

LEARNING DEPENDENT SOURCES USING MIXTURES OF DIRICHLET: APPLICATIONS ON HYPERSPECTRAL UNMIXING

José M. P. Nascimento

Instituto Superior de Engenharia de Lisboa
and Instituto de Telecomunicações
Email: zen@isel.pt

José M. Bioucas-Dias

Instituto de Telecomunicações and
Instituto Superior Técnico,
Technical University of Lisbon
Email: bioucas@lx.it.pt

ABSTRACT

This paper is an elaboration of the DECA algorithm [1] to blindly unmix hyperspectral data. The underlying mixing model is linear, meaning that each pixel is a linear mixture of the endmembers signatures weighted by the correspondent abundance fractions. The proposed method, as DECA, is tailored to highly mixed mixtures in which the geometric based approaches fail to identify the simplex of minimum volume enclosing the observed spectral vectors. We resort then to a statistical framework, where the abundance fractions are modeled as mixtures of Dirichlet densities, thus enforcing the constraints on abundance fractions imposed by the acquisition process, namely non-negativity and constant sum.

With respect to DECA, we introduce two improvements: 1) the number of Dirichlet modes are inferred based on the minimum description length (MDL) principle; 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. The effectiveness of the proposed algorithm is illustrated with simulated and read data.

Index Terms— Blind Hyperspectral Unmixing, Minimum Description Length (MDL), Mixtures of Dirichlet Densities, Augmented Lagrangian Methods, Dependent Sources.

1. INTRODUCTION

Hyperspectral imaging sensors collect two dimensional spatial images from the Earth's surface over many contiguous bands of high spectral resolution covering the visible, near-infrared, and shortwave infrared (wavelengths between $0.3 \mu\text{m}$ and $2.5 \mu\text{m}$), in hundreds of narrow (on the order of 10 nm) contiguous spectral bands. These radiances, collected in spectral vectors, are mixtures of spectra from the substances (also called endmembers) present in the respective pixel coverage. The linear mixing assumption is widely used

This work was supported by Instituto de Telecomunicações, Fundação para a Ciência e Tecnologia, and fellowship SFRH/BPD/39475/2007.

to describe the observed hyperspectral vectors. According to this assumption, a mixed pixel is a linear combination of endmember signatures (endmember spectra) weighted by the correspondent abundance fractions. Under this model, the observations from a scene are in a simplex whose vertices correspond to the endmembers [2].

The geometric perspective just referred to has been exploited by many algorithms which we classify into two classes:

1. **Pure pixel**- The data set contains at least one pure pixel of each endmember. The *vertex component analysis* (VCA), [3], the *automated morphological endmember extraction* (AMEE) [4], the *pixel purity index* (PPI), [5], and the N-FINDR [6] are representative algorithms of this class.
2. **Non-pure pixel**- The data set does not contain pure pixels for some (or for none) endmembers. The *iterated constrained endmembers* (ICE), [7], the *minimum volume simplex* (MVSA) [8], and the *simplex identification by variable splitting and augmented Lagrangian tools* (SISAL) are representative algorithms of this class.

Non-pure pixel based algorithms are conceived for non-pure pixel scenarios. For a correct identification of the mixing matrix, there must be at least $p - 1$ (p stands for the number of endmembers) spectral vectors on each facet of the simplex. However, in highly mixed data sets, this assumption is not realistic, what has fostered research in the statistical front [1, 9].

Hyperspectral unmixing is a blind source separation problem, where abundance fractions can be interpreted as sources. Independent component analysis (ICA) comes naturally to mind and have, in fact, been used in hyperspectral applications. However, as shown in [10], the applicability of canonical ICA to hyperspectral data is compromised by the statistical dependence existing among abundances.

The *dependent component analysis* (DECA) algorithm [1] was introduced to unmix highly mixed hyperspectral data

sets, which places the unmixing problem beyond the reach of the geometrical based methods. In DECA, the abundance fractions are modeled as mixtures of Dirichlet densities, thus automatically enforcing source nonnegativity and constant sum constraints. The resulting algorithm implements a generalized expectation-maximization (GEM) iterative scheme to, simultaneously, estimate the endmember signatures (mixing matrix), density parameters, and the abundance fractions.

In this paper, we introduce two improvements in DECA:

1) the number of Dirichlet modes are inferred based on the minimum description length (MDL) principle as proposed in [11]; 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 [12].

