NONLINEAR UNMIXING OF HYPERSPECTRAL DATA WITH PARTIALLY LINEAR LEAST-SQUARES SUPPORT VECTOR REGRESSION

Jie Chen[⋆†]*, Cédric Richard*[⋆]*, André Ferrari*[⋆]*, Paul Honeine*[†]

[⋆] Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur, France
[†]Université de Technologie de Troyes, CNRS, France
E-mail: {jie.chen, paul.honeine}@utt.fr, {cedric.richard, andre.ferrari}@unice.fr

ABSTRACT

In recent years, nonlinear unmixing of hyperspectral data has become an attractive topic in hyperspectral image analysis, because nonlinear models appear as more appropriate to represent photon interactions in real scenes. For this challenging problem, nonlinear methods operating in reproducing kernel Hilbert spaces have shown particular advantages. In this paper, we derive an efficient nonlinear unmixing algorithm based on a recently proposed linear mixture/nonlinear fluctuation model. A multi-kernel learning support vector regressor is established to determine material abundances and nonlinear fluctuations. Moreover, a low complexity locally-spatial regularizer is incorporated to enhance the unmixing performance. Experiments with synthetic and real data illustrate the effectiveness of the proposed method.

Index Terms— Nonlinear unmixing, hyperspectral image, support vector regression, multi-kernel learning, spatial regularization.

1. INTRODUCTION

Hyperspectral imagery provides data with a wide spectral range and high spectral resolution. These characteristics are desirable for detection and classification of surfaces and chemical elements in real images. Hyperspectral image processing has become a continuously growing area, and now receives considerable attention.

Due to multiple factors, including the possible low spatial resolution of some hyperspectral-imaging devices, the diversity of materials, the reflections of photons onto several objects, etc., problems with mixed pixels can occur and be critical for proper interpretation of images. Spectral unmixing is an important issue to analyze hyperspectral data. It consists of decomposing each mixed pixel into pure endmember spectra, and estimating the fractional abundance of each endmember [1].

Linear unmixing methods have been extensively considered within the area of remote sensing. In the recent years, researchers have progressively focused on nonlinear unmixing method, as there are many situations in which the classical linear mixture model may be insufficient [2]. Recent nonlinear unmixing approaches dedicated to hyperspectral images can be divided into the following classes: 1) Direct methods: Some techniques mimic the nonlinear interactions between endmember spectra by incorporating them into the regression model in an explicit way. For instance, the authors in [3, 4] extended the collection of endmembers by adding artificial crossed spectra to model photon interactions. 2) Manifold learning methods: These approaches exploit low dimensional manifolds embedded in high-dimensional sensed data [5]. Algorithms based on Isomap and LLE, which assume the local linearity of data, were recently used. 3) Kernel-based methods: These approaches can model complex interactions among material signatures by the virtue of kernel trick. In [6], we introduced a new framework for estimating abundances involved in a nonlinear mixture of hyperspectral data. It overcomes the drawbacks of former heuristic kernel unmixing methods [7]. 4) Bayesian methods: In [8, 9], Bayesian inference was used to estimate the fraction of abundances within the context of nonlinear mixing scenarios, using generated samples distributed according to the posterior distribution. 5) Neural networks: This family refers to techniques that intend to establish, based on a collection of training data, an input-output relation between some observed data and their abundances via black-box modeling [10, 11].

Among the above mentioned methods, kernel-based algorithms have demonstrated great generality and excellent performance for nonlinear unmixing [6, 12, 13]. In this paper, the inherent spectral mixing mechanism is modeled by a partial linear model, consisting of linear mixture components and a nonlinear fluctuation term defined in a reproducing kernel Hilbert space (RKHS). A multi-kernel support vector regression algorithm is derived to estimate the material abundances, and to automatically adjust the proportion between the linear and the nonlinear components. Furthermore, spatial correlation is also considered by a local regularization term which does not introduce extra computational complexity. This work is a follow-on work of the state-of-the-art kernel unmixing work [6, 13]. Benefiting from the specific structure of the unmixing problem, and from the spatial information, the proposed approach provides a reliable estimation of fractional abundances. Its advantages are illustrated via comparative experiments.

