Matrix Cofactorization for Joint Spatial–Spectral Unmixing of Hyperspectral Images

Adrien Lagrange, *Student Member, IEEE*, Mathieu Fauvel, *Senior Member, IEEE*, Stéphane May, and Nicolas Dobigeon, *Senior Member, IEEE*

***Abstract*— Hyperspectral unmixing aims at identifying a set of elementary spectra and the corresponding mixture coefficients for each pixel of an image. As the elementary spectra correspond to the reflectance spectra of real materials, they are often very correlated, thus yielding an ill-conditioned problem. To enrich the model and reduce ambiguity due to the high correlation, it is common to introduce spatial information to complement the spectral information. The most common way to introduce spatial information is to rely on a spatial regularization of the abundance maps. In this article, instead of considering a simple but limited regularization process, spatial information is directly incorporated through the newly proposed context of spatial unmixing. Contextual features are extracted for each pixel, and this additional set of observations is decomposed according to a linear model. Finally, the spatial and spectral observations are unmixed jointly through a cofactorization model. In particular, this model introduces a coupling term used to identify clusters of shared spatial and spectral signatures. An evaluation of the proposed method is conducted on synthetic and real data and shows that results are accurate and also very meaningful since they describe both spatially and spectrally the various areas of the scene.**

***Index Terms*— Cofactorization, hyperspectral imaging, image analysis, spectral unmixing.**

I. Introduction

OVER the past decades, the huge potential of Earth observation has pushed the scientific community to develop automatic methods to extract information from the acquired data. Hyperspectral imaging is a specific image modality proposing very rich information in the spectral domain. Each pixel is indeed a dense sampling of the reflectance spectrum of the underlying area with usually hundreds of measurements from visible to infrared domains. The particularities of hyperspectral images have led to the development of specific interpretation methods in order to fully benefit from this spectral information. In particular, spectral unmixing methods [1] are based on the assumption that the reflectance spectrum of a pixel is the result of the mixture of a reduced set of elementary spectra called endmembers. Each of these endmembers is the reflectance spectrum corresponding to a specific material present in the scene. An unmixing method aims at estimating the existing endmembers and recovering the proportions of each material in a given pixel, collected in a so-called abundance vector. These abundance vectors allow, for example, the end user to build abundance maps displaying the distribution of materials over the observed scene.

As hyperspectral images contain rich spectral information, many unmixing methods focus on exploiting it and often neglect the spatial information. Many well-established methods process pixels without taking into consideration the basic idea that neighboring pixels are often very similar. The only shared information between pixels is a common endmember matrix [2], [3]. Nevertheless, advanced methods have been proposed to perform spatially informed spectral unmixing [4]. The most direct approach is to consider the local spatial regularization of the abundance maps. Several works, such as SUnSAL-TV [5] or S2WSU [6], proposed to use the TV-norm regularization to achieve this goal. Identifying clusters of spectrally similar pixels, gathered in homogeneous groups, has also been used to impose spatial smoothing of the abundances (see [7]–[9]). In a different way, other works used the local neighborhood to identify the subset of endmembers present in the neighborhood. It is especially useful when dealing with a large number of endmembers [10], [11]. Finally, to a lesser extent, spatial information has also been used to help the extraction of endmembers. Indeed, endmember extraction is often performed before estimating the abundance vectors. Some preprocessing steps were proposed to ease the extraction and the identification of pure pixels as the averaging of spectra over superpixels [12] or the use of spatial homogeneity scalar factors [13].

Overall, it is noticeable that all these approaches tend to exploit the very simple idea that neighboring pixels should be spectrally similar. However, the spatial information contained

This work was supported in part by the Centre National d'Études Spatiales (CNES), Occitanie Region, EU FP7, through the ERANETMED JC-WATER Program under Project ANR-15-NMED-0002-02 MapInvPlnt, in part by the ANR-3IA Artificial and Natural Intelligence Toulouse Institute (ANITI), and in part by the European Research Council under Grant ERC FACTORY-CoG-681839. *(Corresponding author: Adrien Lagrange.)*

Adrien Lagrange is with the University of Toulouse, IRIT/INP-ENSEEIHT, 31000 Toulouse, France (e-mail: adrien.lagrange@enseeiht.fr).
Mathieu Fauvel is with CESBIO, University of Toulouse, CNES/CNRS/INRA/IRD/UPS, 31400 Toulouse, France (e-mail: mathieu.fauvel@inra.fr).
Sté hane May is with the Centre National d'Études Spatiales (CNES), DCT/SI/AP, 31400 Toulouse, France (e-mail: stephane.may@cnes.fr).
Nicolas Dobigeon is with the University of Toulouse, IRIT/INP-ENSEEIHT, 31000 Toulouse, France, and also with the Institut Universitaire de France (IUF), France
(e-mail:

Digital Object Identifier 10.1109/TGRS.2020.2968541

in remote sensing images is richer than this simple statement. This article attempts to show that the conventional hyperspectral unmixing approach can leverage the spatial information to help for spectral discrimination. It relies on the hypothesis that very spectrally similar pixels can be discriminated by analyzing their spatial contexts. For instance, vegetated areas generally lead to challenging unmixing tasks due to high correlations between signatures associated with distinct vegetation types. However, it may generally be easier to discriminate these types of vegetation by analyzing their respective spatial contexts, even extracted from a grayscale panchromatic image. As an example, hardwoods are expected to exhibit canopies different from conifers, thus resulting in different spatial textures. Similarly, crops are arranged with specific spatial patterns different from those characterizing grassland. Exploiting spatial patterns and textures descriptors is, thus, expected to be helpful to the unmixing process. To exploit this assumption, this article proposes a model based on a cofactorization task to jointly infer common spatial and spectral signatures from the image.

Cofactorization methods, sometimes referred to as coupled dictionary learning, have been implemented with success in many application fields, e.g., for text mining [14], music source separation [15] and image analysis [16], [17]. The main idea is to define an optimization problem relying on two factorizing models supplemented by a coupling term enforcing a dependence between the two models. The method proposed in this article jointly considers a spectral unmixing model and a decomposition of contextual features computed from a panchromatic image of the same scene. The coupling term is interpreted as clustering identifying groups of pixels sharing similar spectral signatures and spatial contexts. This method exhibits two major advantages: 1) it provides very competitive results even though the method is unsupervised (i.e., it estimates both endmember signatures and abundance maps) and 2) it provides very insightful results since the scene is partitioned into areas characterized by spectral and spatial signatures.

The remaining of this article is organized as follows. Section II defines the spectral and spatial models and further discusses the joint cofactorization problem. Section III details the optimization scheme developed to solve the resulting nonconvex nonsmooth minimization problem. An evaluation of the proposed joint model is conducted first on synthetic data in Section IV and then on real data in Section V. Finally, Section VI concludes this article and presents some research perspectives to this article.

II. Toward Spatial–Spectral Unmixing

The main goal of this section is to introduce a model capable of spectrally and spatially characterizing a hyperspectral image. In particular, instead of incorporating prior spatial information as a regularization [5], the concept of spatial unmixing, detailed in Section II-B, is introduced alongside a conventional spectral unmixing model in order to propose a new joint framework of spatial–spectral unmixing.

A. Spectral Mixing Model

Spectral unmixing aims at identifying the elementary spectra and the proportion of each material in a given pixel [1]. Each of the P pixels $\mathbf{y}_p$ is a d_1-dimensional measurement of a reflectance spectrum and is assumed to be a combination of R_1 elementary spectra $\mathbf{m}_r$, called endmembers, with $R_1 \ll d_1$. The so-called abundance vector $\mathbf{a}_p \in \mathbb{R}^{R_1}$ refers to the corresponding mixing coefficients in this pixel. In a general case, where no particular assumption is made on the observed scene, the conventional linear mixture model (LMM) is widely adopted to describe the mixing process. It assumes that the observed mixtures are linear combinations of the endmembers. Within an unsupervised framework, i.e., when both endmember signatures and abundances should be recovered, linear spectral unmixing can be formulated as the following minimization problem:

$$\min_{\mathbf{M},\mathbf{A}} \ \|\mathbf{Y} - \mathbf{M}\mathbf{A}\|_{\mathrm{F}}^2 + \iota_{\mathbb{R}_+^{d_1 \times R_1}}(\mathbf{M}) + \iota_{\mathbb{S}_{R_1}^P}(\mathbf{A}) \tag{1}$$

where the matrices $\mathbf{Y} \in \mathbb{R}^{d_1 \times P}$ gather all the observed pixels, $\mathbf{M} \in \mathbb{R}^{d_1 \times R_1}$ are the endmembers, $\mathbf{A} \in \mathbb{R}^{R_1 \times P}$ are the abundance vectors, and $\iota_{\mathbb{R}_+^{d_1 \times R_1}}(\cdot)$ and $\iota_{\mathbb{S}_{R_1}^P}(\cdot)$ are, respectively, indicator functions on the nonnegative quadrant and the column-wise indicator function on the R_1-dimensional probability simplex denoted by $\mathbb{S}_{R_1}$. The nonnegative constraint over $\mathbf{M}$ is justified by the fact that endmember signatures are reflectance spectra and, thus, nonnegative. The second indicator function enforces nonnegative and sum-to-one constraints on the abundance vectors $\mathbf{a}_p$ $(p = 1, \ldots, P)$ in order to interpret them as proportion vectors. It is worth noting that the sum-to-one constraint is sometimes disregarded since it has been argued that relaxing this constraint out offers a better adaptation to possible changes of illumination in the scene [18]. Due to the usual ill-conditioning of the endmember matrix $\mathbf{M}$, the objective function underlying (1) is often granted with additional regularizations promoting expected properties of the solution. In particular, numerous works exploited the expected spatial behavior of the mixing coefficients to introduce spatial regularizations enforcing piecewise constant [5], [8] or smoothly varying [2], [3] abundance maps, possibly driven by external knowledge [19]. Conversely, this article does not consider spatial information as a prior knowledge but rather proposes a decomposition model dedicated to the image spatial content, paving the way toward the concept of spatial unmixing. This contribution is detailed in Sections II-B–II-D.

