

An Investigation of Likelihoods and Priors for Bayesian Endmember Estimation

Alina Zare and Paul Gader

CISE Dept., University of Florida, Gainesville, FL 32611
(*azare@cise.ufl.edu, pgader@cise.ufl.edu*)

Abstract. A Gibbs sampler for piece-wise convex hyperspectral unmixing and endmember detection is presented. The standard linear mixing model used for hyperspectral unmixing assumes that hyperspectral data reside in a single convex region. However, hyperspectral data is often non-convex. Furthermore, in standard unmixing methods, endmembers are generally represented as a single point in the high dimensional space. However, the spectral signature for a material varies as a function of the inherent variability of the material or environmental conditions. Therefore, it is more appropriate to represent each endmember as a full distribution to incorporate the variability and utilize this information during spectral unmixing. A Gibbs sampler that searches for several sets of endmember distributions, i.e. a piece-wise convex representation, is presented. The hyperspectral data is partitioned among the sets of endmember distributions using a Dirichlet process prior that also estimates the number of needed sets. The proposed likelihood follows from a convex combination of normal endmember distributions with a Dirichlet prior on the abundance values. A normal distribution is also applied as a prior for the mean values of the endmember distributions.

The Gibbs sampler that is presented partitions the data into convex regions, determines the number of convex regions required and determines endmember distributions and abundance values for all convex regions. Results are presented on hyperspectral data that indicate the ability of the method to effectively estimate endmember distributions and the number of sets of endmember distributions.

Keywords: Hyperspectral, Unmixing, Endmember, Gibbs, Metropolis-in-Gibbs, Dirichlet Process

INTRODUCTION

Hyperspectral images are three-dimensional data cubes containing both spatial and spectral information about the image scene. A hyperspectral image can be viewed as a stack of two-dimensional images collected over a range of narrow, contiguous wavelengths. Given a hyperspectral image, the task of *endmember* detection is to estimate the “pure” spectral signatures of the materials found in a hyperspectral scene. Furthermore, the task of *spectral unmixing* determine the *proportion* (or *abundance*) of each endmember found in every hyperspectral pixel. The standard model applied during hyperspectral unmixing is the linear mixing model [1]. In the linear mixing model, the spectral signatures in a hyperspectral scene are modeled as convex combinations of the endmember signatures. This model can be written as follows.

$$\mathbf{x}_i = \sum_{k=1}^M p_{ik} \mathbf{e}_k + \boldsymbol{\varepsilon}_i \quad (1)$$

where N is the number of pixels, M is the number of endmembers, $\boldsymbol{\varepsilon}_i$ is an error term, p_{ik} is the abundance of endmember k in pixel i , and \mathbf{e}_k is the k^{th} endmember. The abundances

of this model satisfy the following constraints.

$$p_{ik} \geq 0 \quad \forall k = 1, \dots, M \quad (2)$$

$$\sum_{k=1}^M p_{ik} = 1 \quad (3)$$

Given the input hyperspectral data, \mathbf{X} , endmember detection estimates the spectral signatures of the endmembers, \mathbf{E} , and spectral unmixing determines the values of the proportions (or abundances) \mathbf{P} of each endmember in every hyperspectral data point. By following this model, finding endmembers amounts to estimating the spectral signatures whose convex hull enclose the hyperspectral data.

Several methods have been developed for endmember detection based on the linear mixing model. These include methods that rely on the pixel purity assumption and assume the endmembers can be found within the data set [2, 3, 4, 5]. Methods have also been developed based on Non-Negative Matrix Factorization [6, 7, 8, 9], Independent Components Analysis [10, 11] and others [12, 13, 14]. However, all of these methods search for a single set of endmembers and, therefore, a single convex region to describe a hyperspectral scene. Since these algorithms assume a single convex region, they cannot find appropriate endmembers for non-convex data sets.

