# Sampling Piece-wise Convex Unmixing and Endmember Extraction

Alina Zare, *Member, IEEE,* Paul Gader, *Fellow, IEEE,* and George Casella *Fellow, American Statistics Association, Institute of Mathematical Statistics, International Statistics Institute*

***Abstract*—A Metropolis-within-Gibbs Sampler for Piece-wise Convex hyperspectral Unmixing and Endmember extraction (S-PCUE), is presented. The standard linear mixing model used for hyperspectral unmixing assumes that hyperspectral data reside in a single convex region. However, hyperspectral data are often non-convex. Furthermore, in standard endmember extraction and 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 and environmental conditions. Therefore, it is more appropriate to represent each endmember as a full distribution and use this information during spectral unmixing. The proposed method searches for several sets of endmember distributions. By using several sets of endmember distributions, a piece-wise convex mixing model is applied and, given this model, the proposed method performs spectral unmixing and endmember estimation given this non-linear representation of the data. Each set represents a random simplex. The vertices of the random simplex are modeled by the endmember distributions. The hyperspectral data is partitioned into sets associated with each of the extracted sets of endmember distributions using a Dirichlet process prior. The Dirichlet process prior also estimates the number of sets. Thus, the Metropolis-within-Gibbs sampler partitions the data into convex regions, estimates the required number of convex regions, estimates endmember distributions, and abundance values for all convex regions. Results are presented on real hyperspectral and simulated data that indicate the ability of the method to effectively estimate endmember distributions and the number of sets of endmember distributions.**

***Index Terms*—Hyperspectral, endmember, unmixing, Markov chain monte carlo, sampling, piece-wise convex, spectral variation.**

## I. Introduction

THE Linear Mixing Model ($LMM$) represents the spectral signatures in a hyperspectral scene as convex combinations of endmembers. As stated in [1], endmembers are often defined as the spectral signatures of the distinct substances in a hyperspectral data set. The equation and constraints defining the $LMM$ are as follows.

$$\mathbf{x}_j = \sum_{k=1}^{M} p_{jk}\mathbf{e}_k + \epsilon_j \tag{1}$$

A. Zare is with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO, 65211 USA e-mail: zarea@missouri.edu

P. Gader is with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611 USA email: pgader@cise.ufl.edu

G. Casella is with the Department of Statistics, University of Florida, Gainesville, FL 32611 USA email: casella@ufl.edu

Manuscript received October 31, 2010, Revised September 10, 2011, Revised February 12, 2012, Accepted June 24, 2011.

Here $N$ is the number of pixels, $M$ is the number of endmembers, $\epsilon_j$ is an error term, $p_{jk}$ is the proportion of endmember $k$ in pixel $j$, and $\mathbf{e}_k$ is the $k^{th}$ endmember. The proportions satisfy the following constraints.

$$p_{jk} \geq 0 \quad \forall k = 1, \ldots, M \tag{2}$$

$$\textstyle\sum_{k=1}^{M} p_{jk} = 1 \tag{3}$$

Mathematically, the $LMM$ assumes endmembers are vertices of a simplex that approximately enclose the spectra, or data points, present in an image. The approximation is represented by the error term.

Several methods have been developed for unmixing based on the linear mixing model. These include methods, such as Vertex Component Analysis [2], that rely on the pixel purity assumption and assume the endmembers can be found within the data set [2]–[5]. Methods have also been developed based on Non-Negative Matrix Factorization [6]–[11], Independent Components Analysis [12]–[14] and others [15]–[18]. All 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.

If a scene contains multiple distinct regions that do not share common materials and if each region contains linear mixtures of materials, then the set of all image spectra will consist of a union of simplices. A single simplex is convex but the union of simplices is unlikely to be convex. Therefore, a piece-wise convex model is a more appropriate model than a single convex region. Moreover, the extremal points of the individual convex sets may appear to be interior points in the convex hull of all image pixels. Hyperspectral images often exhibit these characteristics.

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 sets of spectra from 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.

Furthermore, the spectral signatures for a material vary within a 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, endmembers are represented as full distributions in S-PCUE rather than a single point. In the current implementation, each endmember is as a Gaussian distribution. In this case, the representation of a random simplex is the same as the representation given by the Normal Compositional Model [19], [20]. In [20], the Normal Compositional model is applied to perform spectral unmixing given a known set of endmembers. In contrast, the proposed method estimates the endmembers in addition to performing spectral unmixing. Furthermore, the proposed method estimates the number of random simplices needed to describe the data set.

