\boldsymbol{h}_n^* = \frac{1}{\zeta+1}\left(\boldsymbol{M}^\top\boldsymbol{\beta}_n^* + \boldsymbol{\gamma}_n^* + \zeta\boldsymbol{\xi}_n^{(k)}\right).$$

If necessary, note that $\boldsymbol{h}_n^*$ can be normalized afterward by writing $\boldsymbol{h}_n^* = \theta_n^*\,\boldsymbol{\alpha}_n^*$, with $\boldsymbol{\alpha}_n^*$ being the vector of what would correspond to fractional abundances and $\theta_n^* = \mathbf{1}^\top\boldsymbol{h}_n^*$ being the scaling factor. Following [12], i.e., using equivalence between optimization problems as explained in [44, p. 130], it can be shown that θ_n^* and $\boldsymbol{\alpha}_n^*$ are the solutions of

$$\begin{aligned}
\boldsymbol{\alpha}_n^*, \theta_n^*, \psi_n^* = \underset{\boldsymbol{\alpha}_n,\theta_n,\psi_n,\boldsymbol{e}_n}{\arg\min}\ & \frac{1}{2}\Bigg(\|\theta_n\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|_{\mathcal{H}}^2 + \frac{1}{\mu}\sum_{\ell=1}^{L} e_{n,\ell}^2 \\
&\qquad + \zeta\left\|\theta_n\boldsymbol{\alpha}_n - \boldsymbol{\xi}_n^{(k)}\right\|^2\Bigg) \\
\text{subject to}\quad & e_{n,\ell} = r_{n,\ell} - \left(\theta_n\boldsymbol{\alpha}_n^\top\boldsymbol{m}_{\lambda_\ell} + \psi_n(\boldsymbol{m}_{\lambda_\ell})\right) \\
& \boldsymbol{\alpha}_n \succeq 0 \qquad \boldsymbol{\alpha}_n^\top\mathbf{1}_R = 1 \qquad \theta_n > 0.
\end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank Prof. Bioucas-Dias and his colleagues for their generous assistance in providing most of the data used in this paper.

REFERENCES

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

[2] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," in *Proc. SPIE*, 2009, vol. 7477, p. 74 770I.

[3] N. Raksuntorn and Q. Du, "Nonlinear spectral mixture analysis for hyperspectral imagery in an unknown environment," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 836–840, Oct. 2010.

[4] 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.

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

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

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

[8] F. Bovolo, L. Bruzzone, and L. Carlin, "A novel technique for subpixel image classification based on support vector machine," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2983–2999, Nov. 2010.

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

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

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

[12] 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.

[13] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral images with multi-kernel learning," in *Proc. IEEE WHISPERS*, 2012, pp. 1–12.

[14] 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*, 2013, pp. 2174–2178.

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

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

[17] D. M. Rogge, B. Rivard, J. Zhang, A. Sanchez, J. Harris, and J. Feng, "Integration of spatial-spectral information for the improved extraction of endmembers," *Remote Sens. Environ.*, vol. 110, no. 3, pp. 287–303, Oct. 2007.

[18] M. Zortea and A. Plaza, "Spatial preprocessing for endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2679–2693, Aug. 2009.

[19] 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.

[20] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[21] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.

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

[23] 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.

[24] 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.

[25] 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.

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

[27] T. A. Warner and M. C. Shank, "Spatial autocorrelation analysis of hyperspectral imagery for feature selection," *Remote Sens. Environ.*, vol. 60, no. 1, pp. 58–70, Apr. 1997.

[28] Q. Du, L. Zhang, and N. Raksuntorn, "Improving the quality of extracted endmembers," in *Proc. WHISPERS*, 2009, pp. 1–4.

[29] B. Schölkopf, J. C. Burges, and A. J. Smola, *Advances in Kernel Methods*. Cambridge, MA, USA: MIT Press, 1999.

[30] K. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–201, Mar. 2001.

[31] F. Abdallah, C. Richard, and R. Lengellé, "An improved training algorithm for nonlinear kernel discriminants," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2798–2806, Oct. 2004.

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

[33] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
[34] J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1–3, pp. 293–318, Apr. 1992.
[35] A. M. Baldridge, S. J. Hook, C. I. Grove, and G. Rivera, "The ASTER spectral library version 2.0," *Remote Sens. Environ.*, vol. 113, no. 4, pp. 711–715, Apr. 2009.
[36] D. C. Heinz and C.-I. Chang, "Fully constrained least squares linear mixture analysis for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.
[37] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[38] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. Suppl. 1, pp. S110–S122, Sep. 2009.
[39] N. Acito, M. Diani, and G. Corsini, "Signal-dependent noise modeling and model parameter estimation in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 2957–2971, Aug. 2011.
[40] J. Meola, M.-T. Eismann, R.-L. Moses, and J.-N. Ash, "Modeling and estimation of signal-dependent noise in hyperspectral imagery," *Appl. Opt.*, vol. 50, no. 21, pp. 3829–3846, Jul. 2011.
[41] L. Alparone, M. Selva, B. Aiazzi, S. Baronti, F. Butera, and L. Chiarantini, "Signal-dependent noise modeling and estimation of new-generation imaging spectrometers," in *Proc. IEEE WHISPERS*, 2009, pp. 1–4.
[42] B. Aiazzi, L. Alparone, S. Baronti, A. Barducci, P. Marcoionni, I. Pippi, and M. Selva, "Noise modeling and estimation of hyperspectral data from airborne imaging spectrometer," *Ann. Geophys.*, vol. 49, no. 1, pp. 1–9, Feb. 2006.
[43] 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.
[44] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

Jie Chen (S'12) was born in Xi'an, China, in 1984. He received the B.S. degree in information and telecommunication engineering in 2006 from the Xi'an Jiaotong University, Xi'an, and the Dipl.Ing. and the M.S. degrees in information and telecommunication engineering in 2009 from the University of Technology of Troyes (UTT), Troyes, France, and from the Xi'an Jiaotong University, respectively. In 2013, he received the Ph.D. degree in systems optimization and security from the UTT.

He is currently a Postdoctoral Researcher with the Côte d'Azur Observatory, University of Nice Sophia-Antipolis, Nice, France. His current research interests include adaptive signal processing, kernel methods, hyperspectral image analysis, and supervised and unsupervised learning.

Cédric Richard (S'98–M'01–SM'07) was born on January 24, 1970 in Sarrebourg, France. He received the Dipl.Ing. and M.S. degrees in electrical and computer engineering and the Ph.D. degree in electrical and computer engineering from the University of Technology of Compiègne (UTC), Compiègne, France, in 1994 and 1998, respectively.

He joined the Côte d'Azur Observatory, University of Nice Sophia-Antipolis, Nice, France, in 2009, where he is currently a Professor of electrical engineering. From 1999 to 2003, he was an Associate Professor with the University of Technology of Troyes (UTT), Troyes, France. From 2003 to 2009, he was a Professor with the UTT and the supervisor of a group consisting of 60 researchers and Ph.D. students. In winter 2009 and autumns 2010 and 2011, he was a Visiting Researcher with the Department of Electrical Engineering, Federal University of Santa Catarina (UFSC), Florianòpolis, Brazil. He is the author of over 200 papers. His current research interests include statistical signal processing and machine learning.

Dr. Richard served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2006 to 2010 and of Elsevier Signal Processing since 2009. In 2009, he was nominated liaison local officer for The European Association for signal Processing (EURASIP) and member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society. In 2013, has was also elected member of the Machine Learning for Signal Processing Technical Committee of this society. He is a junior member of the Institut Universitaire de France since October 2010. He was the General Chair of the XXIth francophone conference Groupement de Recherche en Traitement du Signal et des Images (GRETSI) on Signal and Image Processing that was held in Troyes in 2007 and of the IEEE Statistical Signal Processing Workshop that was held in Nice in 2011. Since 2005, he has been a member of the board of the federative CNRS research group Information, Signal, Images et ViSion (ISIS) on Information, Signal, Images and Vision. He is a member of the GRETSI association board and of the EURASIP society. He received, with Paul Honeine, the Best Paper Award for "Solving the pre-image problem in kernel machines: a direct method" at the 2009 IEEE Workshop on Machine Learning for Signal Processing.

Paul Honeine (M'07) was born in Beirut, Lebanon, on October 2, 1977. He received the Dipl.Ing. degree in mechanical engineering and the M.Sc. degree in industrial control from the Faculty of Engineering, Lebanese University, Beirut, Lebanon, in 2002 and 2003, respectively. In 2007, he received the Ph.D. degree in systems optimization and security from the University of Technology of Troyes, Troyes, France, where he was also a Postdoctoral Research Associate with the Systems Modeling and Dependability Laboratory from 2007 to 2008.

Since 2008, he has been an Assistant Professor with the University of Technology of Troyes. His research interests include nonstationary signal analysis and classification, nonlinear signal processing, sparse representations, machine learning, and wireless sensor networks.

Dr. Honeine received, with Cédric Richard, the Best Paper Award for "Solving the pre-image problem in kernel machines: a direct method" at the 2009 IEEE Workshop on Machine Learning for Signal Processing.

Nonlinear Estimation of Material Abundances in Hyperspectral Images With ℓ_1-Norm Spatial Regularization

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

***Abstract*—Integrating spatial information into hyperspectral unmixing procedures has been shown to have a positive effect on the estimation of fractional abundances due to the inherent spatial–spectral duality in hyperspectral scenes. However, current research works that take spatial information into account are mainly focused on the linear mixing model. In this paper, we investigate how to incorporate spatial correlation into a nonlinear abundance estimation process. A nonlinear unmixing algorithm operating in reproducing kernel Hilbert spaces, coupled with a ℓ_1-type spatial regularization, is derived. Experiment results, with both synthetic and real hyperspectral images, illustrate the effectiveness of the proposed scheme.**

***Index Terms*—Hyperspectral imaging, ℓ_1-norm regularization, nonlinear spectral unmixing, spatial regularization.**

I. Introduction

HYPERSPECTRAL imaging provides 2-D spatial images over many contiguous bands. The high spectral resolution allows to identify and quantify distinct materials from remotely observed data. This area has received considerable attention in the last decade. Applications include land use analysis, mineral detection, environment monitoring, field surveillance, etc. Typically, the observed reflectance at each pixel is a spectral mixture of several material signatures, termed endmembers, due to the limited spatial resolution of certain observation devices and diversity of materials. For this reason, spectral unmixing is an important issue for hyperspectral data processing to investigate information in each pixel [1].