2. NOTATIONS AND NONLINEAR MIXTURE MODEL

Consider that an observed scene consists of R significant endmembers with spectral signature $\boldsymbol{m}_i \in \mathbb{R}^L$, where L denotes the number of spectral bands. Note that, usually, we have $R \ll L$. Let $\boldsymbol{r} \in \mathbb{R}^L$ be an hyperspectral pixel, and $\boldsymbol{\alpha} \in \mathbb{R}^R$ the unknown abundance vector associated to the latter. Let $\boldsymbol{M} = [\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R] \in \mathbb{R}^{L \times R}$ be the matrix of the endmember spectra. For the sake of convenience, the ℓ-th row of $\boldsymbol{M}$ is denoted by $\boldsymbol{m}_{\lambda_\ell}^\top \in \mathbb{R}^L$, that is, $\boldsymbol{m}_{\lambda_\ell}$ is the vector of the endmember signatures at the ℓ-th wavelength band.

Nonlinear mixture usually results from complex interactions of light sources scattered by multiple materials in the scene. Linear mixing model is a reasonable approximation in cases where these interactions are not significant. However, when such effects cannot be neglected, it is more reasonable to assume that mixing consists of a linear trend and a nonlinear fluctuation term, rather than leaving a nonlinear model totally uncontrolled. Based on this point of view, we model the mixing mechanism with respect to each spectral band

$$\psi(\boldsymbol{m}_{\lambda_\ell}) = \theta\, \boldsymbol{\alpha}^\top \boldsymbol{m}_{\lambda_\ell} + \psi_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}) + e_\ell \qquad (1)$$

where θ is an additional factor that affects the linear mixture, which is parameterized by the abundance vector $\boldsymbol{\alpha}$. Function ψ_{nlin} is supposed to describe the nonlinear interactions between spectra, and e_ℓ is the modeling error. In this paper, the function ψ_{nlin} is restricted to be an element of a RKHS, say $\mathcal{H}_{\text{nlin}}$, such that

$$\psi_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}) = \langle \psi, \kappa_{\text{nlin}}(\cdot, \boldsymbol{m}_{\lambda_\ell})\rangle_{\mathcal{H}_{\text{nlin}}}, \quad \forall \psi_{\text{nlin}} \in \mathcal{H}_{\text{nlin}} \tag{2}$$

The corresponding kernel satisfies to the reproducing property

$$\kappa_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_k}) = \langle \kappa_{\text{nlin}}(\cdot, \boldsymbol{m}_{\lambda_\ell}), \kappa_{\text{nlin}}(\cdot, \boldsymbol{m}_{\lambda_k})\rangle \tag{3}$$

Several nonlinear models, for instance the bilinear model and post-nonlinear models, can be expressed or approximated by this model. Moreover, the non-negativity and the sum-to-one constraints are usually considered for abundance $\boldsymbol{\alpha}$ estimation, namely,

$$\boldsymbol{\alpha} \succeq \mathbf{0} \quad \text{and} \quad \mathbf{1}^\top \boldsymbol{\alpha} = 1 \tag{4}$$

It should be noticed that the rationality of imposing the sum-to-one constraint is under discussion among researchers. In this paper, we preserve it because discarding this constraint will only simplify the mathematical problem and the proposed algorithm. They can be easily modified to be adapted to this case.

3. NONLINEAR UNMIXING WITH LEAST-SQUARES SUPPORT VECTOR REGRESSION

We assume that the endmember matrix $\boldsymbol{M}$ has been determined by an endmember extraction algorithm. This approach is usually referred to as supervised unmixing. Pixel unmixing using the model (1) is achieved by estimating θ, $\boldsymbol{\alpha}$, and ψ_{nlin}. In order to further control the balance between the linear trend and the nonlinear fluctuation term, we propose to solve the following multi-kernel least-squares support vector problem, in which these proportions are controlled via functional norms.

