Hyperspectral Unmixing Based on Mixtures of Dirichlet Components

José M. P. Nascimento, *Member, IEEE*, and José M. Bioucas-Dias, *Member, IEEE*

Abstract—This paper introduces a new unsupervised hyperspectral unmixing method conceived to linear but highly mixed hyperspectral data sets, in which the simplex of minimum volume, usually estimated by the purely geometrically based algorithms, is far way from the true simplex associated with the endmembers. The proposed method, an extension of our previous studies, resorts to the statistical framework. The abundance fraction prior is a mixture of Dirichlet densities, thus automatically enforcing the constraints on the abundance fractions imposed by the acquisition process, namely, nonnegativity and sum-to-one. A cyclic minimization algorithm is developed where the following are observed: 1) The number of Dirichlet modes is inferred based on the minimum description length principle; 2) a generalized expectation maximization algorithm is derived to infer the model parameters; and 3) a sequence of augmented Lagrangian-based optimizations is used to compute the signatures of the endmembers. Experiments on simulated and real data are presented to show the effectiveness of the proposed algorithm in unmixing problems beyond the reach of the geometrically based state-of-the-art competitors.

Index Terms—Augmented Lagrangian method of multipliers, blind hyperspectral unmixing, dependent components, generalized expectation maximization (GEM), minimum description length (MDL), mixtures of Dirichlet densities.

I. INTRODUCTION

HYPERSPECTRAL imaging is a remote sensing technology that collects 2-D spatial images from the Earth's surface in hundreds of narrow and contiguous bands of high spectral resolution covering the visible, near-infrared, and shortwave infrared bands. This technique has been used for planetary exploitation [3], [4] and also for a wide range of applications in the fields of environmental monitoring, agriculture, forestry, geology [5], [6], food safety [7], [8], counterfeit drug detection [9], urban geography [10], detection, classification, and surveillance of military target activities [11], [12].

The spatial resolution corresponding to a single pixel of a hyperspectral image depends on the flying height of the aircraft and on the instantaneous field of view of the sensor. Very often, the resolution cells, in an image, contain several substances. Thus, the radiances collected in spectral vectors are mixtures of spectra from the constituent substances (also called endmembers) present in the respective resolution cells [5].

The linear mixing assumption has been widely used to describe the observed hyperspectral vectors. According to this assumption, a mixed pixel is a linear combination of endmember signatures weighted by the corresponding abundance fractions [13]. Due to physical considerations, the abundance fractions are subject to the so-called nonnegativity and a full-additivity (sum-to-one) constraints [14]. Thus, the observed spectral vectors in a given scene are in a simplex whose vertices correspond to the endmembers.

Hyperspectral unmixing is a challenging task, underlying many hyperspectral imagery applications, that decomposes a mixed pixel into a collection of reflectance spectra, called endmember signatures, and the corresponding abundance fractions. This can be handled as a two-procedure task: the endmember extraction, which identifies endmembers present in the scene, and the inversion step, which estimates the proportion of each endmember for each pixel of the image [15]–[17].

Most popular endmember extraction algorithms, such as the *vertex component analysis* (VCA) [18], the *automated morphological endmember extraction* [19], the *pixel purity index* [20], the N-FINDR [21] (see [22] for recently introduced reinterpretations and improvements of N-FINDR), and the *iterative error analysis* [23], exploit the geometric perspective referred to above and assume that the data set contains at least one pure pixel of each endmember, e.g., a pixel containing just a single endmember. These methods are followed by a fully constrained least square estimation [24] or by a maximum likelihood estimation [25] of the abundance fractions to complete the unmixing procedure.

When the data set does not contain pure pixels for some (or for none) endmembers, a popular line of attack introduced by Craig and named *minimum-volume transform* (MVT) [26] is widely applied. MVT finds the smallest simplex that contains all observed spectral vectors. Other methods that follow the same strategy and do the unmixing procedure in one step are the *iterated constrained endmembers* (ICEs) [27], the *sparsity promoting ICE* (SPICE) [28], the *minimum-volume enclosing simplex* (MVES) algorithm [29], the *robust MVES* [30], the *minimum-volume simplex* (MVSA) [31], the *simplex identification by variable splitting and augmented Lagrangian tools* (SISAL) [32], and the *alternating projected subgradients* [33].

A necessary condition for the correct identification of the mixing matrix is the existence of at least $p - 1$ (p stands for the number of endmembers) spectral vectors on each facet of the data simplex. However, in highly mixed data sets, this assumption is not realistic, which has fostered research in the statistical front [1], [34]–[37].

Hyperspectral unmixing is a blind source separation problem, where abundance fractions can be interpreted as sources. Independent component analysis (ICA) [38] comes naturally to mind and have, in fact, been used in hyperspectral applications (see, e.g., references in [39]). However, as shown in [39], the applicability of canonical ICA to hyperspectral data is compromised by the statistical dependence existing among abundances. In addition to this limitation, ICA does not ensure the nonnegativity constraint.

The *nonnegative matrix factorization* (NMF) [40]–[42] was designed to unmix nonnegative sources. Elaborations of these ideas have been applied to hyperspectral unmixing. For example, the works [43]–[46] use regularization terms to unmix hyperspectral data with additivity constraint. However, NMF minimizes a nonconvex function with respect to factor matrices leading to local minima solutions. The Bayesian framework, where constraints can be incorporated directly in the problem formulation and any parameter involved modeled as a random variable, opens the door to highly flexible approaches to unmixing [35], [36], [47], [48].

Dirichlet processes have been used in the *piecewise convex endmember detection* (PCE) [48] algorithm to determine the number of convex regions needed to describe an input hyperspectral image. PCE estimates a set of endmember distributions for each context, rather than a single spectrum. The work [49], closely related with PCE, uses a Bayesian framework to unmix hyperspectral images taking into account possible spatial correlations between pixels. This work assumes that the image can be partitioned into regions to which hidden variables are associated. In these regions, the statistical properties of the abundance fractions are homogeneous, which leads, under given conditions, to a mixture of densities similar to the ones herein considered.

A. Contribution

This paper elaborates on the *dependent component analysis* (DECA) methodology introduced in [1] and [2] to unmix highly mixed hyperspectral data sets, which are beyond the reach of the purely geometrically based methods. DECA uses a mixture of Dirichlet densities as prior for the abundance fractions. Mixtures of densities allow one to model complex distributions in which the mass probability is scattered by a number of clusters inside the simplex. Furthermore, the Dirichlet density automatically enforces the nonnegativity and sum-to-one constraints on the abundance fractions.

The two main improvements to DECA herein introduced are the following: 1) The number of Dirichlet modes is adaptively inferred based on the minimum description length (MDL) principle as proposed in [50]; 2) the generalized expectation maximization (GEM) algorithm we adopt to infer the model parameters is improved by using alternating minimization and

augmented Lagrangian methods to compute the mixing matrix, similarly to those used in [32].

This paper is organized as follows. Section II formulates the unmixing problem. Section III describes the fundamentals of the proposed method. Sections IV and V illustrate aspects of the performance of the proposed approach with experimental data based on USGS laboratory spectra and with real hyperspectral data collected by the AVIRIS sensor, respectively. Section VI concludes with some remarks.

II. PROBLEM STATEMENT

Assuming the linear observation model, each pixel $\mathbf{y} \in \mathbb{R}^L$ of a hyperspectral image (L is the number of bands) is given by

$$\mathbf{y} = \mathbf{M}\mathbf{s} + \mathbf{n} \qquad (1)$$

where $\mathbf{M} \equiv [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p]$ is an $L \times p$ mixing matrix ($\mathbf{m}_j$ denotes the jth endmember signature), p is the number of endmembers present in the covered area, $\mathbf{s} = [s_1, s_2, \ldots, s_p]^{\mathrm{T}}$ is the abundance vector containing the fractions of each endmember, and $\mathbf{n}$ is the additive noise vector (notation $(\cdot)^{\mathrm{T}}$ stands for vector transposed).

To be physically meaningful [14], abundance fractions are subject to nonnegativity and sum-to-one constraints, i.e., abundance fractions are in the $p - 1$ probability simplex $\{\mathbf{s} \in \mathbb{R}^p : s_j \geq 0, \sum_{j=1}^{p} s_j = 1\}$. Note that only $p - 1$ components of $\mathbf{s}$ are free, i.e., $s_p = 1 - \sum_{j=1}^{p-1} s_j$. Therefore, the spectral vectors are in a $(p - 1)$-dimensional simplex in $\mathbb{R}^L$.

Usually, the number of endmembers is much lower than the number of bands ($p \ll L$), and thus, it is advantageous, in terms of signal-to-noise ratio (SNR), memory usage, and computational complexity, to represent the spectral vectors in a signal subspace basis [51]. In most hyperspectral applications, the SNR is sufficiently large, meaning that noise can be safely neglected after the observed data projection onto the signal subspace. Nevertheless, we evaluate the robustness of the proposed method to noise in Section IV.

Let $\mathbf{E}_p \equiv [\mathbf{e}_1, \ldots, \mathbf{e}_p]$ be a matrix $L \times p$, with p orthonormal directions spanning the signal subspace. This matrix is determined by a recent method, termed as *hyperspectral signal identification by minimum error* (HySime) [51], a fully automatic and unsupervised algorithm. HySime starts by estimating the noise covariance matrix, and then, it infers the number of endmembers and selects the subset of eigenvectors that best represent the signal subspace in the least squared error sense.

Let $(\mathbf{E}_p \mathbf{E}_p^{\mathrm{T}})\mathbf{y}$ denote the orthogonal projection of $\mathbf{y} \in \mathbb{R}^L$ onto the subspace spanned by the columns of $\mathbf{E}_p$, and let $\mathbf{x} \equiv \mathbf{E}_p^{\mathrm{T}}\mathbf{y} \in \mathbb{R}^p$ denote the respective coordinates with respect to the columns of $\mathbf{E}_p$, i.e., with respect to the orthonormal basis defined by the columns of $\mathbf{E}_p$. Owing to spectral variability from pixel to pixel, topographic modulation, and modeling errors, and even assuming negligible noise, the observed vectors $\mathbf{y}$, and therefore $\mathbf{x}$, are not in a simplex set [39]. Since our approach, as many other, assumes that the observed spectral vectors belong to a simplex set, we identify the affine set containing this simplex set and project vectors $\mathbf{x} \in \mathbb{R}^p$ onto it. Fig. 1 shows this
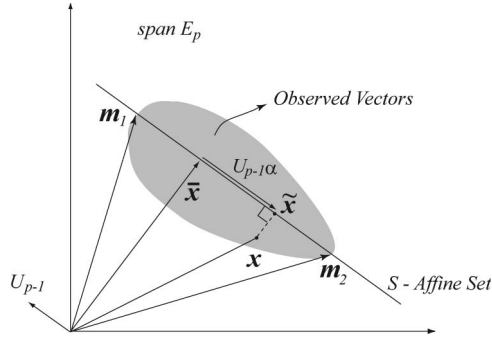
Fig. 1. Illustration of the affine projection. S represents the affine set. Gray cloud represents the set $\{\mathbf{x}^{(i)}, i = 1, \ldots, N\}$. $\tilde{\mathbf{x}}^{(i)}$ is the affine projection of $\mathbf{x}^{(i)}$.

procedure. The vector $\mathbf{x} \in \mathbb{R}^p$ is projected onto the identified simplex set S, yielding the projected vector $\tilde{\mathbf{x}} \in \mathbb{R}^p$. Given a point on S, for example, $\overline{\mathbf{x}}$, we may write $\tilde{\mathbf{x}} = \overline{\mathbf{x}} + \boldsymbol{\Delta}\mathbf{x}$. Since $\boldsymbol{\Delta}\mathbf{x} \in \mathbb{R}^{p-1}$, it can always be written as $\mathbf{U}_{p-1}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \in \mathbb{R}^{p-1}$ and $\mathbf{U}_{p-1}$ is a $p \times (p-1)$ orthonormal matrix spanning the subspace $\{\mathbf{z} - \overline{\mathbf{x}} : \mathbf{z} \in S\}$.

With this scheme, we have two projections: The first projects the observed spectral vectors onto the identified subspace signal, and the second projects the already projected vectors onto the identified affine set. We stress that the effect of the two projections could not be obtained just with the second one, at least in the case of nonwhite noise. This is so because the optimal signal subspace estimate depends on the noise correlation matrix, as shown in [51].