B. Spatial Mixing Model

As previously mentioned, this article proposes to complement the conventional linear unmixing problem (1) with an additional data-fitting term accounting for spatial information already contained in the hyperspectral image. To do so, for sake of generality, we assume that the scene of interest is characterized by vectors of spatial features $\mathbf{s}_p \in \mathbb{R}^{d_2}$ describing the context around the corresponding hyperspectral pixel indexed by p. The features can be extracted from the hyperspectral image directly or from any other available

images of any modality of the same scene, with possibly better spatial resolution. A common choice for designing these features will be discussed later. To capture common spatial patterns, akin to a so-called spatial unmixing, these P d_2-dimensional spatial feature vectors $\mathbf{s}_p$ gathered in a matrix $\mathbf{S} \in \mathbb{R}^{d_2 \times P}$ are linearly decomposed and recovered from the optimization problem

$$\min_{\mathbf{D},\mathbf{U}} \ \|\mathbf{S} - \mathbf{D}\mathbf{U}\|_{\mathrm{F}}^2 + \iota_{\mathbb{R}_+^{d_2 \times R_2}}(\mathbf{D}) + \iota_{\mathbb{S}_{R_2}^P}(\mathbf{U}) \tag{2}$$

where $\mathbf{D} \in \mathbb{R}^{d_2 \times R_2}$ is a dictionary matrix and $\mathbf{U} \in \mathbb{R}^{R_2 \times P}$ is the corresponding coding matrix.

The spatial model underlying (2) can be interpreted as a dictionary-based representation learning task. It means that the image in the considered feature space can be decomposed as a sum of elementary patterns collected in the matrix $\mathbf{D}$ of spatial signatures. The corresponding coding coefficients are gathered in $\mathbf{U}$. The nonnegativity constraints are imposed to ensure an additive decomposition similar to what is done in the context of nonnegative matrix factorization (NMF) [20]. Finally, without any constraint on the norms of $\mathbf{U}$ and $\mathbf{D}$, the problem would suffer from a scaling ambiguity between $\mathbf{U}$ and $\mathbf{D}$. To cope with this issue, additional sum-to-one constraints are imposed on the columns of $\mathbf{U}$. This choice, which somehow brings some loss of generality when compared to normalizing the rows of $\mathbf{U}$, leads to possibly amplitude-varying atoms in $\mathbf{D}$. In other words, to describe similar spatial patterns of different amplitudes, additional atoms should be included in the spatial dictionary $\mathbf{D}$. However, normalizing the columns in $\mathbf{U}$ has the advantage of leading to spatial and spectral representation vectors of the same unit norm, which prevents any unbalances in the coupling process introduced in Section II-C.

It is worth noting that a model similar to (2) was implicitly assumed in [21]–[23], where a single-band image acquired by scanning transmission electron microscopy is linearly unmixed by the principal component analysis [24], independent component analysis [25], N-FINDR [26], or deep convolutional neural networks. In these works, the spatial feature space is defined by the magnitude of a sliding 2-D-discrete Fourier transform, which unlikely ensures the additivity, or at least linear separability, assumptions underlying the mixtures. As an alternative, the strategy adopted in this article relies on a patch-based representation of the image, as popularized by several seminal contributions in the literature. Combined with a linear decomposition underlying (2), this representation can be easily motivated by the intrinsic property of image self-similarity. It has already shown its practical interest for various image processing tasks, including classification and denoising. As archetypal examples, patch-based dictionary learning methods leverage on linear models similar to (2) to capture spatial redundancies [27]–[30]. This self-similarity property is also a key assumption to motivate linear aggregation steps in many nonlocal denoising techniques such as NL-means [31] and BM3D [32]. Its simplicity makes this model popular in many application fields, including medical imaging [33] and photography [34]. Thus, in the numerical experiments reported in Sections IV and V, the spatial features will be chosen as the elementary patches extracted from the virtual panchromatic image computed by averaging the hyperspectral bands.

C. Coupling Spatial and Spectral Mixing Models

Sections II-A and II-B have defined two matrix factorization problems associated with two unmixing tasks considered independently. To mutually benefit from spectral and spatial information brought by the image, these two tasks should be considered jointly. A natural approach consists in coupling the two tasks by relating the coding factors involved in the two matrix decompositions. A hard coupling between the coding matrices, which would consist in constraining spectral and spatial abundance maps to be equal (i.e., $\mathbf{A} = \mathbf{U}$), will be shown to be not sufficiently flexible to account for complex interactions between spatial and spectral information (see Section IV). Conversely, the proposed approach relies on the simple yet sound assumption that the pixels in the scene obey a small number of spectral and spatial behaviors that can be clustered. This implicitly means that a given elementary material, uniquely represented by a single endmember spectrum, is expected to be present in the scene with a small number of distinct spatial configurations. Thus, the coupling is designed as a clustering task which aims at recovering common behaviors jointly exhibited by the spatial and spectral abundance maps. More precisely, this clustering task can be formulated as another NMF problem

$$\min_{\mathbf{B},\mathbf{Z}} \ \left\| \begin{pmatrix} \mathbf{A} \\ \mathbf{U} \end{pmatrix} - \mathbf{B}\mathbf{Z} \right\|_{\mathrm{F}}^2 + \frac{\lambda_z}{2}\mathrm{Tr}(\mathbf{Z}^T\mathbf{V}\mathbf{Z}) \tag{3}$$

$$+ \ \iota_{\mathbb{R}_+^{(R_1+R_2)\times K}}(\mathbf{B}) + \iota_{\mathbb{S}_K^P}(\mathbf{Z}) \tag{4}$$

with $\mathbf{V} = \mathbf{1}_K\mathbf{1}_K^T - \mathbf{I}_K$, where $\mathbf{I}_K$ is the $K \times K$ identity matrix, $\mathbf{1}_K$ is the $K \times 1$ vector of ones, and $\mathrm{Tr}(\cdot)$ is the trace operator. The two coding matrices $\mathbf{A}$ and $\mathbf{U}$ are concatenated, and the clustering is conducted on the columns of the resulting whole coding matrix. The matrix $\mathbf{Z} \in \mathbb{R}^{K \times P}$ describes the assignments to the clusters, where $\mathbf{z}_p$ gathers the probabilities of belonging to each of the clusters; hence, the nonnegativity and sum-to-one constraint enforced on it. It is accompanied with a specific regularization [see the second term in (3)]. This penalty promotes orthogonality over the lines of $\mathbf{Z}$ since it can be rewritten as $\mathrm{Tr}(\mathbf{Z}^T\mathbf{V}\mathbf{Z}) = \sum_{k_1 \neq k_2} \langle \mathbf{z}_{k_1,:} | \mathbf{z}_{k_2,:} \rangle$, where $\langle \cdot | \cdot \rangle$ stands for the scalar product. Due to nonnegativity of the elements of $\mathbf{Z}$, this term becomes minimal when the assignments to clusters obey a hard decision, i.e., when one component of $\mathbf{z}_p$ is equal to 1 and the others are set to 0. Thus, a strict orthogonality constraint would make the clustering problem (3) equivalent to a k-means problem [35]. Centroids of the K clusters define the columns of the matrix $\mathbf{B} \in \mathbb{R}^{(R_1+R_2)\times K}$. Interestingly, each centroid is then the concatenation of a mean spectral abundance and a mean spatial abundance. In particular, it means that the pixels of a given cluster share the same spectral properties and a similar spatial context. Indeed, these centroids, when combined with the spectral and spatial signature matrices $\mathbf{M}$ and $\mathbf{D}$, provide a compact representation of the spectral and spatial contents of each cluster. More precisely, each column of the

$(d_1 + d_2) \times K$-matrix defined by

$$\begin{pmatrix} \bar{\mathbf{M}} \\ \bar{\mathbf{D}} \end{pmatrix} = \begin{pmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} \mathbf{B} \tag{5}$$

can be interpreted as the spatial–spectral signature of each cluster, resulting from the concatenation of a mean spectral signature $\bar{\mathbf{m}}_k$ and a mean spatial signature $\bar{\mathbf{d}}_k$ $(k = 1, \ldots, K)$.

D. Joint Spatial–Spectral Unmixing Problem

Given the spectral mixing model recalled in Section II-A, the spatial mixing model introduced in Section II-B, and their coupling term proposed in Section II-C, we propose to conduct spatial–spectral unmixing jointly by considering the overall minimization problem

$$\begin{aligned} \min_{\mathbf{M},\mathbf{A},\mathbf{D},\mathbf{U},\mathbf{B},\mathbf{Z}} \; & \frac{\lambda_0}{2}\|\mathbf{Y} - \mathbf{M}\mathbf{A}\|_{\mathrm{F}}^2 + \iota_{\mathbb{R}_+^{d_1 \times R_1}}(\mathbf{M}) + \iota_{\mathbb{S}_{R_1}^P}(\mathbf{A}) \\ & + \frac{\lambda_1}{2}\|\mathbf{S} - \mathbf{D}\mathbf{U}\|_{\mathrm{F}}^2 + \iota_{\mathbb{R}_+^{d_2 \times R_2}}(\mathbf{D}) + \iota_{\mathbb{S}_{R_2}^P}(\mathbf{U}) \\ & + \frac{\lambda_2}{2}\left\| \begin{pmatrix} \mathbf{A} \\ \mathbf{U} \end{pmatrix} - \mathbf{B}\mathbf{Z} \right\|_{\mathrm{F}}^2 + \frac{\lambda_z}{2}\mathrm{Tr}(\mathbf{Z}^T\mathbf{V}\mathbf{Z}) \\ & + \iota_{\mathbb{R}_+^{(R_1+R_2)\times K}}(\mathbf{B}) + \iota_{\mathbb{S}_K^P}(\mathbf{Z}) \end{aligned} \tag{6}$$

where λ_0, λ_1, and λ_2 adjust the respective contribution of the various fitting terms. It is worth noting that because of the sum-to-one constraints enforced on the spectral abundance vectors $\mathbf{a}_p$ and spatial abundance vectors $\mathbf{u}_p$, all these coding vectors have the same unitary ℓ_1-norm. It has the great advantage of avoiding a reweighing of $\mathbf{A}$ and $\mathbf{U}$ in the coupling term regardless of the number of endmembers and dictionary atoms. Section III describes the optimization scheme adopted to solve the joint spatial–spectral unmixing problem (6).

III. Optimization Scheme

A. Proximal Alternating Linearized Minimization (PALM) Algorithm

The cofactorization problem (6) is a nonconvex, nonsmooth optimization problem. For these reasons, the problem remains very challenging to solve and requires the use of advanced optimization tools. The choice has been made to resort to the PALM algorithm [36]. The core concept of PALM is to update each block of variables alternatively according to a proximal gradient descent step. This algorithm has the advantage of ensuring convergence to a critical point of the objective function even in the case of a nonconvex, nonsmooth problem.