The Sampling Piece-wise Convex Unmixing and Endmember Extraction (S-PCUE) method is a fully stochastic unmixing algorithm based on a piece-wise convex model. The use of a piece-wise convex representation was first presented in [21], [22] and [23]. The algorithms represented there will be referred to as PCE. The S-PCUE method differs from PCE because S-PCUE fully uses a complete Bayesian inference strategy whereas PCE does not. Although PCE uses the Dirichlet Process for partitioning the data points into endmember sets, the endmembers and proportions were estimated by maximizing an objective function. Therefore, PCE is a Stochastic EM algorithm. The S-PCUE algorithm provides a fully stochastic extension of these methods by using a Gibbs sampling approach to sample all desired parameters. The S-PCUE algorithm estimates several sets of endmembers, the abundances for each data point, and the number of endmembers sets needed to represent a hyperspectral image. The proposed S-PCUE algorithm improves on the stochastic EM-type algorithms in both time complexity and convergence guarantees. Since the S-PCUE algorithm does not require a maximization step each iteration, the time per iteration is drastically reduced. Furthermore, since the proposed algorithm satisfies MCMC convergence properties, the proposed method inherits MCMC convergence guarantees.

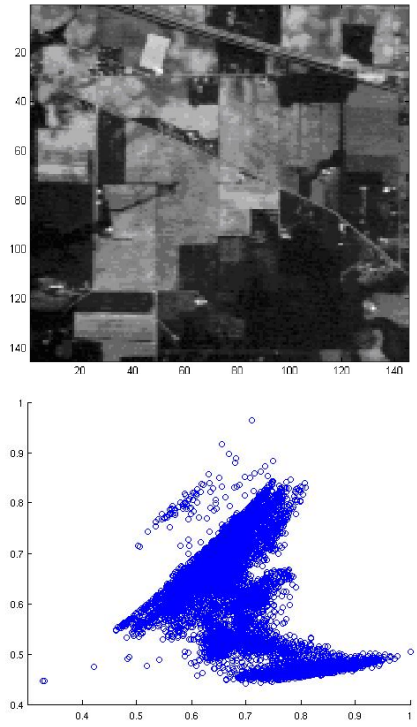



Fig. 1. The June 1992 AVIRIS Indian Pines "Scene 4" data set [24]. These data were collected over the Indian Pines test site in an agricultural area of northern Indiana. The image has $145 \times 145$ pixels with 220 spectral bands. The data contains approximately two-thirds agricultural land and one-third forest and other elements [25]. 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 $\mu$m). (b) The AVIRIS Indian Pines hyperspectral data set after applying Maximum Noise Fraction dimensionality reduction to two dimensions [26]. This illustrates that the Indian Pines hyperspectral data set is not convex but, instead, appears to be piece-wise convex.

## II. The Sampling Piece-wise Convex Unmixing and Endmember Extraction Method

The S-PCUE method represents endmembers as random vectors and uses a Metropolis-within-Gibbs sampling technique to extract their values. Representing endmembers as random vectors provides the capability of representing the spectral variability associated with a particular material or environmental condition.

### *A. Model Description*

Each set of endmember distributions represents the materials that mix together in some subset of the image scene. Each pixel in the scene is assumed to be a mixture of endmembers from one and only one set of endmember distributions. Therefore, the set of sets of endmember distributions defines a partition of the image pixels that is, $X$ is a disjoint union, $X = \cup_{r=1}^{R} \Gamma_r$. Each subset $\Gamma_r$ is assumed to satisfy the linear mixing model with endmembers that are distributions. We, therefore, refer to them as convex sets. The number of endmember distribution sets is estimated by sampling from a Dirichlet Process. In the current implementation, the endmembers are modeled using Gaussian distributions with a fixed isotropic diagonal covariance, $\hat{\mathbf{e}}_{m,r} \sim N(\mathbf{e}_{m,r}, \mathbf{S}_{m,r})$ where $\mathbf{e}_{m,r}$ is the mean value for the $m^{th}$ endmember distribution in the $r^{th}$ endmember distribution set and $\mathbf{S}_{m,r}$ is the covariance for the $m^{th}$ endmember distribution in the $r^{th}$ convex set. In the current implementation, all endmember distributions are given the same fixed isotropic diagonal covariance, $\mathbf{S}$. Therefore, each of these covariance matrices are proportional to the identity matrix.

