Dependent Component Analysis: A Hyperspectral Unmixing Algorithm

José M. P. Nascimento[1] and José M. B. Dias[2]

[1] Instituto Superior de Engenharia de Lisboa and Instituto de Telecomunicações,
R. Conselheiro Emídio Navarro N 1, edifício DEETC, 1950-062 Lisboa Portugal
Tel.:+351.21.8317237, Fax:+351.21.8317114,
zen@isel.pt
[2] Instituto Superior Técnico and Instituto de Telecomunicações,
Av. Rovisco Pais, Torre Norte, Piso 10, 1049-001 Lisboa Portugal
Tel.: +351.21.8418466, Fax: +351.21.841472,
bioucas@lx.it.pt
⋆

Abstract. Linear unmixing decomposes a hyperspectral image into a collection of reflectance spectra of the materials present in the scene, called endmember signatures, and the corresponding abundance fractions at each pixel in a spatial area of interest.
This paper introduces a new unmixing method, called Dependent Component Analysis (DECA), which overcomes the limitations of unmixing methods based on Independent Component Analysis (ICA) and on geometrical properties of hyperspectral data.
DECA models the abundance fractions 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. The performance of the method is illustrated using simulated and real data.

1 Introduction

Spaceborn and airborne hyperspectral sensors acquire images of ground surface radiances in hundreds of narrow contiguous bands simultaneously [1]. The radiances, collected in a spectral vector, are mixtures of spectra from the substances present in the respective pixel coverage.

Given a set of mixed hyperspectral vectors, linear unmixing aims at estimating the number of reference substances, also called endmembers, their spectral signatures, and their abundance fractions.

Linear spectral unmixing considers that a mixed pixel is a linear combination of endmember signatures weighted by the correspondent abundance fractions [2]. Under this model, the observations from a scene are in a simplex whose vertices correspond to the endmembers.

⋆ 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.

Several approaches such as *vertex component analysis* (VCA) [3], *pixel purity index* (PPI) [4], and N-FINDR [5] 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.

Independent Component Analysis (ICA) has recently been proposed as a tool to blindly unmix hyperspectral data [6–8]. However, ICA applicability is compromised by the statistical dependence existing among abundances [9]. This dependence results from the constant sum constraint imposed on the abundance fractions by the acquisition process [2]. In ICA jargon, sources are not independent. Thus, the central assumption of ICA is not satisfied.

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, thus automatically enforcing source nonnegativity and constant sum constraints. DECA is in the vein of works [10, 11] replacing independent sources represented by Mixtures of Gaussians (MOGs) with mixtures of Dirichlet (MODs) sources. The mixing matrix is inferred by a generalized expectation-maximization (GEM) type algorithm. 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 2 describes the fundamentals of the proposed method. Sections 3 and 4 illustrate aspects of the performance of DECA approach with experimental data and real data, respectively. Section 5 ends the paper by presenting a few concluding remarks.

2 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 a $L \times p$ 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 [2], 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 [12]. 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 $\mathbf{s} = \mathbf{W}\mathbf{x}$.

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 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

$$\begin{aligned}\mathcal{L}_N(\mathbf{W}, \boldsymbol{\theta}) &\equiv \frac{1}{N} \log\, p_X(\mathcal{X}|\mathbf{W}, \boldsymbol{\theta}) \\ &= \mathbb{T}\left[\log p_X(\mathbf{x}|\mathbf{W}, \boldsymbol{\theta})\right] \\ &= \mathbb{T}\left[\log p_S(\mathbf{s}|\boldsymbol{\theta})\right] + \log|\det \mathbf{W}|, \end{aligned} \tag{2}$$

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 $\mathbf{x}$.

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

$$p_S(\mathbf{s}|\boldsymbol{\theta}) = \sum_{q=1}^{K} \epsilon_q \underbrace{\frac{\Gamma(\sum_{j=1}^{p} \theta_{qj})}{\prod_{j=1}^{p} \Gamma(\theta_{qj})} \prod_{j=1}^{p} s_j^{\theta_{qj}-1}}_{D(\mathbf{s}|\boldsymbol{\theta}_q)}, \tag{3}$$

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.*, $\boldsymbol{\theta} = \{\epsilon_1, \ldots, \epsilon_K, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K\}$. Replacing expression (3) in to (2), it follows that

$$\mathcal{L}_N(\mathbf{W}, \boldsymbol{\theta}) = \mathbb{T}\left[\log \sum_{q=1}^{K} \epsilon_q D(\mathbf{s}|\boldsymbol{\theta}_q)\right] + \log(|\det \mathbf{W}|). \tag{4}$$