In order to obtain these convergence results, the objective function has to ensure a specific set of properties. First, the various terms of the objective function have to be separable in a sum of one smooth term $g(\cdot)$ and a set of independent nonsmooth terms. Then, each of the independent nonsmooth term has to be a proper, lower semicontinuous function $f_i : \mathbb{R}^{n_i} \to (-\infty, +\infty]$, where n_i is the input dimension of f_i. Finally, a sufficient condition is that the smooth term is a $\mathcal{C}^2$-continuous function and its partial gradients are globally Lipschitz with respect to the derivative variable. Further details are available in [36].

Algorithm 1 PALM

Initialize variables $\mathbf{M}^0$, $\mathbf{A}^0$, $\mathbf{D}^0$, $\mathbf{U}^0$, $\mathbf{B}^0$ and $\mathbf{Z}^0$;
Set $\alpha > 1$;
while *stopping criterion not reached* **do**
  $\mathbf{M}^{k+1} \in \mathrm{prox}^{\alpha L_{\mathbf{M}}}_{\iota_{\mathbb{R}_+^{d_1\times R_1}}}(\mathbf{M}^k - \frac{1}{\alpha L_{\mathbf{M}}}\nabla_{\mathbf{M}} g(\mathbf{M}^k, \mathbf{A}^k, \mathbf{D}^k, \mathbf{U}^k, \mathbf{B}^k, \mathbf{Z}^k))$;
  $\mathbf{A}^{k+1} \in \mathrm{prox}^{\alpha L_{\mathbf{A}}}_{\iota_{\mathbb{S}_{R_1}^P}}(\mathbf{A}^k - \frac{1}{\alpha L_{\mathbf{A}}}\nabla_{\mathbf{A}} g(\mathbf{M}^{k+1}, \mathbf{A}^k, \mathbf{D}^k, \mathbf{U}^k, \mathbf{B}^k, \mathbf{Z}^k))$;
  $\mathbf{D}^{k+1} \in \mathrm{prox}^{\alpha L_{\mathbf{D}}}_{\iota_{\mathbb{R}_+^{d_2\times R_2}}}(\mathbf{D}^k - \frac{1}{\alpha L_{\mathbf{D}}}\nabla_{\mathbf{D}} g(\mathbf{M}^{k+1}, \mathbf{A}^{k+1}, \mathbf{D}^k, \mathbf{U}^k, \mathbf{B}^k, \mathbf{Z}^k))$;
  $\mathbf{U}^{k+1} \in \mathrm{prox}^{\alpha L_{\mathbf{U}}}_{\iota_{\mathbb{S}_{R_2}^P}}(\mathbf{U}^k - \frac{1}{\alpha L_{\mathbf{U}}}\nabla_{\mathbf{U}} g(\mathbf{M}^{k+1}, \mathbf{A}^{k+1}, \mathbf{D}^{k+1}, \mathbf{U}^k, \mathbf{B}^k, \mathbf{Z}^k))$;
  $\mathbf{B}^{k+1} \in \mathrm{prox}^{\alpha L_{\mathbf{B}}}_{\iota_{\mathbb{R}_+^{(R_1+R_2)\times K}}}(\mathbf{B}^k - \frac{1}{\alpha L_{\mathbf{B}}}\nabla_{\mathbf{B}} g(\mathbf{M}^{k+1}, \mathbf{A}^{k+1}, \mathbf{D}^{k+1}, \mathbf{U}^{k+1}, \mathbf{B}^k, \mathbf{Z}^k))$;
  $\mathbf{Z}^{k+1} \in \mathrm{prox}^{\alpha L_{\mathbf{Z}}}_{\iota_{\mathbb{S}_K^P}}(\mathbf{Z}^k - \frac{1}{\alpha L_{\mathbf{Z}}}\nabla_{\mathbf{Z}} g(\mathbf{M}^{k+1}, \mathbf{A}^{k+1}, \mathbf{D}^{k+1}, \mathbf{U}^{k+1}, \mathbf{B}^{k+1}, \mathbf{Z}^k))$;
end
return $\mathbf{M}^{end}$, $\mathbf{A}^{end}$, $\mathbf{D}^{end}$, $\mathbf{U}^{end}$, $\mathbf{B}^{end}$, $\mathbf{Z}^{end}$

In problem (6), the smooth term $g(\cdot)$ is composed of the three quadratic terms and the orthogonality-promoting regularization. All these terms verify the gradient Lipschitz and $\mathcal{C}^2$-continuous properties. Moreover, the nonsmooth terms f_i are separable into independent terms. Moreover, since they are all indicators functions on convex sets, their proximal operators are well-defined and, more specifically, are defined as the projection on the corresponding convex set. The projection on the nonnegative quadrant is a simple thresholding of the negative values, and the projection on the probability simplex can be achieved by a simple sort followed by a thresholding as described in [37].

A summary of the overall optimization scheme is given in Algorithm 1, where $L_{\mathbf{X}}$ stands for the Lipschitz constant of the gradient of $g(\cdot)$ considered as a function of $\mathbf{X}$. Partial gradients and Lipschitz moduli are all provided in the Appendix. Additional details regarding the implementation are discussed in Algorithm 1.

B. Implementation Details

1) Initialization and Convergence: As explained, the PALM algorithm only ensures convergence to a critical point of the objective function. Hence, it is important to have a good initialization of the variables to be estimated. In the following experiments, the initial endmember matrix $\mathbf{M}^0$ has been chosen as the output of the vertex component analysis (VCA) [38]. Abundance matrix is then initialized by solving the fully constrained least-squares (FCLS) problem $\min_{\mathbf{A}\in\mathbb{S}_{R_1}^P} \|\mathbf{Y} - \mathbf{M}\mathbf{A}\|_{\mathrm{F}}^2$. Finally, $\mathbf{D}^0$ and $\mathbf{U}^0$ are initialized by performing a k-means algorithm on columns of $\mathbf{S}$. Similarly, $\mathbf{B}^0$ and $\mathbf{Z}^0$ are initialized by a k-means on the concatenation of $\mathbf{U}^0$ and $\mathbf{A}^0$.

As stated in Algorithm 1, a criterion is needed to monitor the convergence of the optimization algorithm. In the following experiments, the residual error of the objective function is computed at each iteration, and when the relative gap between the two last iterations is below a given threshold (10^{-4} for these experiments), the algorithm is stopped.

2) Hyperparameters: Several weighting coefficients $\lambda_\cdot$ have been introduced in problem (6) to adjust the respective contribution of each term. In the following experiments, some of these coefficients have been renormalized to take in consideration the respective dimensions and dynamics of the matrices defining each term, thus yielding:

$$\begin{cases} \lambda_0 = \dfrac{1}{d_1 \|\mathbf{Y}\|_\infty^2} \tilde{\lambda}_0 \\ \lambda_1 = \dfrac{1}{d_2 \|\mathbf{S}\|_\infty^2} \tilde{\lambda}_1. \end{cases} \tag{7}$$

IV. Experiments Using Simulated Data

Performance of the proposed spatial–spectral unmixing method has been assessed by conducting experiments on both synthetic and real data. The use of synthetic data makes quantitative validation possible, whereas it is not possible with real data since there is no reference data.

A. Data Generation

In order to properly evaluate the relevance of the proposed model, two synthetic images referred to as Image 1 and Image 2 have been generated such that they incorporate consistent spatial and spectral information. Note that the proposed simulation framework, detailed hereafter, does not directly rely on the forward model underlying the joint spectral–spatial unmixing problem in (6).

Each image is supposed to be composed of J homogeneous regions with common spatial and spectral characteristics. These regions have been randomly generated according to a Potts–Markov random field with J classes [39]. Each region is characterized by specific (yet nonpiecewise constant) spatial and spectral contents whose generation processes are described in the following. According to the linear mixing model, the observed pixel spectra in each region are assumed to result from the linear combination of R_1 endmembers gathered in the common matrix $M = [\mathbf{m}_1, \ldots, \mathbf{m}_{R_1}]$

$$\forall p \in \mathcal{P}_j, \quad \mathbf{y}_p = \mathbf{M}\mathbf{a}_p^{(j)} \tag{8}$$

where $\mathcal{P}_j$ denotes the set of pixels belonging to the jth region and $\mathbf{a}_p^{(j)} \in \mathbb{S}_{R_1}$ is the abundance vectors of the pth pixel in the jth region. The endmember signatures in M have been extracted from the ASTER library [40]. Each of these regions is characterized by a particular texture. To ensure realistic spatial patterns, the J textures associated with the J regions are extracted from real panchromatic images. They are depicted in Fig. 1 for Image 1 and Image 2. The intensity of the pth pixel of the grayscale image texture associated with the jth region is denoted by $t_p^{(j)} \in (0, 1)$ ($p \in \mathcal{P}_j$). The key ingredient is the appropriate design of the abundance vectors $\mathbf{a}_p^{(j)}$ ($p \in \mathcal{P}_j$) that jointly encode the spatial and

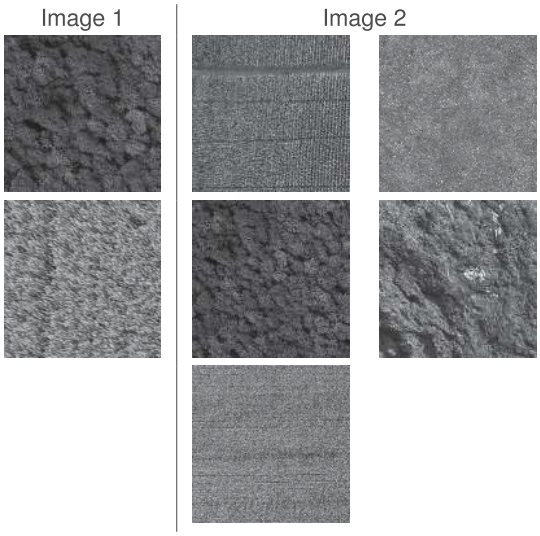

Fig. 1. Synthetic data set: textures (forest and wheat) for Image 1 and (corn, grass, forest, rock, and wheat) for Image 2.

