# Robust Collaborative Nonnegative Matrix Factorization for Hyperspectral Unmixing

Jun Li, *Member, IEEE*, José M. Bioucas-Dias, *Member, IEEE*, Antonio Plaza, *Fellow, IEEE*, and Lin Liu

***Abstract*—Spectral unmixing is an important technique for remotely sensed hyperspectral data exploitation. It amounts to identifying a set of pure spectral signatures, which are called *endmembers*, and their corresponding fractional, *draftrulesabundances* in each pixel of the hyperspectral image. Over the last years, different algorithms have been developed for each of the three main steps of the spectral unmixing chain: 1) estimation of the number of endmembers in a scene; 2) identification of the spectral signatures of the endmembers; and 3) estimation of the fractional abundance of each endmember in each pixel of the scene. However, few algorithms can perform all the stages involved in the hyperspectral unmixing process. Such algorithms are highly desirable to avoid the propagation of errors within the chain. In this paper, we develop a new algorithm, which is termed robust collaborative nonnegative matrix factorization (R-CoNMF), that can perform the three steps of the hyperspectral unmixing chain. In comparison with other conventional methods, R-CoNMF starts with an overestimated number of endmembers and removes the redundant endmembers by means of collaborative regularization. Our experimental results indicate that the proposed method provides better or competitive performance when compared with other widely used methods.**

***Index Terms*—Endmember extraction, hyperspectral imaging, robust collaborative nonnegative matrix factorization (R-CoNMF), spectral unmixing.**

## I. Introduction

SPECTRAL unmixing is an important task for remotely sensed hyperspectral data exploitation [1]. The linear mixture model (LMM) is a widely used technique for spectral unmixing, which is based on the principle that each captured pixel in a hyperspectral image can be represented as the linear combination of a finite set of spectrally pure constituent spectra or *endmembers*, weighted by an abundance factor that establishes the proportion of each endmember in the pixel under inspection [2].

Within the LMM paradigm, three main types of unmixing algorithms can be identified: geometrical, statistical, and sparse regression based [3]. Geometrical unmixing algorithms work under the assumption that the endmembers of a hyperspectral image are the vertices of a simplex of minimum volume enclosing the data set (i.e., the set of hyperspectral vectors, which are also termed as the hyperspectral image) [4] or of a simplex of maximum volume contained in the convex hull of the data set [5], [6]. Among minimum-volume algorithms, we highlight the seminal minimum-volume-constrained nonnegative matrix factorization (MVC-NMF) method, which is also based on NMF principles. As their name suggests, statistical methods [3] are based on analyzing mixed pixels by means of statistical principles, such as Bayesian approaches [7], [8]. Finally, sparse-regression-based algorithms [9]–[12] are based on expressing each mixed pixel in a scene as a linear combination of a finite set of pure spectral signatures that are known *a priori* and available in a library. Although each of these methods exhibits their own advantages and disadvantages, the fact is that geometrical approaches have been most frequently used by the hyperspectral research community up to now [3], [6], [13]–[15]. This is mainly due to their reduced—although still quite high—computational cost when compared with the other types of unmixing algorithms, as well as to the fact that they represent a straightforward interpretation of the LMM.

To fully unmix a given hyperspectral image by means of a geometrical method, the majority of the state-of-the-art approaches is based on dividing the whole process into three concatenated steps [3]: 1) estimation of the number of endmembers in a scene; 2) identification of the spectral signatures of these endmembers; and 3) estimation of the endmember abundances in each pixel of the scene. In the last few years, several techniques have been developed for addressing each part of the chain, with particular emphasis on the identification of endmembers (with and without assuming the presence of pure spectral signatures in the input hyperspectral data [2], [3], [16]). However, there are very few spectral unmixing algorithms that can address all the stages involved in the hyperspectral unmixing process [17]. Although this general processing chain has been proven effective for unmixing certain types of hyperspectral images, it has some drawbacks.

1) The first drawback comes from the fact that the output of each stage is the input of the following stage, which may favor the propagation of errors within the chain.

Manuscript received July 16, 2015; revised December 13, 2015 and April 14, 2016; accepted May 26, 2016. Date of publication July 7, 2016; date of current version August 11, 2016. This work was supported by the Portuguese Science and Technology Foundation under Project UID/EEA/50008/2013 and the Chinese National Science Foundation under Project 41431178.

J. Li is with the Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: lijun48@mail.sysu.edu.cn).

J. M. Bioucas-Dias is with the Instituto de Telecomunicações, Instituto Superior Técnico, Universidade Técnica de Lisboa, 1049-1 Lisbon, Portugal (e-mail: bioucas@lx.it.pt).

A. Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10071 Cáceres, Spain.

L. Liu is with the Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China, and also with the Department of Geography, University of Cincinnati, Cincinnati, OH 45221-0131 USA.

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

Digital Object Identifier 10.1109/TGRS.2016.2580702

2) The second more relevant problem is given by the huge variability of the results obtained when estimating the number of endmembers of a hyperspectral scene with different state-of-the-art algorithms or, in some cases, with the same algorithm but with different initialization parameters. For instance, if the number of endmembers is obtained by means of computing the virtual dimensionality (VD) [18] of the image, very different results may be obtained depending on the technique employed for estimating the VD itself [19] or the parameters used by this technique. The hyperspectral signal identification by minimum error (HySime) algorithm [20] is another popular approach for estimating the number of endmembers, which generally provides results that differ from those achieved with VD. This adds more uncertainties to the process of accurately estimating the number of endmembers of a hyperspectral image. Other techniques have been developed for estimating the number of endmembers, such as eigenvalue likelihood maximization [21], outlier detection methods [22], or a technique for geometry-based estimation of the number of endmembers based on the convex hull in [23], providing different results.
3) A third issue is the computational complexity of the whole process, which makes highly desirable the development of a unified and fast technique that is able to address the different parts involved in the hyperspectral unmixing chain.

The challenge related with the correct identification of the number of endmembers is crucial for unmixing algorithms in general and for geometric-based algorithms in particular [3]. The estimation of the number of endmembers by means of subspace identification algorithms can be considered generally effective when the hyperspectral images are well approximated by the LMM [18], [20]. However, this part of the chain is more challenging when the endmembers have a similar shape or the mixing process has nonlinear components [24]. Let $p$ denote the true number of endmembers in a hyperspectral image, and let $q$ denote the number of endmembers estimated by a certain algorithm. In general, $q$ fluctuates around $p$, but three situations are possible.

- If $q=p$, we have an ideal situation, and most existing endmember identification algorithms rely on this assumption.
- If $q < p$, then the number of endmembers is underestimated. This situation is generally easy to identify as the reconstruction of the original hyperspectral image based, for instance, on the well-known LMM contains errors, thus allowing a trained analyst to identify that more endmembers are needed for the model to work properly.
- If $q > p$, then the number of endmembers is overestimated. This is a difficult problem as compared with the underestimation situation, since the reconstruction error using more endmembers than needed would provide similar or better values than those obtained when $q = p$. This is particularly critical for scenarios in which there are no pure pixels existing in the observed image and, at the same time, for endmember identification algorithms designed without assuming that the true endmembers are present in the input hyperspectral data [4], [7], [25]–[31]. In this paper, we particularly address this case and provide an algorithm that can work in situations in which the number of endmembers is overestimated *a priori*.

Our contribution in this work is a new algorithm, which is termed robust collaborative nonnegative matrix factorization (R-CoNMF), that can perform the three steps of the hyperspectral unmixing chain. Although the proposed algorithm can be easily adapted to a scenario in which the number of endmembers is known in advance, the proposed method assumes that the number of endmembers corresponds to an overestimated value. Our method uses the $\ell_{2,1}$ mixed-norm regularizer [32] applied to the abundance matrix, which promotes sparsity among the lines of that matrix. As seen below, this is equivalent to imposing sparsity among the endmembers simultaneously (collaboratively) for all pixels such that it is guaranteed that only the true endmembers have contributions to the estimated abundances. The proposed R-CoNMF uses another regularization term to promote minimum volume by pushing the endmembers toward the mean value of the data set or bring the endmembers to the real solution quadratically regularized by a given simplex, other than the minimum volume. Endowed with these two regularizers, R-CoNMF does not require a prior step to estimate the number of endmembers and can be effectively used in scenarios in which the number of endmembers is not known *a priori*.

The remainder of this paper is organized as follows. Section II describes the proposed approach. Section III describes the optimization algorithm used to implement the proposed approach. Sections IV and V describe the obtained experimental results and a detailed intercomparison with other state-of-the-art approaches. Our experimental results, conducted using both synthetic and real hyperspectral data, demonstrate that the proposed method exhibits very good performance in noisy scenarios, without the need to know the number of endmembers in advance. Comparisons to other state-of-the-art algorithms covering the full hyperspectral unmixing chain are also given. Section VI concludes this paper with some remarks and hints at plausible future research.

## II. Proposed Approach

Let $\mathbf{Y} \equiv [\mathbf{y}_1, \ldots, \mathbf{y}_n] \in \mathbb{R}^{d \times n}$ be the matrix representation of a hyperspectral image with $n$ spectral vectors and $d$ spectral bands. Under the LMM, we have [3]

$$\begin{aligned} &\mathbf{Y} = \mathbf{M}\mathbf{S} + \mathbf{N} \\ \text{s.t.}: \quad &\mathbf{S} \geq 0,\ \mathbf{1}_p^T \mathbf{S} = \mathbf{1}_n^T \end{aligned} \tag{1}$$

where $\mathbf{M} \equiv [\mathbf{m}_1, \ldots, \mathbf{m}_p] \in \mathbb{R}^{d \times p}$ is a so-called mixing matrix containing $p$ endmembers; $\mathbf{m}_i$ denotes the $i$th endmember signature; $\mathbf{S} \equiv [\mathbf{s}_1, \ldots, \mathbf{s}_n] \in \mathbb{R}^{p \times n}$ is the abundance matrix containing the endmember fractions $\mathbf{s}_i$, for pixels $i = 1, \ldots, n$; $\mathbf{S} \geq 0$ is the abundance nonnegativity constraint [2] to be understood in the component-wise sense; $\mathbf{1}_p^T \mathbf{S} = \mathbf{1}_n^T$ is the sum-to-one constraint that stems from a physical interpretation of the abundance vector; and $\mathbf{1}_p = [1, 1, \ldots, 1]^T$ is a column vector of size $p$ (the notation $[\cdot]^T$ stands for vector or matrix transpose). Finally, $\mathbf{N}$ collects the errors that may affect the measurement process (e.g., noise).

In this paper, we address the estimation of $p$, i.e., the number of endmembers, of mixing matrix $\mathbf{M}$, and of abundance matrix $\mathbf{S}$. Although $p$ is not known beforehand, we assume that we have access to an overestimate thereof. That is, we are given a number $q$ such that $q \geq p$. This way, we account for a common situation in which an overestimate of the number of endmembers is easy to compute, which is not usually the case regarding the true number of endmembers.

We tackle the estimation of $p$, $\mathbf{M}$, and $\mathbf{S}$ by seeking a solution for the following NMF optimization:

$$\begin{aligned} \min_{\mathbf{A},\mathbf{X}} \quad & \left(\frac{1}{2}\right)\|\mathbf{Y}-\mathbf{A}\mathbf{X}\|_F^2 + \alpha\|\mathbf{X}\|_{2,1} + \frac{\beta}{2}\|\mathbf{A}-\mathbf{P}\|_F^2 \\ \text{s.t.}: \quad & \mathbf{X} \in \mathcal{S}_{q-1} \quad \mathbf{A} \in \mathcal{A}_{q-1} \end{aligned} \tag{2}$$

where $\|\cdot\|_2$ and $\|\cdot\|_F$ stand for the Euclidean and Frobenius norms, respectively; $\mathbf{A} \equiv [\mathbf{a}_1, \ldots, \mathbf{a}_q] \in \mathbb{R}^{d\times q}$ and $\mathbf{X} \in \mathbb{R}^{q\times n}$ are optimization variables linked with the mixing matrix and the abundance matrix, respectively; $\|\mathbf{X}\|_{2,1} \equiv \sum_{i=1}^{q} \|\mathbf{x}^i\|_2$ denotes the $\ell_{2,1}$ (see, e.g., [33]) mixed norm of matrix $\mathbf{X}$ ($\mathbf{x}^i$ denotes the $i$th row of $\mathbf{X}$); $\mathbf{P} \equiv [\mathbf{y}_{i_1}, \ldots, \mathbf{y}_{i_q}]$ is a set of $q$ spectrally observed vectors inferred with a pure-pixel algorithm and, thus, that are *close* to the extremes of the simplex; $\alpha$ and $\beta$ are regularization parameters; $\mathcal{S}_{q-1}$ is the collection of $q \times n$ matrices whose columns belong to the probability simplex of dimension $q-1$; and $\mathcal{A}_{q-1}$ is the collection of matrices of size $d \times q$ whose columns belong to the affine set of dimension $q-1$ that best represents the data $\mathbf{Y}$ in the mean square error sense [29]. The introduction of this constraint removes the shortcomings associated violations to the sum-to-one constraints usually observed in real data sets.