$$\begin{aligned}
\psi^*,\theta^*,u^* = \underset{\psi,\theta,u}{\arg\min}\ & \frac{1}{2}\left(\frac{\|\psi_{\text{lin}}\|^2_{\mathcal{H}'_{\text{lin}}}}{u} + \frac{\|\psi_{\text{nlin}}\|^2_{\mathcal{H}'_{\text{nlin}}}}{1-u}\right) + \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 \\
\text{subject to}\quad & e_\ell = r_\ell - \psi(\boldsymbol{m}_{\lambda_\ell}) \\
\text{and}\quad & \psi = \psi_{\text{lin}} + \psi_{\text{nlin}} \quad \text{with} \quad \psi_{\text{lin}}(\boldsymbol{m}_{\lambda_\ell}) = \theta\, \boldsymbol{\alpha}^\top \boldsymbol{m}_{\lambda_\ell} \\
& \boldsymbol{\alpha} \succeq \mathbf{0} \qquad \mathbf{1}^\top \boldsymbol{\alpha} = 1 \\
& \theta \in \mathbb{R}_+^* \quad 0 \le u \le 1
\end{aligned}$$

where $\mathcal{H}_{\text{lin}}$ and $\mathcal{H}_{\text{nlin}}$ are RKHS. This problem is not convex, and is difficult to solve as formulated. Fortunately, it can be transformed into an equivalent convex problem in the sense of [14, p. 130]. Consider the variable change $\boldsymbol{h} = \theta\, \boldsymbol{\alpha}$. The cost function can be directly reformulated as a function of $\boldsymbol{h}$. The linear function is expressed by

$$\psi_{\text{lin}}(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{h}^\top \boldsymbol{m}_{\lambda_\ell}, \tag{5}$$

and the two constraints over $\boldsymbol{\alpha}$ become

$$(\boldsymbol{h} \succeq \mathbf{0}) \qquad \text{and} \qquad (\mathbf{1}^\top \boldsymbol{h} = \theta \quad \text{with} \quad \theta \in \mathbb{R}_+^*)$$

The latter constraint can be eliminated because it is trivial due to the former one. Because $\boldsymbol{h} = \theta\, \boldsymbol{\alpha}$ and $\mathbf{1}^\top \boldsymbol{\alpha} = 1$, we have $\boldsymbol{\alpha}^* = \boldsymbol{h}^*/\theta^*$ with $\theta^* = \mathbf{1}^\top \boldsymbol{h}^*$. This leads us to the following problem

$$\min_u J(u) \qquad \text{subject to} \qquad 0 \le u \le 1 \tag{6}$$

where $J(u)$ is defined by

$$J(u) = \begin{cases} \min_\psi F(u,\psi) = \frac{1}{2}\left(\frac{\|\psi_{\text{lin}}\|^2_{\mathcal{H}_{\text{lin}}}}{u} + \frac{\|\psi_{\text{nlin}}\|^2_{\mathcal{H}_{\text{nlin}}}}{1-u}\right) + \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 \\ \quad \text{subject to} \quad e_\ell = r_\ell - \psi(\boldsymbol{m}_{\lambda_\ell}) \\ \qquad\qquad \psi = \psi_{\text{lin}} + \psi_{\text{nlin}} \\ \qquad\qquad \psi_{\text{lin}}(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{h}^\top \boldsymbol{m}_{\lambda_\ell} \qquad \boldsymbol{h} \succeq \mathbf{0} \end{cases} \tag{7}$$

Appropriate definitions for $\mathcal{H}_{\text{lin}}$ and $\mathcal{H}_{\text{nlin}}$, and continuity considerations, are used for defining $J(u)$ in $u = 0$ and 1. See [6] for details. It can be shown that the problem (6)-(7) is a convex optimization problem by virtue of the convexity of the so-called perspective function defined by $f(u,\psi) = \|\psi\|^2_{\mathcal{H}_{\text{lin/nlin}}}/u$ over $\mathbb{R}_+ \times \mathcal{H}_{\text{lin/nlin}}$ [14]. This allows to formulate a two-stage optimization procedure, with respect to ψ and u successively.