spectral contents in the jth region. To include consistent spatial information, these abundance vectors are assumed to be convex combinations of two predefined extreme spectral behaviors $\boldsymbol{\psi}_i^{(j)} \in \mathbb{S}_{R_1}$ ($i \in \{1, 2\}$), that is

$$\mathbf{a}_p^{(j)} = t_p^{(j)} \boldsymbol{\psi}_1^{(j)} + \left(1 - t_p^{(j)}\right) \boldsymbol{\psi}_2^{(j)} \tag{9}$$

where the grayscale intensity $t_p^{(j)} \in (0, 1)$ of the jth texture modulates the spectral content in the pth pixel of the jth region. This supports the idea that a texture can be seen as small spatial variations of the proportions of elementary components between extreme spectral signatures. In particular, a white pixel in the jth texture (i.e., $t_p^{(j)} = 1$) leads to a spectral content only driven by $\boldsymbol{\psi}_1^{(j)}$, and the corresponding pixel in jth region of the generated hyperspectral image is $\mathbf{y}_p = \mathbf{M}\boldsymbol{\psi}_1^{(j)}$. Conversely, a black pixel in the jth texture (i.e., $t_p^{(j)} = 0$) leads to a spectral content only driven by $\boldsymbol{\psi}_2^{(j)}$, and the corresponding pixel is $\mathbf{y}_p = \mathbf{M}\boldsymbol{\psi}_2^{(j)}$. Of course, gray pixels in the texture lead to mixed behaviors with $\mathbf{y}_p = t_p^{(j)}\mathbf{M}\boldsymbol{\psi}_1^{(j)} + (1 - t_p^{(j)})\mathbf{M}\boldsymbol{\psi}_2^{(j)}$. The generated abundance maps are shown in Fig. 2. Again, note that this simulation protocol does not rely on the cofactorization formalism underlying the proposed algorithm.

Two images have been generated according to this process. Image 1 is a 200×200-pixel image with 385 spectral bands composed of $R_1 = 4$ endmembers and $J = 2$ regions. Image 2 is a 300×300-pixel image with 385 spectral bands with $R_1 = 9$ endmembers and $J = 5$ regions. Note that for these two images, the texture-based modulating intensity $t_p^{(j)}$ never reaches the extreme values 0 and 1 ($\forall p, \forall j$). In addition, from the two hyperspectral images, corresponding panchromatic images have been generated by first dividing each band individually by its empirical mean over the pixels and then summing all these normalized bands for each pixel. The contrast has been finally adjusted such that the minimum and maximum values of the image are 0 and 255, respectively,

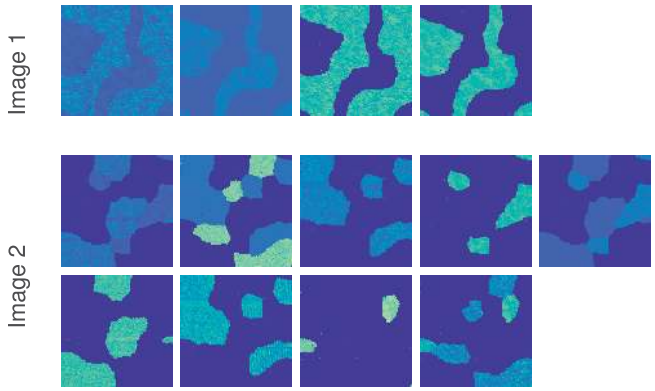

Fig. 2. Synthetic data set: abundance maps.

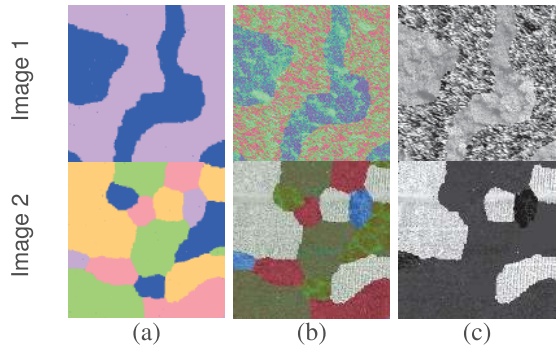

Fig. 3. Synthetic data set. (a) Segmentation map. (b) Color composition of the hyperspectral image. (c) Panchromatic image.

to obtain an 8-bit image. The generated hyperspectral and panchromatic images are shown in Fig. 3.

B. Compared Methods

In order to assess the performance of the proposed spatial–spectral unmixing model, referred to as SP2U, the unmixing results have been compared with those of several well-established methods. First, the result of the initialization method has been used as a baseline. This method is conventional [2] and consists in extracting endmembers using the VCA method [38] and then solving an FCLS problem. This first method is referred to as VCA + FCLS hereafter.

The second compared method uses again an FCLS method to estimate the abundance vectors but uses an alternative endmember extraction algorithm. This method, called SISAL [41], tries to estimate the minimum volume simplex containing the observed hyperspectral data by solving a nonconvex problem using a splitting augmented Lagrangian technique.

The third compared method relies on a similar linear mixing model assumed by VCA + FCLS and SISAL + FCLS. However, instead of estimating the endmember signatures and abundances sequentially, it performs a joint estimation, thus yielding an NMF task with an additional sum-to-one constraint. This method referred to as NMF in the following is a depreciated version of the SP2U problem (6), where $\lambda_1 = \lambda_2 = \lambda_z = 0$, and has been solved and initialized similarly.

The fourth method SUnSAL-TV was introduced in [5] and proposes to solve a conventional linear unmixing problem with an additional spatial regularization term to incorporate spatial information. The regularization term is chosen as a total variation applied to the abundance maps $\mathbf{A}$. It promotes in particular similarity of abundance vectors of neighboring pixels. In this case, the local information is used, whereas SP2U method relates pixels sharing the same spatial context, akin to a nonlocal framework. It is important to note that this method does not estimate the endmember matrix, which is estimated beforehand using VCA or SISAL.

The fifth method, denoted by n-SP2U, is a naive counterpart of the proposed SP2U method. Instead of using the coupling term introduced in Section II-C, the abundance matrix $\mathbf{A}$ and the coding coefficients $\mathbf{U}$ are directly considered equal, thus yielding the following problem:

$$\min_{\mathbf{M},\mathbf{A},\mathbf{D}} \frac{\lambda_0}{2}\|\mathbf{Y}-\mathbf{M}\mathbf{A}\|_{\mathrm{F}}^2 + \iota_{\mathbb{R}_+^{d_1\times R_1}}(\mathbf{M}) + \frac{\lambda_1}{2}\|\mathbf{S}-\mathbf{D}\mathbf{A}\|_{\mathrm{F}}^2 + \iota_{\mathbb{R}_+^{d_2\times R_2}}(\mathbf{D}) + \iota_{\mathbb{S}_{R_1}^P}(\mathbf{A}). \quad (10)$$

This method is considered for comparison since it may seem natural when willing to couple factorizations associated with spatial and spectral unmixing. However, it actually appears very unlikely to perform well in real scenarios. Indeed, the naive model n-SP2U actually enforces the same size of dictionary for the spatial mixing and the spectral mixing. In realistic cases, we expect to have less spectral signatures than spatial signatures. Indeed, a given elementary material can be presented with several spatial patterns, whereas a given spatial pattern would be unlikely associated with distinct spectral signatures. To account for the case of a given spectral signature associated with several spatial signatures, the n-SP2U model would need a larger endmember matrix $\mathbf{M}$ with several equal columns, each one associated with a specific spatial signature in $\mathbf{D}$. The resulting estimated abundance matrix $\mathbf{A}$ would need to follow a fusion process to sum abundance maps corresponding to the same spectral signatures.

The last compared method, denoted by c-SPU, is another simplified version of SP2U, where the spatial data fitting term has been removed. It corresponds to an NMF-based spectral unmixing method combined with a clustering of the spectral abundance vectors. The c-SPU method solves the problem

$$\min_{\mathbf{M},\mathbf{A},\mathbf{B},\mathbf{Z}} \frac{\lambda_0}{2}\|\mathbf{Y}-\mathbf{M}\mathbf{A}\|_{\mathrm{F}}^2 + \iota_{\mathbb{R}_+^{d_1\times R_1}}(\mathbf{M}) + \iota_{\mathbb{S}_{R_1}^P}(\mathbf{A}) + \frac{\lambda_2}{2}\|\mathbf{A}-\mathbf{B}\mathbf{Z}\|_{\mathrm{F}}^2 + \frac{\lambda_z}{2}\mathrm{Tr}(\mathbf{Z}^T\mathbf{V}\mathbf{Z}) + \iota_{\mathbb{R}_+^{(R_1+R_2)\times K}}(\mathbf{B}) + \iota_{\mathbb{S}_K^P}(\mathbf{Z}). \quad (11)$$

Comparing the results provided by c-SPU with those obtained by the proposed method SP2U aims at demonstrating the interest of considering the spatial data fitting term. At a lesser extent, it will also measure the benefit of introducing the clustering term, which is expected to act as a regularizer for the spectral abundance matrix $\mathbf{A}$.