There have been significant efforts during the past decade to solve linear unmixing problems, and several methods have been successfully derived within the context of hyperspectral imaging. Nevertheless, linear models can only accurately capture simple interactions between elements, e.g., in situations where the mixing of materials is not intimate and multiple scattering is negligible [1]. Recently, several researchers have started aggressively exploring nonlinear unmixing techniques. In [2] and [3], the authors extended the collection of endmembers by explicitly adding nonlinear artificial cross-terms of pure spectral signatures and used a conventional linear unmixing algorithm over this new dictionary. In [4], a Bayesian inference algorithm dedicated to the generalized bilinear mixture model was proposed. In [5], the same strategy was applied to estimate the fractional abundances of a polynomial post nonlinear mixing model. Manifold learning techniques were investigated in [6] and [7], and unmixing algorithms using geodesic distances or manifold-related regularization terms were proposed. Moreover, algorithms operating in reproducing kernel Hilbert spaces were considered to process hyperspectral data. In [8], a fuzzy-input fuzzy-output support vector machine was designed for subpixel image classification. Kernel-based nonlinear unmixing approaches were investigated in [9] and [10]. These algorithms were mainly obtained by replacing each inner product between endmember spectra, in the cost functions to be optimized, by a kernel function. This can be viewed as a nonlinear distortion map applied to the spectral signature of each material, which is of little physical interest in solving the unmixing problem because the nonlinear nature of mixing may also be characterized by nonlinear interactions of the materials. Physically inspired kernel-based models were introduced in [11] to circumvent this drawback, where we modeled each mixed pixel by a linear mixture of endmember spectra coupled with an additive nonlinear interaction term. The latter was used to model nonlinear effects of photon interactions and was defined in a reproducing kernel Hilbert space. In [12]–[14], a more complete and sophisticated theory and new methods were derived to automatically adjust the balance between linear and nonlinear components. A postnonlinear mixing model was also introduced in [15].

Although several linear and nonlinear state-of-the-art unmixing techniques have shown interesting performance for this problem, they have been mostly focused on exploiting spectral information available in the hyperspectral space. These approaches consist of considering pixel vectors as if they were independent from their neighboring pixels. However, one of the distinguishing properties of remotely sensed data is that they convey multivariate information into a 2-D pictorial representation [16]. Subsequently, instead of disregarding spatial contextual information, hyperspectral analysis techniques should benefit from the inherent spatial–spectral duality in

Manuscript received June 10, 2012; revised December 6, 2012 and March 11, 2013; accepted May 10, 2013. This work was supported by the Agence Nationale pour la Recherche, France, (Hypanema project, ANR-12-BS03-003).

J. Chen and C. Richard are with the Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur, 06108 Nice, France (e-mail: jie.chen@unice.fr; cedric.richard@unice.fr).

P. Honeine is with the Institut Charles Delaunay, Centre National de la Recherche Scientifique (CNRS), Université de Technologie de Troyes, 10010 Troyes Cedex, France (e-mail: paul.honeine@utt.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2264392

hyperspectral scenes. Following this idea, researchers have attempted to exploit spatial information in hyperspectral image analysis. Spatial preprocessing techniques were investigated for endmember determination in [17]–[19]. Spatial correlation was incorporated into hyperspectral image classification algorithms in [20] and [21]. Concerning the unmixing problem, a nonnegative matrix factorization-type problem regularized by the ℓ_1-norm of differences between neighboring pixels was proposed in [22]. A projected subgradient method was used to solve this problem. In [23], a Markov random field was proposed to model the spatial dependence of the pixels within classes. Bayesian inference was then used to estimate the model parameters. In [24], total variation was employed for spatial regularization to enhance the unmixing performance. The alternating direction method of multipliers was used to solve the regularized problem. Some other works also showed that incorporating spatial information can have a positive impact on the hyperspectral unmixing process [25], [26].

To the best of our knowledge, spatial regularization has not yet been extensively studied in a nonlinear unmixing process. As nonlinear unmixing is itself an important but challenging issue, it appears difficult to address these two problems simultaneously. A new nonlinear model was proposed in our work [12], where we assumed that a nonlinear mixture can be decomposed into a linear trend and an additive nonlinear fluctuation term in a reproducing kernel Hilbert space to model nonlinear effects. Based on this promising advance within the area of nonlinear unmixing, in this paper, we take spatial information into account in the unmixing process using ℓ_1-norm spatial regularization. An optimization method based on split-Bregman iterations is proposed to deal with this problem that suffers from the nonlinearity of the model and the nonsmoothness of the regularization term. Experiments with both synthetic and real data are conducted to validate the proposed method.

II. Formulation of the Problem

A. Notations

Suppose that the hyperspectral image under study has 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 facilitate the presentation, we transform this 3-D image into a $L \times N$ matrix, with $N = w \times h$ the total number of pixels. Then, let

- $n \in \{1, \ldots, N\}$ be the sequential index of pixels in the image.
- $\boldsymbol{r}_n = [r_{n1}, r_{n2}, \ldots, r_{nL}]^\top$ be the $(L \times 1)$ observed reflectance vector for the pixel n, which consists of a mixture of R endmember spectra.
- $\boldsymbol{M} = [\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_R]$ be the $(L \times R)$ endmember target matrix, where each column $\boldsymbol{m}_i$ is an endmember spectral signature. For expository convenience, we denote by $\boldsymbol{m}_{\lambda_\ell}^\top$ the ℓ-th row of $\boldsymbol{M}$, namely, the vector of the R endmember signatures at the ℓ-th wavelength band.
- $\boldsymbol{\alpha}_n = [\alpha_{n1}, \alpha_{n2}, \ldots, \alpha_{nR}]^\top$ be the $(R \times 1)$ abundance vector associated to the pixel n.
- $\boldsymbol{A} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N]$ be the matrix composed of all the abundance vectors.
- $\mathbf{1}_N$ be the $(N \times 1)$ all-one vector and $\mathbf{I}_N$ be the $(N \times N)$ identity matrix.

Other symbols will be defined in the context where they are used.

B. Formulation of the Nonlinear Unmixing Problem With Spatial Regularization

Suppose that the materials in a scene have been determined by some endmember extraction algorithm. The unmixing problem boils down to estimating the abundance vector associated to each pixel. One way to exploit spatial information is to define an appropriate criterion to be optimized, e.g., by considering extra penalty terms to promote the similarity of fractional abundances between neighboring pixels. The rationale is that homogeneous regions within which correlation among neighboring pixels is high potentially exist in real images. This suggests that increasing spatial homogeneity should tend to increase the accuracy of the representation of spectral objects and to suppress high-spatial-frequency artifacts [27], although there are risks to remove small features.

Taking the spatial relationship among pixels into consideration, the unmixing problem can then be solved by minimizing a general cost function with respect to the matrix $\boldsymbol{A}$

$$J(\boldsymbol{A}) = J_{\mathrm{err}}(\boldsymbol{A}) + \eta J_{\mathrm{sp}}(\boldsymbol{A}) \tag{1}$$

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

$$\boldsymbol{A} \succeq \mathbf{0}$$
$$\boldsymbol{A}^\top \mathbf{1}_R = \mathbf{1}_N.$$

Recent works have raised the question of relaxing the sum-to-one constraint. Indeed, poor estimates of the endmember signatures may affect the performance of the unmixing process. This constraint is maintained in the following as a baseline for comparison with existing approaches. However, in Appendix, we also briefly consider the case where the sum-to-one constraint is removed from the model. In the general expression (1), the function J_{err} represents the modeling error, and J_{sp} is a regularization term to promote the similarity of the fractional abundances within neighboring pixels. The nonnegative parameter η controls the tradeoff between data fidelity and pixel similarity. In [22] for instance, the ℓ_1-norm of the differences of abundance vectors in the neighborhood of each pixel was used as the spatial regularizer. In [24], anisotropic total variation norm was considered for linear sparse unmixing. In [26], a so-called fuzzy local information proportion was used for incorporating spatial information. In [21] and [23], a Markov random field was proposed to model spatial dependence. All these approaches have shown a substantial advantage of using spatial information for hyperspectral data unmixing, although there are risks to remove small and significant features. Preprocessing can be conducted to alleviate such hazards by separating these features in advance [28].