3.1. Solving with respect to ψ

By the strong duality property, we shall now derive a dual problem that has the same solution $J(u) = F(u,\psi^*)$ as the primal problem (7). Let us introduce the Lagrange multipliers β_ℓ and γ_r. The Lagrange function associated with the problem (7) can be written as

$$\begin{aligned} G = & \frac{1}{2}\left(\frac{1}{u}\|\boldsymbol{h}\|^2 + \frac{1}{1-u}\|\psi_{\text{nlin}}\|^2_{\mathcal{H}_{\text{nlin}}}\right) + \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 \\ & - \sum_{\ell=1}^{L} \beta_\ell \left(e_\ell - r_\ell + \psi(\boldsymbol{m}_{\lambda_\ell})\right) - \sum_{r=1}^{R} \gamma_r\, h_r \end{aligned} \tag{8}$$

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

$$\begin{cases} \boldsymbol{h}^* = u\left(\sum_{\ell=1}^{L} \beta_\ell^*\, \boldsymbol{m}_{\lambda_\ell} + \boldsymbol{\gamma}^*\right) \\ \psi^*_{\text{nlin}} = (1-u)\sum_{\ell=1}^{L} \beta_\ell^*\, \kappa_{\text{nlin}}(\cdot, \boldsymbol{m}_{\lambda_\ell}) \\ e_\ell^* = \mu\, \beta_\ell^* \end{cases} \tag{9}$$

Substituting (9) into (8) leads us to the dual problem (10), see next page, where $\boldsymbol{K}_{\text{nlin}}$ is the Gram matrix defined by

$$[\boldsymbol{K}_{\text{nlin}}]_{\ell k} = \kappa_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_k}) \tag{11}$$

Problem (10) is a standard quadratic programming problem which can be solved by a generic solver. Once $\boldsymbol{\beta}^*$ and $\boldsymbol{\gamma}^*$ are determined, pixel reconstruction can be performed using

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

with $\psi^*(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{m}_{\lambda_\ell}^\top \boldsymbol{h}^* + \psi^*_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell})$ defined in equation (9). Finally, using the relation between $\boldsymbol{h}$ and $\boldsymbol{\alpha}$, the estimated abundance vector is given by

$$\boldsymbol{\alpha}^* = \frac{\boldsymbol{M}^\top \boldsymbol{\beta}^* + \boldsymbol{\gamma}^*}{\mathbf{1}^\top (\boldsymbol{M}^\top \boldsymbol{\beta}^* + \boldsymbol{\gamma}^*)} \tag{12}$$

3.2. Solving with respect to u

In the general framework for multi-kernel learning described in [15], the kernel weights were updated by the reduced gradient method. In [6, 13], a gradient descent was also used to update u. However, gradient based updates usually have low convergence rates. An optimal step size search strategy is often considered to accelerate the convergence rate, with a an additional computational cost. We shall now show that there is a closed-form expression for the optimum u^*. Notice that the function

$$f_{p,q}(u) = \frac{p}{u} + \frac{q}{1-u} \quad \text{with} \quad p, q \ge 0 \tag{13}$$

$$J(u) = \begin{cases} \max\limits_{\boldsymbol{\beta},\boldsymbol{\gamma}} G'(u,\boldsymbol{\beta},\boldsymbol{\gamma}) = -\frac{1}{2}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix}^\top \left(\begin{array}{c|c} \boldsymbol{K}_u + \mu\boldsymbol{I} & u\boldsymbol{M} \\ \hline u\boldsymbol{M}^\top & u\boldsymbol{I}\end{array}\right)\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix} + \begin{pmatrix}\boldsymbol{r}\\ \boldsymbol{0}\end{pmatrix}^\top \begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix} \\ \qquad \text{subject to} \quad \boldsymbol{\gamma} \succeq \boldsymbol{0} \\ \qquad \text{with} \quad \boldsymbol{K}_u = u\boldsymbol{M}\boldsymbol{M}^\top + (1-u)\boldsymbol{K}_{\text{nlin}} \end{cases} \tag{10}$$