TABLE I

Image 1: QUANTITATIVE RESULTS (AVERAGED OVER TEN TRIALS)

Model	aSAM($\mathbf{M}$)	RE	RMSE($\mathbf{A}$)	Time (s)
VCA+FCLS	0.180 ($\pm 1.1 \times 10^{-2}$)	6.86×10^{-3} ($\pm 6.3 \times 10^{-3}$)	0.150 ($\pm 1.9 \times 10^{-2}$)	19 (± 11)
SISAL+FCLS	0.151 ($\pm 3.4 \times 10^{-3}$)	2.81×10^{-3} ($\pm 3.5 \times 10^{-6}$)	0.114 ($\pm 3.9 \times 10^{-3}$)	23 (± 0.1)
NMF	0.175 ($\pm 5.6 \times 10^{-3}$)	3.86×10^{-3} ($\pm 9.8 \times 10^{-4}$)	0.151 ($\pm 2.1 \times 10^{-2}$)	27 (± 29)
VCA+SUnSAL-TV	0.180 ($\pm 1.1 \times 10^{-2}$)	7.61×10^{-3} ($\pm 4.5 \times 10^{-3}$)	0.132 ($\pm 3.2 \times 10^{-2}$)	27 (± 0.1)
SISAL+SUnSAL-TV	0.151 ($\pm 2.9 \times 10^{-3}$)	4.6×10^{-3} ($\pm 1.1 \times 10^{-4}$)	$\mathbf{0.0989}$ ($\pm 4.1 \times 10^{-3}$)	28 (± 0.3)
n-SP2U	0.188 ($\pm 1.5 \times 10^{-2}$)	28.1×10^{-3} ($\pm 1.2 \times 10^{-3}$)	0.192 ($\pm 9.6 \times 10^{-3}$)	93 (± 14)
c-SPU	0.178 ($\pm 3.8 \times 10^{-3}$)	10.9×10^{-3} ($\pm 8.0 \times 10^{-4}$)	0.139 ($\pm 9.1 \times 10^{-3}$)	**8** (± 0.5)
SP2U	$\mathbf{0.108}$ ($\pm 2.2 \times 10^{-2}$)	6.88×10^{-3} ($\pm 3.5 \times 10^{-4}$)	0.166 ($\pm 7.2 \times 10^{-2}$)	409 (± 38)

C. Performance Criteria

Performance of all methods has been assessed in terms of endmember estimation using the average spectral angle mapper (aSAM)

$$\text{aSAM}(\mathbf{M}) = \frac{1}{R_1}\sum_{r=1}^{R_1} \arccos\left(\frac{\langle \mathbf{m}_r^{(\text{ref})} | \mathbf{m}_r \rangle}{\|\mathbf{m}_r^{(\text{ref})}\|_2 \|\mathbf{m}_r\|_2}\right) \quad (12)$$

and also in terms of abundance estimation using the root-mean-square error (RMSE)

$$\text{RMSE}(\mathbf{A}) = \sqrt{\frac{1}{PR_1}\|\mathbf{A}^{(\text{ref})} - \mathbf{A}\|_{\text{F}}^2} \quad (13)$$

where $\mathbf{m}_r^{(\text{ref})}$ and $\mathbf{A}$ are the rth actual endmember signature and the actual abundance matrix, respectively.

Two additional information have also been included in the results: the processing time, which includes the initialization, the endmember extraction, and the abundances estimation, and the reconstruction error which measures how the model fits to the observed data

$$\text{RE} = \sqrt{\frac{1}{\text{Pd}_1}\|\mathbf{Y} - \mathbf{MA}\|_{\text{F}}^2}. \quad (14)$$

It is worth noting that RE should not be understood as a criterion of unmixing performance. It rather measures the ability of a given model to fit the observations. Thus, a very low RE may not be systematically suitable since it could be explained by overfitting. Conversely, a high RE may help to diagnose a model unreliability or issues in algorithmic convergences. However, when comparing a given set of methods, REs of the same order of magnitude ensure that all methods are able to describe the data with similar accuracy, and the criteria for unmixing performance (RMSE and SAM) can be compared fairly.

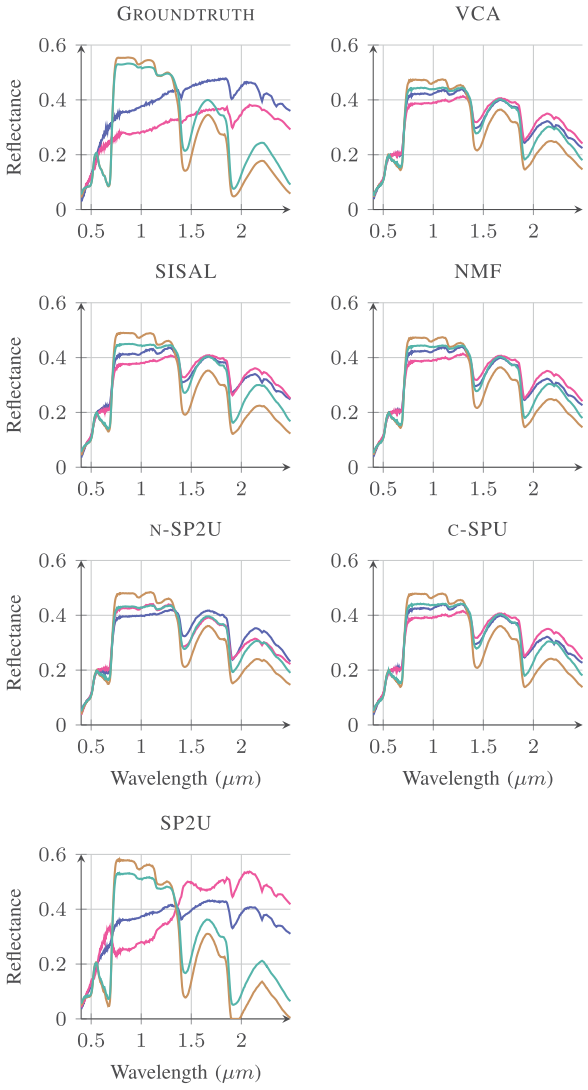

Fig. 4. Image 1: estimated endmembers.

D. Results

As stated in Section II-B, the spatial feature matrix $\mathbf{S}$ has been extracted from the panchromatic image. For each pixel, the spatial feature vector $\mathbf{s}_p$ ($p \in \{1, \ldots, P\}$) is obtained by concatenating the values of the pixels in an 11×11-pixel patch centered on the considered pixel. This choice may seem very naive, but patch-based image decompositions have proven their interest for many tasks. This choice has also the advantage of offering a direct interpretation of the spatial content and cluster centroids as small 11×11 patches. In addition, designing the most appropriate spatial feature is out of the scope of this article; its main objective is to introduce the concept of spatial–spectral unmixing. Moreover, for these experiments, the actual number of endmembers has been assumed known, and thus, $R_1 = 4$ for Image 1 and $R_1 = 9$ for Image 2. The number of dictionary atoms and clusters has been empirically adjusted and set such that $R_2 = 20$ and $K = 30$ for Image 1

TABLE II

Image 2: QUANTITATIVE RESULTS (AVERAGED OVER TEN TRIALS)

Model	aSAM($\mathbf{M}$)	RE	RMSE($\mathbf{A}$)	Time (s)
VCA+FCLS	0.176 ($\pm 5.8 \times 10^{-3}$)	8.80×10^{-3} ($\pm 2.2 \times 10^{-3}$)	0.246 ($\pm 4.2 \times 10^{-3}$)	100 ($\pm$27)
SISAL+FCLS	0.187 ($\pm 1.7 \times 10^{-2}$)	4.61×10^{-3} ($\pm 5.0 \times 10^{-6}$)	0.145 ($\pm 2.3 \times 10^{-2}$)	57 ($\pm$0.5)
NMF	0.178 ($\pm 5.9 \times 10^{-3}$)	4.87×10^{-3} ($\pm 6.3 \times 10^{-3}$)	0.246 ($\pm 4.2 \times 10^{-3}$)	109 ($\pm$26)
VCA+SUnSAL-TV	0.176 ($\pm 5.8 \times 10^{-3}$)	9.48×10^{-3} ($\pm 6.4 \times 10^{-4}$)	0.229 ($\pm 3.6 \times 10^{-3}$)	81 ($\pm$0.7)
SISAL+SUnSAL-TV	0.189 ($\pm 9.6 \times 10^{-3}$)	4.74×10^{-3} ($\pm 5.4 \times 10^{-5}$)	0.131 ($\pm 1.2 \times 10^{-2}$)	81 ($\pm$2)
n-SP2U	0.190 ($\pm 1.8 \times 10^{-2}$)	35.3×10^{-3} ($\pm 4.1 \times 10^{-3}$)	0.212 ($\pm 3.0 \times 10^{-2}$)	518 ($\pm$77)
c-SPU	0.177 ($\pm 6.1 \times 10^{-3}$)	18.0×10^{-3} ($\pm 7.6 \times 10^{-4}$)	0.232 ($\pm 1.1 \times 10^{-2}$)	**44** ($\pm$13)
SP2U	$\mathbf{0.155}$ ($\pm 1.4 \times 10^{-2}$)	9.74×10^{-3} ($\pm 4.3 \times 10^{-4}$)	$\mathbf{0.125}$ ($\pm 3.9 \times 10^{-2}$)	1174 ($\pm$62)

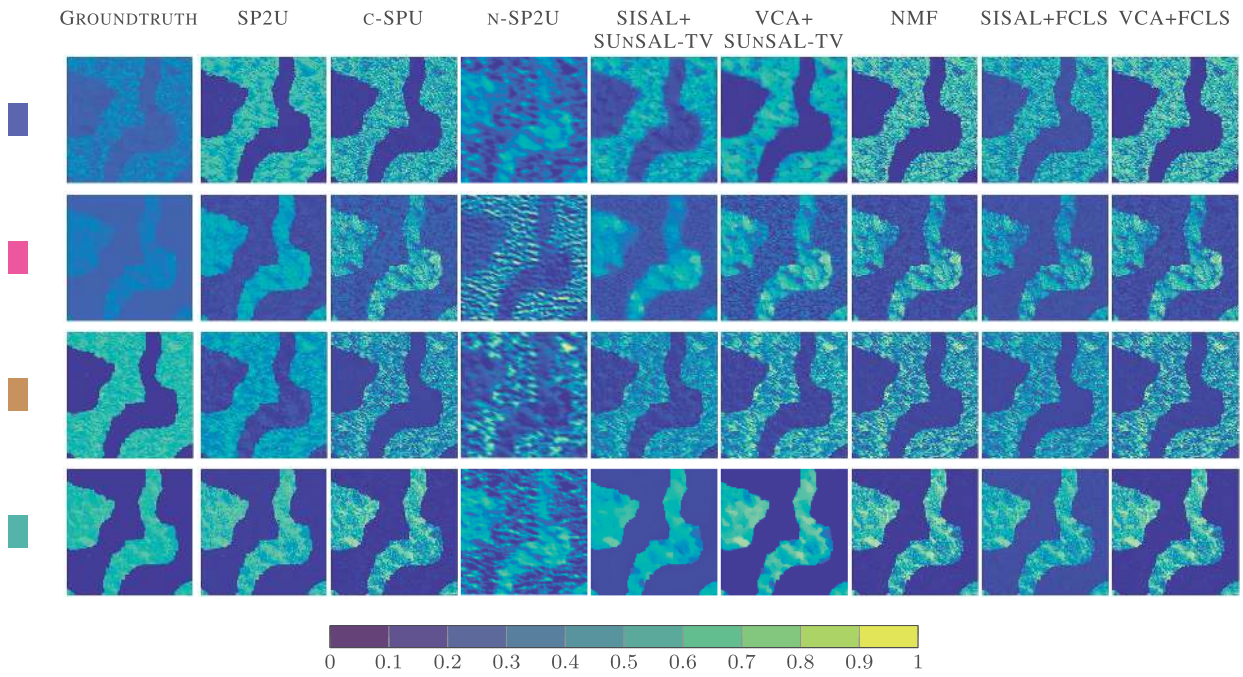

Fig. 5. Image 1: abundance maps (the colored squares refer to the colors used to plot endmembers in Fig. 4).

and $R_2 = 30$ and $K = 40$ for Image 2. It is worth noting that increasing these two parameters tends to improve the performance up to a certain point, where a slow decreasing can be observed. Hence, the choice of these values is not critical as long as they are high enough. It can be explained by the fact that a sufficient number of atoms and centroids are needed to explain the data. However, beyond a certain value, increasing these parameters reduces the regularization induced by the clustering. In a more general case, using features more robust to rotation and translation deformation would likely allow reducing the number of needed clusters and dictionary atoms. Moreover, the weighting terms of the various methods have been adjusted manually using a grid search algorithm in order to obtain consistent results. In particular, weighting coefficients of SP2U method have been set to $\tilde{\lambda}_0 = \tilde{\lambda}_1 = \lambda_2 = 1.0$ and $\lambda_z = 0.1$.