When examining many hyperspectral images, it is found that they are often non-convex and groups of pixels in the image are convex combinations of subsets of the endmembers in the scene. Consider the image shown in Figure 1. This real hyperspectral data set is non-convex and would be better represented with a piece-wise convex representation of the data. By examining these non-convex hyperspectral images, endmembers may appear within the convex hull defined by the other endmembers in the scene. These *interior endmembers* cannot be recovered using methods based on the standard linear mixing model. However, methods based on a piece-wise convex representation are able to recover interior endmembers. The use of a piece-wise convex representation was first presented in [15] and [16]. The proposed method differs from the algorithms in [15] and [16] in that a fully stochastic endmember detection and spectral unmixing method is presented. Previous methods also rely on the Dirichlet Process for partitioning the data, however, these methods rely on a stochastic EM-type algorithm in which the partitioning of data points into endmember sets are sampled using a Dirichlet Process whereas endmember and proportions were estimated by maximizing an objective function. The proposed algorithm provides a fully stochastic extension of these methods by using a Gibbs sampling approach to sample all desired parameters. This proposed algorithm, Sampling Piece-wise Convex Endmember Detection (S-PCE), estimates several sets of endmembers, the abundances for each data point, and the number of endmembers sets needed to represent a hyperspectral image.

THE SAMPLING PIECE-WISE CONVEX ENDMEMBER DETECTION METHOD

The S-PCE method uses a Metropolis-within-Gibbs sampling technique to estimate sets of endmember distributions, abundance values for each data point, and the number of endmember distribution sets needed to represent the hyperspectral scene. The number

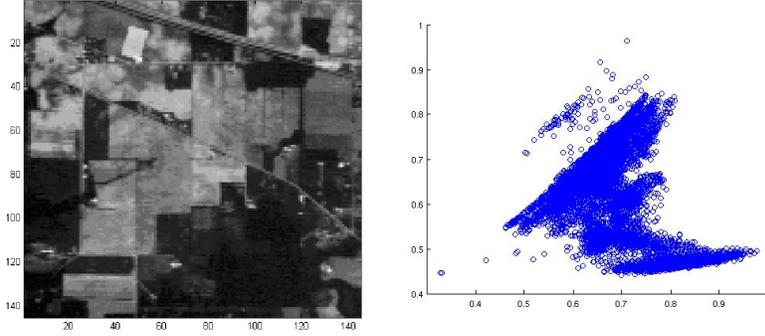


FIGURE 1. The June 1992 AVIRIS Indian Pines “Scene 4” data set [17]. These data were collected over the Indian Pines test site in an agricultural area of northern Indiana. The image has 145×145 pixels with 220 spectral bands. The data contains approximately two-thirds agricultural land and one-third forest and other elements [18]. The crops were at early growth stages and, thus, have approximately 5% crop cover with varying levels of residue from previous crops. (a) Figure showing band 10 (approximately $0.49 \mu\text{m}$) and the ground truth for this data set. (b) The AVIRIS Indian Pines hyperspectral data set after applying Maximum Noise Fraction dimensionality reduction to two dimensions [19]. This illustrates that the Indian Pines hyperspectral data set is not convex but, instead, appears to be piece-wise convex.

of endmember distribution sets is determined by applying a Dirichlet Process prior. Furthermore, hyperspectral endmembers are often represented as single points in a high dimensional space. However, the spectral signatures for a material varies within hyperspectral data collections due to environmental factors such as illumination or atmospheric effects as well as due to the inherent variability of a material. In order to represent this variability, rather than estimating a single point for each endmember, endmembers are represented as full distributions. In the current implementation, the endmembers are modeled using normal distributions with a fixed isotropic diagonal covariance, $\hat{\mathbf{e}}_{k,r} \sim N(\hat{\mathbf{e}}_{k,r} | \mathbf{e}_{k,r}, \mathbf{S}_{k,r})$ where $\mathbf{e}_{k,r}$ is the mean value for the k^{th} endmember distribution in the r^{th} endmember distribution set and $\mathbf{S}_{k,r}$ is the covariance for the k^{th} endmember distribution in the r^{th} partition. In the current implementation, all endmember distributions are given the same fixed isotropic diagonal covariance, \mathbf{S} .