Rarely, if ever, have nonlinear mixing models incorporating spatial information been considered in the literature. In this paper, we intend to formulate J_{err} as in our recent work [12] because this approach has shown excellent performance and low computational cost. For self-containedness, let us briefly review part of this work. Consider the general unmixing process, acting between the entries $r_{n,\ell}$ of the 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 $\psi_{\boldsymbol{\alpha}_n}$ being an unknown nonlinear function to be estimated that defines the interaction between the endmember spectra, in the proportion $\boldsymbol{\alpha}_n$, and with $\boldsymbol{e}_n$ being the estimation error. This requires us to consider a general problem

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

with μ being a positive parameter that controls the tradeoff 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 [12], we defined them by a linear trend parameterized by the abundance vector $\boldsymbol{\alpha}_n$, combined with a nonlinear fluctuation function ψ_n, namely

$$\psi_{\boldsymbol{\alpha}_n}(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{\alpha}_n^\top \boldsymbol{m}_{\lambda_\ell} + \psi_n(\boldsymbol{m}_{\lambda_\ell}) \tag{3}$$

where ψ_n can be any real-valued function of 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 the efficient and accurate resolution of inverse problems of form (2) by exploiting the central idea of this research area, known as the *kernel trick* [29]. This trick has been widely used for solving nonlinear regression and classification problems [30], [31]. We proposed in [12] to conduct data unmixing (2) and (3) by solving the following least square support vector regression (LS-SVR) problem:

$$\begin{aligned} \boldsymbol{\alpha}^*_n, \psi^*_n = \arg\min_{\boldsymbol{\alpha}_n, \psi_n} \frac{1}{2}\left(\|\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|^2_{\mathcal{H}} + \frac{1}{\mu}\|\boldsymbol{e}_n\|^2\right) \\ \text{subject to} \quad \boldsymbol{\alpha}_n \succeq \mathbf{0} \quad \text{and} \quad \mathbf{1}^\top \boldsymbol{\alpha}_n = 1 \end{aligned} \tag{4}$$

where $\boldsymbol{e}_n$ is the $(L \times 1)$ misadjustment error vector with the 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 (2). It can be shown that problem (4) is convex so that it can be solved exactly by the duality theory. This so-called K-Hype method was introduced in [12]. A very efficient extension, called SK-Hype, consisting of relaxing the sum-to-one constraint and automatically adjusting the balance between $\boldsymbol{\alpha}_n^\top \boldsymbol{m}_{\lambda_\ell}$ and $\psi_n(\boldsymbol{m}_{\lambda_\ell})$ via an appropriate parameterization was also presented in this paper. In the following, we restrict our attention to K-Hype, and we maintain the sum-to-one constraint as a baseline for comparison with existing approaches. However, in Appendix, we also briefly address the case where the sum-to-one constraint is removed from the model. Finally, considering all the pixels of the image to process, the modeling error to be optimized writes

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

subject to the nonnegativity and sum-to-one constraints over the abundance vectors. 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 into account the spatial correlation between pixels, we shall use ℓ_1-type regularizers of the form [22], [24] to promote piecewise constant transitions in the fractional abundance of each endmember among neighboring pixels. The regularization function is expressed as

$$J_{\text{sp}}(\boldsymbol{A}) = \sum_{n=1}^{N}\sum_{m \in \mathcal{N}(n)} \|\boldsymbol{\alpha}_n - \boldsymbol{\alpha}_m\|_1 \tag{5}$$

where $\|\ \|_1$ denotes the vector ℓ_1-norm and $\mathcal{N}(n)$ denotes the set of neighbors of the pixel n. Without any loss of generality, in this paper, we define the neighborhood of the pixel n by taking the four nearest pixels $n-1$ and $n+1$ (row adjacency) and $n-w$ 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- and right-hand neighbors, 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 (5) can be rewritten in matrix form as

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

with $\boldsymbol{H}$ being the $(N \times 4N)$ matrix $(\boldsymbol{H}_\leftarrow\ \boldsymbol{H}_\rightarrow\ \boldsymbol{H}_\uparrow\ \boldsymbol{H}_\downarrow)$ and $\|\ \|_{1,1}$ being the sum of the ℓ_1-norms of the columns of a matrix. Obviously, other boundaries for the neighborhood $\mathcal{N}(n)$ may be used by simply defining the appropriate matrix $\boldsymbol{H}$. Finally, note that this regularization function is convex but nonsmooth.

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

$$\begin{aligned} \boldsymbol{A}^*, \boldsymbol{\psi}^* = \arg\min_{\boldsymbol{A}, \boldsymbol{\psi}} \sum_{n=1}^{N}\frac{1}{2}\left(\|\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|^2_{\mathcal{H}} + \frac{1}{\mu}\|\boldsymbol{e}_n\|^2\right) \\ + \eta\|\boldsymbol{A}\boldsymbol{H}\|_{1,1} \\ \text{subject to} \quad \boldsymbol{A} \succeq 0 \quad \text{and} \quad \boldsymbol{A}^\top \mathbf{1}_R = \mathbf{1}_N \end{aligned} \tag{6}$$

where η controls the tradeoff between model fitting in each pixel and similarity among neighboring pixels. The constraints over $\boldsymbol{A}$ define a convex set $\mathcal{S}_A$. For ease of exposition, in the formulation of optimization problems, we will write $\boldsymbol{A} \in \mathcal{S}_A$.

III. Solving the Problem

Although the optimization problem (6) is convex, it cannot be solved easily because it combines an LS-SVR problem with a huge-dimensional nonsmooth regularization term. In order

to overcome this drawback, we rewrite (6) in the following equivalent form:

$$\min_{\boldsymbol{A}\in\mathcal{S}_A,\boldsymbol{\psi}} \quad \sum_{n=1}^{N}\frac{1}{2}\left(\|\boldsymbol{\alpha}_n\|^2+\|\psi_n\|_{\mathcal{H}}^2+\frac{1}{\mu}\|\boldsymbol{e}_n\|^2\right)+\eta\|\boldsymbol{U}\|_{1,1}$$
$$\text{subject to} \quad \boldsymbol{V}=\boldsymbol{A} \quad \text{and} \quad \boldsymbol{U}=\boldsymbol{V}\boldsymbol{H} \tag{7}$$

where two new matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, and two additional constraints, have been introduced. This variable-splitting approach was initially introduced in [32]. Matrix $\boldsymbol{U}$ will allow us to decouple the nonsmooth ℓ_1-norm regularization functional from the constrained LS-SVR problem. Matrix $\boldsymbol{V}$ will make the LS-SVR problem tractable by relaxing connections between pixels.

As studied in [32], 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 (6), the following formulation is obtained:

$$\begin{aligned}&\boldsymbol{A}^{(k+1)},\boldsymbol{\psi}^{(k+1)},\boldsymbol{V}^{(k+1)},\boldsymbol{U}^{(k+1)}\\&=\underset{\boldsymbol{A}\in\mathcal{S}_A,\boldsymbol{\psi},\boldsymbol{V},\boldsymbol{U}}{\arg\min}\sum_{n=1}^{N}\frac{1}{2}\left(\|\boldsymbol{\alpha}_n\|^2+\|\psi_n\|_{\mathcal{H}}^2+\frac{1}{\mu}\|\boldsymbol{e}_n\|^2\right)\\&\quad+\eta\|\boldsymbol{U}\|_{1,1}+\frac{\zeta}{2}\left\|\boldsymbol{A}-\boldsymbol{V}-\boldsymbol{D}_1^{(k)}\right\|_F^2\\&\quad+\frac{\zeta}{2}\left\|\boldsymbol{U}-\boldsymbol{V}\boldsymbol{H}-\boldsymbol{D}_2^{(k)}\right\|_F^2\end{aligned} \tag{8}$$

with

$$\begin{aligned}\boldsymbol{D}_1^{(k+1)}&=\boldsymbol{D}_1^{(k)}+\left(\boldsymbol{V}^{(k+1)}-\boldsymbol{A}^{(k+1)}\right)\\\boldsymbol{D}_2^{(k+1)}&=\boldsymbol{D}_2^{(k)}+\left(\boldsymbol{V}^{(k+1)}\boldsymbol{H}-\boldsymbol{U}^{(k+1)}\right)\end{aligned} \tag{9}$$

where $\|\ \|_F^2$ denotes the matrix Frobenius norm and ζ is a positive parameter. Because of the way that we have split the components 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 that we have to perform are as follows.

1) Step 1—Optimization With Respect to $\boldsymbol{A}$ and $\boldsymbol{\psi}$: The optimization problem (8) reduces to

$$\begin{aligned}\boldsymbol{A}^{(k+1)},\boldsymbol{\psi}^{(k+1)}&=\underset{\boldsymbol{A}\in\mathcal{S}_A,\boldsymbol{\psi}}{\arg\min}\sum_{n=1}^{N}\frac{1}{2}\\&\quad\times\left(\|\boldsymbol{\alpha}_n\|^2+\|\psi_n\|_{\mathcal{H}}^2+\frac{1}{\mu}\|\boldsymbol{e}_n\|^2+\zeta\left\|\boldsymbol{\alpha}_n-\boldsymbol{\xi}_n^{(k)}\right\|^2\right)\end{aligned}$$

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 decomposed into subproblems, each one involving an abundance vector $\boldsymbol{\alpha}_n$. This results from the use of the matrix $\boldsymbol{V}$ in the split iteration algorithm (8).

Let us now solve the local optimization problem

$$\begin{aligned}\boldsymbol{\alpha}_n^{(k+1)},\psi_n^{(k+1)}=\underset{\boldsymbol{\alpha}_n,\psi_n,\boldsymbol{e}_n}{\arg\min}\ &\frac{1}{2}\Bigg(\|\boldsymbol{\alpha}_n\|^2+\|\psi_n\|_{\mathcal{H}}^2+\frac{1}{\mu}\sum_{\ell=1}^{L}e_{n,\ell}^2\\&\qquad+\zeta\left\|\boldsymbol{\alpha}_n-\boldsymbol{\xi}_n^{(k)}\right\|^2\Bigg)\\\text{subject to}\quad & e_{n,\ell}=r_{n,\ell}-\left(\boldsymbol{\alpha}_n^\top\boldsymbol{m}_{\lambda_\ell}+\psi_n(\boldsymbol{m}_{\lambda_\ell})\right)\\&\boldsymbol{\alpha}_n\succeq 0\\&\boldsymbol{\alpha}_n^\top\mathbf{1}_R=1.\end{aligned} \tag{10}$$

