Hyperspectral Unmixing Algorithm via Dependent Component Analysis

José M. P. Nascimento
Instituto Superior de Engenharia de Lisboa
and Instituto de Telecomunicações
R. Conselheiro Emídio Navarro,
N. 1, edifício DEETC,
1959-007 Lisboa, Portugal
Tel.:+351.21.8317237, Fax:+351.21.8317114,
Email: zen@isel.pt

José M. Bioucas-Dias
Instituto de Telecomunicações
Instituto Superior Técnico
Technical University of Lisbon
Av. Rovisco Pais, Torre Norte, Piso 10
1049-001 Lisboa, Portugal
Tel.: +351.21.8418466, Fax:+351.21.8418472,
Email: bioucas@lx.it.pt

***Abstract*—This paper introduces a new method to blindly unmix hyperspectral data, termed *dependent component analysis* (DECA). This method decomposes a hyperspectral images into a collection of reflectance (or radiance) spectra of the materials present in the scene (endmember signatures) and the corresponding abundance fractions at each pixel.**

DECA assumes that each pixel is a linear mixture of the endmembers signatures weighted by the correspondent abundance fractions. These abudances 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. The mixing matrix is inferred by a generalized expectation-maximization (GEM) type algorithm. This method overcomes the limitations of unmixing methods based on Independent Component Analysis (ICA) and on geometrical based approaches. The effectiveness of the proposed method is illustrated using simulated data based on U.S.G.S. laboratory spectra and real hyperspectral data collected by the AVIRIS sensor over Cuprite, Nevada.

I. 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 $0.3\mu m$ and $2.5\mu m$), in hundreds of narrow (on the order of $10nm$) 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. A commonly approach to analyze hyperspectral data is the linear spectral unmixing, which considers that 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 [1].

Several approaches such as *vertex component analysis* (VCA), [2], *pixel purity index* (PPI), [3], and N-FINDR [4] have exploited geometric features of hyperspectral mixtures to determine the smallest simplex containing the data. Those methods assume the presence in the data of at least one pure pixel of each endmember. This is a strong requisite that may not hold in some data sets.

Iterative constrained endmembers (ICE) [5] combines a model based on convex geometry with statistical procedures about the errors to obtain an objective function. The minimization of this function provides the endmember's signatures without the assumption of pure-pixels in the data. ICE is a fast method but it is not fully automated since an estimation of three initial parameters is needed.

Independent Component Analysis (ICA) has recently been proposed as a tool to blindly unmix hyperspectral data (see ref. in [6]). However, ICA applicability is compromised by the statistical dependence existing among abundances. This dependence results from the constant sum constraint imposed on the abundance fractions by the acquisition process. In ICA jargon, sources are not independent. Thus, the central assumption of ICA is not satisfied [6].

The separation of sources that show some degree of dependence was attempted with MaxNG [7] and uMaxEnt [8] methods. The former method use a non-Gaussianity measure based on the L^2-Euclidian distance between probability distributions, whereas, the latter method is based on the maximum entropy principle.

This paper proposes a new method to blindly unmix hyperspectral data, termed *dependent component analysis* (DECA), where abundance fractions are modelled by a mixture of Dirichlet densities (MOD), thus automatically enforcing source nonnegativity and constant sum constraints. The mixing matrix is inferred by a generalized expectation-maximization (GEM) type algorithm. DECA processes simultaneously the endmembers signatures and the abundance fraction estimation. Compared with the geometric based approaches, the advantage of DECA is that there is no need to have pure pixels in the observations. The paper is organized as follows.

Section II describes the fundamentals of the proposed method. Sections III and IV illustrate aspects of the performance of DECA approach with experimental data based on U.S.G.S. laboratory spectra and real hyperspectral data collected by the AVIRIS sensor over Cuprite Nevada, respectively. Section V ends the paper by presenting a few concluding remarks.