Then, given the linear mixing model, each data point is a convex combination of these normally-distributed endmembers. This results in the following likelihood for a data point assigned to a single convex region

$$f(\mathbf{x}_j | z_j = r, \mathbf{E}_r, \mathbf{p}_j) \propto \exp \left\{ -\frac{1}{2} (\mathbf{x}_j - \mathbf{p}_j \mathbf{E}_r)^T \left(\sum_{k=1}^{M_r} p_{jk}^2 \mathbf{S}_{k,r} \right)^{-1} (\mathbf{x}_j - \mathbf{p}_j \mathbf{E}_r) \right\} \quad (4)$$

where \mathbf{x}_j is the j^{th} data point, z_j is the label indicating the partition to which the data point \mathbf{x}_j is assigned, r is the indicator variable for the r^{th} set of endmembers, \mathbf{E}_r is the r^{th} set of endmember means, and \mathbf{p}_j is the vector of proportion values associated with the j^{th} data point where p_{jk} is the k^{th} element of this proportion vector, and M_r is the number endmember distributions in the r^{th} partition. Therefore, given all sets of endmembers and all the data points in the scene which are assumed to be independent, the overall likelihood can be written as $\prod_{r=1}^R \prod_{j \in I_r} f(\mathbf{x}_j | z_j = r, \mathbf{E}_r, \mathbf{p}_j)$ where R is the

number of convex sets, $I_r = \{j|z_j = r\} \subset \{1, \dots, N\}$ denotes the set of indices of the data points that are assigned to the r^{th} convex set, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$, $\mathbf{z} = [z_1, \dots, z_N]$, and $\mathbf{E} = \{\mathbf{E}_1, \dots, \mathbf{E}_R\}$ where \mathbf{E} is the set of endmember mean matrices.

For each endmember distribution set, the mean of the endmember distributions is assumed to have a normal prior distribution.

$$\mathbf{e}_{k,r} \sim N(\mathbf{e}_{k,r} | \mu_r, \mathbf{C}_e) \quad (5)$$

where μ_r is the mean vector for the r^{th} set of endmembers and \mathbf{C}_e is the fixed covariance used in the prior for all endmember sets. Using this prior distribution on the mean of the endmembers distributions for partition encourages the endmember distributions for each partition to have a smaller enclosed volume. In other words, the mean endmembers for each endmember set share a prior distribution that encourages the endmember distributions to have a tight fit around the data. Furthermore, the prior on all of the means over all sets, μ , is given by a Normal distribution whose mean is fixed at the mean of the input hyperspectral data and whose covariance, $\mathbf{C}_\mu = \mathbf{I}\sigma_\mu$ where σ_μ is fixed to a large value.

$$\mu_r \sim N\left(\mu_r \left| \frac{1}{N} \sum_{j=1}^N \mathbf{x}_j, \mathbf{C}_\mu \right.\right) \quad (6)$$

The proportion values for all the data point in the image are given a Dirichlet prior with α values fixed to 1. By fixing the alpha values to 1, the endmembers are further encouraged to have a tight fit around the data [14].

$$\mathbf{p}_j | z_j = r \sim D_{M_r}(\alpha_{1,r}, \dots, \alpha_{M_r,r}) \quad (7)$$

where $D_{M_r}(\cdot)$ denote the M_r -Dirichlet distribution whose distribution function is given by $f(\mathbf{p}_j | z_j = r) = \frac{\Gamma(\sum_{k=1}^{M_r} \alpha_k)}{\prod_{k=1}^{M_r} \Gamma(\alpha_k)} \prod_{k=1}^{M_r} p_{jk}^{\alpha_k - 1}$. In the current implementation, all M_r values are fixed to a constant value that is a set parameter.