We now focus on the identification of the affine set S and on the projection of $\mathbf{x}$ onto it. Let $\{\mathbf{y}^{(i)}, i = 1, \ldots, N\}$ and $\{\mathbf{x}^{(i)} = \mathbf{E}_p^T\mathbf{y}^{(i)}, i = 1, \ldots, N\}$ be the observed and projected (on the signal subspace) data sets, respectively. A solution for the projection of vectors $\{\mathbf{x}^{(i)}, i = 1, \ldots, N\}$ on the hyperplane (an affine set) that best represents this set in the least squares sense is given by [29]

$$\tilde{\mathbf{x}}^{(i)} = \overline{\mathbf{x}} + \mathbf{U}_{p-1}\left(\mathbf{x}^{(i)} - \overline{\mathbf{x}}\right) \in \mathbb{R}^p, \qquad i = 1, \ldots, N \quad (2)$$

where $\overline{\mathbf{x}}$ and $\mathbf{U}_{p-1}$ denote, respectively, the sample average and the set of $p - 1$ eigenvectors of the sample covariance of $\{\mathbf{x}^{(i)}, i = 1, \ldots, N\}$. Fig. 1 shows the affine projection introduced earlier. The affine set is represented by the hyperplane S, the set $\{\mathbf{x}^{(i)}, i = 1, \ldots, N\}$ by the gray cloud, and the affine projection of $\mathbf{x}^{(i)}$ by $\tilde{\mathbf{x}}^{(i)}$. From now on, and to lighten the notation, we drop the tilde symbol on the top of $\mathbf{x}$ to denote the projection on the affine set. We have then

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (3)$$

where $\mathbf{s}$ belongs to the $p - 1$ probability simplex and $\mathbf{A} \in \mathbb{R}^{p \times p}$ is the mixing matrix. The representation of the endmembers in the original space $\mathbb{R}^L$ is given by $\mathbf{M} = \mathbf{E}_p\mathbf{A}$.

In the next section, we address the blind estimation of the mixing matrix $\mathbf{A}$ by modeling the fractional abundances $\mathbf{s}$ as a mixture of Dirichlet densities.

III. PROPOSED APPROACH

Let us assume that $\mathbf{W} \equiv \mathbf{A}^{-1}$ exists. Then, we have $\mathbf{s} = \mathbf{W}\mathbf{x}$. As already referred to, we adopt a k-component Dirichlet finite mixture as a prior for the abundance fractions. The rationale underlying this choice is the following.

1) The Dirichlet density automatically enforces the nonnegativity and sum-to-one constraints, and thus, it accounts for statistical dependence usually found in hyperspectral data.
2) As noted in [52], the Dirichlet density is suited to model fractions.
3) Mixtures of densities allow one to model complex distributions in which the mass probability is scattered over a number of clusters inside the simplex.

The abundance fraction density is then written as

$$p_S(\mathbf{s}|\boldsymbol{\theta}) \equiv \sum_{q=1}^{k} \epsilon_q D(\mathbf{s}|\boldsymbol{\theta}_q),$$

$$= \sum_{q=1}^{k} \epsilon_q \frac{\Gamma\left(\sum_{j=1}^{p}\theta_{qj}\right)}{\prod_{j=1}^{p}\Gamma(\theta_{qj})} \prod_{j=1}^{p} s_j^{\theta_{qj}-1} \quad (4)$$

where, for $q = 1, \ldots, k$, ϵ_q and $D(\mathbf{s}|\boldsymbol{\theta}_q)$ denote, respectively, the probability of mode q and its Dirichlet density with parameter $\boldsymbol{\theta}_q \equiv \{\theta_{q1}, \ldots, \theta_{qp}\}$. We denote the complete set of parameters as $\boldsymbol{\theta} \equiv \{\epsilon_1, \ldots, \epsilon_k, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k\}$. Since the spectral vectors $\mathbf{x}$ play now the role of observed data, we must base our inferences on the their density, denoted by p_X, which, in terms of p_S and given the linear and invertible relation $\mathbf{s} = \mathbf{W}\mathbf{x}$ between $\mathbf{s}$ and $\mathbf{x}$, is given by

$$p_X(\mathbf{x}|\mathbf{W}, \boldsymbol{\theta}) = p_S(\mathbf{s} = \mathbf{W}\mathbf{x}|\boldsymbol{\theta})\left|\det(\mathbf{W})\right|. \quad (5)$$

Consider that each vector $\mathbf{x}$ represents one particular outcome of a p-dimensional random variable $\mathbf{X} = [X_1, \ldots, X_p]^T$. Given a set of N independent and identically distributed (i.i.d.) samples $\mathcal{X} \equiv [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}]$, then the log-likelihood of the set of parameters $\boldsymbol{\theta}$ and unmixing matrix $\mathbf{W}$ is

$$\mathcal{L}(\mathbf{W}, \boldsymbol{\theta}) \equiv \log p_X(\mathcal{X}|\mathbf{W}, \boldsymbol{\theta})$$

$$= \sum_{i=1}^{N} \left[\log p_X\left(\mathbf{x}^{(i)}|\mathbf{W}, \boldsymbol{\theta}\right)\right]$$

$$= \sum_{i=1}^{N} \left[\log p_S\left(\mathbf{s}^{(i)}|\boldsymbol{\theta}\right)\right] + N\log|\det\mathbf{W}|$$

$$= \sum_{i=1}^{N} \left[\log \sum_{q=1}^{k} \epsilon_q D\left(\mathbf{s}^{(i)}|\boldsymbol{\theta}_q\right)\right] + N\log|\det\mathbf{W}| \quad (6)$$

where $\mathbf{s}^{(i)} \equiv \mathbf{W}\mathbf{x}^{(i)}$.

The maximum likelihood estimate

$$(\widehat{\mathbf{W}}, \widehat{\boldsymbol{\theta}})_{\text{ML}} \equiv \arg\max_{\mathbf{W}, \boldsymbol{\theta}} \mathcal{L}(\mathbf{W}, \boldsymbol{\theta}) \quad (7)$$

cannot be found analytically [50], [53]. The usual choice for obtaining the ML estimates of the parameters is the *expectation*

maximization (EM) framework [54], which relies on the so-called incomplete data and missing data. In our setup, $\mathcal{X}$ is the incomplete data. The missing data, denoted by $\mathcal{Z} \equiv \{\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(N)}\}$, are a set of N k-dimensional vectors representing which component has produced each sample. Each vector $\mathbf{z}^{(i)} = [z_1^{(i)}, \ldots, z_k^{(i)}]$ is a binary k-vector, where only one component $z_q^{(i)}$ is set to one indicating which mode produced the i-sample. The complete log-likelihood is then

$$\mathcal{L}_C(\boldsymbol{\theta}, \mathbf{W}) = \log\left[p_{X,Z}(\mathcal{X}, \mathcal{Z}|\boldsymbol{\theta})\right] + N \log|\det(\mathbf{W})|$$

$$= \sum_{i=1}^{N}\left[\sum_{q=1}^{k} z_q^{(i)} \log\left(\epsilon_q D\left(\mathbf{s}^{(i)}|\boldsymbol{\theta}_q\right)\right)\right]$$

$$+ N \log|\det(\mathbf{W})|. \tag{8}$$

The EM algorithm iterates between the E-step and the M-step [54], [55].

1) **E-step**: It computes the conditional expectation of the complete log-likelihood, given the samples and the current estimate $\widehat{\boldsymbol{\theta}}^{(t)}$. The result is the so-called Q-function

$$Q\left(\boldsymbol{\theta}, \mathbf{W}; \widehat{\boldsymbol{\theta}}^{(t)}, \widehat{\mathbf{W}}^{(t)}\right) = N \log|\det(\mathbf{W})|$$

$$+ \sum_{i=1}^{N}\left[\sum_{q=1}^{k} \beta_q^{(i,t)} \log\left[\epsilon_q D\left(\mathbf{s}^{(i,t)}|\boldsymbol{\theta}_q\right)\right]\right] \tag{9}$$

where

$$\beta_q^{(i,t)} \equiv \mathbb{E}\left[z_q^{(i)}|\widehat{\boldsymbol{\theta}}_q^{(t)}\right]$$

$$= \frac{\widehat{\epsilon}_q^{(t)} D\left(\mathbf{s}^{(i,t)}|\widehat{\boldsymbol{\theta}}_q^{(t)}\right)}{\sum_{l=1}^{k} \widehat{\epsilon}_l^{(t)} D\left(\mathbf{s}^{(i,t)}|\widehat{\boldsymbol{\theta}}_l^{(t)}\right)} \tag{10}$$

with $\mathbf{s}^{(i,t)} \equiv \widehat{\mathbf{W}}^{(t)} \mathbf{x}^{(i)}$ and $\mathbb{E}(\cdot)$ denoting the mean value operator.

2) **M-step**: It updates the parameter estimates according to

$$\left(\widehat{\boldsymbol{\theta}}^{(t+1)}, \widehat{\mathbf{W}}^{(t+1)}\right) = \arg\max_{\boldsymbol{\theta}, \mathbf{W}} \left\{Q\left(\boldsymbol{\theta}, \mathbf{W}; \widehat{\boldsymbol{\theta}}^{(t)}, \widehat{\mathbf{W}}^{(t)}\right)\right\}. \tag{11}$$

Optimization (11) is still a hard problem. Instead of solving it exactly, we implement alternate minimization with respect to $\boldsymbol{\theta}$ and to $\mathbf{W}$. In the same vein, instead of computing $\boldsymbol{\theta}_q^{(t+1)}$, we maximize the Q-function with respect to θ_{qj}, for $j = 1, \ldots, p$, resulting in the following learning rules for the mixing probabilities and for the mixture of Dirichlet source parameters [56]

$$\epsilon_q^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} \beta_q^{(i,t)}, \tag{12}$$

$$\widehat{\theta}_{qj}^{(t+1)} = \Psi^{-1}\left(\Psi\left(\sum_{l=1}^{p} \widehat{\theta}_{ql}^{(t)}\right) + \frac{\sum_{i=1}^{N}\left[\beta_q^{(i,t)} \log \widehat{s}_j^{(i,t)}\right]}{\sum_{i=1}^{N}\left(\beta_q^{(i,t)}\right)}\right) \tag{13}$$

for $q = 1, \ldots, k$ and $j = 1, \ldots, p$, where $\Psi(x) \equiv d(\log \Gamma(x))/dx$ is the psi$(\cdot)$ function and $\Psi^{-1}(\cdot)$ denotes its inverse.

The resulting algorithm is of the GEM class [54]: The learning rule (12) maximizes Q-function with respect to $\epsilon_q^{(t)}$, whereas (13) assures that the Q-function does not decrease. We note that steps (12) and (13) define a new GEM algorithm to learn the mixture of Dirichlet densities.[1]

Denoting $\mathcal{S} \equiv \mathbf{W}\mathcal{X} = [\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(N)}]$, the optimization with respect to $\mathbf{W}$ amounts to compute

$$\widehat{\mathbf{W}}^{(t+1)} = \arg\max_{\mathbf{W}} \phi(\mathbf{W}\mathcal{X}) + \log|\det(\mathbf{W})|$$

$$\text{s.t.}: \quad \mathbf{W}\mathcal{X} \succeq 0 \quad \mathbf{1}_p^{\mathrm{T}} \mathbf{W}\mathcal{X} = \mathbf{1}_N^{\mathrm{T}} \tag{14}$$

where the constraints $\mathbf{W}\mathcal{X} \succeq 0$ and $\mathbf{1}_p^{\mathrm{T}} \mathbf{W}\mathcal{X} = \mathbf{1}_N^{\mathrm{T}}$ enforce, respectively, nonnegativity and sum-to-one on each abundance fraction vector ($\mathbf{1}_p$ stands for p-dimensional column vector of ones) and

$$\phi(\mathcal{S}) \equiv \sum_{i=1}^{N}\sum_{l=1}^{p} \gamma_l^{(i)} \log\left(s_l^{(i)}\right)$$

with

$$\gamma_l^{(i)} \equiv \frac{1}{N} \sum_{q=1}^{k} \beta_q^{(i,t)}(\widehat{\theta}_{ql} - 1). \tag{15}$$

The optimization problem (14) is very hard to solve owing to the nonconvex term $\log|\det(\mathbf{W})|$. Herein, we use a modified version, the SISAL algorithm introduced in [32], termed as *Dirichlet mixture unmixing via split augmented Lagrangian* (DUSAL), to solve a problem similar to (14). As in SISAL, we exploit variable splitting and augmented Lagrangian methods. The resulting algorithm is presented in Appendix A. The main difference with respect to that in [32, Algorithm 3] is the replacement of function $\|\mathbf{z}\|_h$ by $-\phi(\mathbf{z})$ and of the respective decoupled optimization, which yields now, for each component of $\mathbf{z}$, the larger root of a second-order polynomial. See line 8 of Algorithm 2 in Appendix A.

Having computed $\widehat{\mathbf{W}}$, the signatures of the endmembers are given by

$$\widehat{\mathbf{M}} = \mathbf{E}_p \widehat{\mathbf{W}}^{-1}. \tag{16}$$

A. Number of Dirichlet Modes

The estimation of the number of modes k is a model order selection problem. It is well known that the ML criterion alone cannot be used to do such an inference because of the nested nature of model parameters for increasing model orders.

---

[1] As in the EM case, a GEM algorithm generates a sequence of increasing log-likelihood $\mathcal{L}(\widehat{\boldsymbol{\theta}}^{(t)}, \widehat{\mathbf{W}}^{(t)})$, for $t = 1, 2, \ldots$. However, in the GEM case, instead of maximizing $Q(\boldsymbol{\theta}, \mathbf{W}; \widehat{\boldsymbol{\theta}}^{(t)}, \widehat{\mathbf{W}}^{(t)})$ with respect to $(\boldsymbol{\theta}, \mathbf{W})$, only a log-likelihood increase is ensured, which is usually simpler to compute [54].