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 [13]. 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, $\mathcal{X}$ 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 $\mathbf{z}^{(i)} = \left[z_1^{(i)}, \ldots, z_K^{(i)}\right]$ 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

$$\begin{aligned}\mathcal{L}_C(\mathbf{W}, \boldsymbol{\theta}) &= \frac{1}{N} \log\left[p_{X,Z}(\mathcal{X}, \mathcal{Z}|\boldsymbol{\theta})\right] \\ &= \mathbb{T}\left[\sum_{q=1}^{K} z_q \log \epsilon_q D(\mathbf{s}|\boldsymbol{\theta}_q)\right] + \log\left(|\det \mathbf{W}|\right). \end{aligned} \tag{5}$$

The EM algorithm iterates between the E-step and the M-step [14, 15]:

– **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

$$Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)}) = \mathbb{T}\left[\sum_{q=1}^{K} \underbrace{\mathrm{E}\left[z_q | \, \mathbf{s}, \widehat{\boldsymbol{\theta}}^{(t)}\right]}_{\beta_q^{(t)}} \log\left[\epsilon_q^{(t)} D\left(\mathbf{s}|\boldsymbol{\theta}_q^{(t)}\right)\right]\right] + \log\left(|\det \mathbf{W}|\right), \quad (6)$$

where

$$\beta_q^{(t)}(\mathbf{s}) = \frac{\widehat{\epsilon}_q^{(t)} D\left(\mathbf{s}|\widehat{\boldsymbol{\theta}}_q^{(t)}\right)}{\sum_{l=1}^{K} \widehat{\epsilon}_l^{(t)} D\left(\mathbf{s}|\widehat{\boldsymbol{\theta}}_l^{(t)}\right)}. \quad (7)$$

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

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

Maximization of expression (8) 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 in the following learning rules for the mixing probabilities and for the mixture of Dirichlet source parameters:

$$\epsilon_q^{(t)} = \mathbb{T}\left[\beta_q^{(t)}(\mathbf{s})\right], \quad (9)$$

$$\widehat{\theta}_{qj}^{(t+1)} = \mathrm{psi}^{-1}\left(\mathrm{psi}\left(\sum_{l=1}^{p} \widehat{\theta}_{ql}^{(t)}\right) + \frac{\mathbb{T}\left[\beta_q^{(t)}(\mathbf{s}) \log \widehat{s}_j^{(t)}\right]}{\mathbb{T}\left[\beta_q^{(t)}(\mathbf{s})\right]}\right), \quad (10)$$

for $q = 1 \ldots, K$ and $j = 1 \ldots, p$, respectively[3].
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 (see [16, 17] 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 $\mathbf{W}$:

$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} + \tau^{(t)} \left(\frac{\partial Q}{\partial \mathbf{W}}\right)^{(t)}, \quad (11)$$

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

$$\left(\frac{\partial Q}{\partial \mathbf{w}_j}\right)^{(t)} = \mathbb{T}\left[\sum_{q=1}^{K} \beta_q^{(t)} \left(\frac{(\widehat{\theta}_{qj}^{(t)} - 1)}{\widehat{s}_j} - \frac{(\widehat{\theta}_{qp}^{(t)} - 1)}{\widehat{s}_p}\right) \mathbf{x}^T\right] + \left[\mathbf{W}^{-T}\right]_{j,:} - \left[\mathbf{W}^{-T}\right]_{p,:}, \quad (12)$$

[3] $\mathrm{psi}(\cdot)$ and $\mathrm{psi}^{-1}(\cdot)$ denote the psi function (logarithmic derivative of the Gamma function) and its inverse, respectively.

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 (5) implements a cyclic maximizer, which splits the estimation of $\mathbf{W}$ and $\boldsymbol{\theta}$ into block maximization operations. The resulting scheme is in the vein of works [10, 11], replacing independent sources represented by MOGs with mixture of Dirichlet sources. An approach based on the majorization maximization (MM) [15] perspective leading to a similar algorithm can be found in [18].

3 Evaluation with Simulated Data

In this section DECA is tested in 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.