Sample Proportion Values. The proportion vectors for each data point and each set of endmembers is sampled in the proposed method using a Metropolis-Hastings step. For implementation, a set of proportions is sampled for each data point for each set of endmember distributions. This is done to be able to compute likelihood values using appropriate proportion vectors for each set of endmembers. The Dirichlet prior shown in Equation 7 is used as the proposal distribution. This results in the following acceptance ratio used to accept or reject new proportion vector samples for each data point in each partition.

$$a = \frac{\Pi(\mathbf{p}_j^{new} | \mathbf{X}, \mathbf{E}, z_j = r)}{f(\mathbf{p}_j^{new} | z_j = r)} \frac{f(\mathbf{p}_j^{old} | z_j = r)}{\Pi(\mathbf{p}_j^{old} | \mathbf{X}, \mathbf{E}, z_j = r)} = \frac{f(\mathbf{x}_j | \mathbf{E}, \mathbf{p}_j^{new}, z_j = r)}{f(\mathbf{x}_j | \mathbf{E}, \mathbf{p}_j^{old}, z_j = r)} \quad (8)$$

where

$$\Pi(\mathbf{p}_j | \mathbf{X}, \mathbf{E}, z_j = r) \propto f(\mathbf{x}_j | \mathbf{E}, \mathbf{p}_j, z_j = r) f(\mathbf{p}_j | z_j = r). \quad (9)$$

The final equality is found since the proposal distribution is also the prior on the proportion vectors. In summary, the new proportion sample is found to be $\mathbf{p}_j^{new} = \mathbf{p}_j^{new}\gamma + \mathbf{p}_j^{old}(1 - \gamma)$ where $\gamma = I\left(u < \min\left\{\frac{f(\mathbf{x}_j|\mathbf{E}, \mathbf{P}_j^{new}, r)}{f(\mathbf{x}_j|\mathbf{E}, \mathbf{P}_j^{old}, r)}, 1\right\}\right)$.

Sample Endmember Distribution Values. A Metropolis-Hastings step is also used to sample endmembers. The proposal distribution is a Gaussian mixture centered on the previous endmember value.

$$g(\mathbf{e}_{k,r}^{new}|\mathbf{e}_{k,r}^{old}, r) = w_n N(\mathbf{e}_{k,r}^{new}|\mathbf{e}_{k,r}^{old}, \mathbf{C}_n) + w_b N(\mathbf{e}_{k,r}^{new}|\mathbf{e}_{k,r}^{old}, \mathbf{C}_w) \quad (10)$$

where w_n and w_b are fixed parameters determining the relative frequency sampling from a Gaussian with diagonal covariance whose diagonal covariances are either small or large, respectively. In all experimental results shown here, w_n is set to 0.6 and w_b is set to 0.4. Furthermore, \mathbf{C}_n and \mathbf{C}_w are fixed covariances used to generate the endmember samples. In all experimental results shown here, both \mathbf{C}_n and \mathbf{C}_w are set to an isotropic diagonal covariances.

The acceptance ratio will be $a = \frac{\Pi(\mathbf{e}_{k,r}^{new}|\mathbf{X}, \mathbf{P}, r)}{g(\mathbf{e}_{k,r}^{new})} \frac{g(\mathbf{e}_{k,r}^{old})}{\Pi(\mathbf{e}_{k,r}^{old}|\mathbf{X}, \mathbf{P}, r)}$ where $\Pi(\mathbf{e}_{k,r}|\mathbf{X}, \mathbf{P}, r) \propto f(\mathbf{X}|\mathbf{E}, \mathbf{P}, r)f(\mathbf{e}_{k,r}|\boldsymbol{\mu}_r, r)f(\boldsymbol{\mu}_r|r)$ where $f(\mathbf{e}_{k,r}|\boldsymbol{\mu}_r, r)$ is given in Equation 5 and $f(\boldsymbol{\mu}_r|r)$ is given in Equation 6.

Similarly, the endmember prior means, $\boldsymbol{\mu}_r$ are sampled using a Metropolis-Hastings set with a Gaussian mixture as the proposal distribution. In the current implementation, the same Gaussian mixtures used to generate new endmember samples are used to generate the $\boldsymbol{\mu}_r$ sample where the mixture is centered on the previous $\boldsymbol{\mu}_r$ value.