To estimate the number of mixture components, we adopt the model selection criterion

$$\widehat{k} = \arg\min_{k} \{-\mathcal{L}(\boldsymbol{\theta}(k)) + \mathcal{P}(k)\} \qquad (17)$$

where $\mathcal{L}(\boldsymbol{\theta}(k))$ is given by (6), for k modes, and $\mathcal{P}(k)$ is a penalizing term that increases with k [50]. Herein, we adopt the *MDL* [57], [58] which amounts to compute

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ -\log p_S(\mathcal{S}|\boldsymbol{\theta}) - \log p(\boldsymbol{\theta}) \right.$$
$$\left. + \frac{1}{2}\log(|\mathcal{I}(\boldsymbol{\theta})|) + \frac{c}{2}(1-\log(12)) \right\} \qquad (18)$$

where $\mathcal{I}(\boldsymbol{\theta}) \equiv \mathbb{E}[\partial^2(\log p(\mathcal{S}|\boldsymbol{\theta}))/\partial\boldsymbol{\theta}^2]$ is the Fisher information matrix and $c \equiv k(p+1)$ is the dimension of $\boldsymbol{\theta}$.

For mixtures, the Fisher information matrix cannot be obtained analytically. Herein, we follow the study in [50], where $\mathcal{I}(\boldsymbol{\theta})$ is replaced by the complete-data information matrix $\mathcal{I}_C(\boldsymbol{\theta})$. Assuming *a priori* independence of the parameters and adopting the noninformative Jeffreys' prior $p(\boldsymbol{\theta}_j) \propto \sqrt{|\mathcal{I}(\boldsymbol{\theta}_j)|}$ for each $\boldsymbol{\theta}_j$, the negative log-likelihood in (18) is given by (see [50] and [59] for details)

$$\mathcal{L}(\boldsymbol{\theta}(k), k) = -\log p_S(\mathcal{S}|\boldsymbol{\theta}) + \frac{k(p+1)}{2} + \frac{k}{2}\log\left(\frac{N}{12}\right)$$
$$+ \frac{p}{2}\sum_{q=1}^{k}\log\frac{N\epsilon_q}{12}. \qquad (19)$$

The model order and the respective parameters are given by

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \{\mathcal{L}(\boldsymbol{\theta}(k), k), k = k_{\min}, \dots, k_{\max}\}. \qquad (20)$$

Since we do not know the fractions $\mathcal{S}$, we implement the GEM algorithm described in the previous section where the negative log-likelihood is as in (6) plus the MDL terms, and the optimization with respect to $\boldsymbol{\theta}$ is replaced with the optimization with respect to $\boldsymbol{\theta}(k)$ and k.

B. Complete Algorithm

The pseudocode for the DECA algorithm is presented in Algorithmic 1. DECA implements two nested loops: The outer loop, between lines 11 and 34, iterates over k, the number of Dirichlet modes, from $k_{\min}$ to $k_{\max}$. The inner loop, between lines 12 and 24, implements the GEM algorithm (9)–(13) to compute the ML estimate of $\boldsymbol{\theta}(k)$ and of the mixing matrix $\mathbf{W}$ (line 22), for each value of k. The relative variation of the negative log-likelihood $\mathcal{L}((\boldsymbol{\theta}(k), k)$ given by (19) is used as stopping criterion for the inner loop.

To decrease the number of modes, a component annihilation is performed based on the smallest mixing probability ϵ_i (see lines 32 and 33). Finally, for each k, the parameter values are stored if the negative log-likelihood values are smaller than the minimum found so far (lines 25–31).

Algorithm 1: DECA

1: INPUT: $\mathbf{Y} := [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]$, $k_{\max}$, $k_{\min}$
2: OUTPUT: $\widehat{\mathbf{M}}_{\text{best}}, \widehat{\mathbf{s}}_{\text{best}}, \widehat{k}_{\text{best}}, \widehat{\boldsymbol{\theta}}_{\text{best}}$
3: $\mathbf{E}_p := \text{HySime}(\mathbf{Y})$ {HySime determines the number of endmembers p and estimates the signal subspace}
4: $\mathbf{X} := \mathbf{E}_p^{\text{T}}\mathbf{Y}$ {Observed samples projected onto signal subspace: $\mathbf{x} \in \mathbb{R}^p$}
5: $\mathbf{X} := \text{Affine}(\mathbf{X})$ {Identify the affine set that best fits $\mathbf{X}$ and projects $\mathbf{X}$ on it: $\mathbf{x} \in \mathbb{R}^p$}
6: $\widehat{\mathbf{A}} := \text{DUSAL}(\mathbf{X}, p)$; $\widehat{\mathbf{W}} = \widehat{\mathbf{A}}^{-1}$ {Initial estimates: $\widehat{\mathbf{A}} \in \mathbb{R}^{p \times p}$}
7: $\widehat{\boldsymbol{\theta}} :=$ {Initial estimate using criterion defined in Section IV-A}
8: $\mathcal{L}_{\min} := +\infty$
9: $t := 0$
10: $k := k_{\max}$
11: **while** $k \geq k_{\min}$ **do**
12:   **repeat**
13:     $t := t + 1$
14:     $\widehat{\mathbf{s}} := \widehat{\mathbf{W}}\mathbf{X}$
15:     $\beta_q^{(i)} := \widehat{\epsilon}_q D(\widehat{\mathbf{s}}^{(i)}|\widehat{\boldsymbol{\theta}}_q) / \sum_{l=1}^{k} \widehat{\epsilon}_l D(\widehat{\mathbf{s}}^{(i)}|\widehat{\boldsymbol{\theta}}_l)$.
         {for $i = 1, \dots, N$}
16:     $\widehat{\epsilon}_q := (1/N)\sum_{i=1}^{N} \beta_q^{(i)}$
17:     $\widehat{\theta}_{qj} := \Psi^{-1}(\Psi(\sum_{l=1}^{p} \widehat{\theta}_{ql})$
         $+ (\sum_{i=1}^{N}[\beta_q^{(i)} \log s_j^{(i)}] / \sum_{i=1}^{N}(\beta_q^{(i)})))$,
18:     **if** ANY $\widehat{\epsilon}_q = 0$ **then**
19:       Annihilate mode
20:       $k := k - 1$
21:     **end if**
22:     $\widehat{\mathbf{A}} := \text{DUSAL}(\mathbf{X}, p, \widehat{\mathbf{W}})$; $\widehat{\mathbf{W}} = \widehat{\mathbf{A}}^{-1}$
23:     $\mathcal{L}^{(t)} := -\sum_{i=1}^{N}[\log \sum_{q=1}^{k} \widehat{\epsilon}_q D(\widehat{\mathbf{s}}^{(i)}|\widehat{\boldsymbol{\theta}}_q)] -$
         $N \log |\det \widehat{\mathbf{W}}| + (k(p+1)/2) + (k/2)\log(N/12)$
         $+ (p/2)\sum_{q=1}^{k} \log(N\widehat{\epsilon}_q/12)$
24:   **until** $\mathcal{L}^{(t-1)} - \mathcal{L}^{(t)} < 10^{-5}|\mathcal{L}^{(t-1)}|$
25:   **if** $\mathcal{L}^{(t)} < \mathcal{L}_{\min}$ **then**
26:     $\mathcal{L}_{\min} := \mathcal{L}^{(t)}$
27:     $\widehat{k}_{\text{best}} := k$
28:     $\widehat{\boldsymbol{\theta}}_{\text{best}} := \widehat{\boldsymbol{\theta}}$
29:     $\widehat{\mathbf{M}}_{\text{best}} := \mathbf{E}_p\widehat{\mathbf{A}}$
30:     $\widehat{\mathbf{s}}_{\text{best}} := \widehat{\mathbf{s}}$
31:   **end if**
32:   Annihilate mode with smallest $\widehat{\epsilon}$
33:   $k := k - 1$
34: **end while**

C. Identifiability of the Endmembers

The identifiability of the mixing matrix is a pertinent question: Is it possible that different sets of endmembers explain the same observed data? If no further constraint is enforced, in addition to the positivity and sum-to-one, the answer is yes. The proof is very simple. Suppose that we are given a set of linearly independent vectors, for example, $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_p]$, containing in its convex hull a set of endmembers, for example,

$\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_p]$. Therefore, if the data set belongs to the convex hull of $\mathbf{A}$, it also belongs to the convex hull of $\mathbf{B}$. As a consequence, given $\mathbf{x} \in \mathbb{R}^p$, a spectral vector in the data set, we have $\mathbf{x} = \mathbf{A}\mathbf{s}_a$ and $\mathbf{x} = \mathbf{B}\mathbf{s}_b$, where both $\mathbf{s}_a$ and $\mathbf{s}_b$ satisfy the positivity and the sum-to-one constraints.

The minimum-volume constraint makes the spectral unmixing problem identifiable by choosing the minimum-volume mixing matrix. However, as already explained, this constraint leads to incorrect estimates in highly mixed data sets. The method we are introducing was conceived, precisely to overcome that limitation of the minimum-volume-based approaches. Of course, to be effective, the statistical model underlying DECA must be identifiable. We now give formal and experimental evidence of this identifiability.

Let $\mathbf{A}_0$, $\mathbf{W}_0 \equiv \mathbf{A}_0^{-1}$, and $\boldsymbol{\theta}_0$ denote, respectively, the true mixing matrix, the true unmixing matrix, and the true parameters of the Dirichlet mixture. From the definition of the log-likelihood function $\mathcal{L}(\mathbf{W}, \boldsymbol{\theta})$ introduced in (7) and noting that $\mathbf{x}^{(i)} = \mathbf{W}\mathbf{A}_0\mathbf{s}_0^{(i)}$ (the subscript 0 in $\mathbf{s}_0^{(i)}$ refers to a sample from $p_S(\cdot|\boldsymbol{\theta}_0)$), we have

$$\frac{1}{N}\mathcal{L}_N(\mathbf{W}, \boldsymbol{\theta})$$

$$= \frac{1}{N}\sum_{i=1}^{N} \log\left\{p_S\left(\mathbf{W}\mathbf{x}^{(i)}|\boldsymbol{\theta}\right)|\det \mathbf{W}|\right\} \quad (21)$$

$$= \frac{1}{N}\sum_{i=1}^{N} \log\left\{p_S\left(\mathbf{W}\mathbf{A}_0\mathbf{s}_0^{(i)}|\boldsymbol{\theta}\right)|\det(\mathbf{W}\mathbf{A}_0)|\right\} \quad (22)$$

$$\simeq \int_{S_{p-1}} p_S(\mathbf{s}|\boldsymbol{\theta}_0) \log\left\{p_S(\mathbf{W}\mathbf{A}_0\mathbf{s}|\boldsymbol{\theta})|\det(\mathbf{W}\mathbf{A}_0)|\right\} d\mathbf{s} \quad (23)$$

$$= -\int_{S_{p-1}} p_S(\mathbf{s}|\boldsymbol{\theta}_0) \log \frac{p_S(\mathbf{s}|\boldsymbol{\theta}_0)}{p_S(\mathbf{W}\mathbf{A}_0\mathbf{s}|\boldsymbol{\theta})|\det(\mathbf{W}\mathbf{A}_0)|} d\mathbf{s} + c^{\text{te}}$$

$$(24)$$

where, assuming a large N, we have invoked the strong law of large numbers [60] to obtain (23) and c^{te} in (24) is a term not depending on $(\mathbf{W}, \boldsymbol{\theta})$. The right-hand side of (24) is, apart from c^{te}, the Kullback Libler divergence (KLD) [61] between the densities $p_S(\cdot|\boldsymbol{\theta}_0)$ and $p_S(\mathbf{W}\mathbf{A}_0(\cdot)|\boldsymbol{\theta})|\det(\mathbf{W}\mathbf{A}_0)|$. Therefore, computing the ML estimate of $(\mathbf{W}, \boldsymbol{\theta})$ amounts to minimize the KLD between those densities.

The KLD between two densities is nonnegative and takes the value zero if and only if the densities are equal almost everywhere [61]. As expected, in our case, the parameters $(\mathbf{W}, \boldsymbol{\theta}) = (\mathbf{W}_0, \boldsymbol{\theta}_0)$ are an ML estimate as they yield zero KLD. The relevant question now is whether there is any parameters $(\mathbf{W}', \boldsymbol{\theta}') \neq (\mathbf{W}_0, \boldsymbol{\theta}_0)$ such that $p_S(\mathbf{s}|\boldsymbol{\theta}_0) = p_S(\mathbf{W}\mathbf{A}_0\mathbf{s}|\boldsymbol{\theta})|\det(\mathbf{W}\mathbf{A}_0)|$ almost everywhere in S_{p-1}. The answer is negative. The proof of this results beyond the scope of this paper. We give, however, an informal justification hereinafter.