It is assumed that the linear mixing model holds for the set of data points represented by each set of endmember distributions. Since the endmembers are modeled using Gaussian distributions, the current implementation of S-PCUE follows the Normal Compositional Model [20] in which endmembers are random vectors represented by Gaussian distributions and hyperspectral data points are random vectors distributed according to a convex combination of the associated Gaussian endmember distributions. The identity of the endmember set for each hyperspectral data point is unknown and is represented by a latent variable. Thus, the likelihood for a data

point assigned to a single convex region is given by

$$\mathbf{x}_j|\mathbf{E}_r, \mathbf{p}_j \sim f(\mathbf{x}_j|\mathbf{E}_r, \mathbf{p}_j) = N\left(\mathbf{p}_j\mathbf{E}_r, \sum_{m=1}^{M} p_{jm}^2 \mathbf{S}_{m,r}\right) \quad (4)$$

where $\mathbf{x}_j$ is the $j^{th}$ data point, $r$ is the indicator variable for the $r^{th}$ set of endmembers, $\mathbf{E}_r$ is a matrix whose rows are the endmember means for the $r^{th}$ set of endmembers, and $\mathbf{p}_j$ is the vector of proportion values associated with the $j^{th}$ data point where $p_{jm}$ is the $m^{th}$ element of this proportion vector, and $M$ is the number of endmember distributions in the $r^{th}$ partition. A somewhat subtle point that is noted here is that there is assumed to be only one true set of proportions for each data point. However, there are $M$ estimates of the set of proportions, one for each partition. Let $z_j$ denote the latent variable representing the partition to which the data point $\mathbf{x}_j$ is assigned. Then the overall likelihood can be written as

$$\prod_{r=1}^{R} \prod_{j \in I_r} f(\mathbf{x}_j|\mathbf{E}_r, \mathbf{p}_j)$$

where $R$ is the number of convex sets, $I_r = \{j|z_j = r\} \subset \{1, \ldots, N\}$ denotes the set of indices of the data points that are assigned to the $r^{th}$ convex set, and $\mathbf{E} = \{\mathbf{E}_1, \ldots, \mathbf{E}_R\}$ denotes the set of endmember mean matrices.

For each partition, the means of the associated endmember distributions are assumed to share a Gaussian prior distribution.

$$\mathbf{e}_{m,r} \sim N(\mu_r, \mathbf{C}_r), \quad m = 1, \ldots, M \quad (5)$$

where $\mu_r$ and $\mathbf{C}_r$ are the mean vector and full covariance hyperparameters, respectively, in the Gaussian prior distribution governing the $r^{th}$ set of endmember distribution means. These hyperparameters are also estimated within the S-PCUE algorithm. This prior for the mean of the endmember distributions of a partition was chosen for a number of reasons. First, it 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. Second, estimating a full covariance hyperparameter for each partition allows for a tight fit around data whose collective shape and size differ from partition to partition. Finally, this distribution is a conjugate prior to endmember distributions allowing for an efficient Gibbs sampling step.

The prior on all of the means over all sets, $\mu$, is also given by a Gaussian distribution whose mean is fixed at the mean of the input hyperspectral data and whose covariance is $\mathbf{C}_\mu = \mathbf{I}\sigma_\mu$ where $\sigma_\mu$ is fixed to a large value relative to the spread of the entire input data set. This practice, of using marginal means to estimate hyperparameters, is in the spirit of empirical Bayes analysis [27]

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

The prior used for the $\mathbf{C}_r$ covariance matrices is the Inverse-Wishart distribution. This prior distribution was chosen as it is the conjugate prior of the Gaussian distribution over a full covariance matrix.

The proportion values for all the data points in the image are given a Dirichlet prior.

$$\mathbf{p}_j|z_j = r \sim D_M(\alpha_{1,r}, \ldots, \alpha_{M,r}) \quad (7)$$

where $D_M(\cdot)$ denote the $M$-factor Dirichlet distribution whose density function is given by

$$D_M(\mathbf{p}_j|z_j = r) = \frac{\Gamma\left(\sum_{m=1}^{M} \alpha_{m,r}\right)}{\prod_{m=1}^{M} \Gamma(\alpha_{m,r})} \prod_{m=1}^{M} p_{jm}^{\alpha_{m,r}-1}.$$

By fixing the alpha values to 1, the endmembers are encouraged to have a tight fit around the data [17]. In the current implementation, all alpha values are fixed to 1. However, one could choose to use other values for each $\alpha$. In the current implementation, the number of endmembers, $M$, is a set parameter fixed to a constant value across all partitions.

### *B. Sampling Method*

The S-PCUE algorithm iteratively samples the parameters and hyperparameters of interest using a Metropolis-within-Gibbs sampling method. The algorithm is outlined in the psuedo-code shown in Algorithm 1. In the following, each step of the sampling algorithm, initialization, and the methods used for setting algorithm parameters are described.