Sample Partition Labels. The labels, r , are distributed according to a Dirichlet process. These labels determine the number of endmember distribution sets needed to describe an input hyperspectral data set as well as the partitioning of the data point into the various endmember distribution sets. For all existing partitions, the likelihood is computed using the current E and P matrices. For new partitions, \mathbf{E}^* and \mathbf{P}^* matrices are sampled K times using the same proposal distributions as above.

$$f(z_i = z_j \quad j \neq i | \mathbf{z}_{-i}, \mathbf{x}_i) = C \frac{n_{-i,j}}{\alpha + N - 1} f(\mathbf{x}_i | \mathbf{p}_i, \mathbf{E}_r, z_j = r) f(\mathbf{E}_r) f(\mathbf{p}_i | z_i = r) \quad (11)$$

$$f(z_i \neq z_j \quad \forall j \neq i | \mathbf{z}_{-i}, \mathbf{x}_i) = C \frac{\alpha}{\alpha + N - 1} f(\mathbf{x}_i | \mathbf{p}_i^*, \mathbf{E}^*) f(\mathbf{E}^*) f(\mathbf{p}_i^*) \quad (12)$$

where z_i is the indicator variable for the current data point, \mathbf{x}_i , C is a normalization constant, $n_{-i,j}$ is the number of data points excluding \mathbf{x}_i in partition z_j , N is the total number of data points, K is the number of new endmember distribution sets sampled, and α is the innovation parameter for the Dirichlet process. This method of sampling K new endmember distribution sets follows from the method described in [20]. The number of new sets considered, K , is a parameter that is currently set in the algorithm.

Pseudo-code for Sampling PCE Method.

- 1: Initialize Partitions

```

2: for  $r \leftarrow 1$  to  $R_{initial}$  partitions do
3:   Initialize  $\mathbf{E}_r$  and  $\mathbf{P}_r$ 
4: end for
5: for  $k \leftarrow 1$  to number of total iterations do
6:   Randomly reorder data points in  $\mathbf{X}$ 
7:   for  $r \leftarrow 1$  to number of partitions do
8:     for  $j \leftarrow 1$  to number of data points do
9:       Sample proportions,  $\mathbf{p}_j$ , for  $\mathbf{x}_j$  using an M-H step for each set of endmembers
10:    end for
11:    Randomly reorder endmembers  $E$  in partition  $r$ 
12:    for  $k \leftarrow 1$  to number number of endmembers in partition  $r$  do
13:      Sample  $\mathbf{e}_{k,r}$  in partition  $r$  using M-H step
14:    end for
15:    Sample  $\mu_r$  for each endmember distribution set
16:  end for
17:  for  $k \leftarrow 1$  to  $K$  do
18:    Sample new  $\mathbf{E}^*$  and  $\mathbf{P}^*$  matrices
19:  end for
20:  for  $j \leftarrow 1$  to number of data points do
21:    Remove  $\mathbf{x}_j$  from its current partition
22:    Compute DP partition probabilities for  $\mathbf{x}_j$  using Equations 11 and 12.
23:    Sample a partition for  $\mathbf{x}_j$  based on the DP partition probabilities
24:    if A new partition is sampled then
25:      Add the new endmember distribution set to  $\mathbf{E}$  and assign  $\mathbf{x}_j$  to this set
26:    else
27:      Update the label of  $\mathbf{x}_j$  to the sampled endmember distribution set
28:    end if
29:  end for
30: end for

```

RESULTS

The proposed algorithm was applied to the AVIRIS Indian Pines data set shown in Figure 1. The results of the AVIRIS Indian Pines data set show that the proposed method successfully partitions the hyperspectral data into several convex regions and determines endmembers for each convex set. When compared to results using a single convex region, as shown in Figure 2, it can be seen that the piece-wise convex representation is capable of determining the interior endmembers in this data set, i.e. the non-convex nature of this data is represented with multiple endmember distribution sets. The single convex region results are found using the same proposed method restricted to one convex region. The parameters for the AVIRIS Indian Pines data were set to (the covariances list the constant diagonal element values) $\mathbf{S} = 0.001$, $\mathbf{C}_e = 0.01$, $\mathbf{C}_\mu = 10$, $\mathbf{C}_n = 0.01$, $\mathbf{C}_w = 0.5$, $w_n = 0.6$ (and, hence, $w_w = 1 - w_n = 0.4$), and $\alpha = 10$.