Assume that $\mathbf{W} \neq \mathbf{W}_0$. In this case, the parametric family $p_S(\mathbf{W}\mathbf{A}_0(\cdot)|\boldsymbol{\theta})|\det(\mathbf{W}\mathbf{A}_0)|$ does not belong to the mixture of Dirichlet class, and then, it does not follow that $p_S(\mathbf{s}|\boldsymbol{\theta}_0) = p_S(\mathbf{W}\mathbf{A}_0\mathbf{s}|\boldsymbol{\theta})|\det(\mathbf{W}\mathbf{A}_0)|$ almost everywhere in S_{p-1}. As-

TABLE I
PARAMETERS OF THE DIRICHLET DENSITIES AND THEIR STATISTICS

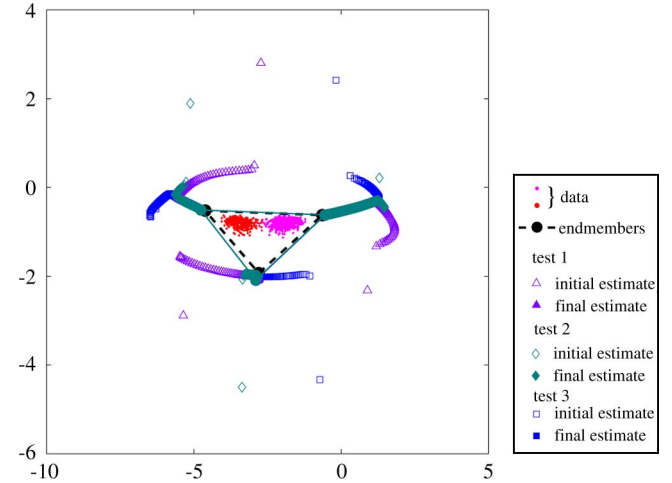| Region	A	B
Number of pixels ($N = 10^4$)	$2N/3$	$N/3$
Parameters values	(6, 25, 9)	(7, 8, 23)
$(\mathbb{E}[\mathbf{s}_1], \ldots, \mathbb{E}[\mathbf{s}_p])$	(0.15, 0.52, 0.22)	(0.18, 0.21, 0.60)
$(\text{var}[\mathbf{s}_1], \ldots, \text{var}[\mathbf{s}_p])$ ($\times 10^{-3}$)	(3.10, 5.70, 4.30)	(3.90, 4.30, 6.10)

Fig. 2. Illustration of the DECA evolution for three different initial matrices $\widehat{\mathbf{A}}$. (Black large dots) True endmembers. (Red and pink dots) Spectral vectors. Initial estimates: (Purple triangles) Test 1, (blue squares) test 2, and (green diamonds) test 3. Final estimates: (Filled purple triangles) Test 1, (filled blue squares) test 2, and (filled green diamonds) test 3.

sume now that $\mathbf{W} = \mathbf{W}_0$ and $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Again, it is not possible to have $p_S(\mathbf{s}|\boldsymbol{\theta}) = p_S(\mathbf{s}|\boldsymbol{\theta}_0)$ almost everywhere in S_{p-1}. We conclude then that the KLD is zero if and only if $(\mathbf{W}, \boldsymbol{\theta}) = (\mathbf{W}_0, \boldsymbol{\theta}_0)$.

The next experiment, based on 50 Monte Carlo runs, gives further evidence of the claimed identifiability of the statistical model. For each run, the data sets are generated according to (3), where three endmember signatures ($p = 3$) are selected from the USGS digital spectral library [62]. These reflectances are observed in $L = 224$ spectral bands ranging from 0.3 to 2.5 μm. The scene is composed by $N = 10^4$ pixels partitioned into two different regions. The size of regions A and B represents 66% and 33%, respectively. The abundance fractions for each region follow a Dirichlet distribution with parameters $\boldsymbol{\theta}_A = [6, 25, 9]$ and $\boldsymbol{\theta}_B = [7, 8, 23]$ for regions A and B, respectively. Table I summarizes, for each region, the scene parameters and the mean and variance of the abundance fraction statistics.

Fig. 2 shows a scatterplot of the evolution of DECA for three initializations, which differ from an endmember rotation. Black large dots represent the true endmembers, and red and pink dots represent the observed spectral vectors for each region. Purple triangles, blue squares, and green diamonds represent the initial estimated matrix $\widehat{\mathbf{A}}$ and the successive DECA estimates along the iterations, and filled triangles, squares, and diamonds represent the final estimates. Notice that, in spite of the large differences among the three initializations, the final estimates of the endmembers are practically equal. This is clearly in line with our claim that DECA statistical model is identifiable.
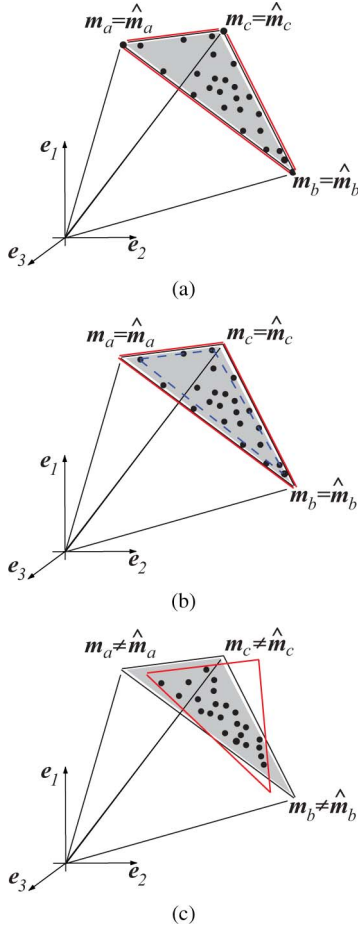
(a)

(b)

(c)

Fig. 3. Illustration of the geometrically based approach problem. Example of mixture: (a) With pure pixels (solid red line—estimated simplex); (b) without pure pixels and with pixels in the facets (solid red line—estimated simplex based on minimum volume; dashed blue line—estimated simplex by VCA); and (c) highly mixed pixels (solid red line—estimated simplex based on minimum volume).

IV. EVALUATION WITH SIMULATED DATA

In this section, the proposed method is tested and quantitatively compared with VCA [18], MVSA, [31], MVES [29], MVC-NMF [44], SPICE [28], and SISAL [32]. The parameters of the algorithms were hand tuned for optimal performance.

Fig. 3 shows three data sets raising different degrees of difficulties in which unmixing is concerned: The data set shown in Fig. 3(a) contains pure pixels, i.e., the spectra corresponding to the simplex vertices are in the data set. This is the easiest scenario with which all the unmixing algorithms cope without problems; the data set shown in Fig. 3(b) does not contain pure pixels, at least for some endmembers. This is a much more challenging task, usually attacked with the minimum-volume-based methods such as MVSA, MVES, MVC-NMF, ICE, SPICE, and SISAL. A necessary condition for this class of algorithms to yield good results is the presence of at least $p-1$ spectral vectors per simplex facet, which is not the case shown in Fig. 3(c), where a highly mixed data set is shown. Unmixing of these data sets is beyond the reach of any of above referred to algorithm and is where our approach exhibits a clear advantage.

The experiments presented in this section are based on 50 Monte Carlo runs. For each run, the data set, generated according to (3), is generated with different endmember signatures, selected from the USGS digital spectral library [62], with different abundance fractions, and with different additive noise levels.

Four experiments with simulated data are presented: In the first experiment, the algorithms are tested in scenarios with different Dirichlet parameters for the noiseless case. This experiment allows us to show the functioning of the algorithm for highly mixed data and thus with the absence of pure pixels. In the second experiment, the performance is measured as a function of the number of endmembers present in the scene. In the third experiment, the number of pixels of the scene is varied. Finally, in the fourth experiment, the algorithms are evaluated with respect to the SNR, defined by

$$SNR \equiv 10 \log_{10} \frac{\mathbb{E}\left[(\mathbf{As})^{\mathrm{T}} \mathbf{As}\right]}{\mathbb{E}\left[\mathbf{n}^{\mathrm{T}} \mathbf{n}\right]} \qquad (25)$$

where $\mathbf{n}$ is additive zero-mean Gaussian noise which is assumed to be i.i.d.

To evaluate the performance of the different algorithms, the abundance fractions and the signature estimates are compared with the true ones. Based on the mean square error (MSE), we define the spectral mean error (SME) and the abundance mean error (AME)

$$SME \equiv \frac{1}{pL} \|\mathbf{M} - \widehat{\mathbf{M}}\|_F^2 \qquad (26)$$

$$AME \equiv \frac{1}{pN} \|\mathbf{S} - \widehat{\mathbf{S}}\|_F^2 \qquad (27)$$

where the columns of $\widehat{\mathbf{M}} = \mathbf{E}_p \widehat{\mathbf{W}}^{-1}$ are endmember signature estimates, $\mathbf{S} \equiv [\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \dots, \mathbf{s}^{(N)}]$ is a $p \times N$ matrix with the endmember abundance fractions of each pixel, and $\widehat{\mathbf{S}} \equiv [\widehat{\mathbf{s}}^{(1)}, \widehat{\mathbf{s}}^{(2)}, \dots, \widehat{\mathbf{s}}^{(N)}]$ holds the abundance fraction estimates (notation $\| \cdot \|_F$ stands for Frobenius norm).

A common performance metric is the spectral angle distance, which measures the angle between a signature $\mathbf{m}_i$ and its estimate $\widehat{\mathbf{m}}_i$ [13]. Based on this metric, we define a spectral mean angle error (SMAE), given by

$$SMAE \equiv \sqrt{\frac{1}{p} \sum_{i=1}^{p} \left[\arccos \left(\frac{\mathbf{m}_i^{\mathrm{T}} \widehat{\mathbf{m}}_i}{\|\mathbf{m}_i\| \|\widehat{\mathbf{m}}_i\|} \right) \right]^2}. \qquad (28)$$

It is clear that the performance of the algorithms increases as SME, AME, and SMAE indices approach zero. Notice, however, that the estimates of $\mathbf{M}$ and $\mathbf{S}$ are up to a permutation matrix; thus, a simple algorithm based on the Hungarian method [63], [64] has been designed to infer the permutation matrix leading to the best index of performance.

A. Experiment I

In this section, data sets are generated as in Section III-C, i.e., three endmember signatures ($p = 3$) were selected from the USGS digital spectral library. The scene is composed by
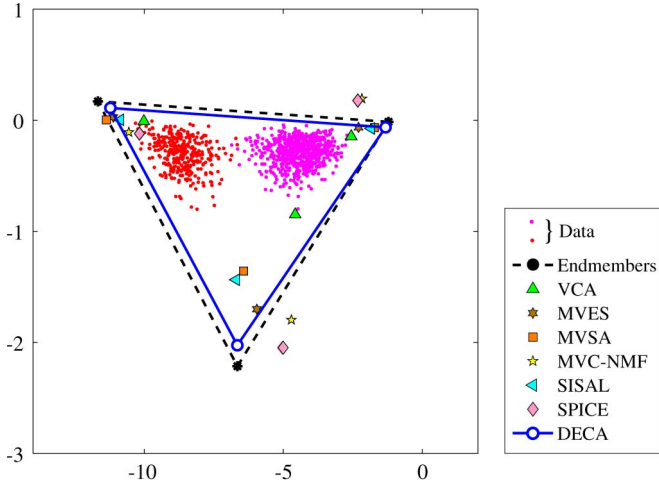
Fig. 4. Scatterplot of the three-endmember mixture. (Large dots) True endmembers. (Squares) MVSA estimate. (Triangles up) VCA estimate. (Hexagrams) MVES estimate. (Pentagrams) MVC-NMF. (Diamonds) SPICE estimate. (Triangles left) SISAL estimate. (Circles) Proposed method.

TABLE II
PARAMETERS OF THE DIRICHLET DENSITIES AND ITS ESTIMATES FOR EACH REGION OF THE SCENE

Region	A	B
Number of pixels ($N = 10^4$)	$2N/3$	$N/3$
Mixing probabilities (ϵ_q)	0.664	0.336
Parameter values	(6, 25, 9)	(7, 8, 23)
Parameter estimates	(5.5, 25.9, 9.1)	(6.5, 6.9, 23.0)

$N = 10^4$ pixels partitioned into two different regions (the size of regions A and B represents 66% and 33%, respectively). The abundance fractions for each region follow a Dirichlet distribution with $\boldsymbol{\theta_A} = [6, 25, 9]$ and $\boldsymbol{\theta_B} = [7, 8, 23]$. Table I summarizes the scene parameters and the abundance fraction statistics for each region. Notice that, for this setting, the spectral vectors are highly mixed. Fig. 4 shows a scatterplot for a simulated scene, where dots represent the observed spectral vectors and large dots represent the true endmembers. The two clouds correspond to the two regions in the scene.

In this experiment, the number of modes varies from $k_{\max} = 5$ to $k_{\min} = 1$, the Dirichlet parameters are randomly initialized, and the mixing probabilities are set to $\epsilon_q = 1/k$, for $q = 1, \ldots, k$. This setting reflects a situation in which no knowledge of the size and the number of regions in the scene exists.

The parameters of the remaining methods were hand tuned for optimal performance, namely, the MVC-MNF regularization parameter is $\tau = 10^{-4}$; the convergence tolerance for MVES method is 10^{-6}; SISAL regularization parameters are set to $\lambda = 10$, $\tau = 1$, and $\mu = 10^{-4}$; SPICE regularization parameter is set to $\mu = 10^{-3}$; the sparsity parameter is $\Gamma = 0.5$; and the stopping criterion is 10^{-6}.