The objective function shown on the right-hand side of (2) has three terms: a data fidelity term $\|\mathbf{Y}-\mathbf{A}\mathbf{X}\|_F^2$, which promotes solutions with low reconstruction error; the $\ell_{2,1}$ mixed-norm $\|\mathbf{X}\|_{2,1}$, which promotes row sparsity on $\mathbf{X}$ [34], that is, it promotes solutions with complete rows $\mathbf{x}^i$ set to zero; and the term $(\beta/2)\|\mathbf{A}-\mathbf{P}\|_F^2$, which pushes the columns of $\mathbf{A}$ toward a given solution $\mathbf{P}$. For example, $\mathbf{P}$ can be a solution provided by a pure-pixel-based algorithm, which, jointly with the constraint $\mathbf{X} \in \mathcal{S}_{q-1}$, pushes the endmembers to the extremes of the simplex defined by the data $\mathbf{Y}$. Another choice is $\mathbf{P} \equiv [\overline{\mathbf{y}}, \ldots, \overline{\mathbf{y}}]$, where $\overline{\mathbf{y}}$ is the sample mean vector. This choice has a minimum-volume flavor. Fig. 1 graphically shows the behavior of the third regularizer term: (a) $\mathbf{P}$ is a solution given by VCA, and (b) $\mathbf{P}$ is defined by the mass center, for a simulated problem with three endmembers, 4000 pixels, and SNR $= 30$ dB, in which we set $\alpha = 1e-5$. The solution represented using dots in brown to red in Fig. 1 corresponds to the endmembers obtained by R-CoNMF with $\beta$ from Inf to 0. It is clear in Fig. 1 that the simplex regularizer strongly depends on $\mathbf{P}$. For the considered problem, when $\beta$ is well set, the final solutions are very similar and close to the real ones.

As shown in the optimization (2), we conclude that the proposed optimization aims at finding a couple $(\mathbf{A}, \mathbf{X})$ yielding low reconstruction error, low number of nonzero rows of $\mathbf{X}$, and a simplex defined by the columns of $\mathbf{A}$ constrained by a given solution or minimum volume. The balance between these three terms is set by the regularization parameters $\alpha$ and $\beta$.

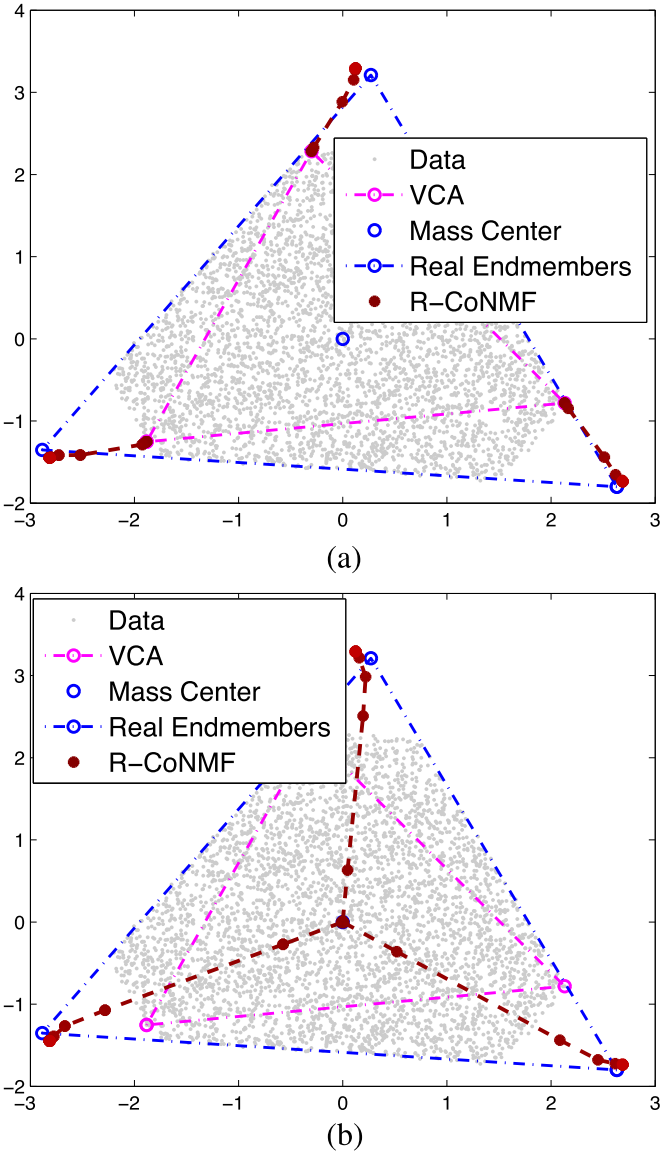



Fig. 1. Graphical illustration of the behavior of the third regularizer term in our proposed optimization. (a) $\mathbf{P}$ is a solution given by VCA. (b) $\mathbf{P}$ is defined by the mass center.

Let $\Lambda$ stand for a set of naturals, $\mathbf{A}_{|\Lambda} \equiv [\mathbf{a}_i, i \in \Lambda]$ and $\mathbf{X}^{|\Lambda} \equiv [(\mathbf{x}^i)^T, i \in \Lambda]^T$, i.e., $\mathbf{A}_\Lambda$ is a matrix containing the columns of $\mathbf{A}$ with indexes in set $\Lambda$, and $\mathbf{X}^\Lambda$ is a matrix containing the rows of $\mathbf{X}$ with indexes in set $\Lambda$. In addition, let $\text{rowsupp}(\mathbf{X}) \equiv \{i, : \mathbf{x}^i \neq \mathbf{0}\}$, i.e., $\text{rowsupp}(\mathbf{X})$ denotes the row support of $\mathbf{X}$. With these definitions in place, we introduce

$$\widehat{\mathbf{M}} \equiv \widehat{\mathbf{A}}_{|\text{rowsupp}(\widehat{\mathbf{X}})} \tag{3}$$

$$\widehat{\mathbf{S}} \equiv \widehat{\mathbf{X}}^{|\text{rowsupp}(\widehat{\mathbf{X}})}. \tag{4}$$

Provided that the mixed-norm term $\alpha\|\mathbf{X}\|_{2,1}$ in (2) drives the rows of $\mathbf{X}$ not corresponding to endmembers to zero, then it is expectable that, apart from a permutation, the data fidelity and the volume terms drive $(\mathbf{A}, \mathbf{X})$ to solutions close to $(\mathbf{M}, \mathbf{S})$ in the sense that $\widehat{\mathbf{M}} \simeq \mathbf{M}$ and $\widehat{\mathbf{S}} \simeq \mathbf{S}$. This scenario is shown in Fig. 2.

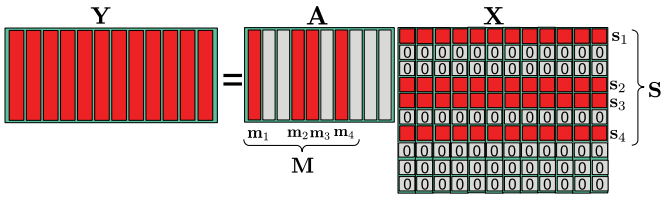



Fig. 2. Illustration of the concept of row (collaborative) sparsity promoted $\ell_{2,1}$ norm under the linear mixing model. The abundance matrix $\mathbf{S}$ is formed by the nonzero rows of $\mathbf{X}$, and the mixing matrix $\mathbf{M}$ is formed by the correspondent columns of $\mathbf{A}$.

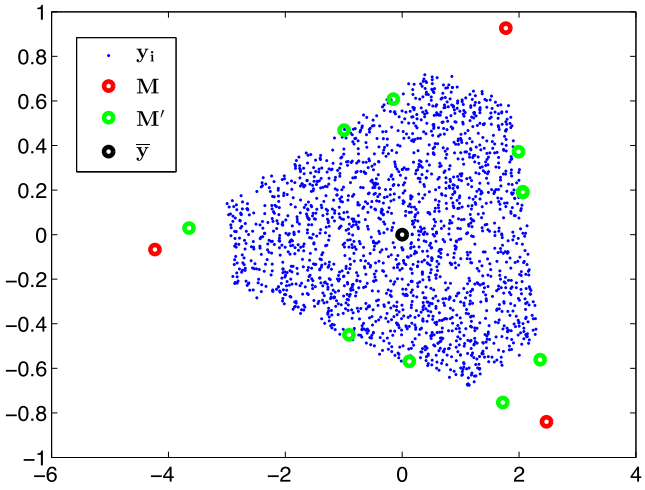



Fig. 3. Projection of the spectral vectors $\overline{\mathbf{y}}_i$, for $i = 1, \ldots, n$ (blue); of the endmember signatures $\mathbf{m}_i$, for $i = 1, \ldots, p$ (red); and of the columns of $\mathbf{A}$, which are not endmembers (green). The spectral mean value is shown in black. See text for more details.

The quality of estimate $(\widehat{\mathbf{M}}, \widehat{\mathbf{S}})$ given by (3) and (4) relies primarily on the number of endmembers detected, which is given by $\widehat{p} = |\text{rowsupp}(\widehat{\mathbf{X}})|$. If $\widehat{p}$ considerably departs from the true number $p$, there is little hope in obtaining a reasonable estimate of the couple $(\mathbf{M}, \mathbf{S})$. To shed light on this issue, we built a simulated data set using a matrix $\mathbf{M} \in \mathbb{R}^{224\times 3}$ (i.e., $d = 224$, and $p = 3$) sampled from the United States Geological Survey (USGS) library[1] and an abundance matrix $\mathbf{S} \in \mathbb{R}^{3\times 2000}$ (i.e., $n = 2000$) uniformly sampled on the simplex. The abundance vectors $\mathbf{s}_i$ such that $\max(\mathbf{s}_i) > 0.8$ were discarded to model the absence of pure pixels.

Fig. 3 shows a scatterplot of the vectors $\mathbf{y}_i$, for $i = 1, \ldots, n$, projected onto the affine set defined by the columns of $\mathbf{M}$ centered at $\overline{\mathbf{y}}$, which is plotted in black. The signal-to-noise ratio (SNR) defined as

$$\text{SNR} = 10 \log_{10} \left( \frac{\|\mathbf{MS}\|_2^2}{\|\mathbf{N}\|_2^2} \right)$$

was set to 30 dB. Fig. 3 also shows the projection of the matrix $\mathbf{A} = [\mathbf{M}, \mathbf{M}'] \in \mathbb{R}^{d\times q}$, where $\mathbf{M}'$ contains $q - p = 10$ spectral vectors lying on the facets of the simplex defined by $\mathbf{M}$. The projections of $\mathbf{M}$ and $\mathbf{M}'$ are in red and green, respectively.

[1]http://speclab.cr.usgs.gov/spectral-lib.html

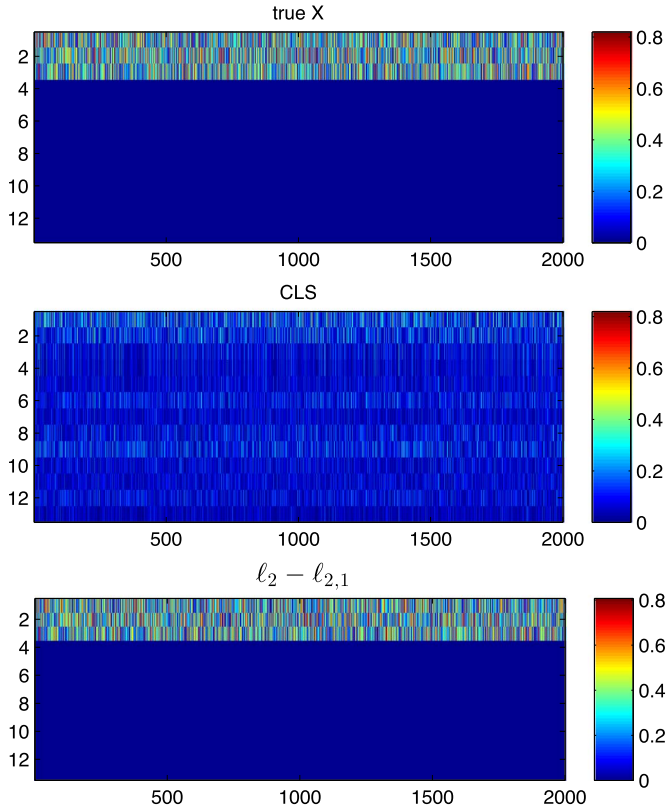



Fig. 4. Solution of the optimization problem (2) with respect to $\mathbf{X}$ for $\mathbf{A}$ known and represented in Fig. 3. (Top) True $\mathbf{X}$. (Middle) CLS solution ($\alpha = 0$). (Bottom) $\ell_2 - \ell_{2,1}$ solution ($\alpha > 0$).