II. Statistical Modelling and Unmixing Algorithm

Assuming the linear observation model, each pixel $\mathbf{r}$ 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 $\mathbf{r} = \mathbf{M}\mathbf{s}$, where $\mathbf{M} \equiv [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p]$ is an $L \times p$ mixing matrix ($\mathbf{m}_i$ denotes the ith endmember signature), p is the number of endmembers present in the covered area, and $\mathbf{s} = [s_1, s_2, \ldots, s_p]^T$ 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 ($p \ll L$). Thus, the observed spectral vectors can be projected onto the signal subspace. The identification of the signal subspace improves the SNR, allows a correct dimension reduction, and thus yields gains in computational time and complexity [9]. Let $\mathbf{E}_p$ be a matrix, with orthonormal columns, spanning the signal subspace. The coordinates of the spectral vector $\mathbf{r}$ with respect to $\mathbf{E}_p$ are

$$\begin{aligned} \mathbf{x} &\equiv \mathbf{E}_p^T \mathbf{r} \\ &= \mathbf{E}_p^T \mathbf{M}\mathbf{s} \\ &= \mathbf{A}\mathbf{s}, \end{aligned} \tag{1}$$

where $\mathbf{A}$ is a $p \times p$ square mixing matrix and $\mathbf{x} = [x_1, x_2, \ldots, x_p]^T$ is a $p \times 1$ vector. Let's assume that $\mathbf{W} \equiv \mathbf{A}^{-1}$ exists. Then, we have $\mathbf{s} = \mathbf{W}\mathbf{x}$.

Assume that the abundance fractions follow a K-component Dirichlet finite mixture given by

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

where the complete set of parameters $\boldsymbol{\theta}$ needed to specify the mixture contains the mixing probabilities $\epsilon_1, \ldots, \epsilon_K$ and the q-component Dirichlet parameters $\boldsymbol{\theta}_q = \{\theta_{q1}, \ldots, \theta_{qp}\}$, for $q = 1, \ldots, K$, *i.e.*, $\boldsymbol{\theta} = \{\epsilon_1, \ldots, \epsilon_K, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K\}$.

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 i.i.d. samples $\mathcal{X} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\}$, then, we may write the likelihood of the unmixing matrix $\mathbf{W}$ and of the set of parameters $\boldsymbol{\theta}$ as

$$\begin{aligned} \mathcal{L}_N(\mathbf{W}, \boldsymbol{\theta}) &\equiv \frac{1}{N} \log p_X(\mathcal{X}|\mathbf{W}, \boldsymbol{\theta}) \\ &= \mathbb{T}\left[\log p_X(\mathbf{x}|\mathbf{W}, \boldsymbol{\theta})\right] \\ &= \mathbb{T}\left[\log p_S(\mathbf{s}|\boldsymbol{\theta})\right] + \log|\det \mathbf{W}| \\ &= \mathbb{T}\left[\log \sum_{q=1}^{K} \epsilon_q D(\mathbf{s}|\boldsymbol{\theta}_q)\right] + \log\left(|\det \mathbf{W}|\right), \end{aligned} \tag{3}$$

where we have used the fact that $p_X(\mathbf{x}) = p_S(\mathbf{s})|\det(\mathbf{W})|$ and $\mathbb{T}[\mathbf{x}] \equiv 1/N \sum_{i=1}^{N} \mathbf{x}^{(i)}$, *i.e.*, $\mathbb{T}[\mathbf{x}]$ is the sample average of $\mathbf{x}$.

The maximum likelihood (ML) estimate $\left(\widehat{\mathbf{W}}, \widehat{\boldsymbol{\theta}}\right) = \arg\max_{\mathbf{W}, \boldsymbol{\theta}} \mathcal{L}_N(\mathbf{W}, \boldsymbol{\theta})$ can not be found analytically [10]. The usual choice for obtaining the ML estimates of the parameters is the EM framework [11], which relies on the so-called incomplete data and missing data. In our setup, $\mathcal{X}$ denotes the incomplete data and $\mathcal{Z} = \left\{\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(N)}\right\}$ the missing data, a set of N labels indicating which component has produced each sample. Each label $\mathbf{z}^{(i)} = [z_1^{(i)}, \ldots, z_K^{(i)}]$ 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

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

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 $\widehat{\theta}^{(t)}$. The result is the so-called Q-function

$$\begin{aligned} Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)}) &= \mathbb{T}\left[\sum_{q=1}^{K} \beta_q^{(t)} \log\left[\epsilon_q^{(t)} D\left(\mathbf{s}|\boldsymbol{\theta}_q^{(t)}\right)\right]\right] \\ &\quad + \log\left(|\det \mathbf{W}|\right), \end{aligned} \tag{5}$$

  where

$$\beta_q^{(t)}(\mathbf{s}) = \widehat{\epsilon}_q^{(t)} D\left(\mathbf{s}|\widehat{\boldsymbol{\theta}}_q^{(t)}\right) \Big/ \sum_{l=1}^{K} \widehat{\epsilon}_l^{(t)} D\left(\mathbf{s}|\widehat{\boldsymbol{\theta}}_l^{(t)}\right). \tag{6}$$