This paper is organized as follows. Section 2 describes the fundamentals of the proposed method. Section 3 presents the method to infer the number of components modes of the Dirichlet mixtures. Sections 4 and 5 illustrate aspects of the performance of the proposed approach with experimental data based on U.S.G.S. laboratory spectra and with real hyperspectral data collected by the AVIRIS sensor, respectively. Section 6 concludes with some remarks.

2. STATISTICAL MODELLING

Assuming the linear observation model, each pixel \mathbf{y} of an hyperspectral image can be represented as a spectral vector in \mathbb{R}^L (L is the number of bands) and is given by $\mathbf{y} = \mathbf{M}\mathbf{s}$, where $\mathbf{M} \equiv [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_p]$ is an $L \times p$ mixing matrix (\mathbf{m}_j denotes the j th endmember signature), p is the number of endmembers present in the covered area, and $\mathbf{s} = [s_1, s_2, \dots, s_p]^T$ is the abundance vector containing the fractions of each endmember (notation $(\cdot)^T$ stands for vector transposed).

To be physically meaningful, abundance fractions are subject to nonnegativity and constant sum constraints, *i.e.*, $\{\mathbf{s} \in \mathbb{R}^p : s_j \geq 0, \sum_{j=1}^p s_j = 1\}$. 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, its is advantageous, in terms of SNR, memory usage, and computational complexity, to represent the spectral vectors in a signal subspace basis [13].

Let \mathbf{E}_p be a matrix, with orthonormal columns, spanning the signal subspace. The coordinates of the spectral vector \mathbf{y} with respect to \mathbf{E}_p are

$$\begin{aligned} \mathbf{x} &\equiv \mathbf{E}_p^T \mathbf{y} \\ &= \mathbf{A}\mathbf{s}, \end{aligned} \quad (1)$$

where $\mathbf{A} = \mathbf{E}_p^T \mathbf{M}$ is a $p \times p$ square mixing matrix and $\mathbf{x} =$

$[x_1, \dots, x_p]^T$ is a $p \times 1$ vector. Let's assume that $\mathbf{W} \equiv \mathbf{A}^{-1}$ exists. Then, we have $\mathbf{s} = \mathbf{W}\mathbf{x}$.

We assume that the abundance fractions follow a k -component Dirichlet finite mixture given by

$$p_S(\mathbf{s}|\boldsymbol{\theta}) = \sum_{q=1}^k \epsilon_q \underbrace{\frac{\Gamma(\sum_{j=1}^p \theta_{qj})}{\prod_{j=1}^p \Gamma(\theta_{qj})} \prod_{j=1}^p s_j^{\theta_{qj}-1}}_{D(\mathbf{s}|\boldsymbol{\theta}_q)}, \quad (2)$$

where the complete set of parameters $\boldsymbol{\theta}$ needed to specify the mixture contains the mixing probabilities $\epsilon_1, \dots, \epsilon_k$ and the q -component Dirichlet parameters $\boldsymbol{\theta}_q = \{\theta_{q1}, \dots, \theta_{qp}\}$, for $q = 1, \dots, k$, *i.e.*, $\boldsymbol{\theta} = \{\epsilon_1, \dots, \epsilon_k, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k\}$. Since the observed data are the spectral vectors \mathbf{x} , we must base our inferences on the density $p_X(\mathbf{x})$, which, in terms of p_S is given by

$$p_X(\mathbf{x}) = p_S(\mathbf{s} = \mathbf{W}\mathbf{x}|\boldsymbol{\theta}) |\det(\mathbf{W})|.$$

Consider that each vector \mathbf{s} represents one particular outcome of a p -dimensional random variable $\mathbf{S} = [S_1, \dots, S_p]^T$. Given a set of N i.i.d. samples $\mathcal{S} = \{\mathbf{s}^{(1)}, \dots, \mathbf{s}^{(N)}\}$, then, we may write the log-likelihood of the set of parameters $\boldsymbol{\theta}$ as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &\equiv \log p_S(\mathcal{S}|\boldsymbol{\theta}) \\ &= \sum_{i=1}^N \left[\log p_S(\mathbf{s}^{(i)}|\boldsymbol{\theta}) \right] \\ &= \sum_{i=1}^N \left[\log \sum_{q=1}^k \epsilon_q D(\mathbf{s}^{(i)}|\boldsymbol{\theta}_q) \right]. \end{aligned} \quad (3)$$

The maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{ML} = \arg \max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$ can not be found analytically [11]. The usual choice for obtaining the ML estimates of the parameters is the EM framework [14], which relies on the so-called incomplete data and missing data. In our setup, \mathcal{S} denotes the incomplete data and $\mathcal{Z} \equiv \{\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(N)}\}$, a set of N k -dimensional vectors representing which component has produced each sample, the missing data: each vector $\mathbf{z}^{(i)} = [z_1^{(i)}, \dots, z_k^{(i)}]$ is a binary k -vector, where only one component $z_q^{(i)}$ is set to one indicating which mode produced the i -sample. Defining $\mathcal{X} \equiv \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$, the complete log-likelihood is then

$$\begin{aligned} \mathcal{L}_C(\boldsymbol{\theta}, \mathbf{W}) &= \log [p_{X,Z}(\mathcal{X}, \mathcal{Z}|\boldsymbol{\theta})] + N \log |\det(\mathbf{W})| \\ &= \sum_{i=1}^N \left[\sum_{q=1}^k z_q^{(i)} \log \left(\epsilon_q D(\mathbf{s}^{(i)}|\boldsymbol{\theta}_q) \right) \right] \\ &\quad + N \log |\det(\mathbf{W})|. \end{aligned} \quad (4)$$

The EM algorithm iterates between the E-step and the M-step:

- **E-step:** Computes the conditional expectation of the complete log-likelihood, given the samples and the

current estimate $\hat{\boldsymbol{\theta}}^{(t)}$. The result is the so-called Q -function

$$Q(\boldsymbol{\theta}, \mathbf{W}; \hat{\boldsymbol{\theta}}^{(t)}, \widehat{\mathbf{W}}^{(t)}) = N \log |\det(\mathbf{W})| + \sum_{i=1}^N \left[\sum_{q=1}^k \beta_q^{(i,t)} \log \left[\epsilon_q D(\mathbf{s}^{(i)} | \boldsymbol{\theta}_q) \right] \right] \quad (5)$$

where

$$\beta_q^{(i,t)} = \frac{\hat{\epsilon}_q^{(t)} D(\mathbf{s}^{(i,t)} | \hat{\boldsymbol{\theta}}_q^{(t)})}{\sum_{l=1}^k \hat{\epsilon}_l^{(t)} D(\mathbf{s}^{(i,t)} | \hat{\boldsymbol{\theta}}_l^{(t)})}. \quad (7)$$

and $\mathbf{s}^{(i)} \equiv \mathbf{W} \mathbf{x}^{(i)}$ and $\mathbf{s}^{(i,t)} \equiv \widehat{\mathbf{W}}^{(t)} \mathbf{x}^{(i)}$

- **M-step:** Updates the parameter estimates according to

$$(\hat{\boldsymbol{\theta}}^{(t+1)}, \widehat{\mathbf{W}}^{(t+1)}) = \arg \max_{\boldsymbol{\theta}, \mathbf{W}} \left\{ Q(\boldsymbol{\theta}, \mathbf{W}; \hat{\boldsymbol{\theta}}^{(t)}, \widehat{\mathbf{W}}^{(t)}) \right\}. \quad (8)$$

Optimization (8) 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 $\boldsymbol{\theta}^{(t+1)}$, we maximize Q with respect to θ_j , for $j = 1, \dots, p$, resulting in the following learning rules for the mixing probabilities and for the mixture of Dirichlet source parameters:

$$\epsilon_q^{(t+1)} = \frac{1}{N} \sum_{i=1}^N \beta_q^{(i,t)}, \quad (9)$$

$$\hat{\theta}_{qj}^{(t+1)} = \Psi^{-1} \left(\Psi \left(\sum_{l=1}^p \hat{\theta}_{ql}^{(t)} \right) + \frac{\sum_{i=1}^N [\beta_q^{(i,t)} \log \hat{s}_j^{(t)}]}{\sum_{i=1}^N \beta_q^{(i,t)}} \right), \quad (10)$$

for $q = 1 \dots, K$ and $j = 1 \dots, p$, where $\Psi(\cdot)$ and $\Psi^{-1}(\cdot)$ denote the psi function and its inverse, respectively.