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 to (10) is written as

$$\begin{aligned}\mathcal{L}_n=&\frac{1}{2}\left(\|\boldsymbol{\alpha}_n\|^2+\|\psi_n\|_{\mathcal{H}}^2+\frac{1}{\mu}\sum_{\ell=1}^{L}e_{n,\ell}^2+\zeta\left\|\boldsymbol{\alpha}_n-\boldsymbol{\xi}_n^{(k)}\right\|^2\right)\\&-\sum_{\ell=1}^{L}\beta_\ell\left(e_{n,\ell}-r_{n,\ell}+\boldsymbol{\alpha}_n^\top\boldsymbol{m}_{\lambda_\ell}+\psi_n(\boldsymbol{m}_{\lambda_\ell})\right)\\&-\sum_{r=1}^{R}\gamma_r\alpha_{n,r}+\lambda_n\left(\boldsymbol{\alpha}_n^\top\mathbf{1}_R-1\right)\end{aligned} \tag{11}$$

with $\gamma_{n,r}\geq 0$. The conditions for the optimality of $\mathcal{L}_n$ with respect to the primal variables lead us to

$$\begin{cases}\boldsymbol{\alpha}_n^*=\frac{1}{\zeta+1}\left(\sum_{\ell=1}^{L}\beta_{n,\ell}^*\boldsymbol{m}_{\lambda_\ell}+\boldsymbol{\gamma}_n^*-\lambda_n^*\mathbf{1}+\zeta\boldsymbol{\xi}_n^{(k)}\right)\\\psi_n^*=\sum_{\ell=1}^{L}\beta_{n,\ell}^*\kappa(\cdot,\boldsymbol{m}_{\lambda_\ell})\\e_{n,\ell}^*=\mu\beta_{n,\ell}^*\end{cases} \tag{12}$$

where κ is the reproducing kernel of $\mathcal{H}$. By substituting (12) into (11), we get the dual problem

$$\begin{aligned}&\max_{\boldsymbol{\beta}_n,\boldsymbol{\gamma}_n,\lambda_n}\mathcal{L}_n'(\boldsymbol{\beta}_n,\boldsymbol{\gamma}_n,\lambda_n)\\&=-\frac{\rho}{2\zeta}\begin{pmatrix}\boldsymbol{\beta}_n\\\hline\boldsymbol{\gamma}_n\\\hline\lambda_n\end{pmatrix}^\top\left(\begin{array}{c|c|c}\boldsymbol{K}_\psi&\boldsymbol{M}&-\boldsymbol{M}\mathbf{1}_R\\\hline\boldsymbol{M}^\top&\mathbf{I}&-\mathbf{1}_R\\\hline-\mathbf{1}_R^\top\boldsymbol{M}^\top&-\mathbf{1}_R^\top&R\end{array}\right)\\&\quad\times\begin{pmatrix}\boldsymbol{\beta}_n\\\hline\boldsymbol{\gamma}_n\\\hline\lambda_n\end{pmatrix}+\begin{pmatrix}\boldsymbol{r}_n-\rho\boldsymbol{M}\boldsymbol{\xi}_n^{(k)}\\\hline-\rho\boldsymbol{\xi}_n^{(k)}\\\hline\rho\boldsymbol{\xi}_n^{(k)\top}\mathbf{1}_R-1\end{pmatrix}^\top\begin{pmatrix}\boldsymbol{\beta}_n\\\hline\boldsymbol{\gamma}_n\\\hline\lambda_n\end{pmatrix}\\&\text{subject to}\quad\boldsymbol{\gamma}_n\succeq\mathbf{0}\\&\text{with}\quad\boldsymbol{K}_\psi=\frac{1}{\zeta}(\boldsymbol{K}+\mu\mathbf{I})+\boldsymbol{M}\boldsymbol{M}^\top\quad\text{and}\quad\rho=\frac{\zeta}{1+\zeta}\end{aligned} \tag{13}$$

where $\boldsymbol{K}$ is the Gram matrix defined as $[\boldsymbol{K}]_{\ell p}=\kappa(\boldsymbol{m}_{\lambda_\ell},\boldsymbol{m}_{\lambda_p})$. The problem (13) is a convex quadratic programming problem with respect to the dual variables.

TABLE I
NONLINEAR UNMIXING WITH SPATIAL REGULARIZATION ALGORITHM

Initialization
Choose the kernel κ_{nlin} and calculate the kernel matrix $\boldsymbol{K}_{\text{nlin}}$.
Choose the regularization parameters μ and η in problem (6).
Initialize $\boldsymbol{U}^{(0)}$, $\boldsymbol{V}^{(0)}$, $\boldsymbol{D}_1^{(0)}$ and $\boldsymbol{D}_2^{(0)}$.
Repeat
Calculate primal variables:
- Calculate $\boldsymbol{A}^{(k+1)}$ by solving the QP problem (13) for each pixel.
- Calculate $\boldsymbol{V}^{(k+1)}$ using (15), and $\boldsymbol{U}^{(k+1)}$ using (17).

Update $\boldsymbol{D}_1^{(k+1)}$ and $\boldsymbol{D}_2^{(k+1)}$ using (9).
until stopping criterion is satisfied
Estimate Use the final $\boldsymbol{A}^{(k+1)}$ as the fractional abundances.

Finally, provided that the optimal dual variables $\boldsymbol{\beta}_n^*$, $\boldsymbol{\gamma}_n^*$, and λ_n^* have been determined, the vector of fractional abundances $\boldsymbol{\alpha}_n^{(k+1)}$ is estimated by

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

This process has to be repeated for $n = 1, \ldots, N$ in order to get $\boldsymbol{A}^{(k+1)}$. Matrices $\boldsymbol{A}^{(k+1)}$ and $\boldsymbol{V}^{(k+1)}$, whose calculation is presented hereafter, allow to evaluate $\boldsymbol{D}_1^{(k+1)}$ using (9).

2) Step 2—Optimization With Respect to $\boldsymbol{V}$: The optimization problem (8) now reduces to

$$\boldsymbol{V}^{(k+1)} = \arg\min_{\boldsymbol{V}} \left\|\boldsymbol{A}^{(k+1)} - \boldsymbol{V} - \boldsymbol{D}_1^{(k)}\right\|_F^2 + \left\|\boldsymbol{U}^{(k)} - \boldsymbol{V}\boldsymbol{H} - \boldsymbol{D}_2^{(k)}\right\|_F^2. \quad (14)$$

Equating to zero the derivative of the above expression with respect to $\boldsymbol{V}$ leads to

$$\left(\boldsymbol{A}^{(k+1)} - \boldsymbol{V} - \boldsymbol{D}_1^{(k)}\right) - \left(\boldsymbol{U}^{(k)} - \boldsymbol{V}\boldsymbol{H} - \boldsymbol{D}_2^{(k)}\right)\boldsymbol{H}^\top = 0$$

whose solution is then given by

$$\boldsymbol{V}^{(k+1)} = \left(\boldsymbol{A}^{(k+1)} - \boldsymbol{D}_1^{(k)} + \left(\boldsymbol{U}^{(k)} - \boldsymbol{D}_2^{(k)}\right)\boldsymbol{H}^\top\right)(\mathbf{I} + \boldsymbol{H}\boldsymbol{H}^\top)^{-1}. \quad (15)$$

3) Step 3—Optimization With Respect to $\boldsymbol{U}$: Finally, the optimization problem that we have to consider is as follows:

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

Its solution can be expressed via the well-known soft threshold function

$$\boldsymbol{U}^{(k+1)} = \text{Thresh}\left(\boldsymbol{V}^{(k+1)}\boldsymbol{H} + \boldsymbol{D}_2^{(k)}, \frac{\eta}{\zeta}\right) \quad (17)$$

where $\text{Thresh}(\cdot, \tau)$ denotes the componentwise application of the soft threshold function defined as [33]

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

4) Discussion: To conclude, problem (7) is solved by iteratively applying (8) and (9) until some stopping criterion is satisfied. It can be shown that, if the problem (8) has a solution $\boldsymbol{A}^*$ given any $\zeta > 0$, then the generated sequence $\boldsymbol{A}^{(k)}$ converges to the optimum $\boldsymbol{A}^*$ [34].

Problem (8) is addressed by solving three subproblems with respect to $\boldsymbol{A}$ [see (10)], with respect to $\boldsymbol{V}$ [see (14)], and with respect to $\boldsymbol{U}$ [see (16)], repeatedly. The first subproblem with respect to $\boldsymbol{A}$ is solved by considering the standard quadratic programming problem defined by (13), of size $L + R + 1$, subject to R inequality constraints and one equality constraint, for each pixel. It is equivalent to the K-Hype algorithm described in [12]. Its computational complexity does not depend on the nature of the nonlinearity in the mixture model as it is formulated in a reproducing kernel Hilbert space. The second subproblem with respect to $\boldsymbol{V}$ has an explicit solution (15) that involves the inverse of the matrix $(\boldsymbol{I} + \boldsymbol{H}\boldsymbol{H}^\top)$. The latter can be calculated once the neighborhood relationship is defined. The third subproblem with respect to $\boldsymbol{U}$ has also an explicit solution (17). The computational time required by the two latter stages is almost negligible compared to the resources required by the first stage. Finally, the overall running time of the algorithm is mainly dictated by the number of iterations on $\boldsymbol{A}$, $\boldsymbol{U}$, and $\boldsymbol{V}$. It can be adjusted to compromise between computational time and convergence accuracy. The algorithm is provided in pseudocode form in Table I.

IV. EXPERIMENT RESULTS

In this section, experiments on spatially correlated images are reported to validate the proposed algorithm.

A. Experiments With Synthetic Images

1) Simulation Scenario Settings: Two spatially correlated abundance maps were generated for the following experiments. The endmembers were randomly selected from the spectral library advanced spaceborne thermal emission and reflection radiometer (ASTER) [35]. Each signature of this library has reflectance values measured over 224 spectral bands, uniformly distributed in the interval of 3–12 μm. Two synthetic images identical to that in [24] were used.