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

$$\widehat{\boldsymbol{\theta}}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} \left\{Q\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)}\right)\right\}. \tag{7}$$

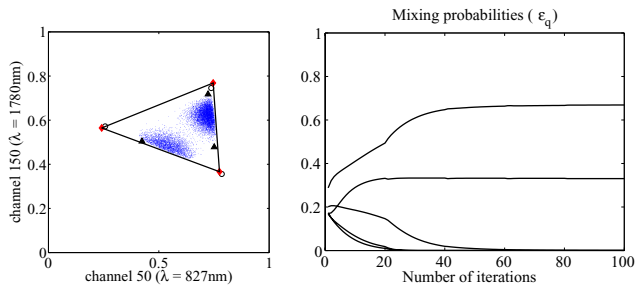
Maximization of expression (7) is still a hard optimization problem. Instead of computing $\boldsymbol{\theta}^{(t+1)}$, we maximize $Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)})$ with respect to θ_j, for $j = 1, \ldots, p$, resulting the following learning rules for the mixing probabilities and for the mixture of Dirichlet source parameters:

$$\epsilon_q^{(t)} = \mathbb{T}\left[\beta_q^{(t)}(\mathbf{s})\right], \tag{8}$$

$$\widehat{\theta}_{qj}^{(t+1)} = \Psi^{-1}\left(\Psi\left(\sum_{l=1}^{p} \widehat{\theta}_{ql}^{(t)}\right) + \frac{\mathbb{T}\left[\beta_q^{(t)}(\mathbf{s}) \log \widehat{s}_j^{(t)}\right]}{\mathbb{T}\left[\beta_q^{(t)}(\mathbf{s})\right]}\right), \tag{9}$$

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) [11]: the learning rule (8) maximizes Q-function with respect to $\epsilon_q^{(t)}$, whereas expression (9)



(a) (b)

Fig. 1: (a) Scatterplot (bands $\lambda = 827nm$ and $\lambda = 1780nm$) of the three endmembers mixture: true endmembers (circles); VCA estimate (triangles); DECA estimate (diamonds); (b) Dirichlet mixing probabilities.

assures that the Q-function does not decrease (see [12] for details).

Since $\partial Q/\partial \mathbf{W} = 0$ is not a linear equation and cannot be solved analytically, an iterative gradient type learning rule is derived for the unmixing matrix $\mathbf{W}$:

$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} + \tau^{(t)} \left(\frac{\partial Q}{\partial \mathbf{W}} \right)^{(t)}, \tag{10}$$

where $\tau^{(t)}$ determines the learning rate on iteration t and

$$\begin{aligned} \left(\frac{\partial Q}{\partial \mathbf{w}_j} \right)^{(t)} &= \mathbb{T} \left[\sum_{q=1}^{K} \beta_q^{(t)} \left(\frac{(\widehat{\theta}_{qj}^{(t)} - 1)}{\widehat{s}_j} - \frac{(\widehat{\theta}_{qp}^{(t)} - 1)}{\widehat{s}_p} \right) \mathbf{x}^T \right] \\ &\quad + \left[\mathbf{W}^{-T} \right]_{j,:} - \left[\mathbf{W}^{-T} \right]_{p,:}, \end{aligned} \tag{11}$$

where $\mathbf{w}_j$, for $j = 1, \ldots, p-1$ denotes the jth row of matrix $\mathbf{W}$ and $\left[\mathbf{W}^{-T} \right]_{j,:}$ denotes the jth row of the inverse of $\mathbf{W}$ transposed.

The algorithm aimed at the maximization of the log-likelihood (4) implements a cyclic maximizer, which splits the estimation of $\mathbf{W}$ and $\boldsymbol{\theta}$ into block maximization operations. An approach based on the majorization maximization (MM) [13] perspective leading to a similar algorithm can be found in [14].

III. Evaluation with Simulated Data

In this section DECA is tested on simulated scenes. The data is generated according to expression (1), where three signatures where selected from the USGS digital spectral library. The scene is composed by 10^5 pixels partitioned into two regions; region A has the half size of the region B. The abundance fractions follow a Dirichlet distribution with $\boldsymbol{\theta}_A = [9,\ 2,\ 9]$ and $\boldsymbol{\theta}_B = [2,\ 15,\ 7]$ for regions A and B of the scene, respectively. Pure pixels were removed from the data set in order to illustrate the robustness of DECA to the absence of pure pixels.