As the solution of the considered problem suffers from a permutation ambiguity inherent to factor models, a reordering of the endmembers is, thus, necessary before any evaluation. In this experiment, this relabeling is performed such that the aSAM is minimum. The quantitative results, averaged over ten trials, have then been computed for Image 1 and Image 2 and are presented, respectively, in Tables I and II.

The first conclusion of these results is that the SP2U method gives the best estimation of the endmember matrix. All other endmember extraction algorithms are clearly behind. In particular, from Fig. 4, we can see that SP2U is the only method identifying that there are two spectra very different from the others which correspond to the two soil spectra. Considering the degraded results obtained by c-SPU, it is clear that the spatial model has a real beneficial influence on the results. Another interesting remark is that the NMF model barely improves the initializing point given by VCA + FCLS. It appears to converge in a few iterations to a local minimum close to initialization. Overall, it seems that including the spatial information allows identifying more clearly the endmembers in particular in the considered case where the pure pixel assumption does not hold.

Then, regarding the estimation of abundances, the evaluation is less straightforward since it depends on the estimation of the endmembers. RMSE is computed after the reordering of the

endmembers and, for Image 1, the best abundance maps are obtained with SISAL + FCLS but they are not associated with the best estimated set of endmembers. The case of Image 2 is easier to discuss since the best abundance maps, obtained by SP2U, are associated with the best set of endmembers. It is also interesting to consider a qualitative evaluation of the obtained abundance maps depicted in Fig. 5. Even if the quantitative results seem to support the quality of the abundance maps retrieved by SUnSAL-TV, the results visually appear overly smooth. On the other hand, abundance maps estimated by SP2U seem visually relevant, but the corresponding RMSE suffers from an overestimation of abundances corresponding to soil spectra.

In addition, even if RE is not a relevant criterion to assess the quality of unmixing results, the values reported in Tables I and II show that most of the models are equally good at recovering mixtures explaining the observed data. The naive counterpart n-SP2U of SP2U exhibits significantly higher REs, which was expected as explained in Section IV-B. Some methods such as SISAL + FCLS get slightly lower REs. However, this can be easily explained by the fact that such a method aims at merely minimizing the RE, which does not necessarily lead to better RMSEs.

Finally, it is interesting to have a look at the computation times. SP2U appears as the slowest method since it inherits from a much richer model. However, the reported computation times should be taken cautiously. Indeed, SUnSAL-TV and SISAL + FCLS were implemented with a fixed number of iterations and are based on Lagrangian augmented splitting methods. Conversely, other methods use a PALM algorithm with a different stopping criterion (see Section III-B).

V. Experiments Using Real Data

A. Real Data Set

The real aerial hyperspectral image used to conduct the following experiment was acquired by AVIRIS in 2013 on a site called Citrus Belt 3, California. The image is composed of 224 spectral bands from 400 to 2500 nm with a spatial resolution of 3 m per pixel. After removing bands corresponding to water absorption, a 751×651-pixel image with $d_1 = 175$ spectral bands has been finally obtained. A panchromatic image of the scene is computed by normalizing and then summing all the spectral bands. The resulting image and a color composition of the scene are presented in Fig. 6. It is possible to state that the scene includes a desert area and several vegetation areas. Thus, several soil and vegetation spectra are expected to be identified.

Fig. 6. AVIRIS image. Color composition of (Left) hyperspectral image and (Right) corresponding panchromatic image.

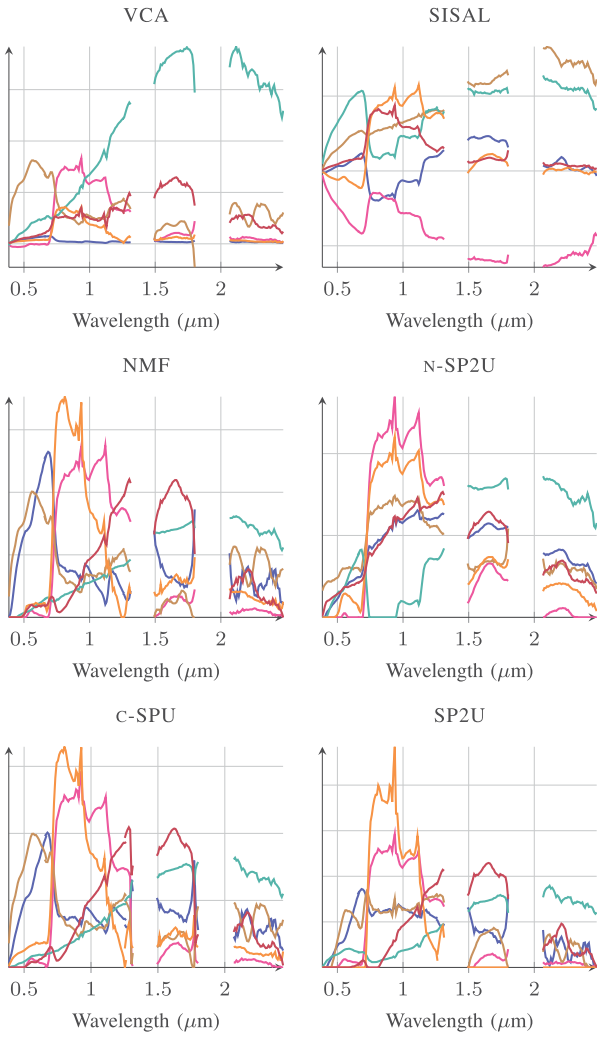

Fig. 7. AVIRIS image: estimated endmembers. Note that endmembers estimated by NMF, n-SP2U, and SP2U have been normalized to avoid scaling ambiguity intrinsic of the estimation method.

B. Compared Methods

As explained in Section II, it is common to consider a sum-to-one constraint for abundance vectors to interpret them as proportion vectors. However, this assumption is not always fulfilled in practical scenarios. In the specific case of the considered AVIRIS image, we decide to drop this constraint due to important illumination variation in the image. For example, the desert area on the upper part of the image is a hill, and the spectrum energy is almost doubled on its sunny side. In order to get a well-defined problem after dropping the sum-to-one constraint, it is necessary to introduce a new constraint such that there is no scaling ambiguity between $\mathbf{M}$ and $\mathbf{A}$. The choice has been made to enforce a unit norm of the endmember spectra. Thus, the initial sum-to-one constraint

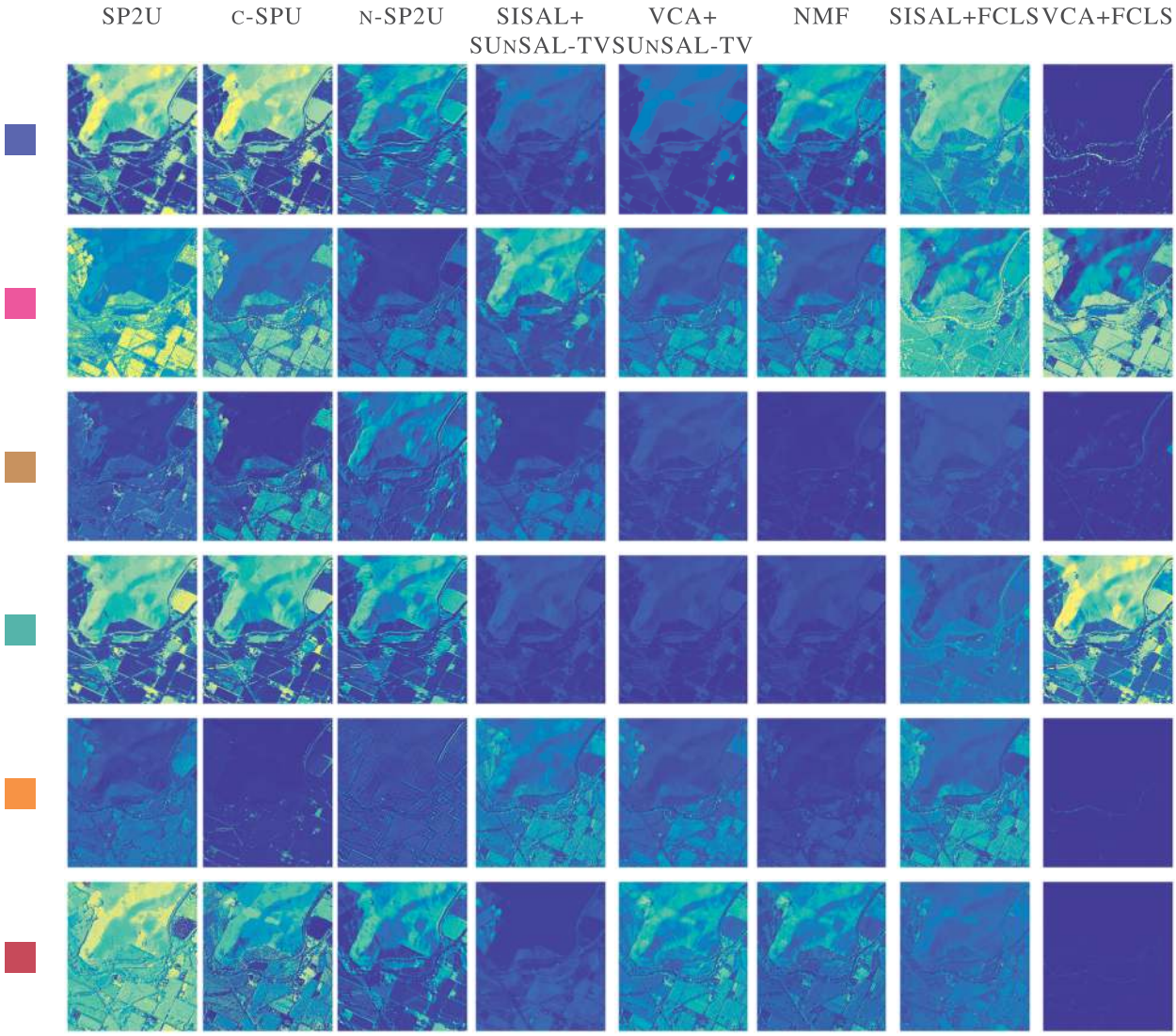

Fig. 8. AVIRIS image: estimated abundance maps. The colored squares refer to the colors used to plot endmembers in Fig. 7. However, no reordering has been performed, i.e., endmembers have no particular relationship between methods.