$$J_{sp}(u) = \begin{cases} \max\limits_{\boldsymbol{\beta},\boldsymbol{\gamma}} G'_{sp}(u,\boldsymbol{\beta},\boldsymbol{\gamma}) = -\frac{1}{2}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix}^\top \left(\begin{array}{c|c} \boldsymbol{K}_u + \mu\boldsymbol{I} & \xi\boldsymbol{M} \\ \hline \xi\boldsymbol{M}^\top & \xi\boldsymbol{I}\end{array}\right)\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix} + \begin{pmatrix}\boldsymbol{r} - \xi\zeta\,\boldsymbol{M}\sum_{i\in\mathcal{N}_n}\omega_i\boldsymbol{h}_i^*\\ -\xi\zeta\sum_{i\in\mathcal{N}_n}\omega_i\boldsymbol{h}_i^*\end{pmatrix}^\top \begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix} \\ \qquad \text{subject to} \quad \boldsymbol{\gamma} \succeq \boldsymbol{0} \\ \qquad \text{with} \quad \boldsymbol{K}_u = \xi\boldsymbol{M}\boldsymbol{M}^\top + (1-u)\boldsymbol{K}_{\text{nlin}} \quad \text{and} \quad \xi = \dfrac{u}{1+u\zeta} \end{cases} \tag{19}$$

is convex over the interval $(0,1)$. The optimum value is achieved at

$$u^* = (1+\sqrt{q/p})^{-1} \tag{14}$$

Now considering the problem (6), and using the stationarity conditions (9), we conclude that the optimum value u^* at each iteration is given by

$$u^* = \left(1 + (1-u^*_{-1})\sqrt{(\boldsymbol{\beta}^{*\top}\boldsymbol{K}_{\text{nlin}}\boldsymbol{\beta}^*)/\|\boldsymbol{h}^*\|^2}\right)^{-1} \tag{15}$$

where u^*_{-1} denotes the optimal u^* obtained at the previous iteration.

4. LOCALLY-SPATIAL REGULARIZATION

One of the distinguishing properties of remotely sensed data is that they convey multivariate information into a 2D pictorial representation [16]. Incorporating spatial information may have a positive impact on hyperspectral unmixing. Exploiting spatial information can be conducted by designing appropriate criteria, usually, by adding a penalizing term to the modeling error that promotes similarity of fractional abundances between one pixel and some others to be defined. Spatial regularization techniques incorporating all the pixels, such as TV norm, lead to a global optimization problem that notably increase the computational burden [17, 18]. To overcome this drawback, we adopt hereafter a local regularization strategy, and we associate it to the nonlinear unmixing algorithm presented before.

Let us identify the current pixel under analysis by the index n. In addition to estimating $\boldsymbol{h}_n$ via the optimization problem (6)-(7), we shall also promote the similarity between $\boldsymbol{h}_n$ and its neighbors by considering the quadratic penalty term

$$J_{\text{reg}}(\boldsymbol{h}_n) = \frac{\zeta}{2}\sum_{i\in\mathcal{N}_n}\omega_i\|\boldsymbol{h}_n - \boldsymbol{h}_i^*\|^2. \tag{16}$$

where the dissimilarities with the neighboring pixels i in $\mathcal{N}_n$ are weighted by ω_i. The abundance vectors $\boldsymbol{h}_i^*$ are considered to have already been estimated. If the image is treated one pixel after another, sequentially along rows, an example of such a neighborhood is defined by $\{n-1, n-w, n-w-1\}$, where w is the width of the image. Although ℓ_1-type regularization has favorable properties in image processing, it adds significant complexity to the nonlinear algorithm. For this reason, we shall employ this ℓ_2 regularization combined with the following strategy

1. Calculate $\widetilde{\omega}_i = \|\boldsymbol{r}_n - \boldsymbol{r}_i\|^2/\|\boldsymbol{r}_n\|^2$, which represents the normalized spectral distance between pixel n and its neighbors;
2. If $\min_i \widetilde{\omega}_i > \nu_0$, with ν_0 a given threshold, the algorithm is conducted without spatial regularization using (10).
3. Otherwise, the weights ω_i are determined from the $\widetilde{\omega}_i$'s by

$$\omega_i = \frac{1/\widetilde{\omega}_i}{\sum_{i\in\mathcal{N}_n} 1/\widetilde{\omega}_i} \tag{17}$$

and the algorithm is conducted with the spatial regularization term (16).