In this experiment the number of modes is set to $K = 5$, 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. Fig. 1a presents a scatterplot (bands $\lambda = 827nm$ and $\lambda = 1780nm$) of the simulated scene, where dots represent the pixels. The two clouds corresponds to the two regions. It is also presented the true endmembers (circles), the endmembers estimation (diamonds), and for comparison purposes the endmembers estimation by VCA (triangles). Estimates provided by the DECA algorithm are close to the true endmembers. The algorithm searches for the smallest simplex that contains all data, whereas VCA finds the most pure pixels in data (see triangles in Fig. 1a). Since there is no pure pixels in data, VCA performs worse than DECA.

Fig 1b), presents the evolution of the Dirichlet mixing probabilities (ϵ_q, for $q = 1, \ldots, K$) as a function of the number of iterations of the algorithm. Note that three modes tend to zero and the remaining modes have the values of 0.65 and 0.33, corresponding to the weight of the region B and region A respectively. The Dirichlet parameters estimate for the two modes are $\widehat{\boldsymbol{\theta}}_A = [9.0,\ 2.2,\ 10.0]$ and $\widehat{\boldsymbol{\theta}}_B = [2.5,\ 14.8,\ 9.7]$ for region A and B, respectively. Although the estimated values are close to the true parameter values ($\boldsymbol{\theta}_A = [9,\ 2,\ 9]$ and $\boldsymbol{\theta}_B = [2,\ 15,\ 7]$), we note that this does not have to happen necessarily, since the same distribution can be modelled with different MODs. We stress that the main purpose of the DECA algorithm is the estimation of the unmixing matrix $\mathbf{W}$ and not of the MOD parameters.

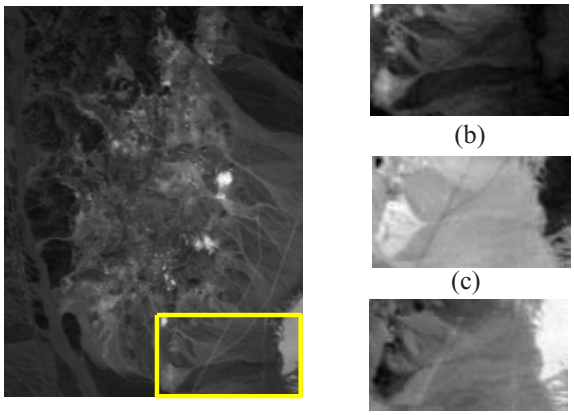
The result of the separation process is illustrated trough the product of the unmixing matrix $\mathbf{W}$ and square mixing matrix $\mathbf{A}$ which is, in an ideal scenario, the identity matrix $\mathbf{I}_p$, apart from a permutation. In this experiment the obtained product is

$$\mathbf{WA} = \begin{bmatrix} \mathbf{0.97} & 0.02 & -0.02 \\ 0.03 & \mathbf{0.93} & -0.02 \\ 0.00 & 0.04 & \mathbf{1.03} \end{bmatrix}. \tag{12}$$

IV. Experiments with Real Hyperspectral Data

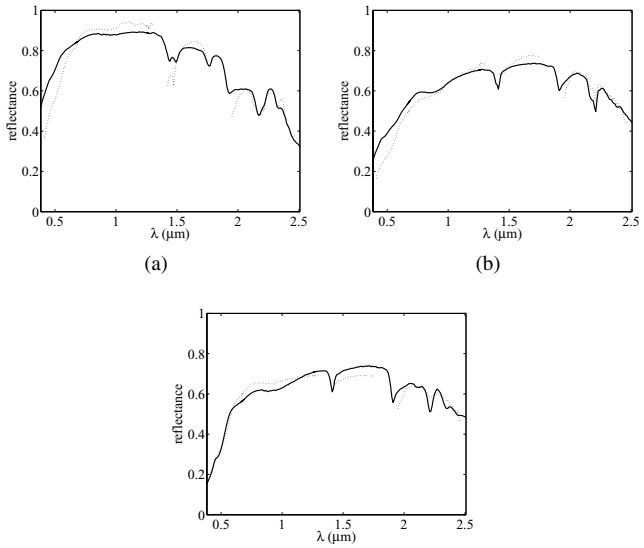
In this section, the proposed method, DECA, is applied to real hyperspectral data collected by the AVIRIS sensor over Cuprite, Nevada. This site has been extensively used for remote sensing experiments over the past years and its geology was previously mapped in detail [15]. This site has become a standard test site for comparison of unmixing and endmember extraction algorithms.