For this data set, the minimum of the negative log-likelihood [see (19)] is reached when the estimated number of modes is $\widehat{k} = 2$ which is the number of regions on the scene. The estimated parameters at this iteration are presented on the last line of Table II. Note that the estimated values are close to the true parameter values. The Dirichlet mixing probability (ϵ_q)
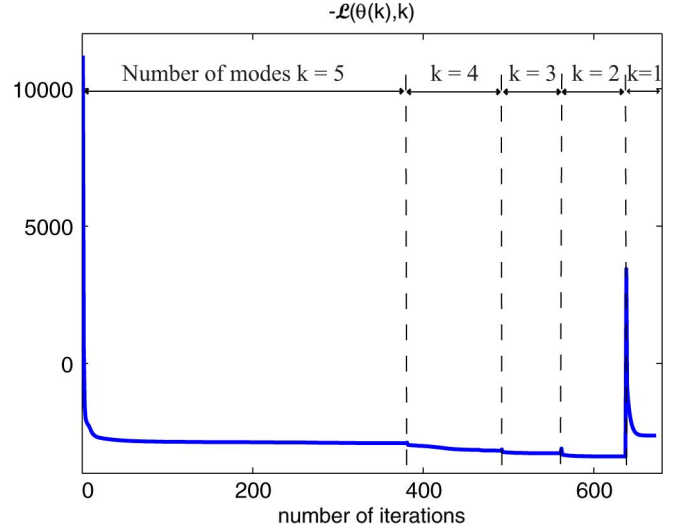

Fig. 5. Negative log-likelihood evolution as a function of the number of iterations. Vertical lines represent the iteration where a component annihilation is performed.

values are 0.664 and 0.336, which correspond to the areas in the scene occupied by regions A and B, respectively. Fig. 5 shows the evolution of the negative log-likelihood as a function of the number of iterations. The minimum of the function occurs at iteration number 637 when $\widehat{k} = 2$. Notice that the negative log-likelihood is not monotonic because DECA finds the optimal model order by computing a sequence of ML estimates $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}(k)$, for $k = k_{\min}, \ldots, k_{\max}$ (see Algorithm 1, lines 12–24). The $\widehat{k}_{\mathrm{best}}$ corresponds to the minimum value of $-\mathcal{L}(\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}(k))$, for $k = k_{\min}, \ldots, k_{\max}$. To make clear the dependence of the negative log-likelihood $-\mathcal{L}(\boldsymbol{\theta}(k), k)$ on the model order k, we show in Fig. 5 vertical lines where a component annihilation is performed.

Fig. 4 shows the true endmembers (large dots), the estimated endmembers by the proposed method (circles), by MVSA (squares), by VCA (triangles up), by MVES (hexagrams), by MVC-NMF (pentagrams), by SPICE (diamonds), and by SISAL (triangles left). Since the endmembers are highly mixed, i.e., there is no spectral vector near the vertices nor near facets, the proposed method gives better estimates than the other methods. Note that the endmember estimates generated by the proposed method are closer to the true endmembers than the estimates provided by the compared algorithms. Table III shows the mean and the variance (in brackets) of the performance results for the different algorithms based on 50 Monte Carlo runs. These results illustrate the advantage of the proposed scheme when compared with the remaining methods.

In order to test the effectiveness of DECA as a function of the initialization parameters, these Dirichlet parameters are initialized in three different ways: Randomly, all parameters set to one and all parameters set to five. The first initialization does not assume any kind of prior information, the second assumes that all samples are uniformly distributed over the simplex, and the last assumes that the data set is highly mixed. Table IV presents the mean and the variance (in brackets) of the performance measures. This test shows that DECA yields similar performance independently of the initialization parameters.

TABLE III
MEAN AND VARIANCE (IN BRACKETS) OF THE EVALUATION
RESULTS FOR THE DIFFERENT ALGORITHMS, BASED ON
50 MONTE CARLO RUNS ($N = 10^4$, $p = 3$, $L = 224$,
$\boldsymbol{\theta_A} = [6, 25, 9]$, $\boldsymbol{\theta_B} = [7, 8, 23]$, AND $SNR = \infty$)

Algorithm	SME	SMAE	AME
DECA	0.0002 (0)	0.0215 (0.0005)	0.0007 (0)
VCA [18]	0.0108 (0)	0.1835 (0.0057)	0.0557 (0.0053)
MVSA [31]	0.0027 (0)	0.1166 (0.0036)	0.0033 (0)
MVES [29]	0.0052 (0)	0.1660 (0.0071)	0.0063 (0.0006)
MVC-NMF [44]	0.0059 (0)	0.1504 (0.0055)	0.0052 (0)
SPICE [28]	0.0069 (0)	0.1611 (0.0067)	0.0066 (0)
SISAL [32]	0.0024 (0)	0.1162 (0.0036)	0.0031 (0)

TABLE IV
MEAN AND VARIANCE (IN BRACKETS) OF THE EVALUATION RESULTS
FOR DECA WITH DIFFERENT INITIALIZATION PARAMETERS,
BASED ON 50 MONTE CARLO RUNS ($N = 10^4$, $p = 3$, $L = 224$,
$\boldsymbol{\theta_A} = [6, 25, 9]$, $\boldsymbol{\theta_B} = [7, 8, 23]$, AND $SNR = \infty$)

Algorithm	SME	SMAE	AME
DECA (θ_i random)	0.0002 (0)	0.0215 (0.0005)	0.0007 (0)
DECA ($\theta_i = 1$)	0.0003 (0)	0.0261 (0.0007)	0.0007 (0)
DECA ($\theta_i = 5$)	0.0001 (0)	0.0191 (0.0001)	0.0001 (0)

TABLE V
EVALUATION RESULTS FOR THE DIFFERENT ALGORITHMS,
BASED ON 50 MONTE CARLO RUNS ($\boldsymbol{\theta} = (1, 1, 1)$,
$N = 10^4$, $p = 3$, $L = 224$, AND $SNR = \infty$)

Algorithm	SME	SMAE	AME
DECA	0.0000 (0.0000)	0.0013 (0.0000)	0.0000 (0.0000)
VCA	0.0000 (0.0000)	0.0101 (0.0000)	0.0001 (0.0000)
MVSA	0.0000 (0.0000)	0.0023 (0.0000)	0.0000 (0.0000)
MVES	0.0003 (0.0000)	0.0285 (0.0073)	0.0032 (0.0000)
MVC-NMF	0.0000 (0.0000)	0.0075 (0.0000)	0.0000 (0.0000)
SPICE	0.0005 (0.0000)	0.0384 (0.0004)	0.0006 (0.0000)
SISAL	0.0000 (0.0000)	0.0017 (0.0000)	0.0000 (0.0000)

TABLE VI
EVALUATION RESULTS FOR THE DIFFERENT ALGORITHMS,
BASED ON 50 MONTE CARLO RUNS ($\boldsymbol{\theta} = (5, 5, 5)$,
$N = 10^4$, $p = 3$, $L = 224$, AND $SNR = \infty$)

Algorithm	SME	SMAE	AME
DECA	0.0001 (0.0000)	0.0141 (0.0000)	0.0001 (0.0000)
VCA	0.0065 (0.0000)	0.1268 (0.0010)	0.0129 (0.0001)
MVSA	0.0011 (0.0000)	0.0576 (0.0003)	0.0005 (0.0000)
MVES	0.0015 (0.0003)	0.0579 (0.0013)	0.0043 (0.0000)
MVC-NMF	0.0060 (0.0000)	0.1282 (0.0009)	0.0055 (0.0000)
SPICE	0.0089 (0.0000)	0.1573 (0.0013)	0.0086 (0.0000)
SISAL	0.0011 (0.0000)	0.0604 (0.0004)	0.0005 (0.0000)

TABLE VII
EVALUATION RESULTS FOR THE DIFFERENT ALGORITHMS,
BASED ON 50 MONTE CARLO RUNS ($\boldsymbol{\theta} = (10, 10, 10)$,
$N = 10^4$, $p = 3$, $L = 224$, AND $SNR = \infty$)

Algorithm	SME	SMAE	AME
DECA	0.0005 (0.0000)	0.0434 (0.0012)	0.0002 (0.0000)
VCA	0.0166 (0.0004)	0.1956 (0.0082)	0.0419 (0.0003)
MVSA	0.0035 (0.0000)	0.1195 (0.0045)	0.0018 (0.0000)
MVES	0.0056 (0.0001)	0.1234 (0.0044)	0.0086 (0.0000)
MVC-NMF	0.0098 (0.0000)	0.1568 (0.0060)	0.0056 (0.0000)
SPICE	0.0149 (0.0000)	0.1928 (0.0069)	0.0104 (0.0000)
SISAL	0.0036 (0.0000)	0.1318 (0.0048)	0.0013 (0.0000)

TABLE VIII
EVALUATION RESULTS FOR THE DIFFERENT ALGORITHMS,
BASED ON 50 MONTE CARLO RUNS ($p = 3$, $N = 10^4$,
$L = 224$, $\boldsymbol{\theta} = [5, \ldots, 5]$, AND $SNR = \infty$)

Algorithm	SME	SMAE	AME
DECA	0.0001 (0.0000)	0.0140 (0.0004)	0.0001 (0.0000)
VCA	0.0050 (0.0001)	0.1190 (0.0027)	0.0066 (0.0012)
MVSA	0.0007 (0.0000)	0.0517 (0.0018)	0.0007 (0.0000)
MVES	0.0012 (0.0000)	0.0774 (0.0034)	0.0021 (0.0000)
MVC-NMF	0.0034 (0.0000)	0.1040 (0.0024)	0.0030 (0.0000)
SPICE	0.0070 (0.0000)	0.1461 (0.0025)	0.0068 (0.0000)
SISAL	0.0007 (0.0000)	0.0607 (0.0019)	0.0004 (0.0000)

Three different experiments are now presented with a scene composed by one single region, with $N = 10^4$ pixels and three endmembers ($p = 3$) and where abundance fractions follow a Dirichlet distribution with $\boldsymbol{\theta} = [1, 1, 1]$, $\boldsymbol{\theta} = [5, 5, 5]$, and $\boldsymbol{\theta} = [10, 10, 10]$. The difficulty of the unmixing problem increases with the value of the Dirichlet parameters: For $\boldsymbol{\theta} = [1, 1, 1]$, the abundances are uniformly distributed over the simplex, and thus, it is very likely to have pixels over, or very close, to the facets of the simplex; for $\boldsymbol{\theta} = [10, 10, 10]$, this is no longer true, since pixels are highly mixed and thus the geometrically based approaches fail. Tables V–VII present the evaluation metrics for the three different setups. As expected, the performance of all algorithms is worse for the highly mixed scene. Note, however, that the proposed method shows much better results in highly mixed scenes, when compared with the other methods.

B. Experiment II

In this section, the different algorithms are evaluated as a function of the number of endmembers present in the scene. The setup for this experiment is composed by an image (with $N = 10^4$) where abundance fractions follow a Dirichlet distribution with $\theta_i = 5$, for $i = 1, \ldots, p$. The number of endmembers takes values in the set $\{3, 5, 10\}$.

Tables VIII–X show the performance results as a function of the number of endmembers. Once more, the proposed algorithm outperforms, by large, the remaining methods, although its performance is better when the number of endmembers is small.