Step 2. allows to disable spatial regularization in the case where the spectral distance between pixel n and its neighbors is significant, which may preserve transient phenomena. This step can obviously be eliminated if preferred, and the definition (17) changed. Finally, unmixing with spatial regularization is performed by adding (16) to the Lagrange function (8). The conditions for optimality yield

$$\begin{cases} \boldsymbol{h}_n^* = \frac{u}{1+u\zeta}\left(\sum_{\ell=1}^L \beta_\ell^*\,\boldsymbol{m}_{\lambda_\ell} + \boldsymbol{\gamma}^* + \zeta\sum_{i\in\mathcal{N}_n}\omega_i\boldsymbol{h}_i^*\right) \\ \psi^*_{\text{nlin}} = (1-u)\sum_{\ell=1}^L \beta_\ell^*\,\kappa_{\text{nlin}}(\cdot,\boldsymbol{m}_{\lambda_\ell}) \\ e_\ell^* = \mu\,\beta_\ell^* \end{cases} \tag{18}$$

Substituting (18) into the Lagrange function leads us to the dual problem (19). It is interesting to note that this optimization problem has the same computational complexity as (10). Once the duality variables are determined, the vector of fractional abundances can be estimated as follows

$$\boldsymbol{\alpha}_n^* = \frac{\left(\boldsymbol{M}^\top\boldsymbol{\beta} + \boldsymbol{\gamma}^* + \zeta\sum_{i\in\mathcal{N}_n}\omega_i\boldsymbol{h}_i^*\right)}{\boldsymbol{1}^\top\left(\boldsymbol{M}^\top\boldsymbol{\beta} + \boldsymbol{\gamma}^* + \zeta\sum_{i\in\mathcal{N}_n}\omega_i\boldsymbol{h}_i^*\right)}. \tag{20}$$

Finally, note that the update equation (15) for u remains unchanged.

5. EXPERIMENT RESULTS

5.1. Experiments with synthetic data

Let us first report some experimental results on synthetic images. The endmembers were randomly selected from the library ASTER where signatures have reflectance values measured over 224 spectral bands, uniformly distributed in the interval $3-12$ μm [19].

5.1.1. Spatially-uncorrelated pixels

Random pixels were first generated to test the performance of the basic nonlinear unmixing algorithm. Pixels with three, five and eight endmembers were generated respectively with the linear model, the

Table 1. Spatially-uncorrelated pixels: RMSE comparison ($\times 10^{-2}$)

	$R=3$			$R=5$			$R=8$		
	linear	bilinear	PNMM	linear	bilinear	PNMM	linear	bilinear	PNMM
FCLS	1.10 ± 0.02	33.2 ± 4.37	21.3 ± 1.31	1.65 ± 0.03	23.6 ± 1.73	14.8 ± 0.72	3.16 ± 0.08	21.3 ± 1.31	12.7 ± 0.49
ExtM	1.29 ± 0.02	6.27 ± 0.43	19.6 ± 1.42	2.31 ± 0.07	6.50 ± 0.29	14.5 ± 0.61	3.98 ± 0.16	7.36 ± 0.31	12.3 ± 0.52
K-Hype	2.61 ± 0.07	3.30 ± 0.13	5.40 ± 0.28	3.37 ± 0.11	3.34 ± 0.13	5.13 ± 0.24	3.65 ± 0.13	3.56 ± 0.14	5.11 ± 0.24
Proposed	1.92 ± 0.04	3.66 ± 0.14	3.21 ± 0.12	3.18 ± 0.11	3.65 ± 0.14	4.99 ± 0.24	3.21 ± 0.11	3.70 ± 0.14	4.95 ± 0.26

generalized bilinear mixture model [8], and the post-nonlinear mixing model (PNMM) defined by $\boldsymbol{r} = (\boldsymbol{M}\boldsymbol{\alpha})^{0.7} + \boldsymbol{n}$. In each scenario, 1000 pixels were generated for evaluating and comparing the performance. All these images were corrupted with an additive white Gaussian noise $\boldsymbol{n}$ with SNR = 30 dB. The Gaussian kernel defined as $\kappa_{\text{nlin}}(\boldsymbol{m}_{\lambda_p}, \boldsymbol{m}_{\lambda_\ell}) = \exp(-\|\boldsymbol{m}_{\lambda_p} - \boldsymbol{m}_{\lambda_\ell}\|^2 / 2\sigma_0^2)$ was used.