In Fig. 4, the top image shows in false color the true $\mathbf{X}$: The horizontal axis denotes pixels, the vertical axis denotes columns of $\mathbf{A}$, and the colors denote abundances. The first three rows of $\mathbf{X}$ display the abundances with respect to $\mathbf{M}$, and the remaining rows display the abundances with respect to $\mathbf{M}'$, which are zero. With this setting, the constrained least squares (CLS) solution for $\mathbf{X}$, which amounts to minimizing (2) with respect to $\mathbf{X}$ with $\alpha = 0$ (absence of the $\ell_{2,1}$ norm), yields the estimate shown in the top right-hand side part of Fig. 4. This estimate is far from the true $\mathbf{X}$ because the columns of $\mathbf{M}'$ are linearly dependent on the columns of $\mathbf{M}$, thus yielding a severely ill-posed inverse problem. The image shown in the bottom part of Fig. 3 is the constrained $\ell_2 - \ell_{2,1}$ solution obtained by minimizing (2) with respect to $\mathbf{X}$ with $\alpha > 0$. This optimization is convex, and we have used the collaborative sparse unmixing by variable splitting and augmented Lagrangian (CLSUnSAL) algorithm [11] to solve it. The recovered solution is very close to the true solution, and, more importantly for our problem, the support of $\mathbf{X}$ was correctly recovered, i.e., $\text{rowsupp}(\widehat{\mathbf{X}}) = \text{rowsupp}(\mathbf{X})$.

An in-depth study of the recovery guarantees in $\ell_2 - \ell_{2,1}$ problems is out of the scope of this paper (see, e.g., [33]), and the toy experiment we have just presented does not provide, of course, any theoretical contribution to this study. Our objective is to clarify the role we expect the collaborative $\ell_{2,1}$ norm will play in the optimization criterion (2). In Sections IV and V, we present a comprehensive array of results providing evidence for the effectiveness of the proposed method.

## III. Optimization Algorithm

The nonconvex data fidelity term $(1/2)\|\mathbf{Y}-\mathbf{A}\mathbf{X}\|_F^2$ present in (2) makes the respective optimization hard. Herein, we adopt the proximal alternating optimization (PAO) [35], which, under suitable conditions, converges to a critical point of the underlying nonconvex objective function.

For notational convenience, we redefine the optimization (2) as follows:

$$\min_{\mathbf{A}\in\mathbb{R}^{d\times q},\mathbf{X}\in\mathbb{R}^{q\times n}} L(\mathbf{A},\mathbf{X})$$

where

$$L(\mathbf{A},\mathbf{X}) \equiv \frac{1}{2}\|\mathbf{Y}-\mathbf{A}\mathbf{X}\|_F^2+\alpha\|\mathbf{X}\|_{2,1}+\frac{\beta}{2}\|\mathbf{A}-\mathbf{P}\|_F^2 + \iota_{\mathcal{S}_{q-1}}(\mathbf{X})+\iota_{\mathcal{A}_{q-1}}(\mathbf{A}) \quad (5)$$

where $\iota_{\mathcal{S}_{q-1}}:\mathbb{R}^{q\times n}\to\mathbb{R}\cup\{+\infty\}$ and $\iota_{\mathcal{A}_{q-1}}:\mathbb{R}^{d\times q}\to\mathbb{R}\cup\{+\infty\}$ denote the indicator function of sets $\mathcal{S}_{q-1}$ and $\mathcal{A}_{q-1}$, respectively; that is, $\iota_{\mathcal{S}_{q-1}}(\mathbf{X})=+\infty$ if $\mathbf{X}\notin\mathcal{S}_{q-1}$, and $\iota_{\mathcal{S}_{q-1}}(\mathbf{X})=0$ otherwise; $\iota_{\mathcal{A}_{q-1}}(\mathbf{A})=+\infty$ if $\mathbf{A}\in\mathcal{A}_{q-1}$, and $\iota_{\mathcal{A}_{q-1}}(\mathbf{A})=0$ otherwise.

Given $(\mathbf{A}_{(0)},\mathbf{X}_{(0)})$, PAO generates the sequence $(\mathbf{A}_{(t)},\mathbf{X}_{(t)})\to(\mathbf{A}_{(t+1)},\mathbf{X}_{(t)})\to(\mathbf{A}_{(t+1)},\mathbf{X}_{(t+1)})$ as follows:

$$\mathbf{A}_{(t+1)}=\arg\min_{\mathbf{A}} L\left(\mathbf{A},\mathbf{X}_{(t)}\right)+\frac{\lambda_t}{2}\left\|\mathbf{A}-\mathbf{A}_{(t)}\right\|_F^2 \quad (6)$$

$$\mathbf{X}_{(t+1)}=\arg\min_{\mathbf{X}} L\left(\mathbf{A}_{(t+1)},\mathbf{X}\right)+\frac{\mu_t}{2}\left\|\mathbf{X}-\mathbf{X}_{(t)}\right\|_F^2 \quad (7)$$

where $\lambda_t,\mu_t$, for $t=0,1,\ldots$, are sequences of positive numbers. We remark that the above procedure can be interpreted as a regularized version of a two-block nonlinear Gauss–Seidel method [36].

Let $\mathcal{A}_{q-1}\equiv\{\mathbf{z}\in\mathbb{R}^d:\mathbf{z}=\overline{\mathbf{y}}+\mathbf{V}\boldsymbol{\alpha},\boldsymbol{\alpha}\in\mathbb{R}^{q-1}\}$, where $\overline{\mathbf{y}}$ is the sample mean observed vector, and $\mathbf{V}\in\mathbf{R}^{d\times(q-1)}$ is a matrix holding the first $(q-1)$ principal components of the data $\mathbf{Y}$. We remark that the columns of $\mathbf{V}$ are orthonormal, i.e., $\mathbf{V}^T\mathbf{V}=\mathbf{I}_{q-1}$, where $\mathbf{I}_{q-1}$ denotes the identity matrix of dimension $q-1$. In addition, we also define $\overline{\mathbf{Y}}\equiv[\overline{\mathbf{y}},\ldots,\overline{\mathbf{y}}]\in\mathbb{R}^{d\times q}$ and write $\mathbf{A}=\overline{\mathbf{Y}}+\mathbf{V}\boldsymbol{\Delta}$. We remark that any $\mathbf{A}\in\mathcal{A}_{q-1}$ admits a unique representation in the form $\mathbf{A}=\overline{\mathbf{Y}}+\mathbf{V}\boldsymbol{\Delta}$, where $\boldsymbol{\Delta}=\mathbf{V}^T(\mathbf{A}-\overline{\mathbf{Y}})$.

With the above definitions in place, the solution to the quadratic optimization (6) is

$$\mathbf{A}_{(t+1)}=\overline{\mathbf{Y}}+\mathbf{V}\boldsymbol{\Delta}_{(t+1)} \quad (8)$$

$$\boldsymbol{\Delta}_{(t+1)}=\mathbf{V}^T\left((\mathbf{Y}-\overline{\mathbf{Y}})\mathbf{X}_{(t)}^T+\beta(\mathbf{P}-\overline{\mathbf{Y}})+\lambda_t\boldsymbol{\Delta}_{(t)}\right) \quad (9)$$

$$\cdot\left(\mathbf{X}_{(t)}\mathbf{X}_{(t)}^T+(\beta+\lambda_t)\mathbf{I}\right)^{-1} \quad (10)$$

where $\mathbf{I}$ denotes the identity matrix of suitable size, and $(\cdot)^T$ denotes the transpose operator.

By defining $\mathbf{Y}'_{(t)}\equiv[\mathbf{Y}^T\sqrt{\mu_t}\mathbf{X}_t^T]^T$ and $\mathbf{A}'_{(t+1)}\equiv[\mathbf{A}_{(t+1)}^T\sqrt{\mu_t}\mathbf{I}]^T$, the optimization (7) can be written as

$$\min_{\mathbf{X}}\frac{1}{2}\left\|\mathbf{Y}'_{(t)}-\mathbf{A}'_{(t+1)}\mathbf{X}\right\|_F^2+\alpha\|\mathbf{X}\|_{2,1}+\iota_{\mathcal{S}_{q-1}}(\mathbf{X}) \quad (11)$$

which is precisely the constrained $\ell_2-\ell_{2,1}$ optimization solved by CLSUnSAL [11], which is a generalization of the sparse unmixing by variable splitting and augmented Lagrangian (SUnSAL) algorithm [37].

The pseudocode for the proposed algorithm, which is termed R-CoNMF, is shown in Algorithm 1.

**Algorithm 1:** Robust Collaborative Nonnegative Matrix Factorization (R-CoNMF)

**Input:** $\mathbf{Y}$ (dataset set), $T\in\mathbb{N}$ (iterations)
$\alpha>0,\beta>0$ (regularization parameters)
$\lambda_t,\mu_t$ (proximal sequences of parameters)
$q>p$ (overestimate of the number of endmembers)
$\delta$ (stopping threshold), $\theta>0$ ( a very small number)
**Output:** $(\widehat{\mathbf{M}},\widehat{\mathbf{S}})$ (mixing and abundances matrices)

1 **Initialization:**
/* Overestimate the signal subspace */
2 $\mathbf{U}\in\mathbb{R}^{d\times q}\leftarrow$ HySime$(\mathbf{Y},q)$ /* orthogonal basis */
3 $\mathbf{Y}\leftarrow\mathbf{U}^T\mathbf{Y}$; $\overline{\mathbf{Y}}\leftarrow\overline{\mathbf{y}}\mathbf{1}_q^T$
4 $\overline{\mathbf{V}}=\text{orth}(\mathbf{Y}*\mathbf{Y}^T/n-\overline{\mathbf{y}}\,\overline{\mathbf{y}}^T,q-1)$ /* $q-1$ principal components */
5 $\mathbf{A}_{(0)}\leftarrow\text{VCA}(\mathbf{Y},q)$ /* estimate $q$ endmembers */
6 $\boldsymbol{\Delta}_{(0)}=\mathbf{V}^T(\mathbf{A}_{(0)}-\overline{\mathbf{Y}})$
7 $\mathbf{P}\leftarrow\mathbf{A}_{(0)}$ /* the VCA solution is used as pure-pixel based solution */
8 $\mathbf{X}_{(0)}\leftarrow\text{SUnSAL}(\mathbf{A}_{(0)},\mathbf{Y})$ /* abundances */
9 $\varepsilon_0=\infty$; $t\leftarrow-1$
10 **begin**
11 **repeat**
12 $t\leftarrow t+1$
/* Optimize with respect to $\mathbf{A}$ */
13 $\boldsymbol{\Delta}_{(t+1)}\leftarrow\mathbf{V}^T\big((\mathbf{Y}-\overline{\mathbf{Y}})\mathbf{X}_{(t)}^T+\beta(\mathbf{P}-\overline{\mathbf{Y}})+\lambda_t\boldsymbol{\Delta}_{(t)}\big)\big(\mathbf{X}_{(t)}\mathbf{X}_{(t)}^T+(\beta+\lambda_t)\mathbf{I}\big)^{-1}$
14 $\mathbf{A}_{(t+1)}\leftarrow\overline{\mathbf{Y}}+\mathbf{V}\boldsymbol{\Delta}_{(t+1)}$
/* Optimize with respect to $\mathbf{X}$ */
15 $\mathbf{Y}'_{(t)}\leftarrow[\mathbf{Y}^T\sqrt{\mu_t}\,\mathbf{X}_t^T]^T$, $\mathbf{A}'_{(t+1)}\equiv[\mathbf{A}_{(t+1)}^T\sqrt{\mu_t}\,\mathbf{I}]^T$
16 $\mathbf{X}_{(t+1)}\leftarrow\text{CLSUnSAL}(\mathbf{A}_{(t+1)},\mathbf{Y}'_{(t)},\alpha)$ /* [11] */
17 $\varepsilon_t\leftarrow\left\|\mathbf{Y}-\mathbf{A}_{(t)}\mathbf{X}_{(t)}\right\|_F$
18 **until** $\big(t\le T$ **and** $|\varepsilon_t-\varepsilon_{t-1}|/\|\mathbf{Y}\|_F>\delta\big)$
/* Detect the support of $\mathbf{X}_{(t)}$ */
19 $\Lambda=\emptyset$
20 **for** $i=1$ **to** $q$ **do**
21 **if** $\|\mathbf{x}_{(t)}^i\|\le\theta$ **then**
22 $\Lambda\leftarrow\Lambda\cup\{i\}$
23 **end**
24 **end**
/* Retrieve the estimates of $\mathbf{M}$ and $\mathbf{S}$ */
25 $\widehat{\mathbf{M}}\leftarrow\mathbf{U}\mathbf{A}_{(t)|\Lambda}$; $\widehat{\mathbf{S}}\leftarrow\mathbf{X}_{(t)|\Lambda}$
26 **end**

In line 2, HySime [20] is used to obtain an overestimate of the signal subspace. HySime can be used in two ways: 1) to estimate the signal subspace, thus including its dimension; or

2) to estimate the subspace of a given order, supplied as input, that minimizes the signal projection error. In this paper, HySime is used in the latter fashion, that is, to estimate a given size of signal subspace, which does not necessarily correspond to the number of endmembers in the data. The reason for this choice is that it often happens that the estimation of the true subspace is a quite challenging problem due to, e.g., poor estimation of the noise statistics or the presence of nonlinear components in the data set. However, the estimation of an overestimate of the signal subspace is an easier problem. Hence, we input a value of $q$ larger than that determined by HySime and let R-CoNMF compute, hopefully, a better estimate, as it has built-in information linked with the LMM not exploited in most signal subspace estimation methods, such as HySime, principal component analysis (PCA), maximum noise fraction (MNF) [38], or VD [18].