C. Experiment III

In this section, the algorithms are evaluated as a function of the number of pixels of the scene. The simulated parameters are as in Experiment II, except for the number of pixels N which takes values in the set $\{500, 1000, 5000\}$ and the number of endmembers is now set to $p = 3$. Tables XI–XIII show the evaluation measures as a function of the number of pixels in the scene. We conclude that the behavior of VCA, MVC-NMF,

TABLE IX
EVALUATION RESULTS FOR THE DIFFERENT ALGORITHMS,
BASED ON 50 MONTE CARLO RUNS ($p = 5$, $N = 10^4$,
$L = 224$, $\boldsymbol{\theta} = [5, \dots, 5]$, AND $SNR = \infty$)

Algorithm	SME	SMAE	AME
DECA	0.0003 (0.0000)	0.0177 (0.0001)	0.0001 (0.0000)
VCA	0.0259 (0.0002)	0.2494 (0.0016)	0.0130 (0.0003)
MVSA	0.0015 (0.0000)	0.0767 (0.0004)	0.0003 (0.0000)
MVES	0.0015 (0.0000)	0.0882 (0.0010)	0.0006 (0.0000)
MVC-NMF	0.0190 (0.0000)	0.1659 (0.0047)	0.0083 (0.0000)
SPICE	0.0224 (0.0001)	0.2160 (0.0095)	0.0090 (0.0000)
SISAL	0.0012 (0.0000)	0.0787 (0.0007)	0.0003 (0.0000)

TABLE X
EVALUATION RESULTS FOR THE DIFFERENT ALGORITHMS,
BASED ON 50 MONTE CARLO RUNS ($p = 10$, $N = 10^4$,
$L = 224$, $\boldsymbol{\theta} = [5, \dots, 5]$, AND $SNR = \infty$)

Algorithm	SME	SMAE	AME
DECA	0.0005 (0.0000)	0.0283 (0.0003)	0.0003 (0.0000)
VCA	0.0429 (0.0001)	0.3344 (0.0025)	0.0563 (0.0000)
MVSA	0.0057 (0.0016)	0.1463 (0.0430)	0.0006 (0.0000)
MVES	0.0020 (0.0005)	0.1069 (0.0006)	0.0001 (0.0000)
MVC-NMF	0.0393 (0.0003)	0.3020 (0.0021)	0.0103 (0.0000)
SPICE	0.0371 (0.0002)	0.3036 (0.0046)	0.0077 (0.0000)
SISAL	0.0022 (0.0000)	0.1058 (0.0006)	0.0001 (0.0000)

TABLE XI
EVALUATION RESULTS FOR THE DIFFERENT ALGORITHMS,
BASED ON 50 MONTE CARLO RUNS ($N = 500$, $p = 3$,
$L = 224$, $\boldsymbol{\theta} = [5, 5, 5]$, AND $SNR = \infty$)

Algorithm	SME	SMAE	AME
DECA	0.0006 (0.0000)	0.0483 (0.0009)	0.0010 (0.0000)
VCA	0.0076 (0.0000)	0.1451 (0.0022)	0.0387 (0.0001)
MVSA	0.0019 (0.0000)	0.0821 (0.0012)	0.0020 (0.0000)
MVES	0.0017 (0.0006)	0.0721 (0.0096)	0.0027 (0.0001)
MVC-NMF	0.0053 (0.0000)	0.1149 (0.0016)	0.0042 (0.0000)
SPICE	0.0091 (0.0000)	0.1508 (0.0027)	0.0099 (0.0000)
SISAL	0.0021 (0.0000)	0.0804 (0.0011)	0.0023 (0.0000)

TABLE XII
EVALUATION RESULTS FOR THE DIFFERENT ALGORITHMS,
BASED ON 50 MONTE CARLO RUNS ($N = 1000$, $p = 3$,
$L = 224$, $\boldsymbol{\theta} = [5, 5, 5]$, AND $SNR = \infty$)

Algorithm	SME	SMAE	AME
DECA	0.0002 (0.0000)	0.0497 (0.0002)	0.0003 (0.0000)
VCA	0.0117 (0.0001)	0.1465 (0.0019)	0.0345 (0.0001)
MVSA	0.0020 (0.0000)	0.0722 (0.0009)	0.0013 (0.0000)
MVES	0.0025 (0.0000)	0.0836 (0.0005)	0.0022 (0.0000)
MVC-NMF	0.0051 (0.0000)	0.1265 (0.0013)	0.0040 (0.0000)
SPICE	0.0053 (0.0000)	0.1475 (0.0018)	0.0077 (0.0000)
SISAL	0.0019 (0.0000)	0.0757 (0.0009)	0.0014 (0.0000)

TABLE XIII
EVALUATION RESULTS FOR THE DIFFERENT ALGORITHMS,
BASED ON 50 MONTE CARLO RUNS ($N = 5000$, $p = 3$,
$L = 224$, $\boldsymbol{\theta} = [5, 5, 5]$, AND $SNR = \infty$)

Algorithm	SME	SMAE	AME
DECA	0.0000 (0.0000)	0.0139 (0.0000)	0.0001 (0.0000)
VCA	0.0099 (0.0002)	0.1208 (0.0033)	0.0221 (0.0003)
MVSA	0.0009 (0.0000)	0.0661 (0.0004)	0.0007 (0.0000)
MVES	0.0009 (0.0001)	0.0597 (0.0025)	0.0035 (0.0000)
MVC-NMF	0.0028 (0.0033)	0.0839 (0.0073)	0.0025 (0.0001)
SPICE	0.0068 (0.0000)	0.1378 (0.0027)	0.0084 (0.0000)
SISAL	0.0005 (0.0000)	0.0477 (0.0005)	0.0004 (0.0000)

TABLE XIV
EVALUATION RESULTS FOR THE DIFFERENT ALGORITHMS,
BASED ON 50 MONTE CARLO RUNS ($SNR = 40$ dB,
$N = 10^4$, $p = 3$, $L = 224$, AND $\boldsymbol{\theta} = [5, 5, 5]$)

Algorithm	SME	SMAE	AME
DECA	0.0000 (0.0000)	0.0102 (0.0001)	0.0001 (0.0000)
VCA	0.0079 (0.0011)	0.1482 (0.0100)	0.0152 (0.0018)
MVSA	0.0012 (0.0000)	0.0597 (0.0017)	0.0007 (0.0000)
MVES	0.0009 (0.0000)	0.0657 (0.0021)	0.0070 (0.0000)
MVC-NMF	0.0062 (0.0000)	0.1315 (0.0026)	0.0049 (0.0000)
SPICE	0.0082 (0.0014)	0.1742 (0.0039)	0.0105 (0.0001)
SISAL	0.0010 (0.0000)	0.0671 (0.0014)	0.0005 (0.0000)

TABLE XV
EVALUATION RESULTS FOR THE DIFFERENT ALGORITHMS,
BASED ON 50 MONTE CARLO RUNS ($SNR = 20$ dB,
$N = 10^4$, $p = 3$, $L = 224$, AND $\boldsymbol{\theta} = [5, 5, 5]$)

Algorithm	SME	SMAE	AME
DECA	0.0001 (0.0000)	0.0240 (0.0007)	0.0004 (0.0000)
VCA	0.0062 (0.0001)	0.1141 (0.0027)	0.0292 (0.0016)
MVSA	0.0010 (0.0000)	0.0745 (0.0008)	0.0009 (0.0000)
MVES	0.0034 (0.0000)	0.1062 (0.0115)	0.0038 (0.0000)
MVC-NMF	0.0035 (0.0000)	0.1060 (0.0013)	0.0043 (0.0000)
SPICE	0.0072 (0.0000)	0.1322 (0.0025)	0.0071 (0.0000)
SISAL	0.0012 (0.0000)	0.0752 (0.0009)	0.0007 (0.0000)

D. Experiment IV

In this section, the robustness to noise is evaluated. The simulated noiseless scenario is generated as in the first experiment with one single region and according to (1), i.e., $N = 10^4$, $p = 3$, $L = 224$, and $\boldsymbol{\theta} = [5, \dots, 5]$. The noisy scenario is obtained by adding i.i.d. zero-mean Gaussian noise to the noiseless data. The different algorithms are tested for different values of SNR (40, 20, and 10 dB). As expected, when comparing the noiseless and noisy data evaluation results (see Tables VI, XIV–XVI), DECA and VCA performance is more accurate when SNR is high. Interestingly, the MVSA, MVES, and SISAL performance slightly worse as the SNR is smaller. This phenomenon occurs because the data are highly mixed, i.e., all data points are concentrated inside the simplex. The addition of noise spreads the data inside the simplex, which brings data points closer to the facets of the simplex, enforcing the MVSA that contains all data closer to the true one.

MVES, and SPICE algorithms is quasi-independent of the number of pixels. The remaining algorithms have slightly better results for large scenes. This is so because it is more likely to have spectral vectors close to the facets as N increases.

TABLE XVI
EVALUATION RESULTS FOR THE DIFFERENT ALGORITHMS,
BASED ON 50 MONTE CARLO RUNS ($SNR = 10$ dB,
$N = 10^4$, $p = 3$, $L = 224$, AND $\boldsymbol{\theta} = [5, 5, 5]$)

Algorithm	SME	SMAE	AME
DECA	0.0056 (0.0000)	0.1461 (0.0136)	0.0037 (0.0000)
VCA	0.0088 (0.0000)	0.1257 (0.0033)	0.0701 (0.0003)
MVSA	0.0014 (0.0015)	0.0719 (0.0059)	0.0036 (0.0000)
MVES	0.0058 (0.0000)	0.1227 (0.0103)	0.0091 (0.0000)
MVC-NMF	0.0070 (0.0000)	0.1207 (0.0024)	0.0076 (0.0000)
SPICE	0.0077 (0.0000)	0.1296 (0.0031)	0.0109 (0.0000)
SISAL	0.0020 (0.0007)	0.0759 (0.0039)	0.0033 (0.0000)

TABLE XVII
COMPUTATION TIME (IN SECONDS) OF EACH UNMIXING METHOD FOR
$N = 10^3$, $p = 3$, $L = 224$, AND $\boldsymbol{\theta} = [5, \ldots, 5]$

Algorithm	Time
DECA	73.77
VCA [18]	0.15
MVSA [31]	1.87
MVES [29]	3.48
MVC-NMF [44]	2.61
SPICE [28]	120.07
SISAL [32]	0.43

The computation time (in seconds) required by each unmixing method for $N = 10^3$, $p = 3$, $L = 224$, and $\boldsymbol{\theta} = [5, \ldots, 5]$ is shown in Table XVII for an nonoptimized MATLAB 2008a 32-bit implementation on a PC equipped with an Intel core duo 2.1-GHz processor with 4 GB of RAM. From this table, one can observe that the computation time spent by VCA (pure-pixel-based unmixing method) is less than that of the other algorithms that do not require pure pixels. Among these algorithms, SISAL is the fastest. Table XVII also indicates that SPICE is the slowest followed by DECA. The complexity of DECA is approximately $O(N(4p^2 + 6\overline{k}p))$, where $\overline{k} = (k_{\min} + k_{\max})/2$.

V. EVALUATION WITH REAL DATA

In this section, the proposed method is applied to real hyperspectral data collected by the AVIRIS sensor. A subset of the Cuprite data set[2] is considered. This site has been extensively used for remote sensing experiments over the past years, and its geology was previously mapped in detail [65].

The data set is composed of 187 spectral bands with 10-nm bandwidth acquired in the 0.4–2.5-μm region (noisy and water absorption bands 1–3, 104–113, 148–167, and 221–224 were removed). The subset contains 50×90 pixels with a ground resolution of 17 m. Fig. 6 shows band 30 (wavelength $\lambda = 647.7$ nm) of the subimage of AVIRIS Cuprite Nevada data set.

For this experiment, the MVC-MNF regularization parameter is set to $\tau = 10^{-5}$; the convergence tolerance for MVES method is 10^{-6}; SISAL regularization parameters are set to $\lambda = 10$, $\tau = 1$, and $\mu = 10^{-5}$; SPICE regularization parameter

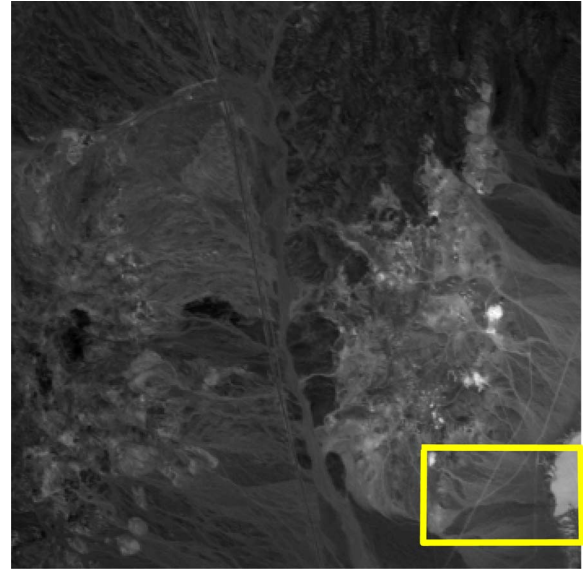[2] Available online at http://aviris.jpl.nasa.gov/html/aviris.freedata.html.

Fig. 6. Band 30 (wavelength $\lambda = 647.7$ nm) of the subimage of AVIRIS Cuprite Nevada data set (rectangle denotes the image fraction used in the experiment).
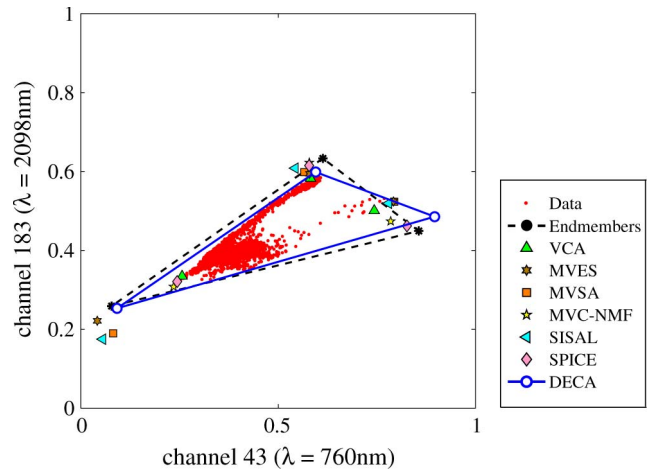

Fig. 7. (Dots) Scatterplot of Cuprite data set. (Circles) Estimated endmembers. (Large dots) USGS signatures. (Squares) MVSA estimate. (Triangles up) VCA estimate. (Hexagrams) MVES estimate. (Pentagrams) MVC-NMF. (Diamonds) SPICE estimate. (Triangles left) SISAL estimate. (Circles) Proposed method.

is set to $\mu = 10^{-4}$; the sparsity parameter is $\Gamma = 0.4$; and the stopping criterion is 10^{-6}.