This data sets consists of 16 classes of data. Of these 16 classes, corn fields comprise one of the largest components. The many corn fields have varying levels of residue from

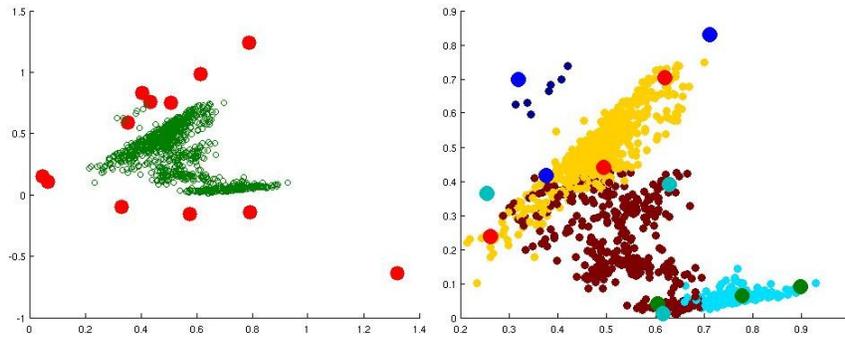


FIGURE 2. A comparison of the results found using a single convex region (a) and multiple convex regions (b) using the Sampling Piece-wise Convex Endmember detection results. The piece-wise convex representaiton results is capable of describing the non-convex nature of the data set. In all of the plots, the small points are the hyperspectral data points and the large points correspond to endmembers. In (b), the colors of the endmembers and data points correspond to the associated partition, i.e. all endmembers of the same color correspond to the same partition and all data points of the same color are assigned to the same partition.

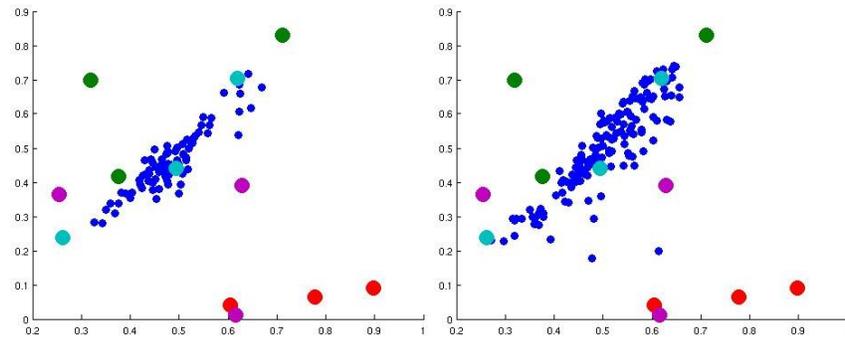


FIGURE 3. Scatterplots of corn class data points in the AVIRIS Indian Pines data set and the endmembers found using the Sampling Piece-wise Convex Endmember detection results. The corn data points are primarily associated with one set of endmembers (light blue) found using S-PCE. In contrast, the corn points would be unmixed with nearly all of the 12 endmembers found when using one region, as shown in Figure 2 (a).

previous crops. Figure 3 shows the scatterplots and the endmembers found for the corn-notill and corn-min classes in the data. Corn-notill contains a large amount of previous crop residue and corn-min contains a moderate amount of previous crop residue. As can be seen, both classes are well represented with one of the convex regions found by S-PCE. In contrast, a large number of the endmembers would contribute to unmixing the corn class with the single convex region results.