In line 2, the columns of $\mathbf{U}$ form an orthogonal basis whose span contains the signal subspace estimated by HySime. As aforementioned, it is very easy to obtain an overestimation of the subspace. Nevertheless, it should be noted that $\mathbf{U}$ can be obtained by any subspace estimation method, such as singular value decomposition, PCA, MNF, or VD. In line 3, the matrix operation $\mathbf{U}^T\mathbf{Y}$ computes the projection of the data set $\mathbf{Y}$ onto the subspace spanned by the columns of $\mathbf{U}$. We remark that this operation reduces the size of the spectral vectors from $d$ to $q$. Since usually $d \gg q$, this projection brings considerable storing and computational advantages.

Matrix $\mathbf{V}$ in line 4 holds the $q-1$ principal components used to represent $\mathbf{A}$ by the affine set $\mathcal{A}_{q-1}$. The initialization of $\mathbf{A}$ and $\mathbf{X}$ is computed in lines 5 and 8 using, respectively, the VCA [39] and SUnSAL [37] algorithms.

Lines 11–18 implement PAO optimizations (6) and (7) and are the core of R-CoNMF. The stopping criterion is based on the relative reconstruction error or maximum number $T$ of iterations. Lines 19–24 detect the support of $\mathbf{X}_{(t)}$, and line 25 retrieves $\widehat{\mathbf{M}}$ and $\widehat{\mathbf{S}}$ according to (3) and (4).

The computational complexity measured in terms of the number of floating-point operations of R-CoNMF is dominated by computations of $\mathbf{A}_{(t+1)}$ in lines 13 and 14 and of $\mathbf{X}_{(t+1)}$ in line 16. The leading term of the computational complexity in the computation of $\mathbf{A}_{(t+1)}$ is $2q^2n$ coming from the matrix multiplications $\mathbf{Y}\mathbf{X}_{(t)}^T$ and $\mathbf{X}_{(t)}\mathbf{X}_{(t)}^T$. The computational complexity of CLSUnSAL is $\mathcal{O}(nq^2N)$, where $N$ is the number of iterations thereof [11]. We conclude therefore that R-CoNMF complexity per iteration is dominated by the computation of $\mathbf{X}_{(t+1)}$ and is approximately given by $\mathcal{O}(nq^2N)$.

### A. Convergence of R-CoNMF

The convergence of alternating optimization (AO) algorithms involving nonconvex and/or nondifferentiable terms is a long-standing research problem to which considerable research efforts have been devoted (see, e.g., [40]). Recently, the exploitation of the Kurdyka–Łojasiewicz inequality (see, e.g., [35], [40], and references therein) has opened the door to a wealth of new AO algorithms and convergence results. Herein and following very closely the work in [35], we prove that R-CoNMF converges to a critical point of $L : \mathbb{R}^{d\times q} \times \mathbb{R}^{q\times n} \to \mathbb{R} \cup \{+\infty\}$.

As previously mentioned, the core of R-CoNMF consists of lines 11–18 of Algorithm 1. This fragment of pseudocode implements the PAO algorithm shown in (6) and (7). The next theorem asserts that the sequence $(\mathbf{A}_{(t)}, \mathbf{X}_{(t)})$, for $t = 0, 1, \ldots$, produced by this algorithm converges to a critical point of $L$. This result is a direct application of [35, Th. 9].

*Theorem 1:* Assume that for some $r_- < r+$, the sequence of stepsizes $\lambda_t, \mu_t$, for $t = 1, \ldots$, belongs to $(r_-, r_+)$, that $L(\cdot, \mathbf{X}_{(0)})$ is proper, and that $\alpha > 0$ and $\beta > 0$, i.e., the regularization parameters shown in (5) are positive. Then, the sequence $(\mathbf{A}_{(t)}, \mathbf{X}_{(t)})$ converges to a critical point of $L$.

*Proof:* Consider the rearrangement of the terms of $L$ in (5) as $L(\mathbf{A}, \mathbf{X}) \equiv f(\mathbf{A}) + Q(\mathbf{A}, \mathbf{X}) + g(\mathbf{X})$, where

$$Q(\mathbf{A}, \mathbf{X}) \equiv \frac{1}{2}\|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2 \tag{12}$$

$$f(\mathbf{A}) \equiv \frac{\beta}{2}\|\mathbf{A} - \mathbf{P}\|_F^2 + \iota_{\mathcal{A}_{q-1}}(\mathbf{A}) \tag{13}$$

$$g(\mathbf{X}) \equiv \alpha\|\mathbf{X}\|_{2,1} + \iota_{\mathcal{S}_{q-1}}(\mathbf{X}). \tag{14}$$

The function $L : \mathbb{R}^{d\times q} \times \mathbb{R}^{q\times n} \to \mathbb{R}$ and its terms $f : \mathbb{R}^{d\times q} \to \mathbb{R}$, $g : \mathbb{R}^{d\times q} \to \mathbb{R} \cup \{+\infty\}$, and $Q : \mathbb{R}^{d\times q} \times \mathbb{R}^{q\times n} \to \mathbb{R}$ have the following properties.

- P1) $\inf_{\mathbb{R}^{d\times q}\times\mathbb{R}^{q\times n}} L > -\infty$.
- P2) $f : \mathbb{R}^{d\times q} \to \mathbb{R} \cup \{+\infty\}$, $g : \mathbb{R}^{d\times q} \to \mathbb{R} \cup \{+\infty\}$ are proper lower semicontinuous.
- P3) $Q : \mathbb{R}^{d\times q} \times \mathbb{R}^{q\times n} \to \mathbb{R}$ is $C^1$.
- P4) $\nabla Q$ is Lipschitz continuous on bounded subsets of $\mathbb{R}^{d\times q} \times \mathbb{R}^{q\times n}$.
- P5) $L$ has the Kurdyka–Łojasiewicz inequality at each point of the domain of $f$.

Property P1) is trivial. Property P2) stems from the fact that $f$ and $g$ have a nonempty domain and are continuous. Property P3) is a consequence of $Q$ being $C^\infty$, which also implies P4) since $\|\nabla^2 Q\|$ is bounded in the spectral norm (i.e., the matrix norm induced by the Euclidean norm) in any bounded subset of $\mathbb{R}^{d\times q} \times \mathbb{R}^{q\times n}$. Property P5) stems from the fact that $L$ is a finite sum of semialgebraic functions (indicator of polyhedral sets, polynomial functions, and Euclidian norms) [35], [40].

From (6) and (7), we have $L(\mathbf{A}_{(t)}, \mathbf{X}_{(t)}) \geq L(\mathbf{A}_{(t+1)}, \mathbf{X}_{(t)}) \geq L(\mathbf{A}_{(t+1)}, \mathbf{X}_{(t+1)})$. That is, the sequence $L(\mathbf{A}_{(t)}, \mathbf{X}_{(t)})$ is decreasing. Given that $\alpha > 0$ and $\beta > 0$, then $L$ is coercive, i.e.,

$$\lim_{\substack{\|\mathbf{A},\mathbf{X}\|_F\to\infty \\ (\mathbf{A},\mathbf{F})\in \mathrm{dom} L}} L(\mathbf{A}, \mathbf{X}) = \infty. \tag{15}$$

Therefore, $\|\mathbf{A}_{(t)}, \mathbf{X}_{(t)}\|_F$ is bounded. Otherwise, $L(\mathbf{A}_{(t)}, \mathbf{X}_{(t)})$ could not be decreasing.

Taking into account properties P1)–P5) and the fact that $\|\mathbf{A}_{(t)}, \mathbf{X}_{(t)}\|_F$ is bounded, then invoking [35, Th. 9], we conclude that $(\mathbf{A}_{(t)}, \mathbf{X}_{(t)})$ converges to a critical point of $L$. □

### B. Applying R-CoNMF

R-CoNMF may be applied assuming that the number of endmembers is either 1) known or 2) unknown. As previously discussed, in the optimization problem (2), the regularizer on the abundance matrix, i.e., $\alpha\|\mathbf{X}\|_{2,1}$, aims at predicting the number of endmembers, whereas the volume regularizer on the mixing matrix, i.e., $(\beta/2)\|\mathbf{A}-\mathbf{P}\|_F^2$, targets for unmixing. Therefore, in the former case, we set $q=p$ and $\beta \gg \alpha$, thus removing the $\ell_{2,1}$ regularizer from the objective function, although keeping the constraint set $\mathcal{S}_{q-1}$. In the latter case, we first apply R-CoNMF to infer the number of endmembers by using $\alpha \gg \beta$. After obtaining the correct number of endmembers, we then apply again R-CoNMF as described in scenario 1).

Let us consider scenario 2). In this case, we run R-CoNMF for a fixed $q > p$, in which we set $\alpha \gg \beta$. We empirically found out that $\beta$ can be very small, $\beta \to 0$, and $\alpha$ can be very large, which would not influence the solution. Let us define $\zeta(i) = \|\mathbf{x}^i\|_2$, for $i = 1, \ldots, q$, as a measure of the sparsity of abundances associated with the corresponding endmembers $\mathbf{a}_i$, for $i = 1, \ldots, q$. Ideally, we should have $\zeta(i) = 0$ if $\mathbf{a}_i$ is not active. Due to the impact of noise and model errors, we relax that criterion as follows: We consider that an endmember $\mathbf{a}_i$ is active, if $\zeta(i) > \xi$, for a small $\xi > 0$. Fig. 5 shows the obtained $\zeta$ and the reconstruction error for a problem with $p = 6$, $n = 4000$, and zero-mean Gaussian independent and identically distributed (i.i.d.) noise with SNR $= 30$ dB. The application of the above criterion with any value of $\xi$ between 0.5 and 4 yields $p = 6$, which is the correct estimate of the number of endmembers. Notice that this number could also be obtained from the plot of the reconstruction error. In more complex scenarios with a lower SNR and a larger number of endmembers, we may devise a strategy to combine both indicators.

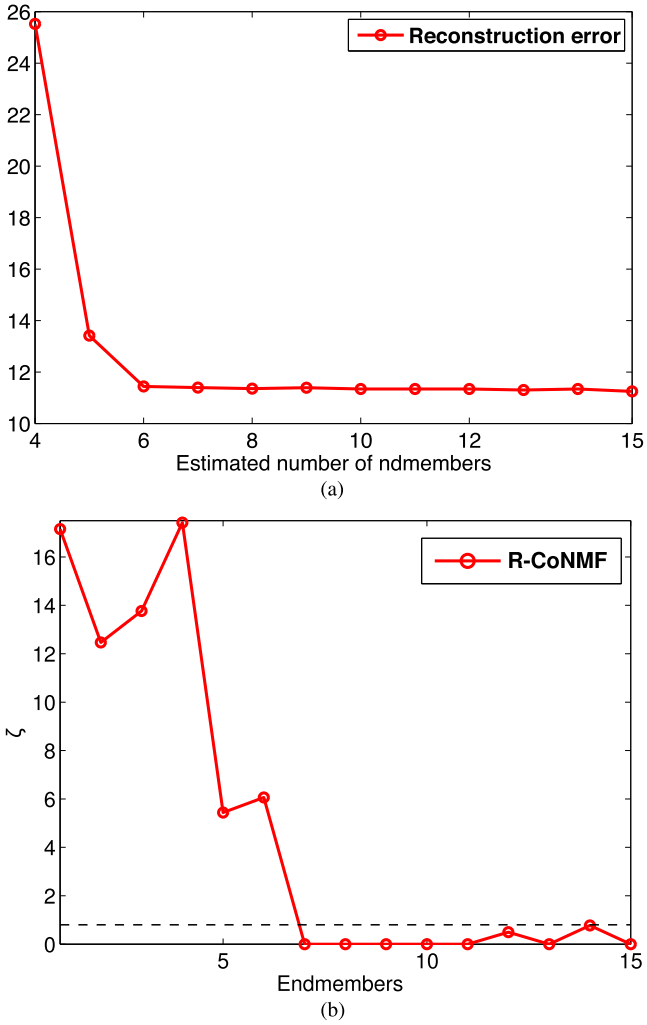



Fig. 5. Simulated problem with $p = 6$, $n = 4000$, and zero-mean Gaussian i.i.d. noise with SNR $= 30$ dB. For the volume regularizer, we use the VCA solution as its boundary. (a) Reconstruction error as a function of the estimated number of endmembers. (b) Degree of sparseness $\zeta$ for $q = 15$.