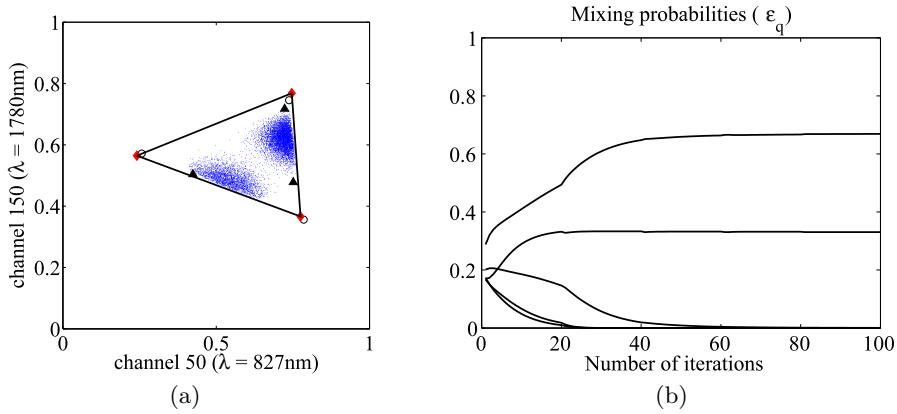

Fig. 1. (a) Scatterplot (bands $\lambda = 827nm$ and $\lambda = 1780nm$) of the three endmembers mixture: true endmembers (circles); VCA estimate (triangles); DECA estimate (diamonds); (b) Dirichlet mixing probabilities.

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. 1(a) 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 A and B, respectively. It is also presented the true endmembers (circles), the endmembers

estimation (diamonds), and, for comparison purposes the endmembers estimation by VCA (triangles). VCA is an iterative algorithm that searches for the vertices of the simplex containing the data. Briefly, VCA projects the spectral vectors onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection. This procedure is repeated until the number of endmembers is exhausted [3].

Estimates provided by the DECA algorithm are very 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. 1(a)). Since there is no pure pixels in data, VCA performs worse than DECA.

Fig 1(b), presents the evolution of the Dirichlet mixing probabilities (ϵ_q, for $q = 1, \ldots, K$) as 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. Table 1 presents the Dirichlet parameters and their estimates of the two modes. Although the estimated values are near from the true parameter values, we note that this does not have to happen necessarily, since the same distribution can be modelled with different MODs. We note that the main purpose of the DECA algorithm is the estimation of the unmixing matrix $\mathbf{W}$ and not the MOD parameters.

Table 1. Estimated Dirichlet parameters.

region	A			B		
$\boldsymbol{\theta}$	9	2	9	2	15	7
$\widehat{\boldsymbol{\theta}}$	9.0	2.2	10.0	2.5	14.8	9.7

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

$$\mathbf{WA} = \begin{bmatrix} \mathbf{0.97} & 0.02 & -0.02 \\ 0.03 & \mathbf{0.93} & -0.02 \\ 0.00 & 0.04 & \mathbf{1.03} \end{bmatrix}. \tag{13}$$

4 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[4]. This site has been extensively used for remote sensing experiments over the past years and its geology was previously mapped in detail [19]. This site has become a standard test site for comparison of unmixing and endmember extraction algorithms.