*1) Sample Proportion Values:* The proportion vectors for each data point and each set of endmembers is sampled in S-PCUE 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 acceptance ratio shown in Equation 8 used to accept or reject new proportion vector samples for each data point in each partition. 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 $\gamma = I\left(u < \min\left\{\frac{f(\mathbf{x}_j|\mathbf{E}, \mathbf{p}_j^{new}, z_j=r)}{f(\mathbf{x}_j|\mathbf{E}, \mathbf{p}_j^{old}, z_j=r)}, 1\right\}\right)$. Here, $I(m) = 1$ when $m$ is true and 0 when m is false and $u$ is randomly drawn from the uniform distribution over $[0, 1]$.

*2) Sample Endmember Distribution Values:* A Metropolis-Hastings step is also used to sample the means of each endmember distribution. The proposal distribution is a Gaussian mixture centered on the previous mean endmember value. The first Gaussian in the mixture has a diagonal covariance with small values and the second Gaussian has a diagonal covariance with large values.

$$g(\mathbf{e}_{m,r}^{new}|\mathbf{e}_{m,r}^{old}) = \quad (10)$$
$$w_n N(\mathbf{e}_{m,r}^{new}|\mathbf{e}_{m,r}^{old}, \quad \mathbf{C}_n) + (1 - w_n)N(\mathbf{e}_{m,r}^{new}|\mathbf{e}_{m,r}^{old}, \mathbf{C}_w)$$

where $w_n$ is a fixed parameter used to determine the relative frequency sampling from a Gaussian with diagonal covariance whose diagonal covariances are either small or large. The covariance matrices, $\mathbf{C}_n$ and $\mathbf{C}_w$ are fixed covariances used to

**Algorithm 1 - Sampling PCUE**: *A Metropolis-within-Gibbs sampler partitions the data into convex regions, estimates the required number of convex regions, estimates endmember distributions, and abundance values for all convex regions. The method used for each step is described in the section shown in parentheses*

```
Set Parameter Values and Initialize Partition (Section II-C)
for r ← 1 to R_initial convex sets do
        Initialize E_r, P_r and C_r (Section II-C)
end for
for k ← 1 to number of total iterations do
        for r ← 1 to number of convex sets do
                for j ← 1 to number of data points do
                        Sample proportions, p_j, for x_j for each set of endmembers (Section II-B1)
                end for
                for k ← 1 to number of endmembers in convex set r do
                        Sample e_{k,r} in convex set r (Section II-B2)
                end for
                Sample μ_r (Section II-B3)
                Sample C_r (Section II-B4)
        end for
        for k ← 1 to K do
                Sample new E*_k and P*_k matrices (Section II-B5)
        end for
        for j ← 1 to number of data points do
                Remove x_j from its current convex set
                Compute DP partition probabilities for x_j using Equations 12 and 13.
                Sample a convex set for x_j based on the DP partition probabilities (Section II-B6)
                if A new convex set is sampled then
                        Add the new endmember distribution set to E and assign x_j to this set
                else
                        Update the label of x_j to the sampled endmember distribution set
                end if
        end for
end for
```

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

where

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

generate the endmember samples. In all experimental results shown here, both $\mathbf{C}_n$ and $\mathbf{C}_w$ are fixed parameters set to isotropic diagonal covariance matrices.

The acceptance ratio will be

$$a = \frac{\Pi(\mathbf{e}_{m,r}^{new}|\mathbf{X},\mathbf{P},\mathbf{z})}{g(\mathbf{e}_{m,r}^{new}|\mathbf{e}_{m,r}^{old})}\frac{g(\mathbf{e}_{m,r}^{old}|\mathbf{e}_{m,r}^{new})}{\Pi(\mathbf{e}_{m,r}^{old}|\mathbf{X},\mathbf{P},\mathbf{z})} \tag{11}$$

where

$$\Pi(\mathbf{e}_{m,r}|\mathbf{X},\mathbf{P},\mathbf{z}) \propto \prod_{j:z_j=r} f\left(\mathbf{x}_j|\mathbf{E}_r,\mathbf{p}_j,z_j=r\right) f(\mathbf{e}_{m,r}|\mu_r,\mathbf{C}_r)$$

with $f(\mathbf{e}_{m,r}|\mu_r,\mathbf{C}_r)$ as the distribution in Equation 5.

As stated in Section II-A, the covariance for each endmember distribution is fixed to a diagonal isotropic matrix in the current implementation.

*3) Sample Endmember Set Means:* The hyperparameters, $\mu_r$, which represent the mean of the Gaussian prior for a set of endmembers, are sampled using a Metropolis-Hastings set. A Gaussian mixture is used as the proposal distribution. This step mimics the previous step described in Section II-B2 for sampling endmember distribution means. In the current implementation, the same Gaussian mixture used to generate new endmember samples in the previous step are used to generate the $\mu_r$ samples. Similarly, the mixture is centered on the previous $\mu_r$ value.