was moved from columns of $\mathbf{A}$ to columns of $\mathbf{M}$. Then, to get abundance maps summing to one, it is possible to normalize the obtained solution *a posteriori*. Similarly, the sum-to-one was removed for SUnSAL-TV, n-SP2U, and NMF. Moreover, similar to the synthetic case, parameters of the problem have been adjusted manually and set to $\tilde{\lambda}_0 = \tilde{\lambda}_1 = \lambda_2 = 1$ and $\lambda_z = 0.1$, $R_1 = 6$, $R_2 = 20$ and $K = 30$.

C. Results

Since no ground truth is available for this data set, only qualitative evaluations of the various methods are performed. First, Fig. 7 shows the endmembers estimated by all the compared methods. As explained in the previous paragraph, endmembers have been normalized except for SISAL and VCA. Regarding SISAL results, it is possible to note that the method estimates endmember signatures taking negative values. Negative endmembers cannot be interpreted as real reflectance spectra, and SISAL, thus, appears the worst compared methods. This method tries to identify a minimum volume simplex containing the observations under the assumption that the observations belong to a $(R_1 - 1)$-dimensional affine set. Thus, these poor results could be explained by a high noise level or nonlinear mixtures. It is difficult to objectively compare the results of the other methods. However, the result obtained by the SP2U method seems consistent with the visual content of the image since we can clearly identify: 1) two vegetation spectra (plotted in pink and orange) with strong absorbance in the visible domain and strong reflectance in the near-infrared domain [42] and 2) two soil spectra (plotted in blue and brown) with an increase of the reflectance from 0.4 to 1 μm [43].

Regarding the abundance maps presented in Fig. 8, it seems again that the maps produced by SP2U are consistent with the actual content of the scene. In particular, they are spatially consistent with natural edges in the image. In addition, SP2U results seem to be sparse in the sense that only a few endmembers are used for a given pixel while other methods

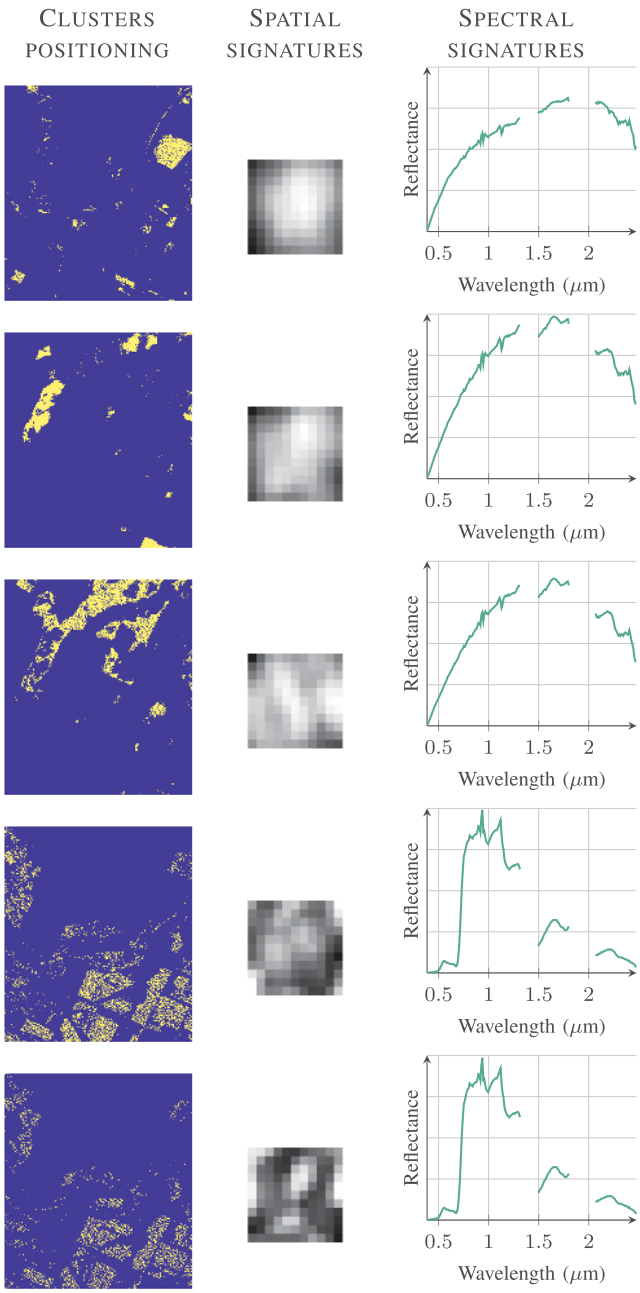

Fig. 9. AVIRIS image. Five particular clusters described by their (Left) spatial positioning, (Middle) mean spatial signature, and (Right) mean spectral signature.

recover very similar abundance maps with all endmembers, see VCA + SUnSAL-TV. From Table III, it seems that ensuring the sum-to-one constraint makes more difficult to fit the observations since VCA + FCLS has the highest RE. As expected, the SP2U method remains the slowest due to the overload of data to manipulate.

In addition, SP2U is not uniquely a spectral unmixing method and provides much richer interpretation. In Fig. 9, the results of the clustering performed by the coupling term are displayed. In particular, Fig. 9 shows the spatial position of the clusters and the mean spatial and spectral signatures

TABLE III
AVIRIS IMAGE: QUANTITATIVE RESULTS

Model	RE	Time (s)
VCA+FCLS	2.8×10^{-3}	12
SISAL+FCLS	0.14×10^{-3}	214
NMF	0.13×10^{-3}	2054
VCA+SUnSAL-TV	0.88×10^{-3}	471
SISAL+SUnSAL-TV	0.15×10^{-3}	455
n-SP2U	1.1×10^{-3}	1347
c-SPU	0.69×10^{-3}	578
SP2U	1.4×10^{-3}	7162

characterizing the clusters, obtained by (5). In this example, the first three clusters correspond to soil areas, whereas the last two are vegetation, more precisely trees. For instance, the recovered spatial patterns associated with soil are smoother when the wooded areas are characterized by variations of higher frequencies.

VI. CONCLUSION

This article proposed a new model to interpret hyperspectral images. This method enriched the traditional spectral unmixing modeling by incorporating a spatial analysis of the data. Two data fitting terms, bringing, respectively, spectral and spatial information, were considered jointly, thus yielding a spatial–spectral unmixing. This coupled learning process was made possible by the introduction of a clustering-driven coupling term linking the two coding matrices. This clustering process identified groups of pixels with similar spectral and spatial behaviors.

The experiments conducted on synthetic and real data showed that the proposed method performed very well both at identifying endmembers and estimating abundances. Moreover, the relevance of this method was not limited to the unmixing results since the outputs of the clustering task were also of high interest. The identified clusters were characterized by their average spectral signature and spatial context.

It is worth noting that in this article, the spatial features were merely chosen as elementary patches directly extracted from a virtual panchromatic image generated from the hyperspectral image. This choice allowed a linear approximation of these features to be motivated, exploiting a well-admitted property of image self-similarity. Since the main objective of this article was to introduce the paradigm of spatial–spectral unmixing, designing the best spatial feature was out of the scope of this article. However, to further explore the relevance of the proposed model, future works should investigate the benefit of using more complex spatial features. For example, resorting to a convolutional representation of the image may be of high interest to identify shift-invariant textured spatial signatures [44], [45].

APPENDIX

This section provides some details regarding the optimization schemes instanced for the proposed cofactorization model. Using notations adopted in Section III, the smooth coupling

term can be expressed as follows:

$$g(\mathbf{M},\mathbf{A},\mathbf{D},\mathbf{U},\mathbf{B},\mathbf{Z}) = \frac{\lambda_0}{2}\|\mathbf{Y}-\mathbf{M}\mathbf{A}\|_{\mathrm{F}}^2 + \frac{\lambda_1}{2}\|\mathbf{S}-\mathbf{D}\mathbf{U}\|_{\mathrm{F}}^2 + \frac{\lambda_2}{2}\left\|\begin{pmatrix}\mathbf{A}\\ \mathbf{U}\end{pmatrix}-\mathbf{B}\mathbf{Z}\right\|_{\mathrm{F}}^2 + \frac{\lambda_z}{2}\mathrm{Tr}(\mathbf{Z}^T\mathbf{V}\mathbf{Z}).$$