Fig. 7 shows a scatterplot of the Cuprite data subset, where dots represent the pixels and circles represent the endmembers estimated by the proposed method. The results provided by VCA, MVSA, MVES, MVC-MNF, SISAL, and SPICE are also shown. Furthermore, we have also plotted the projections of Montmorillonite, Desert Varnish, and Alunite, which are known to dominate the considered subset image [65]. DECA has identified $k = 5$ modes, with parameters $\boldsymbol{\theta}_1 = [1.5, 4.1, 2.9]$, $\boldsymbol{\theta}_2 = [23.4, 51.3, 15.5]$, $\boldsymbol{\theta}_3 = [27.2, 26.6, 4.3]$, $\boldsymbol{\theta}_4 = [17.5, 3.6, 2.5]$, and $\boldsymbol{\theta}_5 = [10.3, 8.0, 7.3]$, and mixing probabilities $\epsilon_1 = 0.04$, $\epsilon_2 = 0.69$, $\epsilon_3 = 0.07$, $\epsilon_4 = 0.10$, and $\epsilon_5 = 0.10$. These parameters determine a highly nonuniform distribution over the simplex as could be inferred from the scatterplot shown in Fig. 7.

TABLE XVIII
EVALUATION RESULTS FOR THE CUPRITE DATA SET

Algorithm	SME	SMAE
DECA	3.8×10^{-3}	0.10
VCA [18]	11.4×10^{-3}	0.20
MVSA [31]	4.5×10^{-3}	0.17
MVES [29]	3.5×10^{-3}	0.15
MVC-NMF [44]	9.2×10^{-3}	0.19
SPICE [28]	9.6×10^{-3}	0.18
SISAL [32]	3.1×10^{-3}	0.14
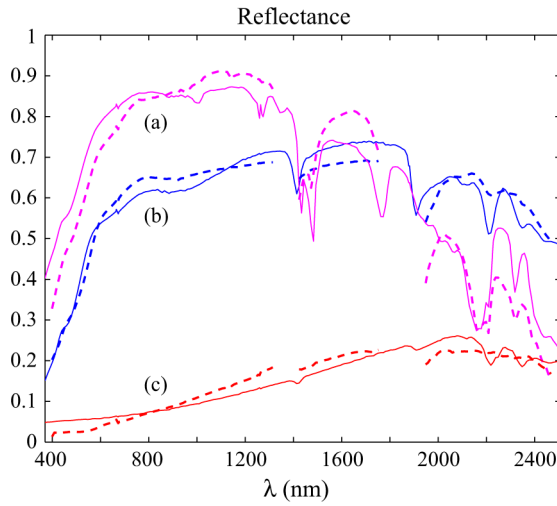
Fig. 8. (Solid line) Endmember signatures and (dashed line) their estimates.
(a) Alunite. (b) Montmorillonite. (c) Desert Varnish.

Fig. 9. Abundance map estimates for DECA. (a) Montmorillonite. (b) Desert Varnish. (c) Alunite.

Table XVIII shows the evaluation metrics SME and SMAE, where the signatures estimated by the different algorithms are compared with the nearest laboratory spectra, i.e., Alunite GDS84 Na03, Desert Varnish GDS78A Rhy, and Montmorillonite + Illite CM37. The proposed method yields the best SMAE and an SME not far from those of SISAL and MVES methods. Fig. 8 shows the endmember signatures and their estimates. The signatures provided by DECA are scaled by a factor in order to minimize the MSE between them and the respective library spectra. The similarity between the plotted curves is evident. Figs. 9–11 show the estimated abundance maps for DECA, VCA, and SISAL methods, respectively. DECA and SISAL abundance fraction estimates are very similar, whereas VCA has a slightly different result, namely, the upper left corner of the scene is mostly Montmorillonite instead of Alunite.

In our setup, the missing data indicate which mode produces each sample. After running the GEM algorithm, we obtain the posterior probability of the modes given the observations, which is, in fact, a soft classification map. Fig. 12(b) shows the image map where each color represents the mode with higher probability computed from (10). Fig. 12(a) shows a scatterplot of the data, where dots represent pixels of the image and their color indicates which mode has produced it. There is a clear relation between the information shown in Fig. 12(a) and (b) and the abundance fractions shown in (11): Modes 3 and 4 are
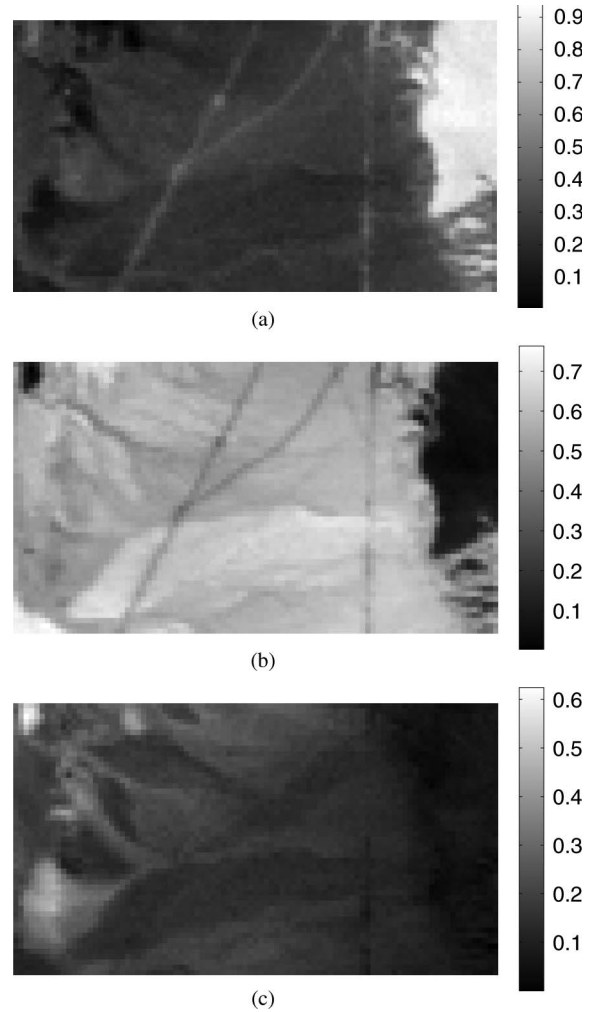
linked with Montmorillonite area, mode 2 is linked with Desert Varnish, and modes 1 and 5 are linked with Alunite.

VI. CONCLUSION

This paper has introduced DECA, a new algorithm to blindly unmix highly mixed hyperspectral data, thus beyond the reach of the geometrically based unmixing algorithms. DECA, developed under the statistical framework, models the abundance fractions as mixtures of Dirichlet densities, thus automatically enforcing the nonnegativity and the sum-to-one constraints. A cyclic minimization algorithm was developed where the following are observed: 1) The number of Dirichlet modes is inferred based on the MDL principle; 2) a GEM algorithm is derived to infer the model parameters; and 3) a sequence of augmented Lagrangian optimizations is used to compute the endmember signatures. The experimental results shown illustrate the potential of the proposed method to unmix hyperspectral data when the observed spectra are highly mixed.

As future research direction, in the vein of works [48] and [49], we intend to extend DECA by introducing a spatial regularization term based on the hidden variables that represent the probability of activating the Dirichlet modes. This
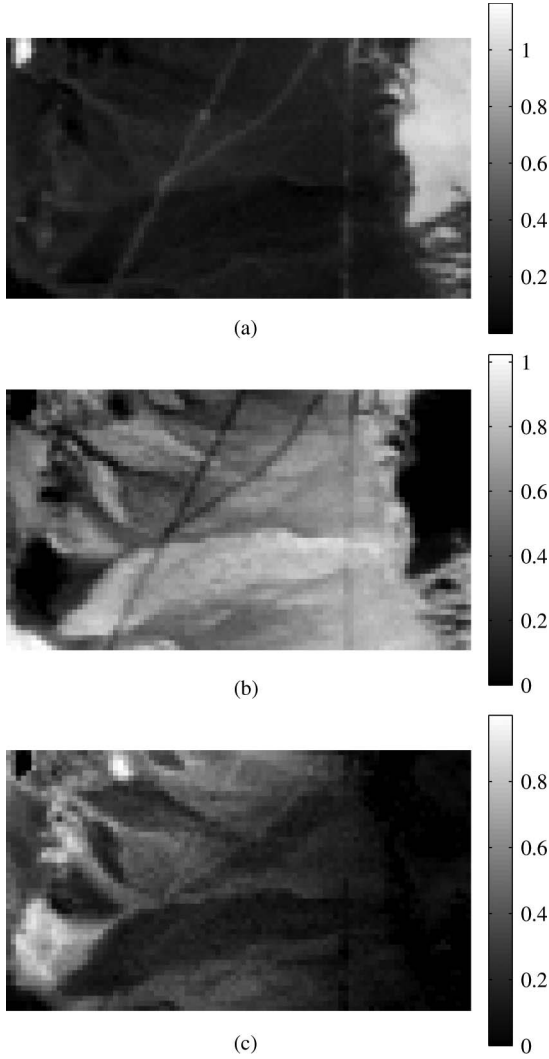
Fig. 10. Abundance map estimates for VCA. (a) Montmorillonite. (b) Desert Varnish. (c) Alunite.

Fig. 11. Abundance map estimates for SISAL. (a) Montmorillonite. (b) Desert Varnish. (c) Alunite.

regularization term enforces neighboring pixels to belong to the same Dirichlet mode.

APPENDIX

This Appendix presents a modification of the SISAL algorithm introduced in [32], termed as DUSAL, to solve the optimization problem that appears in the minimum-volume-type formulations of hyperspectral linear unmixing. SISAL solves a hard nonconvex (due to the presence of $|\det(\mathbf{W})|$) optimization problem by a sequence of nonsmooth convex subproblems using variable splitting and then applying the augmented Lagrangian methods (see [66], [67], and references therein for recent applications of these tools in image processing).

Herein, we apply the same concepts: variable splitting and augmented Lagrangian methods. The optimization with respect to $\mathbf{W}$ in (14) is equivalent to

$$\widehat{\mathbf{W}} = \arg\min_{\mathbf{W}} -\phi(\mathbf{W}\mathcal{X}) - \log|\det(\mathbf{W})|$$
$$\text{s.t.:} \quad \mathbf{W}\mathcal{X} \succeq 0 \quad \mathbf{1}_p^{\mathrm{T}}\mathbf{W} = \mathbf{a}^{\mathrm{T}} \qquad (29)$$

where $\mathbf{a}^{\mathrm{T}} = \mathbf{1}_N \mathcal{X}^{\mathrm{T}}(\mathcal{X}\mathcal{X}^{\mathrm{T}})^{-1}$ (see [32] for the simplification of the equality constraint).

Let $\mathbf{q} \equiv \text{vec}(\mathbf{W})$ denote the operator that stacks the columns of $\mathbf{W}$ in the column vector $\mathbf{q}$. Given that $\text{vec}(\mathbf{W}\mathcal{X}) = (\mathcal{X} \otimes \mathbf{I})\text{vec}(\mathbf{W})$, where $\otimes$ denotes the Kronecker operator, and defining $f(\mathbf{q}) = -\log|\det(\mathbf{W})|$, then (29) can be written as

$$\widehat{\mathbf{q}} = \arg\min_{\mathbf{q}} f(\mathbf{q}) - \phi(\mathbf{A}\mathbf{q})$$
$$\text{s.t.:} \quad \mathbf{B}\mathbf{q} = \mathbf{a} \qquad (30)$$

where $\mathbf{A} = \mathcal{X} \otimes \mathbf{I}$ and $\mathbf{B} = \mathbf{I} \otimes \mathbf{1}_p^{\mathrm{T}}$. We have dropped the inequality $\mathbf{A}\mathbf{q} \succeq 0$, as we define $-\phi(\mathbf{z}) = +\infty$ if some component of $\mathbf{z}$ is negative.