## IV. Experiments With Simulated Data

In this section, we evaluate the proposed R-CoNMF method using synthetic hyperspectral data. The advantage of using synthetic scenes is that they offer a fully controlled analysis scenario in which the properties of our algorithm can be precisely investigated. The synthetic scenes have been generated using the LMM in (1). The scenes comprise $n = 4000$ pixels, and the spectral signatures used for their generation were randomly selected from the USGS digital spectral library. To ensure the difference among the endmembers used for simulation purposes, the spectral angle distance (SAD) between any two spectral signatures is bigger than $10^\circ$. Furthermore, let $p_{\text{mix}}$ be the number of endmembers in one pixel. In real scenarios, it is possible to have a large number of endmembers in a scene, for instance, $p \geq 10$. However, it is unlikely to have a large number of endmembers in one pixel. That is, in general, $p_{\text{mix}}$ is relatively small, e.g., $p_{\text{mix}} \leq 5$. Based on this observation, for the simulated data, if $p \geq 5$, we set $p_{\text{mix}} = 5$. Otherwise, if $p < 5$, we set $p_{\text{mix}} = p$. Finally, to ensure that no pure pixels are present in the synthetic images, we discard all pixels with abundance fractions larger than 0.8, i.e., $\max(\mathbf{s}_i) \leq 0.8$.

In the case of $q = p$, let $\widehat{\mathbf{M}} = \mathbf{A}$ and $\widehat{\mathbf{S}} = \mathbf{X}$ denote the estimates of $\mathbf{M}$ and $\mathbf{S}$, respectively. As performance indicators, we use the relative reconstruction error $\text{RRE} = \|\mathbf{Y} - \widehat{\mathbf{M}}\widehat{\mathbf{S}}\|_F^2 / \|\mathbf{Y}\|_F$, the SAD (in degrees), and two error metrics focused on evaluating the quality of the estimated endmembers, i.e., $\|\widehat{\mathbf{M}} - \mathbf{M}\|_F$, and the quality of the estimated abundances, i.e., $(1/\sqrt{n \times p})\|\widehat{\mathbf{S}} - \mathbf{S}\|_F$.

### A. Parameter Setting

First of all, it should be noted that the observed data $\mathbf{Y}$ is spherized based on the covariance matrix. In all our experiments, we projected the data onto a $q$-dimensional subspace with $q \geq p$. This is based on the fact that hyperspectral data live in a much lower space, and it is very easy to obtain an overestimation of the lower subspace. This can be easily done by PCA [41], retaining, for instance, 99% of the spectral information. In our simulation, we define the affine projection via singular value decomposition by a given $p$.

In addition, there are four parameters: $\alpha$, $\beta$, $\boldsymbol{\lambda} = \{\lambda_0, \lambda_1, \ldots, \lambda_t, \ldots\}$, and $\boldsymbol{\mu} = \{\mu_0, \mu_1, \ldots, \mu_t, \ldots\}$, and the volume regularizer $\mathbf{P}$ involved in R-CoNMF. For these parameters, we use the following settings in our experiments.

- As discussed in Section III-A, R-CoNMF converges for any $\lambda_t > 0$ and $\mu_t > 0$. In the following experiments, we use $\lambda_t = 1$ and $\mu_t = 1$.
- As presented in Section III-B, in all our experiments, we perform two steps for R-CoNMF: a first step to determine the number of endmembers and a second step in which the data are unmixed.
  - In the first step, in which the number of endmembers is assumed to be unknown, i.e., $q > p$, we set $\alpha = 10^{-1}$ and $\beta = 10^{-8}$. This aims at estimating the number of endmembers by enforcing the $\ell_{2,1}$ regularizer, where the estimation is insensitive to $\alpha$.
  - In the second step, where the number of endmembers is assumed to be known, i.e., $q = p$, we have a more difficult situation to set $\beta$. However, we empirically found out that it is easy to obtain a suboptimal setting. Fig. 6 shows the obtained spectral signatures by using different settings of $\beta$: 1) $\mathbf{P}$ is provided by VCA; and 2) $\mathbf{P}$ is defined by the data mass, for a simulated problem by using three endmembers, 4000 pixels, SNR $=$ 30 dB, and $\alpha = 1e-8$, in which $\beta$ changes from 10 to $1e-5$ from black to red. From this example, we can see that it is easy to define a good setting of $\beta$. Therefore, in our experiments, when $q = p$, we set $\beta = 10^{-1}$ and $\alpha = 10^{-8}$.
- In Figs. 1 and 6, we can also observe that R-CoNMF is more robust with the volume regularizer defined by a given solution. Therefore, for the volume regularizer, we use the pure-pixel-based solution provided by the VCA algorithm [39] as its boundary.

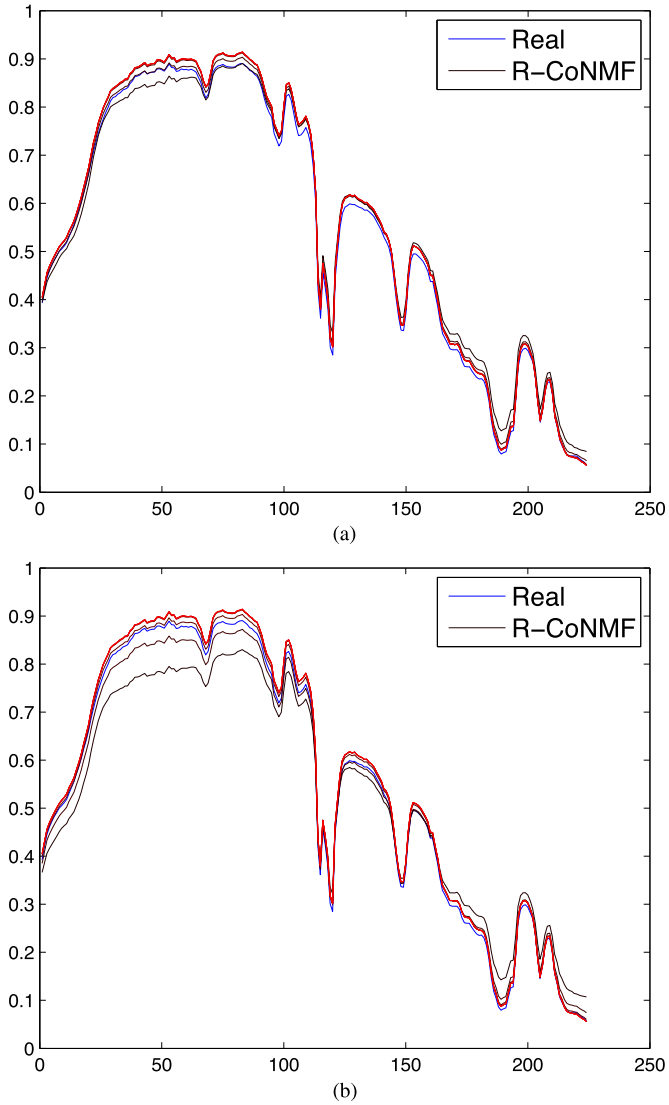



Fig. 6. Obtained spectral signatures by using different settings of $\beta$. (a) $\mathbf{P}$ is provided by VCA. (b) $\mathbf{P}$ is defined by the data mass. The $x$-axis represents spectral band numbers, whereas the $y$-axis represents reflectance values.

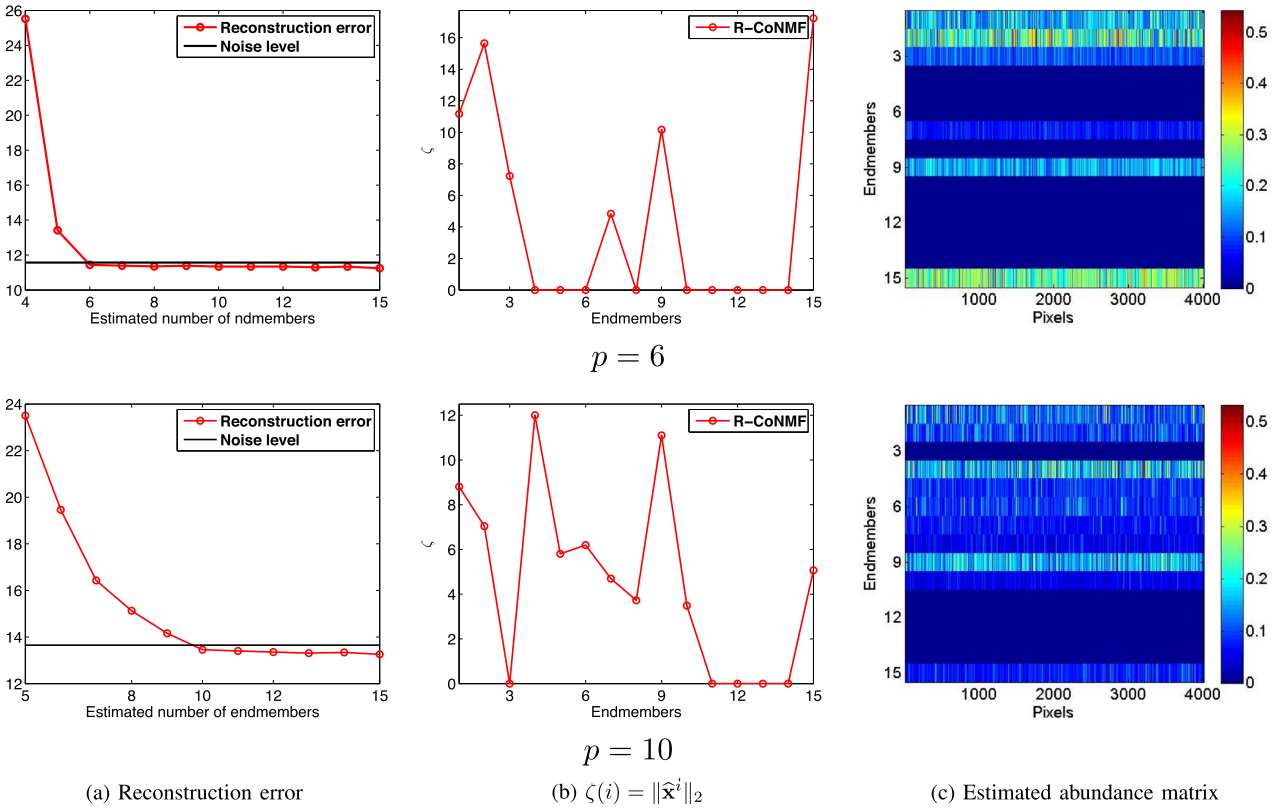
Finally, we also emphasize that all the values reported in the subsequent tables are obtained as the average of 30 independent Monte Carlo runs, whereas the results displayed in figures correspond to one of the 30 Monte Carlo runs conducted.

### *B. Experiment 1: Estimation of the Number of Endmembers*

This experiment aims at evaluating the performance of R-CoNMF for the estimation of the number of endmembers. In the experiments, we set $\alpha \gg \beta$, which mainly focuses on the collaborative constraint. Before describing the results obtained in this experiment, we introduce three new performance discriminators. Let $\ell(q) = \|\mathbf{Y} - \widehat{\mathbf{A}}\widehat{\mathbf{X}}\|_2^2$ be the reconstruction error obtained for $q$ endmembers. We emphasize that, for $q = 1, \ldots, p$, $\ell(q)$ should decrease. Therefore, we should be able to easily find a good estimated number of endmembers if $q$ is equal to (or slightly larger than) $p$. Let $\varepsilon = \|\mathbf{N}\|_F$ be the noise level. Ideally, when $q \geq p$, $\ell(q) \approx \varepsilon$. In this experiment, we use the two aforementioned discriminators, i.e., $\ell(q)$ and $\zeta(i)$, to evaluate the capacity of R-CoNMF in identifying the real number of endmembers in a scene.

Fig. 7 shows the results obtained by the three considered discriminators for two cases: $p = 6$ and $p = 10$. For Fig. 7(a), we can conclude that when $q < p$, the reconstruction error is relatively big, and when $q \geq p$, the reconstruction error significantly decreases, which is similar to the noise level $\varepsilon$. Therefore, it is easy to identify the correct number of endmembers by analyzing the turning point of the reconstruction error. Furthermore, Fig. 7(b) and (c) shows the collaborative regularizer and the estimated abundance matrix, respectively. It can be observed that the number of endmembers can be correctly estimated.