The resulting algorithm is of the generalized expectation-maximization class (GEM) [14]: the learning rule (9) maximizes Q -function with respect to $\epsilon_q^{(t)}$, whereas expression (10) assures that the Q -function does not decrease.

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

$$\begin{aligned} \widehat{\mathbf{W}}^{(t+1)} &= \arg \max_{\mathbf{W}} \phi(\mathbf{W} \mathcal{X}) + \log |\det(\mathbf{W})| \\ \text{sub. to: } \mathbf{W} \mathcal{X} &\geq 0, \quad \mathbf{1}_p^T \mathbf{W} \mathcal{X} = \mathbf{1}_N^T, \end{aligned} \quad (11)$$

where the constraints $\mathbf{W} \mathcal{X} \geq 0$ and $\mathbf{1}_p^T \mathbf{W} \mathcal{X} = \mathbf{1}_N^T$ enforces non-negativity and sum-to-one on each fractional abundance vector ($\mathbf{1}_p$ stands for p -dimensional column vector of ones) and

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

with

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

Problem (11) is very hard to solve owing to the term $\log |\det(\mathbf{W})|$. The recently introduced SISAL algorithm [12] is a quite effective tool to solve a problem similar to (11). Therefore, we apply here the same concepts: variable splitting and augmented Lagrangian methods. See [12] for details. The resulting algorithm is given in Algorithm 3 of [12], with the function $\|\mathbf{z}\|_h$ replaced with $\phi(\mathbf{z})$. The optimization problem in line 5 of Algorithm 3 is decoupled and thus very easy to solve. For each component of \mathbf{z} , it amounts to find the larger root of a second order polynomial.

3. 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 criterium alone cannot be used to do such an inference because of the nested nature of model parameters for increasing model orders.

To estimate the number of mixture components the we adopt the model selection criterium

$$\hat{k} = \arg \min_k \{-\mathcal{L}(\boldsymbol{\theta}(k)) + \mathcal{P}(k)\}, \quad (12)$$

where $\mathcal{L}(\boldsymbol{\theta}(k))$ is given by expression (3), for k modes, and $\mathcal{P}(k)$ is a penalizing term that increases with k [11]. Herein, we adopt the the *minimum description length* (MDL) [15, 16] which amount to compute

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \arg \min_{\boldsymbol{\theta}} \left\{ -\log p_{\mathcal{S}}(\mathcal{S} | \boldsymbol{\theta}) - \log p(\boldsymbol{\theta}) \right. \\ &\quad \left. + \frac{1}{2} \log (|\mathcal{I}(\boldsymbol{\theta})|) + \frac{c}{2} (1 - \log(12)) \right\}, \end{aligned} \quad (13)$$

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

For mixtures, the Fisher information matrix, cannot be obtained analytically. Herein, we follow [11], 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 a noninformative Jeffreys' prior for each $\boldsymbol{\theta}_j$, the objective function in (13) is given by

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}(k), k) &= -\log p_{\mathcal{S}}(\mathcal{S} | \boldsymbol{\theta}) + \frac{k(p+1)}{2} \\ &\quad + \frac{k}{2} \log \left(\frac{N}{12} \right) + \frac{1}{2} \sum_{q=1}^K \log \frac{N \epsilon_q}{12}. \end{aligned} \quad (14)$$

The model order and the respective parameters are given by

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \{ \mathcal{L}(\boldsymbol{\theta}(k), k), k = k_{min}, \dots, k_{max} \}. \quad (15)$$

Table 1: Parameters of the Dirichlet densities, their statistics, and its estimates for each region of the scene.

Region	A	B
Number of pixels	2 000	1 000
Parameters values	6, 9, 25	7, 8, 23
Mean ($\times 10^{-1}$)	1.50, 2.25, 5.25	1.84, 2.10, 6.05
Variance ($\times 10^{-3}$)	3.10, 4.30, 5.70	3.90, 4.30, 6.10
Parameters estimates	5.5, 9.1, 25.9	6.5, 6.9, 23.0

Since, we do not know the fractions \mathcal{S} , we implement the GEM algorithm described in the previous section where the objective function is as in (4) plus the MDL terms and the optimization with respect θ is replaced with the optimization with respect to $\theta(k)$ and k .

The resulting optimization algorithm is an iterative scheme that requires to compute the objective function and to estimate the parameters θ for each value of k .