Our approach was compared with the fully constrained least square method (FCLS) [20], and the extended endmember-matrix method (ExtM) [3], and the K-Hype algorithm [6]. The root mean square error (RMSE) of the abundance vectors was used to compare these algorithms. Preliminary runs were performed with independent data to set their parameters. The results are reported in Table 1. The proposed method had better performance than the other algorithms except for linearly-mixed data. It however outperformed K-Hype in this case because it can automatically balance ψ_{lin} and ψ_{nlin}. This is a key advantage because nonlinear mixing effects can be weak in some situations, in particular if there is a locally-dominant material. Figure 1 illustrates this situation, where material #9 is dominant in the center of the scene. See image DC2 in [17], mixed with a bilinear model.

5.1.2. *Spatially-correlated pixels*

We shall now check the advantage of incorporating spatial information, and validate the proposed locally-spatial regularization. A synthetic image of 75×75 pixels was generated with the bilinear mixture model using five signatures. The pure regions and mixed regions involved between 2 and 5 endmembers, spatially distributed in the form of square regions. See image DC1 in [17] for details. The image was corrupted by a white Gaussian noise with SNR = 25 dB. The regularization parameter ζ was set to 10. The threshold ν_0 was set to 0.01. The true abundances and estimated abundances of the 4-th and 5-th materials are illustrated in Figure 2, with RMSE at the bottom of each image. The effect of spatial regularization can be clearly observed.

5.2. Experiments with real data

This experiment illustrates the performance of the proposed algorithms, when applied to real hyperspectral data. This scene is the well-known image captured on the Cuprite mining district (NV, USA) by AVIRIS. A sub-image of 250×191 pixels was chosen to evaluate the algorithms. This area of interest has $L = 188$ spectral bands. The number of endmembers was first estimated via the virtual dimensionality, and R was accordingly set to 12 [21]. VCA algorithm was then used to extract the endmembers. The proposed algorithms, with/without spatial regularization, were run with the Gaussian kernel. The kernel bandwidth was set to $\sigma_0^2 = 2$. The regularization parameter μ was fixed to $5 \cdot 10^{-3}$. The spatial regularization parameter ζ was set to 10. The unmixing results for three materials are shown in Figure 3. It can be observed that the mineral distributions are consistent with the reference map [22], and the US Highway 95 can be clearly identified whereas the linear unmixing failed to separate it [6].

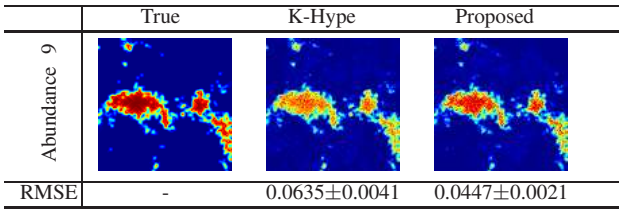

Fig. 1. Illustration of the situation where one material dominates the others.