Finally, we perform a comparison with HySime for estimating the number of endmembers. The proposed R-CoNMF estimated the subspace by including the sum-to-one and nonnegativity constraints, whereas HySime performs estimation just on the linear mixing model. Ideally, R-CoNMF exploits more prior information. However, we observed that when the




(a) Reconstruction error (b) $\zeta(i) = \|\hat{\mathbf{x}}^i\|_2$ (c) Estimated abundance matrix

Fig. 7. (a) Reconstruction error $\|\mathbf{Y} - \widehat{\mathbf{M}}\widehat{\mathbf{S}}\|_2^2$ obtained by R-CoNMF for a simulated problem and SNR = 30 dB, as a function of the estimated number of endmembers $q$. (b) Regularizer of the abundance matrix $\widehat{\mathbf{X}}$. (c) Estimated abundance matrix $\widehat{\mathbf{X}}$.

number of endmembers is small, i.e., $p < 8$, both methods lead to very good performance. The advantage of using more prior information for the proposed R-CoNMF comes when $p$ increases. For $p = 10$, R-CoNMF missed seven times, whereas HySime missed nine, out of 30 runs. When $p = 15$, R-CoNMF only provided the correct estimation six times, whereas HySime provided the correct estimation two times. The other estimation varies in 12–16. This slight difference indicates that with more prior information in hand, R-CoNMF can lead to better estimates of the data subspace.

### C. Experiment 2: $q = p$

In a second experiment, we evaluate the performance of R-CoNMF in the case that the number of estimated endmembers coincides with the number of real endmembers, i.e., $q = p$. Here, we use the MVC-NMF algorithm in [25], MVSA [26], and SISAL [27] for comparison with our method. Table I displays the results obtained by MVC-NMF and our proposed R-CoNMF algorithm for all considered performance discriminators for different values of $q = p = \{4, 6, 8, 10, 15\}$. In all cases, we considered an SNR of 30 dB and reported the results obtained from averaging the results of 30 Monte Carlo runs.

From the results reported in Table I, we can make the following observations. First and foremost, when there are only a few endmembers in the image (e.g., $q = p = 4$), all algorithms obtained very good results. This is expected, since in this case, it is relatively easy to solve the optimization problem. It is interesting to observe that as the number of endmembers increases, R-CoNMF obtained very good performance (note the good performance obtained for the case $q = p = 10$). Even in a very difficult scenario such as $q = p = 15$, the solution provided by R-CoNMF is still useful. It should be noted that in cases with a relatively high number of endmembers (i.e., $q = p \geq 8$), the other algorithms yield useless results. This is because when the number of endmembers increases, most pixels are likely to fluctuate around the facets, which is a situation in which minimum-volume-based algorithms are likely to fail [3]. Even in this difficult scenario, in which MVC-NMF, MVSA, and SISAL could not provide feasible results, the proposed R-CoNMF was able to provide a reasonable solution. Based on this experiment, we can conclude that R-CoNMF is quite robust and has no strong constraints related with the quality of the analyzed data set.

For illustrative purposes, Fig. 8 shows the signatures estimated by R-CoNMF and MVC-NMF. The estimated spectral signatures by R-CoNMF are similar to the real ones, whereas those estimated by MVC-NMF are slightly different. Similar observations can be made from the difference maps between the real and estimated abundance maps, as shown in Fig. 9, where the difference of R-CoNMF is much smaller than that of

TABLE I
EVALUATION OF THE PERFORMANCE OF R-CONMF AND MVC-NMF IN THE UNMIXING OF A SYNTHETIC HYPERSPECTRAL DATA SET, SIMULATED WITH SNR = 30 dB, FOR DIFFERENT NUMBERS OF ENDMEMBERS AND $q = p$, WHERE "-" MEANS NO RESULTS

| $q = p$ | R-CoNMF | | | | MVC-NMF | | | |
|---|---|---|---|---|---|---|---|---|
| | $\|\widehat{\mathbf{M}} - \mathbf{M}\|_F$ | $\frac{1}{\sqrt{n \times p}}\|\widehat{\mathbf{S}} - \mathbf{S}\|_F$ | SAD | RRE | $\|\widehat{\mathbf{M}} - \mathbf{M}\|_F$ | $\frac{1}{\sqrt{n \times p}}\|\widehat{\mathbf{S}} - \mathbf{S}\|_F$ | SAD | RRE |
| 4 | 0.17±0.09 | 0.01±3e-3 | 0.64±0.46 | 0.03±0.0 | 0.84±0.47 | 0.03±0.02 | 2.54±1.73 | 0.20±0.05 |
| 6 | 0.20±0.08 | 0.01±4e-3 | 0.56±0.21 | 0.03±1e-4 | 1.44±1.17 | 0.04±0.02 | 3.48±2.57 | 0.23±0.04 |
| 8 | 0.76±0.23 | 0.02±0.01 | 1.89±0.67 | 0.03±1e-3 | 36.50±154.23 | 0.08±0.05 | 11.97±19.48 | 0.25±0.07 |
| 10 | 1.28±0.50 | 0.03±0.01 | 2.77±1.73 | 0.03±1e-3 | - | - | - | - |
| 15 | 3.13±1.81 | 0.05±0.01 | 5.25±3.87 | 0.04±1e-3 | - | - | - | - |
| $q = p$ | MVSA | | | | SISAL | | | |
| | $\|\widehat{\mathbf{M}} - \mathbf{M}\|_F$ | $\frac{1}{\sqrt{n \times p}}\|\widehat{\mathbf{S}} - \mathbf{S}\|_F$ | SAD | RRE | $\|\widehat{\mathbf{M}} - \mathbf{M}\|_F$ | $\frac{1}{\sqrt{n \times p}}\|\widehat{\mathbf{S}} - \mathbf{S}\|_F$ | SAD | RRE |
| 4 | 0.33±0.20 | 1e-4 | 1.18±0.79 | 0.03 | 0.24±0.14 | 1e-4 | 0.82±0.48 | 0.03 |
| 6 | 1.48±0.42 | 2e-4 | 4.88±4.03 | 0.03 | 1.21±0.32 | 2e-4 | 3.48±1.80 | 0.03 |
| 8 | 3.47±1.29 | 3e-4 | 10.41±5.70 | 0.03 | 2.89±0.77 | 2e-4 | 8.84±4.66 | 0.03 |

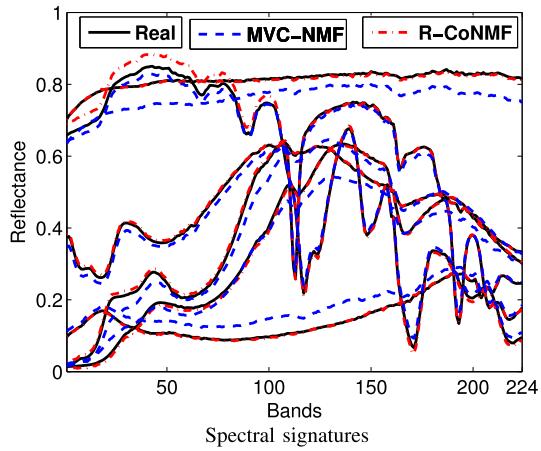



Fig. 8. Spectral signatures of the endmembers extracted by R-CoNMF and MVC-NMF as compared with the reference signatures used for the simulation of a synthetic scene with $q = p = 6$ and SNR = 30 dB.

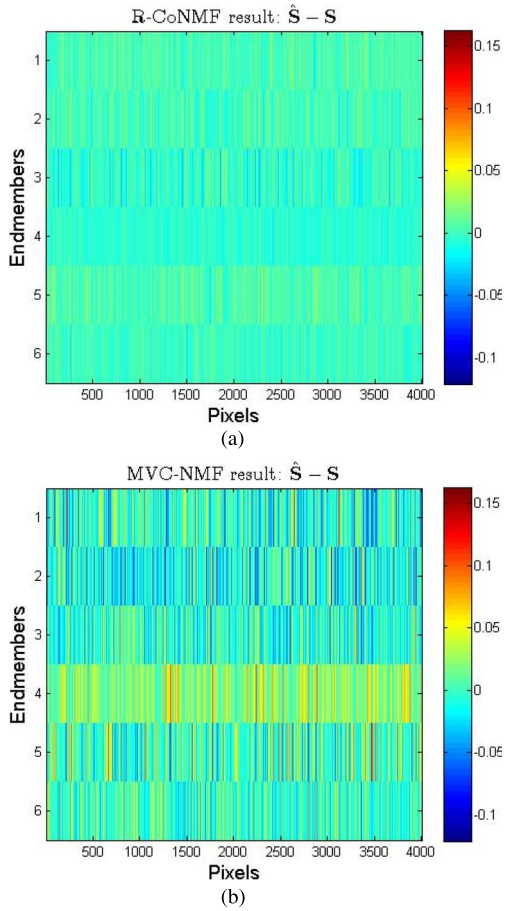



Fig. 9. (a) Difference between real and estimated abundances for the R-CoNMF algorithm. (b) Difference between the real and estimated abundances for the MVC-NMF algorithm.

MVC-NMF. The results in this section indicate that R-CoNMF can perform very accurately in the case that the number of endmembers is known *a priori*, i.e., $q = p$.

### *D. Experiment 3: Robustness to Noise*

In this experiment, we illustrate the performance of the proposed R-CoNMF under different noise conditions with a simulated scenario given by $p = q = 5$. The goal of this experiment is mainly to illustrate the performance of the algorithm with different noise levels. Table II reports the unmixing results obtained by R-CoNMF under different noise levels, considering SNR values of 20, 40, 60, and 80 dB. As shown in Table II, when the noise level is moderately low (i.e., SNR of less than 40 dB), the proposed R-CoNMF can obtain very good unmixing results. In turn, when the level of noise significantly increases (i.e., SNR of 20 dB), the quality of the unmixing results decreases, particularly with regard to the SAD metric. However, it should be noted that 20 dB is not a realistic noise level given the SNR of current imaging spectrometers. As a result, we can conclude that the results reported in Table II indicate very high robustness to noise by the proposed CoNMF.

TABLE II
UNMIXING RESULTS OBTAINED BY CONMF IN A SIMULATED SCENARIO WITH $p = 5$, $q = 5$, AND DIFFERENT NOISE LEVELS, WITH SNR VALUES OF 20, 40, 60, AND 80 dB

| SNR | R-CoNMF | | | | MVC-NMF | | | |
|---|---|---|---|---|---|---|---|---|
| | $\|\widehat{\mathbf{M}} - \mathbf{M}\|_F$ | $\frac{1}{\sqrt{n \times p}}\|\widehat{\mathbf{S}} - \mathbf{S}\|_F$ | SAD | RRE | $\|\widehat{\mathbf{M}} - \mathbf{M}\|_F$ | $\frac{1}{\sqrt{n \times p}}\|\widehat{\mathbf{S}} - \mathbf{S}\|_F$ | SAD | RRE |
| 20 | 0.91±0.39 | 0.03±0.01 | 3.68±3.08 | 0.10±1e-4 | 1.15±0.38 | 0.04±0.01 | 3.80±1.50 | 1.88±0.37 |
| 40 | 0.21±0.07 | 0.01±2e-3 | 0.58±0.30 | 0.01 | 1.31±0.61 | 0.04±0.01 | 3.14±1.22 | 0.03±0.01 |
| 60 | 0.22±0.07 | 0.01±2e-3 | 0.70±0.50 | 1e-3 | 1.14±0.22 | 0.04±0.01 | 2.88±0.64 | 0.01±4e-3 |
| 80 | 0.21±0.07 | 0.01±2e-3 | 0.66±0.30 | 4e-4±1e-4 | 0.97±0.28 | 0.03±0.01 | 2.53±0.84 | 0.01±0.01 |
| SNR | MVSA | | | | SISAL | | | |
| | $\|\widehat{\mathbf{M}} - \mathbf{M}\|_F$ | $\frac{1}{\sqrt{n \times p}}\|\widehat{\mathbf{S}} - \mathbf{S}\|_F$ | SAD | RRE | $\|\widehat{\mathbf{M}} - \mathbf{M}\|_F$ | $\frac{1}{\sqrt{n \times p}}\|\widehat{\mathbf{S}} - \mathbf{S}\|_F$ | SAD | RRE |
| 20 | 2.26±1.08 | 3e-4±1e-4 | 9.11±5.63 | 0.10±1e-4 | 1.67±0.77 | 3e-4±1e-4 | 6.09±3.66 | 0.10±1e-4 |
| 40 | 0.21±0.14 | 0 | 0.57±0.33 | 0.01 | 0.12±0.10 | 0 | 0.29±0.12 | 0.01 |
| 60 | 0.02±0.01 | 0 | 0.06±0.03 | 1e-3 | 0.04±0.02 | 0 | 0.14±0.09 | 1e-3 |
| 80 | 0.02±0.01 | 0 | 0.05±0.03 | 1e-4 | 0.04±0.01 | 0 | 0.14±0.08 | 1e-4 |

## V. EXPERIMENTS WITH REAL DATA

The scene used in our real-data experiments is the well-known Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) Cuprite data set [42], available online in reflectance units.[2] This scene has been widely used to validate the performance of endmember extraction algorithms. The portion used in experiments corresponds to a 250 × 191 pixel subset of the sector labeled f970619t01p02r02 in the online data.[3] The scene comprises 224 spectral bands between 0.4 and 2.5 $\mu$m, with a nominal spectral resolution of 10 nm. Prior to the analysis, bands 1–2, 105–115, 150–170, and 223–224 were removed due to water absorption and low SNR in those bands, leaving a total of 188 spectral bands. The Cuprite site is well understood mineralogically and has several exposed minerals of interest, all included in the USGS library considered in experiments, denoted splib06[4] and released in September 2007. In our experiments, we use spectra obtained from this library (convolved and downsampled to AVIRIS wavelengths) to substantiate the quality of the endmembers derived by R-CoNMF. It should be noted that these data have been extensively used as a benchmark for spectral unmixing applications in previous work [3]. As a result and for space considerations, we only conduct comparisons with MVC-NMF in this work. For a more detailed comparison with other methods, we refer to the review work in [3].