Fig. 2 (a) presents the subimage (50×90 pixels and 224 bands) for this experiment. Due to several degradation mechanisms normally found in hyperspectral applications (namely, signature variability, topography modulation, and noise), the observed data is not in a simplex. To obtain a simplex, a projective projection of data onto a hyperplane $\mathbf{y}^T\mathbf{u} = 1$ is implemented as a pre-processing step (see [2] for more details). A visual comparison between the abundance fractions estimates on the cuprite data set and the ground truth presented in [15] shows that first, second, and third extracted endmembers are predominantly Alunite, Kaolinite, and Montmorillonite, respectively (see Fig. 2 (b)-(d)).



(a) (b) (c) (d)

Fig. 2: (a) Band 30 (wavelength $\lambda = 667.3nm$) of the subimage of AVIRIS cuprite Nevada data set (rectangle denotes the image fraction used in the experiment);(b)-(d) Alunite, Kaolinite, and Montmorillonite abundance fractions;



(a) (b) (c)

Fig. 3: Comparison of the DECA estimated signatures (dotted line) with the U.S.G.S spectral library (solid line): (a) Alunite; (b) Kaolinite; (c) Montmorillonite.

A comparison of the estimated endmember signatures with laboratory spectrum is presented in Fig. 3. The signatures provided by DECA are scaled in order to minimize the mean square error between them and the respective library spectra. The estimated signatures are very close to the laboratory spectra reflectances.

V. Conclusions

Blind hyperspectral linear unmixing aims at estimating the number of endmembers, their spectral signatures, and their abundance fractions at each pixel, using only the observed data (mixed pixels). Geometric approaches have been used whenever pure pixels are present in data. In most cases, however, pure pixels can not be found in data. In such cases, unmixing procedures become a difficult task. ICA has been proposed has a tool to unmix hyperspectral data. However, the source dependence present in hyperspectral data compromises the unmixing results.

In this paper, a new method is proposed to blindly unmix hyperspectral data, where abundance fractions are modelled as Dirichlet sources. This model forces abundance fractions to be nonnegative and to have constant sum on each pixel. The mixing matrix is inferred by an EM type algorithm. The main advantage of this model is that there is no need to have pure pixels in the observations.

The performance of the proposed model is illustrated with simulated and real hyperspectral data. Comparisons with pure pixel estimation methods are conducted. The results achieved shows the effectiveness of DECA on hyperspectral data unmixing. In future work, the proposed algorithm shall be improved in order to account for sensor noise.

Acknowledgements

This work was supported by the FCT, under the projects POSC/EEA-CPS/61271/2004 and PDCTE/CPS/49967/2003 and by IPL under project IPL-5828/2004.

References

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

[2] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.

[3] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *JPL Pub. 93-26, AVIRIS Workshop.*, vol. 1, 1993, pp. 11–14.

[4] M. E. Winter, "N-findr: an algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. of the SPIE conference on Imaging Spectrometry V*, vol. 3753, 1999, pp. 266–275.

[5] 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 Sensing*, vol. 42, no. 10, pp. 2085– 2095, 2004.

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

[7] C. Caiafa and A. Proto, "Separation of statistically dependent sources using an L^2-distance non-gaussianity measure," *Signal Processing (EURASIP)*, vol. 86, no. 11, pp. 3404–3420, 2006.

[8] L. Miao, H. Qi, and H. Szu, "Unsupervised decomposition of mixed pixels using the maximum entropy principle," in *Proc. of the 18th International Conference on Pattern Recognition*, vol. 1, 2006, pp. 1067–1070.

[9] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sensing*, 2007, submmited.

[10] G. McLachlan and D. Peel, *Finite Mixture Models*. John Wiley & Sons, Inc., 2000.

[11] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. John Wiley & Sons, Inc., 1996.

[12] J. M. P. Nascimento and J. M. Bioucas-Dias, *Hyperspectral Data Exploitation: Theory and Applications*. John Wiley & Sons, Inc., 2007, ch. Unmixing Hyperspectral Data: Independent and Dependent Component Analysis.

[13] Optimization, *Kenneth Lange*, 1st ed. Springer, 2004.

[14] T. Minka, "Estimating a dirichlet distribution," M.I.T., Tech. Rep., 2000.

[15] G. Swayze, R. Clark, S. Sutley, and A. Gallagher, "Ground-truthing aviris mineral mapping at cuprite, nevada,," in *JPL Pub. 92, AVIRIS Workshop.*, 1992, pp. 47–49.