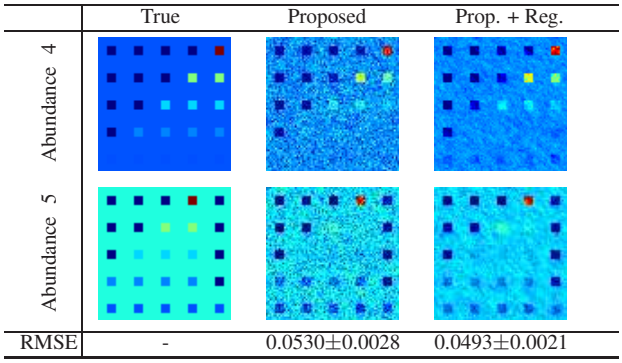

Fig. 2. Effect of locally-spatial regularization

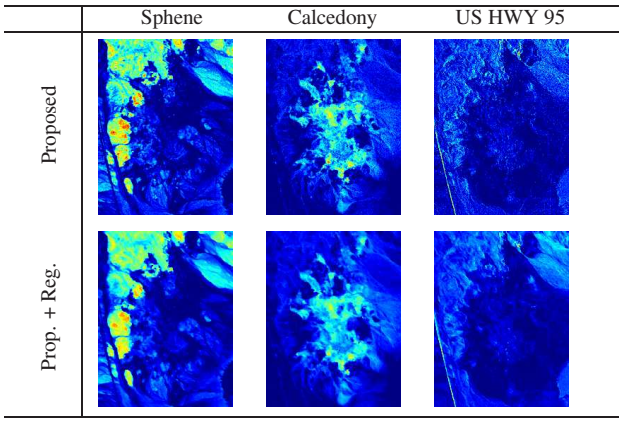

Fig. 3. Unmixing results for the Cuprite district scene

6. CONCLUSION AND PERSPECTIVES

We presented a least-squares support-vector regression approach for hyperspectral data unmixing. The model is partially linear, that is, it combines a linear trend and a nonlinear correction term. Multi-kernel learning was used to tune the balance between the linear and nonlinear components, resulting in a convex optimization problem. A locally-spatial regularizer, which does not introduce extra computational expenses, was proposed to enhance the unmixing performance. Future work will focus on endmember extraction within this nonlinear unmixing framework.

7. REFERENCES

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

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

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

[4] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," in *Proc. of SPIE*, 2009, vol. 7477.

[5] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 441–454, 2005.

[6] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Transactions on Signal Processing*, 2013, to appear.

[7] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Proc. of IEEE International Geoscience and Remote Sensing Symposium (IGARRS)*, 2007.

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

[9] N. Dobigeon Y. Altmann, A. Halimi and J.-Y. Tourneret, "Supervised nonlinear spectral unmixing using a polynomial post nonlinear model for hyperspectral imagery," in *Proc. of IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, 2011.

[10] K. J. Guilfoyle, M. L. Althouse, and C.-I. Chang, "A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis function neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 10, pp. 2314–2318, 2001.

[11] J. Plaza, P. Martínez, R. Pérez, and A. Plaza, "Nonlinear neural network mixture models for fractional abundance estimation in AVIRIS hyperspectral images," in *Proc. AVIRIS workshop*, Pasadena, CA, 2004.

[12] J. Chen, C. Richard, and P. Honeine, "A novel kernel-based nonlinear unmixing scheme of hyperspectral images," in *The 45th Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Pacific Grove (CA), USA, Nov. 2011.

[13] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral images with multi-kernel learning," in *IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Shanghai, China, Jun. 2012.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*, University Press, Cambridge, 2004.

[15] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Granvalet, "SimpleMKL," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.

[16] A. Plaza, Q. Du, J. M. Bioucas-Dias, X. Jia, and F. Kruse, "Foreword to the special issue on spectral unmixing of remotely sensed data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 22, no. 9, pp. 1419–1434, 2011.

[17] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, 2012.

[18] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundances of hyperspectral images with l1-norm spatial regularization," *IEEE Tranactions on Geoscience and Remote Sensing*, 2012, under review.

[19] A. M. Baldridge, S. J. Hook, C. I. Grove, and G. Rivera, "The ASTER spectral library version 2.0," *Remote Sensing of Environment*, vol. 113, no. 711-715, 2009.

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

[21] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex Component Analysis: A fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, April 2005.

[22] R. N. Clark, G. A. Swayze, K. E. Livo, R. F. Kokaly, S. J. Sutley, J. B. Dalton, R. R. McDougal, and C. A. Gent, "Imaging spectroscopy: Earth and planetary remote sensing with the usgs tetracorder and expert systems," *Journal of Geophysical Research*, vol. 108, no. E12, pp. 5131, 2003.