CONCLUSIONS AND FUTURE WORK

The S-PCE algorithm presented here uses a Gibbs sampling approach to estimate sets of endmember distributions, proportion values, and the number of endmember distribution sets needed to describe a non-convex hyperspectral image. Future work to further

develop this method will include extending the algorithm to estimating covariances parameters and the number of endmember distributions per set. Also, incorporating spatial correlations is an area of future work [21].

ACKNOWLEDGMENTS

Research was supported by NSF program Optimized Multi-algorithm Systems for Detecting Explosive Objects Using Robust Clustering and Choquet Integration (CBET-0730484). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of NSF. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon. Thank you to G. Casella for his recommendations on the use of Metropolis-within-Gibbs methods.

REFERENCES

1. N. Keshava, and J. F. Mustard, *IEEE Signal Processing Magazine* **19**, 44–57 (2002).
2. M. E. Winter, “Fast Autonomous Spectral Endmember Determination In Hyperspectral Data,” in *Proceedings of the Thirteenth International Conference on Applied Geologic Remote Sensing*, Vancouver, B.C., Canada, 1999, pp. 337–344.
3. J. Boardman, F. Kruse, and R. Green, “Mapping target signatures via partial unmixing of AVIRIS data,” in *Summaries of the 5th Annu. JPL Airborne Geoscience Workshop*, edited by R. Green, JPL Publ., Pasadena, CA, 1995, vol. 1, pp. 23–26.
4. J. M. P. Nascimento, and J. M. Bioucas-Dias, *IEEE Trans. on Geoscience and Remote Sensing* **43**, 898–910 (2005).
5. A. Plaza, P. Martinez, R. Perez, and J. Plaza, *IEEE Trans. on Geoscience and Remote Sensing* **40**, 2025–2041 (2002).
6. D. Lee, and H. Seung, “Algorithms for Non-Negative Matrix Factorization,” in *Advances in Neural Information Processing Systems 13*, 2000, pp. 556–562.
7. L. Miao, and H. Qi, *IEEE Trans. on Geoscience and Remote Sensing* **45**, 765–777 (2007).
8. V. P. Pauca, J. Piper, and R. J. Plemmons, *Linear Algebra Applications* **416**, 321–331 (2005).
9. S. Jia, and Y. Qian, *IEEE Trans. on Geoscience and Remote Sensing* **47**, 161–173 (2009).
10. T.-M. Tu, *Optical Engineering* **39**, 897–906 (2000).
11. J. Wang, and C.-I. Chang, *IEEE Trans. on Geoscience and Remote Sensing* **44**, 2601–2616 (2006).
12. M. D. Craig, *IEEE Trans. on Geoscience and Remote Sensing* **32**, 542–552 (1994).
13. G. X. Ritter, G. Urcid, and M. S. Schmalz, *Neurocomputing* **72**, 2101–2110 (2009).
14. N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, *IEEE Trans. Signal Processing* **57**, 4355–4368 (2009).
15. A. Zare, *Hyperspectral Endmember Detection and Band Selection using Bayesian Methods*, Ph.D. thesis, University of Florida (2009).
16. A. Zare, and P. Gader, *IEEE Trans. on Geoscience and Remote Sensing* **48**, 2620–2632 (2010).
17. AVIRIS, Free standard data products. (2004), (2004, Sep) Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA. URL <http://aviris.jpl.nasa.gov/html/aviris.freedata.html>.
18. S. B. Serpico, and L. Bruzzone, *IEEE Trans. on Geoscience and Remote Sensing* **39**, 1360–1367 (2001).
19. M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Donne, and J. F. Huntington, *IEEE Trans. on Geoscience and Remote Sensing* **42**, 2085–2095 (2004).
20. R. M. Neal, Markov chain sampling methods for Dirichlet process mixture models, Tech. Rep. 9815, University of Toronto, Toronto, ON, Canada (1998).
21. O. Eches, N. Dobigeon, and J.-Y. Tourneret, *IEEE Trans. on Geoscience and Remote Sensing* (2010), submitted. [Online]. Available: <http://arxiv.org/abs/1002.1059>.