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

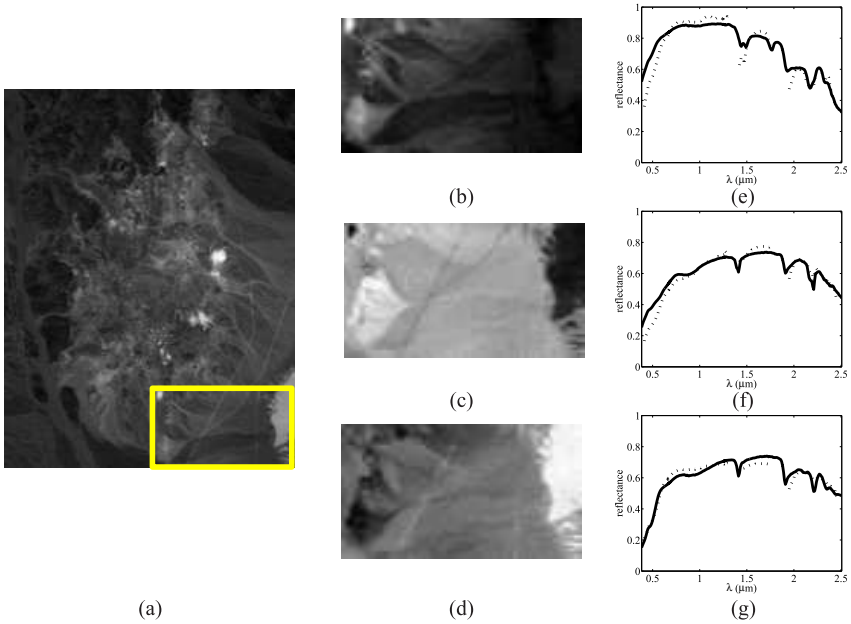

Fig. 2. (a) Band 30 (wavelength $\lambda = 667.3nm$) of the subimage of AVIRIS cuprite Nevada data set (rectangle denotes the image fraction used in the experiment);(b)-(d) Alunite, Kaolinite, and Montmorillonite abundance fractions; (e)-(g) Alunite, Kaolinite, and Montmorillonite spectra (solid line)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

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 on 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 [3] for more details). A visual comparison between the abundance fractions estimates on the cuprite data set and the ground truth presented in [19] shows that first, second, and third extracted endmembers are predominantly Alunite, Kaolinite, and Montmorillonite, respectively (see Fig. 2 (b)-(d)).

A comparison of the estimated endmember signatures with laboratory spectrum is presented in Fig. 2. 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.

5 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.

References

1. Lillesand, T., Kiefer, R., Chipman, J.: Remote Sensing and Image Interpretation. Fifth edn. John Wiley & Sons, Inc. (2004)
2. Keshava, N., Mustard, J.: Spectral unmixing. IEEE Sig. Proc. Mag. **19**(1) (2002) 44–57
3. Nascimento, J., Dias, J.: Vertex component analysis: A fast algorithm to unmix hyperspectral data. IEEE Trans. Geosci. Rem. Sens. **43**(4) (2005) 898–910
4. Boardman, J.: Automating spectral unmixing of AVIRIS data using convex geometry concepts. In: JPL Pub. 93-26, AVIRIS Workshop. Vol. 1. (1993) 11–14
5. Winter, M.: N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. In: Proc. SPIE. Vol. 3753. (1999) 266–275
6. Tu, T.: Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent component analysis approach. Opt. Eng. of SPIE **39**(4) (2000) 897–906
7. Lennon, M., Mouchot, M., Mercier, G., Hubert-Moy, L.: Spectral unmixing of hyperspectral images with the independent component analysis and wavelet packets. In: Proc. of IEEE IGARSS. (2001)
8. Parra, L., Mueller, K., Spence, C., Ziehe, A., Sajda, P.: Unmixing hyperspectral data. Advances in Neural Information Processing Systems **12** (2000) 942–948
9. Nascimento, J., Dias, J.: Does independent component analysis play a role in unmixing hyperspectral data? IEEE Trans. Geosci. Rem. Sens. **43**(1) (2005) 175–187
10. Attias, H.: Independent factor analysis. Neural Computation **11**(4) (1999) 803–851
11. Moulines, E., Cardoso, J., Gassiat, E.: Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In: Proc. of the IEEE ICASSP. Vol. 5. (1997) 3617–3620
12. Nascimento, J., Dias, J.: Signal subspace identification in hyperspectral linear mixtures. In Proc. of 2nd IbPRIA, Springer, LNCS Vol. 3523, (2005) 207–214
13. McLachlan, G., Peel, D.: Finite Mixture Models. John Wiley & Sons, Inc. (2000)
14. McLachlan, G., Krishnan, T.: The EM Algorithm and Extensions. John Wiley & Sons, Inc. (1996)
15. Optimization: Kenneth Lange. First edn. Springer (2004)
16. Dias, J.: An EM algorithm for the estimation of dirichlet parameters. Technical report, Instituto de Telecomunicações, http://www.lx.it.pt/~bioucas/ (2005)
17. Nascimento, J., Dias, J.: Unmixing Hyperspectral Data: Independent and Dependent Component Analysis. In: Hyperspectral Data Exploitation: Theory and Applications. John Wiley & Sons, Inc. (2006) in press.
18. Minka, T.: Estimating a dirichlet distribution. Technical report, M.I.T. (2000)
19. Swayze, G., Clark, R., Sutley, S., Gallagher, A.: Ground-truthing aviris mineral mapping at cuprite, nevada,. In: JPL Pub., AVIRIS Workshop. Vol. 1. (1992) 47–49