For illustrative purposes, Fig. 10(a) shows a mineral map produced in 1995 by USGS, in which the Tricorder 3.3 software product was used to map different minerals present in the Cuprite mining district.[5] The 250 × 190 pixel subscene used in our experiments is shown in Fig. 10. It should be noted that all experiments with this subscene have been performed in a desktop PC with an Intel Core I5 CPU and 4 GB of RAM. Concerning the parameters involved in R-CoNMF, we consequentially follow the settings in the simulated experiments. Regarding the affine projection, we have used the projective projection instead of the affine projection (see [3]), as the former works slightly better in this example. Finally, it should be noted that for the initialization, we run VCA $t = 30$ times and retain the simplex of maximum volume to produce a good initialization, as CoNMF is constrained by the VCA solution. In the case of VCA, this makes sense given the random directions that this algorithm uses to find the extremes of the simplex. For $t = 30$, this procedure takes just 2 s in a standard PC.

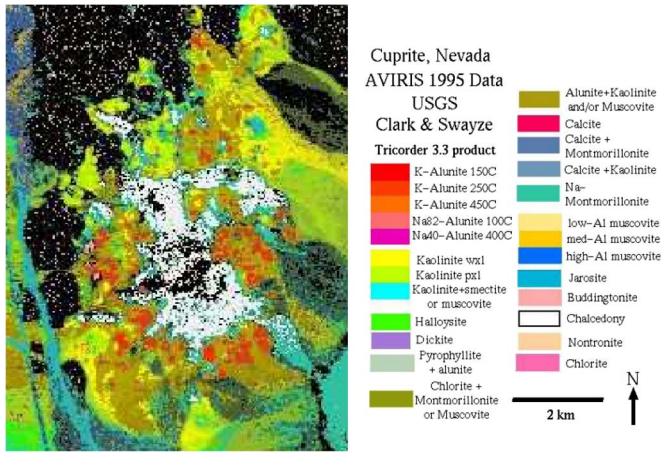



Fig. 10. 250 × 190 pixel subscene used in our experiments, showing the location of different minerals in the Cuprite mining district in Nevada.

### A. Estimation of the Number of Endmembers: $p$

The estimation of the number of endmembers is a challenging issue for the AVIRIS Cuprite scene, due to the strong spectral variability and the large number of mineral alterations present in the scene. According to the official ground truth map generated by the Tricorder 3.3 software and the USGS library, we can identify up to five groups of minerals with 79 endmembers (including different alterations for the same minerals). If we use the VD to estimate the number of endmembers [43], we obtain $p = 14$ endmembers (see [39]) and $p = 9$ for a subset of this image (see [25]). According to the HySime algorithm [20],

[2]http://aviris.jpl.nasa.gov/html/aviris.freedata.html
[3]http://www.lx.it.pt/%7ebioucas/code/
[4]http://speclab.cr.usgs.gov/spectral.lib06
[5]http://speclab.cr.usgs.gov/cuprite95.tgif.2.2um_map.gif

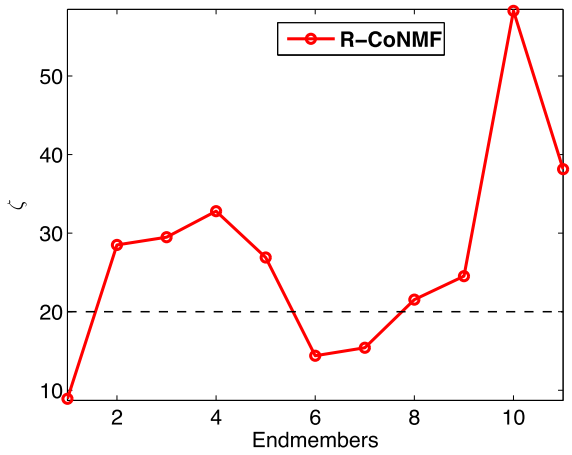



Fig. 11. Regularizer of the abundance matrix $\widehat{\mathbf{X}}$ $\zeta(i) = \|\widehat{\mathbf{x}}^i\|_2$ with $\zeta(i) = \|\widehat{\mathbf{x}}^i\|_2$ for the Cuprite data by using $q = 11$ endmembers.

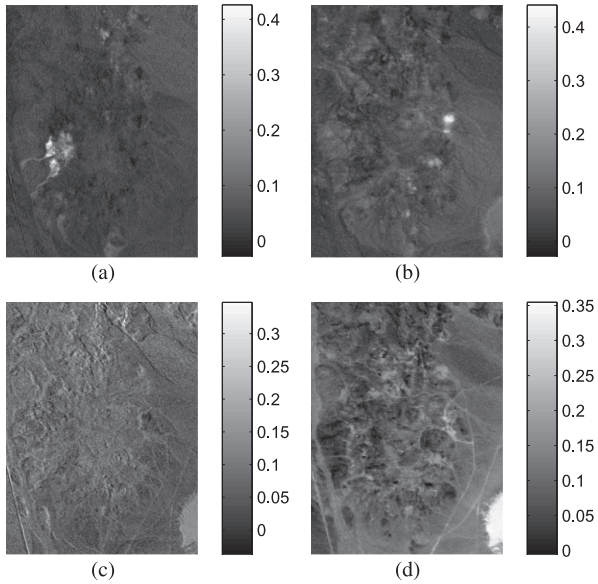



Fig. 12. Abundance maps associated with the endmembers in Fig. 11. (a) Endmember 3. (b) Endmember 5. (c) Endmember 1. (d) Endmember 8.

we can obtain an estimate of $p = 14$ endmembers. However, with $p = 14$, by looking at the results reported in the literature, such as [25], [26], and [29], it can be observed that some of the abundance maps are mostly noise. This is related to the spectral variability and the presence of mineral alterations, which results in spurious endmembers for a given mineral. It should be noted that there is no official agreement regarding the number of endmembers for this data set in the literature, as different algorithms provide different estimations. Therefore, to address the spectral variability and the presence of mineral alterations, we feel that human interpretation is needed for the estimation of the number of endmembers. We adjust the HySime algorithm by setting a threshold $\tau$ controlling the noise and signal level. After interpreting the results, we obtain $p = 10$ endmembers by setting $\tau = 1e - 4$. Notice that we use $p = 10$ as a baseline to evaluate the proposed R-CoNMF.

To evaluate the ability of R-CoNMF to estimate the number of endmembers, we use a scenario in which $q$ is slightly bigger than the $p$ estimated by HySime, i.e., $q = 11$, where Fig. 11 shows the obtained degree of sparseness of the abundances, and Fig. 12 shows the abundance maps corresponding to the three endmembers with lower $\zeta$. It can be observed that (a) corresponds to mineral Buddingtonite, (b) is Muscovite, and (c) is a noisy map. Due to the impact of spectral variability and data complexity, the degree of sparseness of endmember 1 (see Fig. 11) is not zero. However, as shown by its corresponding abundance map [see Fig. 12(c)], the map is noisy in comparison with the other abundance maps. For further analysis, Fig. 12(d) shows the abundance map of Montmorillonite. It is clear that when $q > p$, some of the endmembers are duplicates (alterations) of the real ones, i.e., Fig. 12(c) is a duplicate of Fig. 12(d). Therefore, we can infer that endmember 1 in Fig. 11 can be considered as a redundant endmember, which supports the estimation of ten endmembers provided by our proposed R-CoNMF. This is consistent with the number of endmembers estimated by HySime under the given threshold. It should be noted that, here, we resorted to human interpretation to analyze the result of R-CoNMF, which aims at better understanding the outputs of the endmembers and the abundances. This is due to the fact that the true number of endmembers for the considered data set is an open issue in the literature. Nevertheless, the index in Fig. 11 provides us with an empirical procedure to determine the number of endmembers.

### B. Unmixing

After obtaining the number of endmembers, that is, $p = 10$, we analyze the impact of the mixing matrix to estimate the spectral signatures and abundance fractions. Fig. 13 shows, for each of the endmembers extracted by the R-CoNMF algorithm, a plot of the extracted endmember and its corresponding USGS library signature (along with the SAD between each pair of signatures) and the estimated fractional abundance map. The associations between each endmember and the corresponding USGS library signature were established by using both visual interpretation of the estimated abundances (with regard to the reference map in Fig. 10) and the SAD between the obtained spectral signature and its corresponding library signature. At this point, we emphasize that the analysis scenario is a complex one, in which the same mineral has different alterations present in the library, and there is strong variability for the same kind of mineral. We also emphasize that we are comparing image-derived endmembers with USGS library signatures acquired in perfect conditions and without atmospheric interferers. With the aforementioned considerations in mind, the results reported in Fig. 13 indicate that the estimated endmembers generally provide a good match with regard to the corresponding library signatures.

## VI. Conclusion and Future Lines

In this paper, we have proposed a new algorithm for hyperspectral unmixing termed R-CoNMF. The proposed algorithm is able to perform the three main steps of the spectral

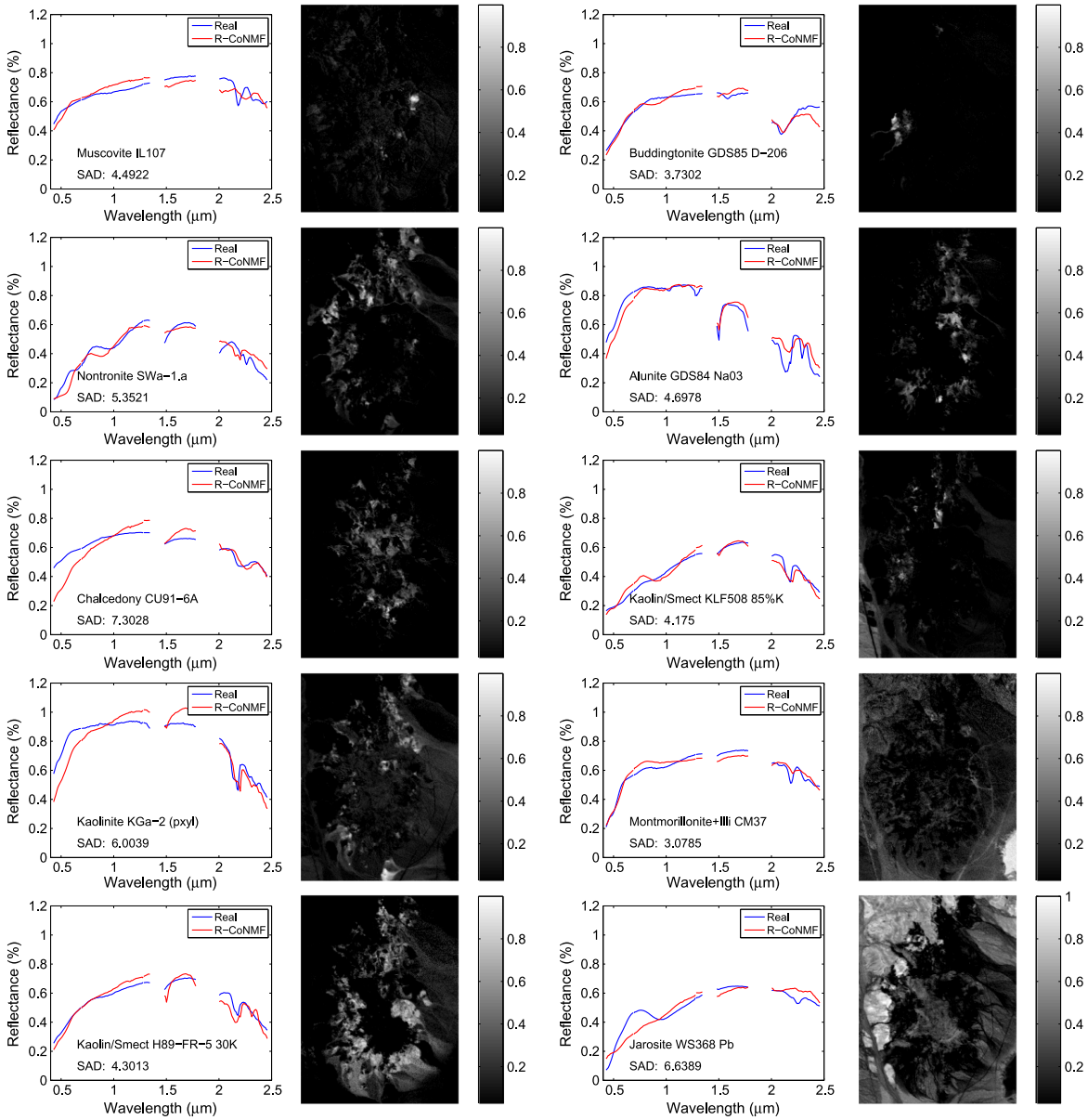



Fig. 13. Endmembers extracted by the R-CoNMF algorithm and their corresponding USGS library signature (along with the SAD between each pair of signatures) and the estimated fractional abundance map for each endmember.