For a practical implementation of PALM, the partial gradients of $g(\cdot)$ and their Lipschitz moduli need to be computed to perform the gradient descent. They are given by

$$\begin{aligned}
\nabla_{\mathbf{M}} g(\mathbf{M},\mathbf{A},\mathbf{D},\mathbf{U},\mathbf{B},\mathbf{Z}) &= \lambda_0(\mathbf{M}\mathbf{A}\mathbf{A}^T - \mathbf{Y}\mathbf{A}^T)\\
\nabla_{\mathbf{A}} g(\mathbf{M},\mathbf{A},\mathbf{D},\mathbf{U},\mathbf{B},\mathbf{Z}) &= \lambda_0(\mathbf{M}^T\mathbf{M}\mathbf{A} - \mathbf{M}^T\mathbf{Y}) + \lambda_2(\mathbf{A}-\mathbf{B}_1\mathbf{Z})\\
\nabla_{\mathbf{D}} g(\mathbf{M},\mathbf{A},\mathbf{D},\mathbf{U},\mathbf{B},\mathbf{Z}) &= \lambda_1(\mathbf{D}\mathbf{U}\mathbf{U}^T - \mathbf{S}\mathbf{U}^T)\\
\nabla_{\mathbf{U}} g(\mathbf{M},\mathbf{A},\mathbf{D},\mathbf{U},\mathbf{B},\mathbf{Z}) &= \lambda_1(\mathbf{D}^T\mathbf{D}\mathbf{U} - \mathbf{D}^T\mathbf{S}) + \lambda_2(\mathbf{U}-\mathbf{B}_2\mathbf{Z})\\
\nabla_{\mathbf{B}} g(\mathbf{M},\mathbf{A},\mathbf{D},\mathbf{U},\mathbf{B},\mathbf{Z}) &= \lambda_2\left(\mathbf{B}\mathbf{Z}\mathbf{Z}^T - \begin{pmatrix}\mathbf{A}\\ \mathbf{U}\end{pmatrix}\mathbf{Z}^T\right)\\
\nabla_{\mathbf{Z}} g(\mathbf{M},\mathbf{A},\mathbf{D},\mathbf{U},\mathbf{B},\mathbf{Z}) &= \lambda_2\left(\mathbf{B}^T\mathbf{B}\mathbf{Z} - \mathbf{B}^T\begin{pmatrix}\mathbf{A}\\ \mathbf{U}\end{pmatrix}\right) + \lambda_z\mathbf{V}\mathbf{Z}
\end{aligned}$$

where $\mathbf{B}_1$ and $\mathbf{B}_2$ correspond to the submatrices of $\mathbf{B}$ defined by the R_1 first rows and R_2 last rows, respectively, such that $\mathbf{B} = \begin{pmatrix}\mathbf{B}_1\\ \mathbf{B}_2\end{pmatrix}$.

All partial gradients are globally Lipschitz as functions of the corresponding partial variables. The following Lipschitz moduli can be explicitly derived as follows:

$$\begin{aligned}
L_{\mathbf{A}}(\mathbf{M}) &= \|\lambda_0\mathbf{M}^T\mathbf{M} + \lambda_2\mathbf{I}_{R_1}\|\\
L_{\mathbf{M}}(\mathbf{A}) &= \|\lambda_0\mathbf{A}\mathbf{A}^T\|\\
L_{\mathbf{U}}(\mathbf{D}) &= \|\lambda_1\mathbf{D}^T\mathbf{D} + \lambda_2\mathbf{I}_{R_2}\|\\
L_{\mathbf{D}}(\mathbf{U}) &= \|\lambda_1\mathbf{U}\mathbf{U}^T\|\\
L_{\mathbf{B}}(\mathbf{Z}) &= \|\lambda_2\mathbf{Z}\mathbf{Z}^T\|\\
L_{\mathbf{Z}}(\mathbf{B}) &= \|\lambda_2\mathbf{B}^T\mathbf{B} + \lambda_z\mathbf{V}\|. \qquad (15)
\end{aligned}$$

REFERENCES

[1] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[2] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. 2nd Workshop Hyperspectral Image Signal Processing: Evolution Remote Sens.*, Jun. 2010, pp. 1–4.

[3] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 525–538, Jan. 2016.

[4] C. Shi and L. Wang, "Incorporating spatial information in spectral unmixing: A review," *Remote Sens. Environ.*, vol. 149, pp. 70–87, Jun. 2014.

[5] M.-D. Iordache, J. M. 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.

[6] S. Zhang, J. Li, H. C. Li, C. Deng, and A. Plaza, "Spectral-spatial weighted sparse regression for hyperspectral image unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3265–3276, Jun. 2018.

[7] X. Wang, Y. Zhong, L. Zhang, and Y. Xu, "Spatial group sparsity regularized nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6287–6304, Nov. 2017.

[8] 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.

[9] O. Eches, J. A. Benediktsson, N. Dobigeon, and J. Tourneret, "Adaptive Markov random fields for joint unmixing and segmentation of hyperspectral images," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 5–16, Jan. 2013.

[10] K. Canham, A. Schlamm, A. Ziemann, B. Basener, and D. Messinger, "Spatially adaptive hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4248–4262, Nov. 2011.

[11] C. Deng and C. Wu, "A spatially adaptive spectral mixture analysis for mapping subpixel urban impervious surface distribution," *Remote Sens. Environ.*, vol. 133, pp. 62–70, Jun. 2013.

[12] D. R. Thompson, L. Mandrake, M. S. Gilmore, and R. Castano, "Superpixel endmember detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4023–4033, Nov. 2010.

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

[14] C. Wang and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2011, pp. 448–456.

[15] J. Yoo, M. Kim, K. Kang, and S. Choi, "Nonnegative matrix partial co-factorization for drum source separation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2010, pp. 1942–1945.

[16] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[17] N. Akhtar and A. Mian, "Nonparametric coupled Bayesian dictionary and classifier learning for hyperspectral classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4038–4050, Sep. 2018.

[18] L. Drumetz, M.-A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3890–3905, Aug. 2016.

[19] T. Uezato, M. Fauvel, and N. Dobigeon, "Hyperspectral image unmixing with LiDAR data-aided spatial regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 4098–4108, Jul. 2018.

[20] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[21] R. K. Vasudevan *et al.*, "Big data in reciprocal space: Sliding fast Fourier transforms for determining periodicity," *Appl. Phys. Lett.*, vol. 106, no. 9, Mar. 2015, Art. no. 091601.

[22] R. K. Vasudevan, M. Ziatdinov, S. Jesse, and S. V. Kalinin, "Phases and interfaces from real space atomically resolved data: Physics-based deep data image analysis," *Nano Lett.*, vol. 16, no. 9, pp. 5574–5581, Sep. 2016.

[23] R. K. Vasudevan *et al.*, "Mapping mesoscopic phase evolution during E-beam induced transformations via deep learning of atomically resolved images," *NPJ Comput. Mater.*, vol. 4, Jun. 2018, Art. no. 30.

[24] I. T. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer-Verlag, 1986.

[25] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, NY, USA: Wiley, 2001.

[26] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *SPIE Imaging Spectrometry V*, vol. 3753, no. 1, M. R. Descour and S. S. Shen, Eds. Bellingham, WA, USA: SPIE, 1999, pp. 266–275.

[27] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[28] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, and F. R. Bach, "Supervised dictionary learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1033–1040.

[29] N. Dobigeon and J.-Y. Tourneret, "Bayesian orthogonal component analysis for sparse representation," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2675–2685, May 2010.

[30] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.

[31] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, Jan. 2005.
[32] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
[33] T. Tong, R. Wolz, P. Coupé, J. V. Hajnal, and D. Rueckert, "Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling," *NeuroImage*, vol. 76, pp. 11–23, Aug. 2013.
[34] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proc. CVPR*, Jun. 2011, pp. 457–464.
[35] F. Pompili, N. Gillis, P.-A. Absil, and F. Glineur, "Two algorithms for orthogonal nonnegative matrix factorization with application to clustering," *Neurocomputing*, vol. 141, pp. 15–25, Oct. 2014.
[36] T. Pock and S. Sabach, "Inertial proximal alternating linearized minimization (iPALM) for nonconvex and nonsmooth problems," *SIAM J. Imag. Sci.*, vol. 9, no. 4, pp. 1756–1787, Jan. 2016.
[37] L. Condat, "Fast projection onto the simplex and the l1 ball," *Math. Program.*, vol. 158, nos. 1–2, pp. 575–585, 2016.
[38] J. Nascimento and J. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
[39] S. Z. Li, *Random Field Modeling in Image Analysis*. London, U.K.: Springer, 2009.
[40] A. Baldridge, S. Hook, C. Grove, and G. Rivera, "The ASTER spectral library version 2.0," *Remote Sens. Environment*, vol. 113, no. 4, pp. 711–715, Apr. 2009.
[41] J. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. 1st Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, Aug. 2009, pp. 1–4.
[42] R. Myneni, F. Hall, P. Sellers, and A. Marshak, "The interpretation of spectral vegetation indexes," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 481–486, Mar. 1995.
[43] M. F. Baumgardner, L. F. Silva, L. L. Biehl, and E. R. Stoner, "Reflectance properties of soils," *Advances in Agronomy*, vol. 38. Amsterdam, The Netherlands: Elsevier, 1986, pp. 1–44.
[44] H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013.
[45] O. Chabiron, F. Malgouyres, J.-Y. Tourneret, and N. Dobigeon, "Toward fast transform learning," *Int. J. Comput. Vis.*, vol. 114, nos. 2–3, pp. 195–216, Sep. 2015.

Adrien Lagrange (Student Member, IEEE) received the Engineering degree in robotics and embedded systems from ENSTA ParisTech, Palaiseau, France, and the M.Sc. degree in machine learning from Paris Saclay University, Saint-Aubin, France, both in 2016. He is currently pursuing the Ph.D. degree with the National Polytechnic Institute of Toulouse, Toulouse, France. He is working on the subject of spectral unmixing and classification of hyperspectral images under the supervision of N. Dobigeon and M. Fauvel.

Mathieu Fauvel (Senior Member, IEEE) received the Ph.D. degree in image and signal processing from the Grenoble Institute of Technology, Grenoble, France, in 2007.

From 2008 to 2010, he was a Post-Doctoral Researcher with the MISTIS Team, National Institute for Research in Computer Science and Control (INRIA). Since 2011, he has been an Associate Professor with the National Polytechnic Institute of Toulouse within the DYNAFOR lab (INRA), Toulouse. His research interests are remote sensing, pattern recognition, and image processing.

Stéphane May received the Engineering degree in telecommunications from the National Institute of Telecommunications, Evry, France, in 1997.

He is currently with the Centre National d'Études Spatiales (French Space Agency), Toulouse, France, where he is developing image processing algorithms and software for the exploitation of earth observation images.

Nicolas Dobigeon (Senior Member, IEEE) received the Ph.D. degree in signal processing from the National Polytechnic Institute of Toulouse, Toulouse, France, in 2012.

He was a Post-Doctoral Researcher with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA, from 2007 to 2008. Since 2008, he has been with the National Polytechnic Institute of Toulouse, where he is currently a Professor. His research interests include statistical signal and image processing.