The first data cube, denoted by IM1 and containing 75×75 pixels, was generated by using five signatures randomly selected from the ASTER library. Pure regions and mixed regions involving between two and five endmembers, distributed spatially in the form of square regions, were generated. The background pixels were defined as mixtures of the same five endmembers with the abundance vector

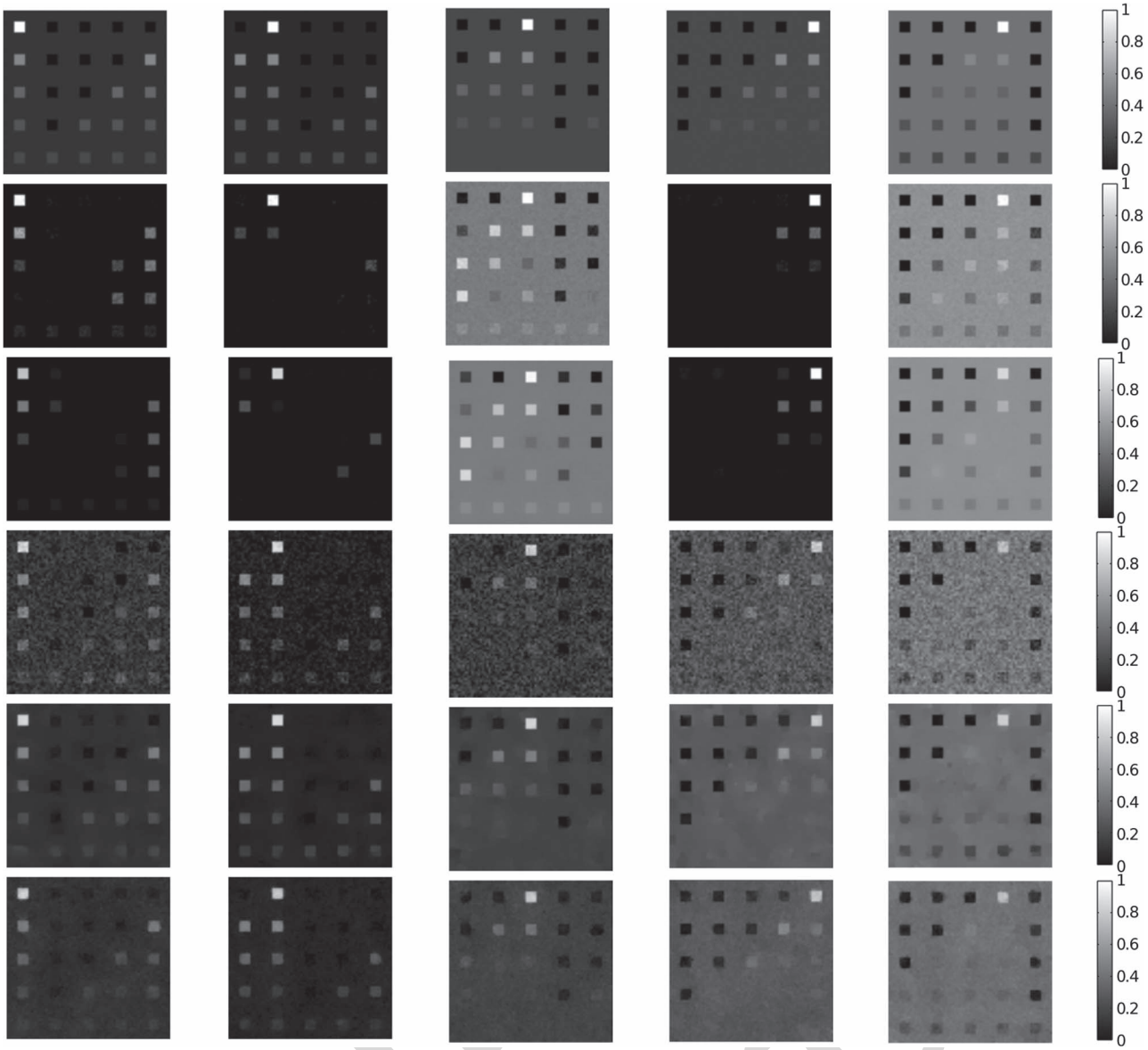

Fig. 1. Estimated abundance maps for IM1 image. For each row, from top to bottom: True abundance maps, FCLS, spatially regularized FCLS, K-Hype, proposed algorithm with four neighbors, and proposed algorithm with eight neighbors.

$[0.1149, 0.0741, 0.2003, 0.2055, 0.4051]^\top$. The first row in Fig. 1 shows the true fractional abundances for each endmember. The reflectance samples were generated with the two nonlinear mixture models defined hereafter, based on the five endmembers, and corrupted by a zero-mean white Gaussian noise $\boldsymbol{v}_n$ with an SNR of 20 dB. The first nonlinear mixture model is the bilinear one defined as

$$\boldsymbol{r}_n = \boldsymbol{M}\boldsymbol{\alpha}_n + \sum_{i=1}^{R}\sum_{j=i+1}^{R} \alpha_{n,i}\alpha_{n,j}\boldsymbol{m}_i \otimes \boldsymbol{m}_j + \boldsymbol{v}_n \qquad (18)$$

with $\otimes$ being the Hadamard product. The second one is a postnonlinear model (PNMM) given by

$$\boldsymbol{r}_n = (\boldsymbol{M}\boldsymbol{\alpha}_n)^\gamma + \boldsymbol{v}_n \qquad (19)$$

with an exponential value $\gamma = 0.7$ applied to the linear mixing model. At the end of this series of experiments with synthetic images, note that we will also consider a signal-dependent noise $\boldsymbol{v}_n$ to conform with conditions that may be experienced with new-generation sensors.

The second data cube, denoted by IM2 and containing 100×100 mixed pixels, was generated using nine endmember signatures. The abundance maps of the endmembers are the same as that for the image DC2 in [24]. Among these nine materials, only the 1st, 3rd, 5th, 8th, and 9th abundances are considered for pictorial illustration in Fig. 2. The first row of this figure depicts the true distribution of these five materials. Spatially homogeneous areas with sharp transitions can be clearly observed. Based on these abundance maps, a hyperspectral data cube was generated with the bilinear model (18) and with the postnonlinear model (19) applied to the nine endmember spectral signatures. The scene was also corrupted by a zero-mean white Gaussian noise $\boldsymbol{v}_n$ with an SNR of 20 dB.

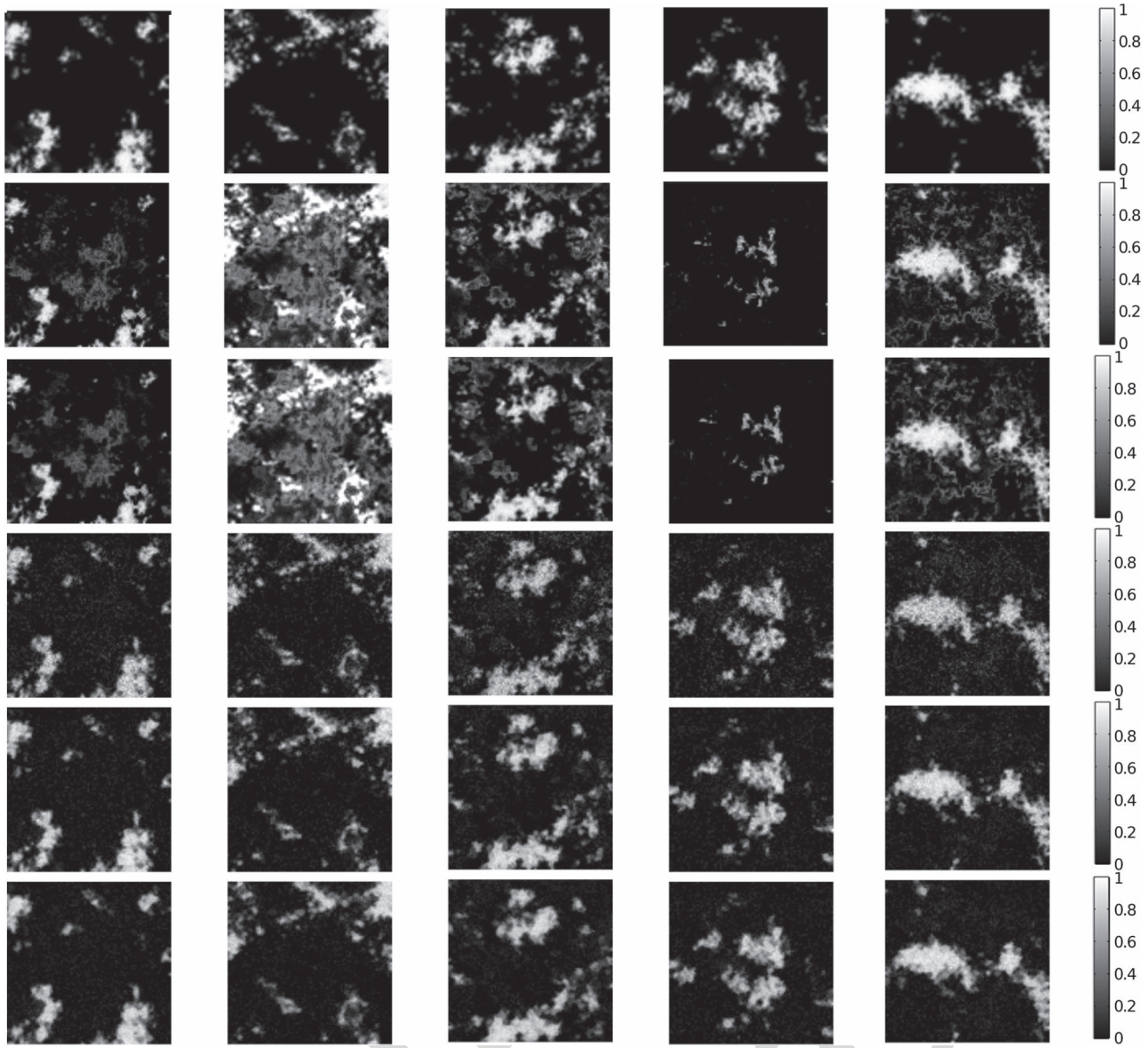

Fig. 2. Estimated abundance maps for IM2 image. For each row, from top to bottom: True abundance maps, FCLS, spatially regularized with four neighbors, FCLS, K-Hype, proposed algorithm with four neighbors, and proposed algorithm with eight neighbors.

2) Comparative Simulations: Several algorithms were tested in order to compare their unmixing performance on these two images. Their tuning parameters were set during preliminary experiments on independent data, via a simple search over the grids defined hereafter.