unmixing chain: 1) estimation of the number of endmembers; 2) identification of the spectral signature of each endmember; and 3) estimation of the corresponding abundances in each pixel of the scene. Our experimental results, conducted using both simulated and real data sets, indicate that the algorithm is able to provide a good estimate of the number of endmembers when this information is not available *a priori* (when the information is available, it can also be effectively used by the algorithm) and generate high-quality endmember signatures and abundance estimations even in scenarios dominated by high-noise conditions. The algorithm is also tested in a popular real scenario, using the well-known AVIRIS Cuprite data set, and compared with other state-of-the-art algorithms that have been routinely used in the hyperspectral community, exhibiting similar or better performance than these algorithms.

As with any new method, there are a few issues that may present challenges over time and deserve future investigation. First and foremost, the setting of parameter $\beta$ involved in the volume regularizer in R-CoNMF (as a well-known difficulty in inverse problems) is still an open issue that deserves further investigation. Furthermore, the complexity of the algorithm can be high for scenarios with a large number of endmembers. However, this complexity is similar to those exhibited by other instances of NMF, which have been shown to be appealing for efficient parallel implementations [43], [44]. Another aspect deserving future attention is a more detailed investigation of the performance of the algorithm in highly mixed scenarios, given the current trend of Earth observation missions to provide a larger coverage of the surface of the Earth, which often leads to coarser spatial resolutions.

## ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the Anonymous Reviewers for their detailed comments and suggestions, which greatly helped improve the technical quality and presentation of this paper.

## REFERENCES

[1] A. Plaza, Q. Du, J. Bioucas-Dias, X. Jia, and F. Kruse, "Foreword to the special issue on spectral unmixing of remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4103–4110, Nov. 2011.

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

[3] J. 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.

[4] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.

[5] M. Winter, "N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE*, 1999, vol. 3753, pp. 266–270.

[6] W.-K. Ma *et al.*, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.

[7] J. M. Bioucas-Dias and J. Nascimento, "Hyperspectral unmixing based on mixtures of Dirichlet components," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, 863–878, Mar. 2011, in press.

[8] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model. Application to hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1–11, Jun. 2010.

[9] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

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

[11] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2013.

[12] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "MUSIC-CSR: Multiple signal classification collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4364–4382, Jul. 2014.

[13] M. Parente and A. Plaza, "Survey of geometric and statistical unmixing algorithms for hyperspectral images," in *Proc. IEEE 2nd WHISPERS*, Reykjavik, Iceland, Jun. 14–16, 2010, pp. 1–4.

[14] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," vol. 47, no. 1, pp. 161–173, Jan. 2009.

[15] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $\ell 1/2$ sparsity-constrained nonnegative matrix factorization," vol. 49, no. 11, pp. 4282–4297, Nov. 2011.

[16] N. Keshava, "A survey of spectral unmixing algorithms," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 55–78, 2003.

[17] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 446–450, Jul. 2007.

[18] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.

[19] J. Harsanyi, W. Farrand, and C.-I. Chang, "Determining the number and identity of spectral endmembers: An integrated approach using Neyman–Pearson eigenthresholding and iterative constrained RMS error minimization," in *Proc. Thematic Conf. Geol. Remote Sens.*, 1993, vol. 1, pp. 1–10.

[20] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[21] B. Luo, J. Chanussot, S. Doute, and L. Zhang, "Empirical automatic estimation of the number of endmembers in hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 24–28, Jan. 2013.

[22] C. Andreou and V. Karathanassi, "Estimation of the number of endmembers using robust outlier detection method," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 247–256, Jan. 2014.

[23] A. Ambikapathi, T.-H. Chan, C.-Y. Chi, and K. Keizer, "Hyperspectral data geometry-based estimation of number of endmembers using p-norm-based pure pixel identification algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2753–2769, May 2013.

[24] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.

[25] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[26] J. Li, A. Agathos, D. Zaharie, J. M. B. Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Sep. 2015.

[27] J. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. 1st IEEE WHISPERS*, Grenoble, France, Aug. 26–28, 2009, pp. 1–4.

[28] T. Chan, C. Chi, Y. Huang, and W. Ma, "Convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.

[29] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.

[30] A. Huck, M. Guillaume, and J. Blanc-Talon, "Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2590–2602, Jun. 2010.

[31] A. Ambikapathi, T.-H. Chan, W.-K. Ma, and C.-Y. Chi, "Chance-constrained robust minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4194–4209, Nov. 2011.

[32] J. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589–602, Mar. 2006.

[33] Y. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Image Process.*, vol. 56, no. 1, pp. 505–519, Jan. 2010.

[34] M. Fornasier and H. Rauhut, "Recovery algorithms for vector-valued data with joint sparsity constraints," *SIAM J. Numerical Anal.*, vol. 46, no. 2, pp. 577–613, 2008.

[35] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Lojasiewicz inequality," *Math. Oper. Res.*, vol. 35, no. 2, pp. 438–457, May 2010.

[36] L. Grippof and M. Sciandrone, "Globally convergent block-coordinate techniques for unconstrained optimization," *Optim. Methods Softw.*, vol. 10, no. 4, pp. 587–637, 1999.

[37] J. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. 2nd WHISPERS*, Jun. 2010, pp. 1–4.

[38] A. Green, M. Berman, P. Switzer, and M. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.

[39] J. Nascimento and J. Bioucas-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.

[40] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[41] X. Jia and J. A. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 538–542, Jan. 1999.

[42] G. Swayze, R. Clark, S. Sutley, and A. Gallagher, "AVIRIS mineral mapping at Cuprite, Nevada," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, vol. 1, pp. 47–49.

[43] S. Robila and L. Maciak, "Considerations on parallelizing nonnegative matrix factorization for hyperspectral data unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 1, pp. 57–61, Jan. 2009.

[44] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 6, pp. 1336–1353, Jun. 2013.

**Jun Li** (M'13) received the B.S. degree in geographic information systems from Hunan Normal University, Changsha, China, in 2004; the M.E. degree in remote sensing from Peking University, Beijing, China, in 2007; and the Ph.D. degree in electrical engineering from the Instituto de Telecomunicações, Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisbon, Portugal, in 2011.

From 2007 to 2011, she was a Marie Curie Research Fellow with the Departamento de Engenharia Electrotécnica e de Computadores and the Instituto de Telecomunicações, IST, Universidade Técnica de Lisboa, in the framework of the European Doctorate for Signal Processing (SIGNAL). She has also been actively involved in the Hyperspectral Imaging Network, which is a Marie Curie Research Training Network involving 15 partners in 12 countries and is intended to foster research, training, and cooperation on hyperspectral imaging at the European level. Since 2011, she has been a Postdoctoral Researcher with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain. She is currently a Professor with Sun Yat-sen University, Guangzhou, China. Her research interests include hyperspectral image classification and segmentation, spectral unmixing, signal processing, and remote sensing.

Dr. Li is an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. She has been a Reviewer of several journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *Pattern Recognition*, *Optical Engineering*, the *Journal of Applied Remote Sensing*, and *Inverse Problems and Imaging*.

**José M. Bioucas-Dias** (S'87–M'95) received the E.E., M.Sc., Ph.D., and "Agregado" degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisbon, Portugal, in 1985, 1991, 1995, and 2007, respectively.

Since 1995, he has been with the Department of Electrical and Computer Engineering, IST, where he was an Assistant Professor from 1995 to 2007 and has been an Associate Professor since 2007. Since 1993, he has been also a Senior Researcher with the Pattern and Image Analysis Group of the Instituto de Telecomunicações, which is a private nonprofit research institution. His research interests include inverse problems, signal and image processing, pattern recognition, optimization, and remote sensing.

Dr. Bioucas-Dias was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (1997–2000) and is an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Guest Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING for the *Special Issue on Spectral Unmixing of Remotely Sensed Data* and of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING for the *Special Issue on Hyperspectral Image and Signal Processing* and is a Guest Editor of the IEEE SIGNAL PROCESSING MAGAZINE for the *Special Issue on Signal and Image Processing in Hyperspectral Remote Sensing*. He was the General Cochair of the 3rd IEEE Geoscience and Remote Sensing Society Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2011) and has been a member of the Program/Technical Committee of several international conferences.

**Antonio Plaza** (M'05–SM'07–F'15) was born in Cáceres, Spain, in 1975. He received the computer engineer degree in 1997, the M.Sc. degree in 1999, and the Ph.D. degree in 2002, all in computer engineering, from the University of Extremadura, Cáceres.

He is currently an Associate Professor (with accreditation for Full Professor) with the Department of Technology of Computers and Communications, University of Extremadura, where he is the Head of the Hyperspectral Computing Laboratory (HyperComp), one of the most productive research groups working on remotely sensed hyperspectral data processing worldwide. He has been the advisor for 12 Ph.D. dissertations and more than 30 Ms.C. dissertations. He was the Coordinator of the Hyperspectral Imaging Network, a European project with a total funding of €2.8 million. He has authored more than 500 publications, including 152 journal papers (more than 100 in IEEE journals), 22 book chapters, and more than 240 peer-reviewed conference proceeding papers (94 in IEEE conferences). He has edited a book on *High-Performance Computing in Remote Sensing* for CRC Press/Taylor and Francis and guest edited nine special issues on hyperspectral remote sensing for different journals. His main research interests comprise hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza is a Fellow of IEEE "for contributions to hyperspectral data processing and parallel computing of Earth observation data." He served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) in 2011–2012 and is currently serving as the President of the Spanish Chapter of IEEE GRSS (since November 2012). He has served as a Proposal Evaluator for the European Commission, the National Science Foundation, the European Space Agency, the Belgium Science Policy, the Israel Science Foundation, and the Spanish Ministry of Science and Innovation. He was also a member of the Steering Committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He is currently serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is also an Associate Editor of IEEE ACCESS and was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER (2011–2012) and the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE (2013). He was a recipient of the recognition of Best Reviewers of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (in 2009) and the recognition of Best Reviewers of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (in 2010), a journal for which he served as an Associate Editor in 2007–2012. He was a recipient of the 2013 Best Paper Award of the JSTARS and the most highly cited paper (2005–2010) in the Journal of Parallel and Distributed Computing. He received best paper awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology. He is a recipient of the Best Ph.D. Dissertation Award at the University of Extremadura, a recognition also received by six of his Ph.D. students. He has reviewed more than 500 manuscripts for more than 50 different journals.

**Lin Liu** received the B.S. and Master's degree from Peking University, Beijing, China, and the Ph.D. degree from The Ohio State University, Columbus, OH, USA.

He is currently a Professor of geography and geographic information science with the University of Cincinnati, Cincinnati, OH, and Sun Yat-sen University, Guangzhou, China. He was the President of the International Association of Chinese Professionals in GIScience (CPGIS) during 1996–1997 and the Chair of the CPGIS Board of Directors during 2008–2010, a former Dean of the School of Geography and Planning with Sun Yat-sen University, and a former Head of Geography and Vice Dean for International Affairs with the University of Cincinnati. He has over 100 publications and 150 presentations. His main research areas are geographic information systems, remote sensing, and their applications, including urban activities, crime, flooding, ecosystems, etc.

Dr. Liu has served as a Conference Chair or a Program Committee Member for 15 international conferences. His research has been funded by the National Science Foundation (NSF), the U.S. Environmental Protection Agency, the National Natural Science Foundation of China (NSFC), and the Ministry of Science and Technology of China. He served on the senior panels of NSF and NSFC.