4. EVALUATION WITH SIMULATED DATA

In this section the proposed method is tested on simulated scenes. The data is generated according to expression (1), where three signatures were selected from the USGS digital spectral library. The scene is composed by 3×10^4 pixels partitioned into two regions; The size of region A and B is 66% and 33%, respectively. The abundance fractions follow a Dirichlet distribution with $\theta_A = [6, 9, 25]$ and $\theta_B = [7, 8, 23]$ for regions A and B of the scene, respectively. Table 1 presents these parameters and their statistics, e.g., mean and variance, for each region. Notice that for this setting, the spectral vectors are highly mixed. Fig. 1 presents a scatterplot of the simulated scene, where dots represent the observed spectral vectors and stars represent the true endmembers. The two clouds corresponds 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 $q = 1, \dots, k$. This setting reflects a situation in which no knowledge of the size and the number of regions in the scene exists.

For this data set the minimum of the cost function (14) is reached when the estimated number of modes is $\hat{k} = 2$ which is the number of regions on the scene. The estimated parameters at this iteration are presented on the last line of Table 1. Note that the estimated values are close to the true parameter values. The Dirichlet mixing probabilities (ϵ_q) values are 0.664 and 0.336, which correspond to the area in the scene occupied by region A and B respectively.

Fig. 1 presents the true endmembers (stars), the endmembers estimation by the proposed method (circles), and for comparison purposes the VCA endmembers estimation (tri-

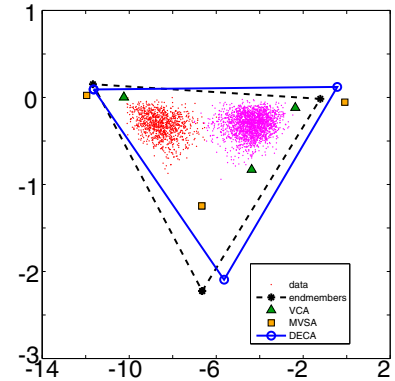


Fig. 1: Scatterplot of the three endmembers mixture: true endmembers (stars); MVSA estimate (squares); VCA estimate (triangles); Proposed method (circles).

angles) and the MVSA estimation (squares). Since the endmembers are highly mixed the proposed method gives better estimates than the others methods. Note that the endmember estimates are closer to the true endmembers, whereas, estimates provided by the MVSA algorithm fits the minimum volume simplex to the data set and VCA finds the most pure pixels in data.

5. 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 set¹ is considered. This site has been extensively used for remote sensing experiments over the past years and its geology was previously mapped in detail [17].

The data set is composed of 187 spectral bands with 10nm bandwidth acquired in the $0.4 - 2.5 \mu\text{m}$ region (noisy and water absorption bands were removed). The subset contains 50×90 pixels with a ground resolution of 17 meters.

Fig. 2 presents a scatterplot of the Cuprite data set, where dots represent the pixels and circles represent the endmembers estimated by the proposed method. For comparison purposes it is also plotted the Montmorillonite (square), Desert Varnish (Diamond), and Alunite (triangle) signature projections, extracted from the USGS library.

Fig. 3 presents the endmembers signatures and their estimates and Fig. 4 presents the estimated abundance maps. A visual comparison between these maps and the ground truth presented in [17] shows that first, second, and third extracted endmembers are predominantly Montmorillonite, Desert Varnish, and Alunite, respectively (see Fig. 4 (a)-(c)).

¹ Available at <http://aviris.jpl.nasa.gov/html/aviris.freedata.html>

6. CONCLUSIONS

In this paper, an elaboration of DECA [1] algorithm is proposed to blindly unmix highly mixed hyperspectral data. Modeling the abundance fractions statistical dependence is a central issue in this framework. Abundance fractions are modeled as mixtures of Dirichlet densities, enforcing the so-called nonnegative and constant sum constraint.

The method herein presented introduces two improvements on DECA: 1) the number of Dirichlet modes are inferred based on the minimum description length (MDL) principle; 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.

The experimental results achieved shows the potential of the proposed method to unmix hyperspectral data when the pure pixel assumption is not fulfilled.