1) The linear unmixing methods [36]: The fully constrained least square method (FCLS) was run with the sum-to-one constraint strictly satisfied for comparability among algorithms. By relaxing the sum-to-one constraint, the nonnegative constrained least square method (NCLS) was also considered.
2) The spatially regularized FCLS/NCLS: In order to compare linear and nonlinear algorithms, we added the spatial regularization term (5) to the FCLS/NCLS algorithms. We conducted the split-Bregman iterations to solve these problems. We varied the spatial regularization parameter η from 0.0025 to 0.01 with an increment of 0.0025.
3) The nonlinear unmixing algorithm K-Hype [12]: Unmixing was performed in this case by solving problem (4). Its nonnegative counterpart obtained by relaxing the sum-to-one constraint (NK-Hype) was also tested. As in [12], the polynomial kernel defined by

$$\kappa(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_\ell}) = \left(1 + \frac{1}{R^2}(\boldsymbol{m}_{\lambda_\ell} - 1/2)^\top (\boldsymbol{m}_{\lambda_\ell} - 1/2)\right)^2 \tag{20}$$

was used. The parameter μ that controls the tradeoff between the misadjustment error and the regularization error was varied in the set $\{0.001, 0.005, 0.01, 0.05, 0.1\}$.
4) The proposed nonlinear algorithms incorporating spatial regularization: K-Hype and its nonnegative counterpart NK-Hype were both considered with spatial regularization. The polynomial kernel (20) was used, and the regularization parameter μ was varied in the same set

TABLE II
PARAMETER SETTINGS FOR THE COMPARATIVE SIMULATIONS ON IM1 AND IM2

	IM1		IM2	
	Bilinear	PNMM	Bilinear	PNMM
FCLS / spatial.-reg. FCLS	$\eta = 0.005$	$\eta = 0.005$	$\eta = 0.005$	$\eta = 0.01$
NCLS / spatial.-reg. NCLS	$\eta = 0.005$	$\eta = 0.0025$	$\eta = 0.005$	$\eta = 0.01$
K-Hype	$\mu = 0.1$	$\mu = 0.1$	$\mu = 0.01$	$\mu = 0.01$
NK-Hype	$\mu = 0.1$	$\mu = 0.1$	$\mu = 0.01$	$\mu = 0.05$
spatial.-reg. K-Hype (proposed)	$\mu = 0.005$, $\eta = 0.5$	$\mu = 0.005$, $\eta = 0.5$	$\mu = 0.005$, $\eta = 0.5$	$\mu = 0.005$, $\eta = 0.5$
spatial.-reg. NK-Hype (proposed)	$\mu = 0.005$, $\eta = 0.5$	$\mu = 0.005$, $\eta = 0.5$	$\mu = 0.005$, $\eta = 0.5$	$\mu = 0.005$, $\eta = 0.5$

TABLE III
RMSE COMPARISON (SIGNAL-INDEPENDENT NOISE): SEE EXPERIMENTS IN SECTION IV-A2

	IM1		IM2		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±0.048	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

TABLE IV
RMSE OF THE PROPOSED METHOD AS FUNCTION OF η: SEE EXPERIMENTS IN SECTION IV-A3

η	0.25	0.50	0.75	1.00
IM1	0.0548±0.0027	0.0456±0.0015	0.0479±0.0024	0.0508±0.0037
IM2	0.0576±0.0040	0.0521±0.0033	0.0530±0.0032	0.0567±0.0037

as above. The parameter ζ was adjusted in an adaptive way based on primal and dual residual norms at each iteration (see [37]). We varied the spatial regularization parameter η from 0.25 to 1 with increments of 0.25. Finally, the optimization algorithm was stopped when both $(\|\boldsymbol{V} - \boldsymbol{A}\|_F/N \times R)$ and $(\|\boldsymbol{U} - \boldsymbol{V}\boldsymbol{H}\|_F/4N \times R)$ became smaller than 10^{-5} or the number of iterations exceeded 10.

The above tests were performed on training images IM1 and IM2 to estimate the best parameter values in the sense that they minimize the estimation error (rmse) defined as

$$E = \sqrt{\frac{1}{NR}\sum_{n=1}^{N} \|\boldsymbol{\alpha}_n - \boldsymbol{\alpha}_n^*\|^2}.$$

These preliminary experiments led to the parameter settings shown in Table II. The performance, with standard deviation, on independent test images IM1 and IM2 is reported in Table III, and the estimated fractional abundances are represented in Figs. 1 and 2. For both images and both nonlinear mixture models, it can be observed that, when applied to nonlinearly mixed data, the linear method FCLS has large estimation errors. The abundance maps appear quite correct visually, but they are severely biased due to the nonlinearity of the mixing model. Relaxing the sum-to-one constraint with the NCLS algorithm allowed to improved the performance in some cases, particularly for IM2 with the bilinear model. The spatially regularized FCLS and NCLS algorithms offer limited performance improvement. Nonlinear methods notably reduce this error in the mean sense, except for IM2 with the bilinear model. In this case, note that most of the areas in the image are characterized by a dominant element with fractional abundance almost equal to one. Mixing phenomena with the bilinear model are thus limited, and the nonlinearity of the unmixing model supported by K-Hype-based algorithms suffers from this situation. They, however, provide high-resolution maps elsewhere. Finally, the proposed spatially regularized methods have lower reconstruction errors than the other proposed algorithms and clearer abundance maps.

3) Influence of the Parameter η: The penalty term η controls the tradeoff between data fitting and similarity among neighboring pixels. In the case of $\eta = 0$, the algorithm reduces to the original K-Hype that only considers spectral information at each pixel. The larger the η, the flatter the image is. In order to illustrate this intuition, we varied parameter η from 0.25 to 1 for IM1 and IM2 generated via the bilinear model, with $\mu = 0.005$. Note that, according to Table IV, the optimal value of η is 0.5 in both cases. To illustrate this experiment, the results are represented for IM1 in Fig. 3.

4) Influence of the Neighborhood: In the above experiments, we have used the four nearest neighbors to construct the difference matrix $\boldsymbol{H}_4$, with the subscript (4) to specify the size of the neighborhood. Any other neighborhood could be considered, provided that the matrix $\boldsymbol{H}$ is properly defined. For illustration purpose, we also considered the larger matrix of the eight nearest neighbors of each pixel defined as $\boldsymbol{H}_8 = (\boldsymbol{H}_\nwarrow\, \boldsymbol{H}_\nearrow\, \boldsymbol{H}_4\, \boldsymbol{H}_\swarrow\, \boldsymbol{H}_\searrow)$, with $\boldsymbol{H}_\nwarrow$, $\boldsymbol{H}_\nearrow$, $\boldsymbol{H}_\swarrow$, $\boldsymbol{H}_\searrow$ being the four diagonal adjacency matrices. The estimation errors of the abundance fractions are reported in Table V. The spatial regularization parameters were set to half of the values in Table II, as twice the number of neighboring pixels is used. Abundance maps for the bilinear scenario provided by the proposed algorithm are shown in the last row of Figs. 1 and 2.

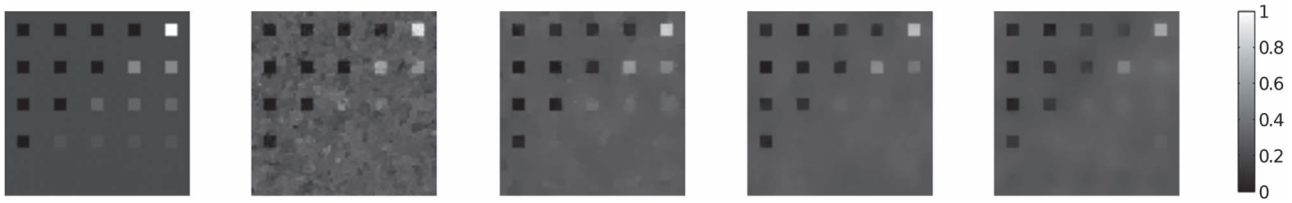

Fig. 3. Influence of spatial regularization parameter η. From left to right: True, $\eta = 0.25$, $\eta = 0.5$ (optimum), $\eta = 0.75$, and $\eta = 1$.

TABLE V
RMSE WITH EIGHT NEIGHBORING PIXELS (SIGNAL-INDEPENDENT NOISE): SEE EXPERIMENTS IN SECTION IV-A4

	IM1		IM2	
	Bilinear	PNMM	Bilinear	PNMM
spatial.-reg. K-Hype (proposed)	0.0509±0.0032	0.0570±0.0073	0.0557±0.0036	0.0916±0.0049
spatial.-reg. NK-Hype (proposed)	0.0568±0.0054	0.0564±0.0067	0.0701±0.0034	0.0858±0.0046

TABLE VI
RMSE COMPARISON (SIGNAL-DEPENDENT NOISE): SEE EXPERIMENTS IN SECTION IV-A5

	IM1		IM2	
	Bilinear	PNMM	Bilinear	PNMM
FCLS	0.1729±0.0091	0.1319±0.0053	0.1679±0.0265	0.1444±0.0096
NCLS	0.1346±0.0134	0.1473±0.0070	0.0790±0.0079	0.1373±0.0134
spatial.-reg. FCLS	0.1729±0.0091	0.1310±0.0053	0.1678±0.0265	0.1382±0.0072
spatial.-reg. NCLS	0.1154±0.0033	0.1474±0.0069	0.0690±0.0057	0.1229±0.0096
K-Hype	0.0784±0.0051	0.0887±0.0070	0.0928±0.0097	0.1194±0.0085
NK-Hype	0.0775±0.0056	0.0878±0.0067	0.0911±0.0080	0.1105±0.0092
spatial.-reg. K-Hype (proposed)	0.0445±0.0016	0.0485±0.0051	0.0517±0.0032	0.0842±0.0043
spatial.-reg. NK-Hype (proposed)	0.0492±0.0023	0.0476±0.0044	0.0640±0.0032	0.0762±0.0042

No significant improvement can be observed in these two cases, but obviously, the proper definition of a neighborhood is closely related to the structure of the images and must be driven by application needs. Fine spatial resolution can greatly improve scene understanding by magnifying subtle details in some cases but may lead to misleading interpretation in some other situations [38].