Using the variable splitting concept, the optimization problem (30) is equivalent to

$$\widehat{\mathbf{q}} = \arg\min_{\mathbf{q},\mathbf{z}} f(\mathbf{q}) - \phi(\mathbf{z})$$
$$\text{s.t.:} \quad \mathbf{B}\mathbf{q} = \mathbf{a} \quad \mathbf{A}\mathbf{q} = \mathbf{z} \qquad (31)$$

where variable $\mathbf{q}$ was split into the pair $(\mathbf{q}, \mathbf{z})$ and linked through the constraint $\mathbf{A}\mathbf{q} = \mathbf{z}$. The so-called augmented Lagrangian for this problem with respect to the constraint $\mathbf{A}\mathbf{q} = \mathbf{z}$ is

$$\mathcal{L}(\mathbf{q},\mathbf{z},\mathbf{d}) \equiv f(\mathbf{q}) - \phi(\mathbf{z}) + \boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{A}\mathbf{q} - \mathbf{z}) + \tau\|\mathbf{A}\mathbf{q} - \mathbf{z}\|^2$$
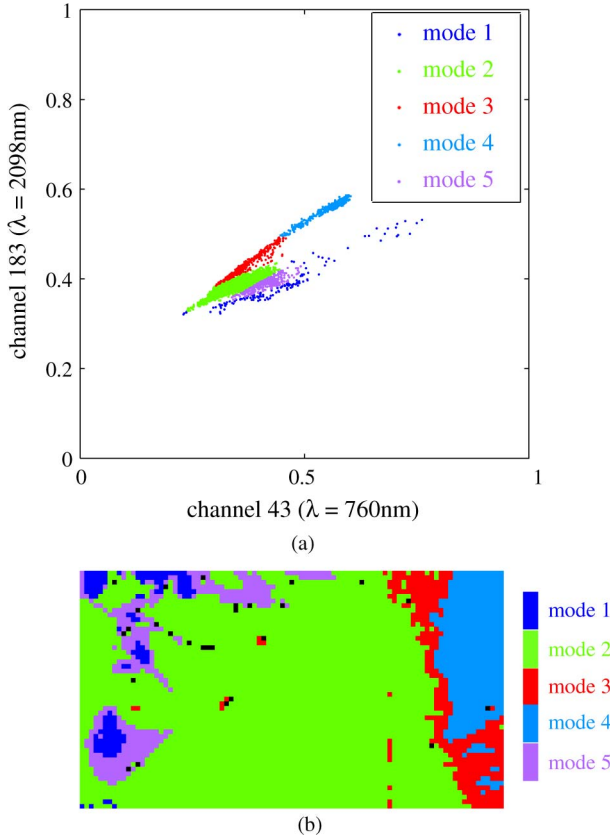$$= f(\mathbf{q}) - \phi(\mathbf{z}) + \tau\|\mathbf{A}\mathbf{q} - \mathbf{z} - \mathbf{d}\|^2 + c \qquad (32)$$

Fig. 12. (a) Scatterplot of Cuprite data set. (b) Image map: Each color is assigned to one Dirichlet mode.

where $\boldsymbol{\alpha}$ holds the Lagrange multipliers, $\mathbf{d} = \boldsymbol{\alpha}/(2\tau)$, and c is an irrelevant constant. The solution of the optimization with respect to $(\mathbf{q}, \mathbf{z})$ is decoupled into a block minimization with respect to $\mathbf{q}$ and with respect to $\mathbf{z}$, thus very easy to solve, where a quadratic approximation $f(\mathbf{q}) = \mathbf{g}^{\mathrm{T}}\mathbf{q} + \mu\|\mathbf{q} - \mathbf{q}_k\|^2$ is used. For each component of $\mathbf{z}$, it amounts to find the larger root of a second-order polynomial.

The pseudocode for the modified SISAL algorithm is presented in Algorithm 2, where lines 5–8 show the block minimization with respect to $\mathbf{q}$ and to $\mathbf{z}$ and line 9 updates $\mathbf{d}$.

Algorithm 2: DUSAL

1: INPUT $\mathbf{Y} \equiv [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N], p, \mathbf{W}$
2: OUTPUT $\widehat{\mathbf{q}}$,
3: $t := 0$
4: **repeat**
5:   $t := t + 1$
6:   $\mathbf{q}^{(t+1)} := \arg\min_{\mathbf{q}}(\mathbf{g}^{\mathrm{T}}\mathbf{q} + (\mu/2)\|\mathbf{q} - \mathbf{q}_k\|^2 + \tau\|\mathbf{A}\mathbf{q} - \mathbf{z}^{(t)} - \mathbf{d}^{(t)}\|^2)$
7:   s.t.: $\mathbf{B}\mathbf{q} = \mathbf{a}$,
8:   $\mathbf{z}^{(t+1)} := \arg\min_{\mathbf{z}}((1/2)\|\mathbf{A}\mathbf{q}^{(t+1)} - \mathbf{z}^{(t)} - \mathbf{d}^{(t)}\|^2 + 1/\tau\phi(\mathbf{z}^{(t)}))$
9:   $\mathbf{d}^{(t+1)} := \mathbf{d}^{(t)} - \mathbf{A}\mathbf{q}^{(t+1)} - \mathbf{z}^{(t+1)}$
10: **until** stopping criterion is satisfied

REFERENCES

[1] J. M. P. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing algorithm via dependent component analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2007, pp. 4033–4036.
[2] J. M. P. Nascimento and J. M. Bioucas-Dias, "Learning dependent sources using mixtures of dirichlet: Applications on hyperspectral unmixing," in *Proc. 1st IEEE Workshop Hyperspectral Image Signal Process.*, 2009, pp. 1–5.
[3] R. B. Singer and T. B. McCord, "Mars: Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance," in *Proc. 10th Lunar and Planetary Sci. Conf.*, 1979, pp. 1835–1848.
[4] G. Filacchione, "Planetary exploration with VIS-NIR imaging spectometers: The VIRTIS-M case," in *Proc. 1st IEEE GRSS Workshop Hyperspectral Image Signal Process.-WHISPERS*, 2009, pp. 1–4.
[5] T. M. Lillesand, R. W. Kiefer, and J. W. Chipman, *Remote Sensing and Image Interpretation*, 5th ed. Hoboken, NJ: Wiley, 2004.
[6] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral Remote Sensing: Principles and Applications*, 1st ed. New York: Taylor & Francis, 2007.
[7] A. M. Lefcourt, "Hyperspectral imaging and food safety," in *Proc. 1st IEEE GRSS Workshop Hyperspectral Image Signal Process.-WHISPERS*, 2009.
[8] B. Park, K. C. Lawrence, W. R. Windham, and D. P. Smith, "Performance of hyperspectral imaging system for poultry surface fecal contaminant detection," *J. Food Eng.*, vol. 75, no. 3, pp. 340–348, Aug. 2006.
[9] M. B. Lopes, J.-C. Wolff, J. M. Bioucas-Dias, and M. A. Figueiredo, "Determination of the composition of counterfeit Heptodin(TM) tablets by near infrared chemical imaging and classical least squares estimation," *Anal. Chim. Acta*, vol. 641, no. 1/2, pp. 46–51, 2009.
[10] C.-I. Chang, Ed., *Hyperspectral Data Exploitation: Theory and Applications*. Hoboken, NJ: Wiley, 2007.
[11] X. Briottet, Y. Boucher, A. Dimmeler, A. Malaplate, A. Cini, M. Diani, H. Bekman, P. Schwering, T. Skauli, I. Kasen, I. Renhorn, L. Klasén, M. Gilmore, and D. Oxford, "Military Applications of Hyperspectral Imagery," in *Proc. SPIE*, W. R. Watkins and D. Clement, Eds., 2006, vol. 6239, no. 1, p. 62 390B.
[12] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging aplications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.
[13] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
[14] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using linear mixing model," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1392–1409, Jul. 2001.
[15] N. Keshava, J. Kerekes, D. Manolakis, and G. Shaw, "An algorithm taxonomy for hyperspectral unmixing," in *Proc. SPIE AeroSense Conf. Algorithms Multispectral Hyperspectral Imag. VI*, 2000, vol. 4049, pp. 42–63.
[16] N. Keshava, "A survey of spectral unmixing algorithms," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 55–78, 2003.
[17] J. M. Bioucas-Dias and A. Plaza, "Hyperspectral unmixing: Geometrical, statistical, and sparse regression-based approaches," in *Proc. SPIE*, L. Bruzzone, Ed., 2010, vol. 7830, no. 1.
[18] J. M. P. Nascimento and J. M. 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.
[19] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2025–2041, Sep. 2002.
[20] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. Summaries 4th Annu. JPL Airborne Geosci. Workshop, JPL Pub. 93-26, AVIRIS Workshop*, 1993, vol. 1, pp. 11–14.
[21] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE Conf. Imag. Spectrom. V*, 1999, vol. 3753, pp. 266–275.

[22] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Rem. Sens.*, vol. 49, no. 11, pp. 4177–4193, Nov. 2011.

[23] R. Neville, K. Staenz, J. Lefebvre, and P. Hauff, "Automatic endmember extraction from hyperspectral data for mineral exploitation," in *Proc. 21st Can. Symp. Remote Sens.*, 1999, pp. 21–24.

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

[25] J. J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1045–1046, Jul. 1996.

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

[27] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "ICE: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.

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

[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. 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.

[31] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2008, vol. 3, pp. III-250–III-253.

[32] J. M. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. 1st IEEE GRSS Workshop Hyperspectral Image Signal Process.-WHISPERS*, 2009, pp. 1–4.

[33] A. Zymnis, S.-J. Kim, J. Skaf, M. Parente, and S. Boyd, "Hyperspectral image unmixing via alternating projected subgradients," in *Proc. 41st Asilomar Conf. Signals, Syst. Comput.*, 2007, pp. 1164–1168.

[34] N. Dobigeon, J.-Y. Tourneret, and C.-I. Chang, "Semi-supervised linear spectral unmixing using a hierarchical bayesian model for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, Jul. 2008.

[35] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.

[36] M. Arngren, M. N. Schmidt, and J. Larsen, "Bayesian nonnegative matrix factorization with volume prior for unmixing of hyperspectral images," in *Proc. IEEE Workshop MLSP*, Sep. 2009, pp. 1–6.

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

[38] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Hoboken, NJ: Wiley, 2001.

[39] J. M. P. Nascimento and J. M. Bioucas-Dias, "Does independent component analysis play a role in unmixing hyperspectral data?" *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 175–187, Jan. 2005.

[40] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, Jun. 1994.

[41] 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.

[42] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Non-negative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*. Hoboken, NJ: Wiley-Blackwell, 2009.

[43] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, Jan. 2009.

[44] 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.

[45] P. V. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra Appl.*, vol. 416, no. 1, pp. 29–47, Jul. 2006.

[46] P. Sajda, S. Du, and L. C. Parra, "Recovery of constituent spectra using non-negative matrix factorization," in *Proc. SPIE*, M. A. Unser, A. Aldroubi, and A. F. Laine, Eds., 2003, vol. 5207, no. 1, pp. 321–331.

[47] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, and C. Carteret, "Bayesian separation of spectral sources under non-negativity and full additivity constraints," *Signal Process.*, vol. 89, no. 12, pp. 2657–2669, Dec. 2009.

[48] A. Zare and P. Gader, "PCE: Piecewise convex endmember detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2620–2632, Jun. 2010.

[49] 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.

[50] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 381–396, Mar. 2002.

[51] 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.

[52] T. Minka, "Estimating a Dirichlet distribution," MIT, Cambridge, MA, Tech. Rep., 2000.

[53] G. McLachlan and D. Peel, *Finite Mixture Models*. Hoboken, NJ: Wiley, 2000.

[54] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. Hoboken, NJ: Wiley, 1996.

[55] K. Lange, *Optimization*, 1st ed. New York: Springer-Verlag, 2004.

[56] J. M. P. Nascimento, "Unsupervised hyperspectral unmixing," Ph.D. dissertation, Instituto Superior Técnico, Technical Univ. Lisbon, Lisbon, Portugal, 2006.

[57] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.

[58] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, Sep. 1978.

[59] D. M. Titterington, A. F. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*. Hoboken, NJ: Wiley, 1986, ser. Probability and Statistics Series.

[60] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*. New York: Springer-Verlag, 2004.

[61] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 2006.

[62] R. N. Clark, G. A. Swayze, A. Gallagher, T. V. King, and W. M. Calvin, "The U.S. Geological Survey Digital Spectral Library: Version 1: 0.2 to 3.0 μm," U.S. Geological Survey, Sunrise Valley Drive Reston, VA, Open File Rep. 93-592, 1993.

[63] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. New York: Dover, 2000, ch. 11.

[64] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quart.*, vol. 2, no. 1/2, pp. 83–97, Mar. 1955.

[65] G. Swayze, R. Clark, S. Sutley, and A. Gallagher, "Ground-truthing AVIRIS mineral mapping at Cuprite, Nevada," in *Proc. Summ. 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 47–49.

[66] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[67] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.

José M. P. Nascimento (S'03–M'06) received the B.S. degree in electronics and telecommunications engineering and the E.E. degree from the Instituto Superior de Engenharia de Lisboa, Politechnic Institute of Lisbon, Lisbon, Portugal, in 1993 and 1995, respectively, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from the Instituto Superior Técnico, Technical University of Lisbon, Lisbon, in 2000 and 2006, respectively.

He is currently a Professor with the Department of Electronics, Telecommunications and Computer Engineering, Instituto Superior de Engenharia de Lisboa. He is also currently a Researcher with the Instituto de Telecomunicações, Lisbon. His research interests include remote sensing, and signal and image processing.

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 Technical University of Lisbon, Lisbon, Portugal, in 1985, 1991, 1995, and 2007, respectively.

Since 1995, he has been with the Department of Electrical and Computer Engineering, Instituto Superior Técnico, Technical University of Lisbon. He is also currently a Senior Researcher with the Pattern and Image Analysis group, Instituto de Telecomunicações, Lisbon, which is a private not-for-profit research institution. His research interests include signal and image processing, pattern recognition, optimization, and remote sensing. He was and is involved in several national and international research projects and networks, including the Marie Curie Actions "Hyperspectral Imaging Network (Hyper-i-net)" and the "European Doctoral Program in Signal Processing (SIGNAL)."

Dr. Bioucas-Dias is an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, and he was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS. He is a Guest Editor of a special issue of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and of a special issue of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He has been a member of program/technical committees of several international conferences. He was a General Cochair of the 3rd IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing-Whispers'2011.