7. REFERENCES

- [1] J. M. P. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing algorithm via dependent component analysis," in *Proc. of the IEEE IGARSS, 2007*, pp. 4033–4036.
- [2] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Processing Mag.*, vol. 19, no. 1, pp. 44–57, 2002.
- [3] 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, 2005.
- [4] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/spectral end-member extraction by multidimensional morphological operations," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2025–2041, 2002.
- [5] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Sum. of the IV Annual JPL Airborne Geosc. Workshop, JPL Pub. 93-26*, vol. 1, 1993, pp. 11–14.
- [6] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. of the SPIE*, vol. 3753, 1999, pp. 266–275.
- [7] M. Berman, H. Kiveri, 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, 2004.
- [8] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data," in *Proc. of the IEEE IGARSS*, vol. 3, 2008, pp. 250–253.
- [9] N. Dobigeon, J.-Y. Tourneret, and C.-I. Chang, "Semi-supervised linear spectral unmixing using a hierarchical bayesian model for hyperspectral imagery," *IEEE Trans. Sig. Proc.*, vol. 56, no. 7, pp. 2684–2695, 2008.
- [10] 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, 2005.
- [11] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE PAMI.*, vol. 44, no. 3, pp. 381–396, 2002.
- [12] J. M. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *First IEEE WHISPERS*, 2009.
- [13] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, 2008.
- [14] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. John Wiley & Sons, Inc., 1996.
- [15] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.
- [16] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
- [17] G. Swayze, R. Clark, S. Sutley, and A. Gallagher, "Ground-truthing aviris mineral mapping at cuprite, nevada,," in *Sum. of the III Annual JPL Airborne Geosc. Workshop*, 1992, pp. 47–49.

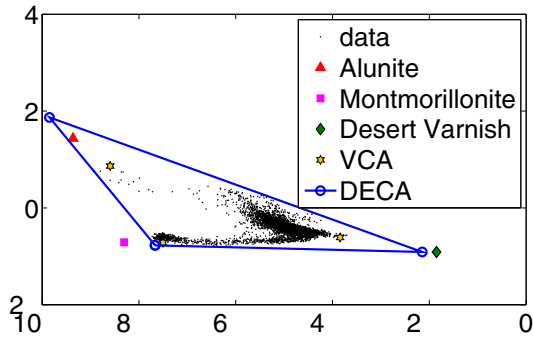


Fig. 2: Scatterplot of Cuprite data set (dots); estimated endmembers (circles); USGS signatures: Montmorillonite (square); Desert Varnish (Diamond), and Alunite (triangle).

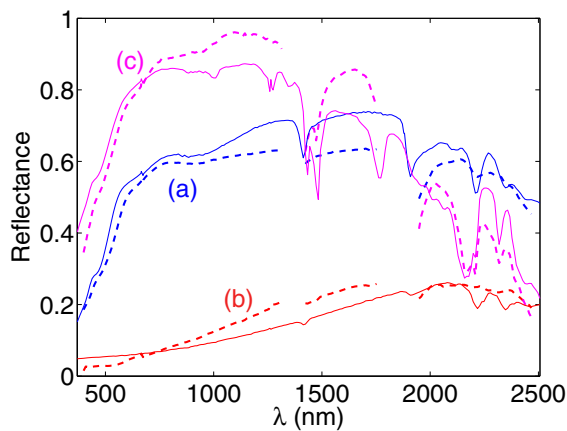


Fig. 3: Endmembers signatures (solid line) and their estimates (dashed line): (a) Montmorillonite; (b) Desert Varnish ; (c) Alunite.

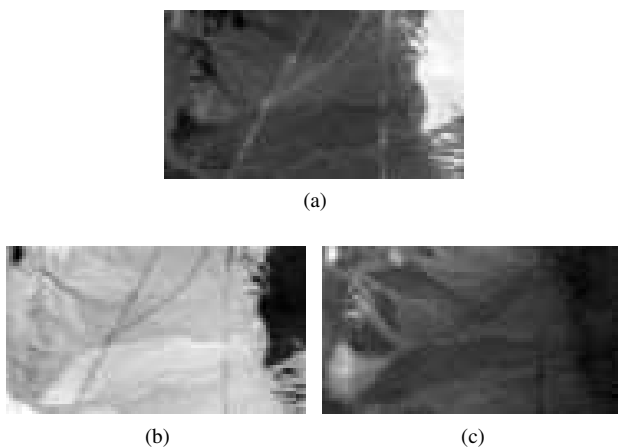


Fig. 4: Abundance maps estimates: (a) Montmorillonite; (b) Desert Varnish ; (c) Alunite.