5) Test With a Signal-Dependent Noise: Unmixing algorithms proposed in the literature have usually been tested on images corrupted by an independent and identical distributed (i.i.d.) additive Gaussian noise. Due to the improved sensitivity of electronic components, this assumption may not be appropriate for data collected with new-generation sensors. Noise modeling and estimation in hyperspectral images has recently become an active subject of research. It is admitted that cameras provide images corrupted by two independent sources of noise, a signal-dependent noise and a signal-independent one [39]–[42]. The former results from the stochastic nature of the photon arrival/detection process. The latter results from sensor electronics and quantization process. We compared the unmixing algorithms on images IM1 and IM2 corrupted by the signal-dependent noise defined as

$$v_{n,\ell} = \tilde{r}_{n,\ell}^{\gamma} v_{n,\ell}^{(1)} + v_{n,\ell}^{(2)}$$

where $\tilde{r}_{n,\ell}$ is the ℓ-th wavelength band of the noise-free reflectance, $v_{n,\ell}^{(1)}$ and $v_{n,\ell}^{(2)}$ are two i.i.d. zero-mean Gaussian noises, and $0 \leq \gamma \leq 1$. Parameter γ was set to $(1/2)$, and the noise variances were set so that $\sigma^2_{v^{(1)}} = \sigma^2_{v^{(2)}}$ and the resulting SNR is 20 dB. Table VI gives the performance of the algorithms. Note that they were not notably affected by this noise setting. The proposed algorithm still exhibits the best performance.

TABLE VII
CLASSIFICATION ACCURACIES AFTER APPLYING SVM TO THREE DIFFERENT TYPES OF FEATURES (FCLS, K-HYPE, AND PROPOSED ALGORITHM)

	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

B. Experiments With AVIRIS Data

This part provides unmixing results for the proposed algorithm when applied on real hyperspectral data. The major difficulty in evaluating the performance of unmixing algorithms is that there are few existing ground-truth references for this purpose. However, in this section, we shall adopt an indirect strategy to circumvent this problem via the unmixing-based classification.

Supervised classification of hyperspectral images is a very challenging but important goal because it generally involves a limited number of training data with unusually high-dimensional patterns. Several feature extraction techniques have been recommended throughout the literature, including principal component analysis (PCA) and independent component analysis (ICA). In [43], the authors explored an alternative strategy consisting of using spectral unmixing for feature extraction, prior to classification. They considered different unmixing-based processes to evaluate the feasibility of this strategy and to perceive the necessity of extracting pure spectral endmembers for classification purposes. The so-called unmixing chain #4 in [43] was found to be the most efficient one. It simply consists of averaging the training samples in each labeled class and uses these spectral signatures to unmix

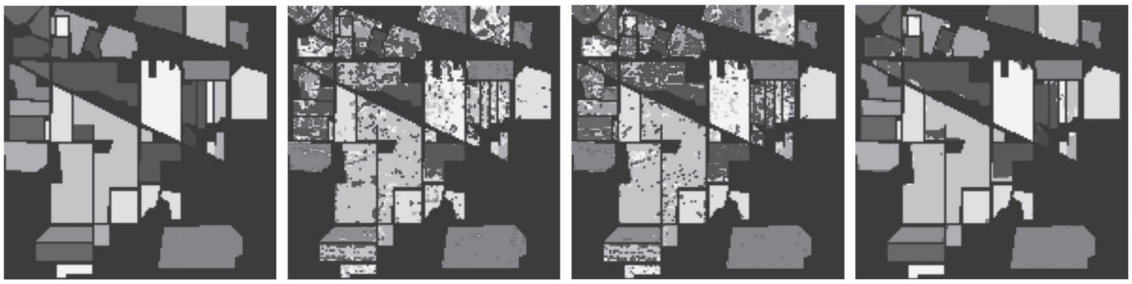

Fig. 4. Indian Pines classification map. From left to right: Ground truth, FCLS (61.36%), K-Hype (71.39%), and proposed algorithm (96.80%).

the original image. The features resulting from the unmixing of training samples are used to train an support vector machine (SVM) classifier. The latter is tested using the remaining labeled samples.

The scene used in our experiment is the well-known data set captured on the Indian Pines region by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). The scene comprises 145 $\times$ 145 samples, consisting of 220 contiguous spectral bands that cover wavelengths ranging from 0.40 to 2.5 μm, with spectral resolution approximately equal to 0.01 μm. Prior to analysis, noisy and water absorption bands were removed, yielding a total of 202 available spectral bands. The ground-truth data contains 16 mutually exclusive classes. The number of pixels in the smallest class is 20, while it is equal to 2468 pixels in the largest class. This widely used benchmark data set is known to be dominated by mixed pixels, even if ground-truth information assigns each pixel a unique class. In this experiment, we used FCLS, K-Hype, and the proposed algorithm with four nearest neighbors for unmixing-based feature extraction. A one-against-all multiclass SVM with Gaussian kernel was applied to these data. We constructed five training sets by randomly selecting 5%, 10%, and 15% of the ground-truth pixels. All the required parameters were optimized by a grid search procedure and fivefold cross-validation. The regularization parameter μ was set to 0.05 for K-Hype and for the proposed algorithm. In addition, for the latter, the spatial regularization parameter η was set to 0.5.

Table VII summarizes the classification accuracies of SVM operating on features extracted with FCLS, K-Hype, and the proposed algorithm. Fig. 4 presents these results in the case of SVM trained with 10% of the available samples per class. It appears that the two nonlinear unmixing algorithms are more effective than the linear one for feature extraction. This clearly means that our nonlinear unmixing model provides less confusing features between the heavily mixed-pixel classes that characterize the Indian Pines benchmark. Finally, we observe that spatial regularization allows to greatly improve the classification accuracy. Spatial homogeneity is a significant prior information for this problem, which allows to substantially improve the quality of the unmixing process.

V. Conclusion

Hyperspectral image unmixing can benefit from both spectral information and spatial information. In this paper, we presented a nonlinear abundance estimation algorithm for the hyperspectral unmixing task. The proposed algorithm integrates the spatial information using ℓ_1-norm regularization into a constrained LS-SVR problem. Split-Bregman iterations were used to solve this optimization problem. Experiments showed the advantage of introducing this spatial regularization into the nonlinear unmixing problem. Spatial correlation within the context of our nonlinear unmixing framework may take various forms. One can promote the similarity of abundance vectors between neighboring pixels, as considered in this paper, or the similarity of the nonlinear fluctuation functions in future works. We may also promote the similarity of the nonlinearity degrees as defined in [12] and derive more localized or adaptive solution strategies to reduce the computational complexity of unmixing algorithms that jointly consider spatial and spectral information.

Appendix

Recent works have raised the question of relaxing the sum-to-one constraint because poor estimates of the endmember signatures or misadjustment of the model may affect the performance of the unmixing process. The interested reader is addressed to [12] for a more detailed discussion on the subject. In this paper, we maintained this constraint for comparison purpose with existing approaches. We shall now provide the main result in the case where this constraint over $\boldsymbol{\alpha}_n$ is relaxed in problem (7). For clarity, let us denote by $\boldsymbol{h}_n$ the nonnormalized vector of abundances. The Lagrange function (11) becomes

$$\mathcal{L}_n = \frac{1}{2}\left(\|\boldsymbol{h}_n\|^2 + \|\psi_n\|_{\mathcal{H}}^2 + \frac{1}{\mu}\sum_{\ell=1}^{L} e_{n,\ell}^2 + \zeta \left\| \boldsymbol{h}_n - \boldsymbol{\xi}_n^{(k)} \right\|^2 \right) - \sum_{\ell=1}^{L} \beta_\ell \left(e_{n,\ell} - r_{n,\ell} + \boldsymbol{h}_n^\top \boldsymbol{m}_{\lambda_\ell} + \psi_n(\boldsymbol{m}_{\lambda_\ell}) \right) - \sum_{r=1}^{R} \gamma_r h_{n,r}$$

with $\gamma_{n,r} \geq 0$. The conditions for the optimality of $\mathcal{L}_n$ with respect to the primal variables lead us to

$$\begin{cases} \boldsymbol{h}_n^* = \frac{1}{\zeta+1}\left(\sum_{\ell=1}^{L} \beta_{n,\ell}^* \boldsymbol{m}_{\lambda_\ell} + \boldsymbol{\gamma}_n^* + \zeta \boldsymbol{\xi}_n^{(k)} \right) \\ \psi_n^* = \sum_{\ell=1}^{L} \beta_{n,\ell}^* \kappa(\cdot, \boldsymbol{m}_{\lambda_\ell}) \\ e_{n,\ell}^* = \mu \beta_{n,\ell}^*. \end{cases}$$

By substituting these conditions into the primal problem, we get the dual problem (21)

$$\begin{aligned}
&\max_{\boldsymbol{\beta}_n,\boldsymbol{\gamma}_n,\lambda_n} \mathcal{L}'_n(\boldsymbol{\beta}_n,\boldsymbol{\gamma}_n,\lambda_n) \\
&\quad = -\frac{\rho}{2\zeta}\begin{pmatrix}\boldsymbol{\beta}_n\\ \boldsymbol{\gamma}_n\end{pmatrix}^\top \left(\begin{array}{c|c}\boldsymbol{K}_\psi & \boldsymbol{M}\\ \hline \boldsymbol{M}^\top & \mathbf{I}\end{array}\right)\begin{pmatrix}\boldsymbol{\beta}_n\\ \boldsymbol{\gamma}_n\end{pmatrix} \\
&\qquad + \begin{pmatrix}\boldsymbol{r}_n - \rho\boldsymbol{M}\boldsymbol{\xi}_n^{(k)}\\ -\rho\boldsymbol{\xi}_n^{(k)}\end{pmatrix}^\top \begin{pmatrix}\boldsymbol{\beta}_n\\ \boldsymbol{\gamma}_n\end{pmatrix} \\
&\text{subject to} \quad \boldsymbol{\gamma}_n \succeq \mathbf{0} \\
&\qquad \text{with} \quad \boldsymbol{K}_\psi = \frac{1}{\zeta}(\boldsymbol{K}+\mu\mathbf{I}) + \boldsymbol{M}\boldsymbol{M}^\top \quad \text{and} \quad \rho = \frac{\zeta}{1+\zeta}.
\end{aligned} \tag{21}$$

Provided that the optimal dual variables $\boldsymbol{\beta}_n^*$ and $\boldsymbol{\gamma}_n^*$ have been determined, the solution $\boldsymbol{h}_n^*$ is given by

$$\boldsymbol{h}_n^* = \frac{1}{\zeta+1}\left(\boldsymbol{M}^\top\boldsymbol{\beta}_n^* + \boldsymbol{\gamma}_n^* + \zeta\boldsymbol{\xi}_n^{(k)}\right).$$

If necessary, note that $\boldsymbol{h}_n^*$ can be normalized afterward by writing $\boldsymbol{h}_n^* = \theta_n^*\,\boldsymbol{\alpha}_n^*$, with $\boldsymbol{\alpha}_n^*$ being the vector of what would correspond to fractional abundances and $\theta_n^* = \mathbf{1}^\top\boldsymbol{h}_n^*$ being the scaling factor. Following [12], i.e., using equivalence between optimization problems as explained in [44, p. 130], it can be shown that θ_n^* and $\boldsymbol{\alpha}_n^*$ are the solutions of

$$\begin{aligned}
\boldsymbol{\alpha}_n^*, \theta_n^*, \psi_n^* = \underset{\boldsymbol{\alpha}_n,\theta_n,\psi_n,\boldsymbol{e}_n}{\arg\min}\ & \frac{1}{2}\Bigg(\|\theta_n\boldsymbol{\alpha}_n\|^2 + \|\psi_n\|_{\mathcal{H}}^2 + \frac{1}{\mu}\sum_{\ell=1}^{L} e_{n,\ell}^2 \\
&\qquad + \zeta\left\|\theta_n\boldsymbol{\alpha}_n - \boldsymbol{\xi}_n^{(k)}\right\|^2\Bigg) \\
\text{subject to}\quad & e_{n,\ell} = r_{n,\ell} - \left(\theta_n\boldsymbol{\alpha}_n^\top\boldsymbol{m}_{\lambda_\ell} + \psi_n(\boldsymbol{m}_{\lambda_\ell})\right) \\
& \boldsymbol{\alpha}_n \succeq 0 \qquad \boldsymbol{\alpha}_n^\top\mathbf{1}_R = 1 \qquad \theta_n > 0.
\end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank Prof. Bioucas-Dias and his colleagues for their generous assistance in providing most of the data used in this paper.

REFERENCES

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

[2] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," in *Proc. SPIE*, 2009, vol. 7477, p. 74 770I.

[3] N. Raksuntorn and Q. Du, "Nonlinear spectral mixture analysis for hyperspectral imagery in an unknown environment," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 836–840, Oct. 2010.

[4] 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.

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

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

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

[8] F. Bovolo, L. Bruzzone, and L. Carlin, "A novel technique for subpixel image classification based on support vector machine," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2983–2999, Nov. 2010.

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

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

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

[12] 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.

[13] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral images with multi-kernel learning," in *Proc. IEEE WHISPERS*, 2012, pp. 1–12.

[14] 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*, 2013, pp. 2174–2178.

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

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

[17] D. M. Rogge, B. Rivard, J. Zhang, A. Sanchez, J. Harris, and J. Feng, "Integration of spatial-spectral information for the improved extraction of endmembers," *Remote Sens. Environ.*, vol. 110, no. 3, pp. 287–303, Oct. 2007.

[18] M. Zortea and A. Plaza, "Spatial preprocessing for endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2679–2693, Aug. 2009.

[19] 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.

[20] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[21] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.

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

[23] 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.

[24] 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.

[25] 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.

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

[27] T. A. Warner and M. C. Shank, "Spatial autocorrelation analysis of hyperspectral imagery for feature selection," *Remote Sens. Environ.*, vol. 60, no. 1, pp. 58–70, Apr. 1997.

[28] Q. Du, L. Zhang, and N. Raksuntorn, "Improving the quality of extracted endmembers," in *Proc. WHISPERS*, 2009, pp. 1–4.

[29] B. Schölkopf, J. C. Burges, and A. J. Smola, *Advances in Kernel Methods*. Cambridge, MA, USA: MIT Press, 1999.

[30] K. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–201, Mar. 2001.

[31] F. Abdallah, C. Richard, and R. Lengellé, "An improved training algorithm for nonlinear kernel discriminants," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2798–2806, Oct. 2004.

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

[33] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
[34] J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1–3, pp. 293–318, Apr. 1992.
[35] A. M. Baldridge, S. J. Hook, C. I. Grove, and G. Rivera, "The ASTER spectral library version 2.0," *Remote Sens. Environ.*, vol. 113, no. 4, pp. 711–715, Apr. 2009.
[36] D. C. Heinz and C.-I. Chang, "Fully constrained least squares linear mixture analysis for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.
[37] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[38] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. Suppl. 1, pp. S110–S122, Sep. 2009.
[39] N. Acito, M. Diani, and G. Corsini, "Signal-dependent noise modeling and model parameter estimation in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 2957–2971, Aug. 2011.
[40] J. Meola, M.-T. Eismann, R.-L. Moses, and J.-N. Ash, "Modeling and estimation of signal-dependent noise in hyperspectral imagery," *Appl. Opt.*, vol. 50, no. 21, pp. 3829–3846, Jul. 2011.
[41] L. Alparone, M. Selva, B. Aiazzi, S. Baronti, F. Butera, and L. Chiarantini, "Signal-dependent noise modeling and estimation of new-generation imaging spectrometers," in *Proc. IEEE WHISPERS*, 2009, pp. 1–4.
[42] B. Aiazzi, L. Alparone, S. Baronti, A. Barducci, P. Marcoionni, I. Pippi, and M. Selva, "Noise modeling and estimation of hyperspectral data from airborne imaging spectrometer," *Ann. Geophys.*, vol. 49, no. 1, pp. 1–9, Feb. 2006.
[43] 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.
[44] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

Jie Chen (S'12) was born in Xi'an, China, in 1984. He received the B.S. degree in information and telecommunication engineering in 2006 from the Xi'an Jiaotong University, Xi'an, and the Dipl.Ing. and the M.S. degrees in information and telecommunication engineering in 2009 from the University of Technology of Troyes (UTT), Troyes, France, and from the Xi'an Jiaotong University, respectively. In 2013, he received the Ph.D. degree in systems optimization and security from the UTT.

He is currently a Postdoctoral Researcher with the Côte d'Azur Observatory, University of Nice Sophia-Antipolis, Nice, France. His current research interests include adaptive signal processing, kernel methods, hyperspectral image analysis, and supervised and unsupervised learning.

Cédric Richard (S'98–M'01–SM'07) was born on January 24, 1970 in Sarrebourg, France. He received the Dipl.Ing. and M.S. degrees in electrical and computer engineering and the Ph.D. degree in electrical and computer engineering from the University of Technology of Compiègne (UTC), Compiègne, France, in 1994 and 1998, respectively.

He joined the Côte d'Azur Observatory, University of Nice Sophia-Antipolis, Nice, France, in 2009, where he is currently a Professor of electrical engineering. From 1999 to 2003, he was an Associate Professor with the University of Technology of Troyes (UTT), Troyes, France. From 2003 to 2009, he was a Professor with the UTT and the supervisor of a group consisting of 60 researchers and Ph.D. students. In winter 2009 and autumns 2010 and 2011, he was a Visiting Researcher with the Department of Electrical Engineering, Federal University of Santa Catarina (UFSC), Florianòpolis, Brazil. He is the author of over 200 papers. His current research interests include statistical signal processing and machine learning.

Dr. Richard served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2006 to 2010 and of Elsevier Signal Processing since 2009. In 2009, he was nominated liaison local officer for The European Association for signal Processing (EURASIP) and member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society. In 2013, has was also elected member of the Machine Learning for Signal Processing Technical Committee of this society. He is a junior member of the Institut Universitaire de France since October 2010. He was the General Chair of the XXIth francophone conference Groupement de Recherche en Traitement du Signal et des Images (GRETSI) on Signal and Image Processing that was held in Troyes in 2007 and of the IEEE Statistical Signal Processing Workshop that was held in Nice in 2011. Since 2005, he has been a member of the board of the federative CNRS research group Information, Signal, Images et ViSion (ISIS) on Information, Signal, Images and Vision. He is a member of the GRETSI association board and of the EURASIP society. He received, with Paul Honeine, the Best Paper Award for "Solving the pre-image problem in kernel machines: a direct method" at the 2009 IEEE Workshop on Machine Learning for Signal Processing.

Paul Honeine (M'07) was born in Beirut, Lebanon, on October 2, 1977. He received the Dipl.Ing. degree in mechanical engineering and the M.Sc. degree in industrial control from the Faculty of Engineering, Lebanese University, Beirut, Lebanon, in 2002 and 2003, respectively. In 2007, he received the Ph.D. degree in systems optimization and security from the University of Technology of Troyes, Troyes, France, where he was also a Postdoctoral Research Associate with the Systems Modeling and Dependability Laboratory from 2007 to 2008.

Since 2008, he has been an Assistant Professor with the University of Technology of Troyes. His research interests include nonstationary signal analysis and classification, nonlinear signal processing, sparse representations, machine learning, and wireless sensor networks.

Dr. Honeine received, with Cédric Richard, the Best Paper Award for "Solving the pre-image problem in kernel machines: a direct method" at the 2009 IEEE Workshop on Machine Learning for Signal Processing.