*4) Sample Endmember Set Covariance:* The covariance for each endmember set, $\mathbf{C}_r$ is assumed to have an Inverse-Wishart prior, $\mathbf{C}_r \sim InvWishart(\mathbf{\Psi},t)$. Given that the Inverse-Wishart is the conjugate prior to the Gaussian distribution with a full covariance, the covariance for each set is sampled directly using a Gibbs step from the posterior Inverse-Wishart distribution resulting from the product of the likelihood of all of the endmember distribution means for a partition, $\prod_{m=1}^{M} N(\mathbf{e}_{m,r}|\mu_r,\mathbf{C}_r)$, and the $InvWishart(\mathbf{\Psi},t)$

prior.

*5) Sample Endmembers and Proportion for Potential New Partitions:* Prior to sampling a partition label for each data point given updated endmember and proportion values, $K$ new sets of endmember distributions and associated proportion values are sampled. These new sets allow the for the addition of new convex sets during the update of the partition labels. This method of sampling $K$ new endmember distribution sets follows from the method described in [28]. The number of new sets considered, $K$, is a fixed parameter in the implementation.

The proportion values for each data point and each of the $K$ new endmember sets is drawn from the Dirichlet prior for the proportion values as defined in Equation 7. The $K$ new sets of endmembers are drawn following the method defined in Algorithm 2.

**Algorithm 2 - Sampling New Endmember Sets**

```
for k ← 1 to K do
    Sample μ_{R+k} from the Gaussian defined in Equation 6
    Sample C_{R+k} from the Inverse-Wishart prior
    for m ← 1 to M do
        Sample e_{m,R+k} from the Gaussian centered at μ_{R+k} with covariance C_{R+k}
    end for
end for
```

*6) 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 points into the various endmember distribution sets. The likelihood of a data point belonging to a convex region is computed for each existing $\mathbf{E}$ and $\mathbf{P}$ set and all potential new endmember sets that were sampled in the previous step as shown in Equations 12 and 13 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 convex set $z_j$, $N$ is the total number of data points, $K$ is the number of new endmember distribution sets sampled, and $\alpha$ is the innovation parameter for the Dirichlet process.

### C. Initialization and Parameter Settings

S-PCUE requires several parameters to be set prior to running the algorithm. For all experimental results shown, initialization for the algorithm and parameters are determined using the following methods.

*1) Initialization:* The initial convex sets and labels for each data point are determined by fitting the data to a Gaussian mixture using Expectation-Maximization (EM) with random initialization. During each iteration of the EM algorithm, Gaussian components are checked to ensure that the covariance matrix for the corresponding Gaussian is not becoming singular. If that is the case, the Gaussian is removed and the associated points are reassigned to the remaining Gaussians. Following the application of the EM algorithm, endmembers are initialized for each convex set using the VCA algorithm [2]. Finally, proportions for each data point are initialized randomly by drawing from a uniform Dirichlet distribution.

*2) Parameters Settings:* As described in Section II-A, the parameters for the Gaussian prior on the mean of the endmember distributions are to set to the mean of the data and variance of the data, $\sigma_\mu = var(\mathbf{X})$.

Section II-B2 states that the means of the endmember distributions are sampled from a mixtures of two Gaussians, one with a small co-variance and another with a large co-variance. Both Gaussians are centered on the previous endmember mean value. The covariances are set using an order weighted averages. To compute these values, the pairwise Euclidean distances between each pair of input data points is computed and then sorted. The small co-variance, $\mathbf{C}_n$, is set to the mean of smallest pairwise distance for each data point. The large co-variance, $\mathbf{C}_w$, is set to mean of the median pairwise distance for each data point. For all runs of the algorithm, $w_n$ was set to 0.9 such that the majority of new endmember means are drawn close to the previous value.

As described in Section II-B6, the sample labels are drawn from a Dirichlet process. The parameters required for this step are the number of new partitions considered, $K$, and the innovation parameter, $\alpha$. The primary consideration in setting the $K$ parameter is the time complexity per iteration. Larger values of $K$ require longer running time per iteration. $K = 5$ was found to be a good balance of running time and the number of new partitions considered. The innovation parameter is set to $\alpha = \frac{K}{N}$ where $N$ is the number of input data points.

The covariance governing each partition, $\mathbf{C}_r$, is sampled from an inverse Wishart distribution as described in Section II-B4. The parameters for the inverse Wishart are set to $\mathbf{\Psi} = \mathbf{I}$ and $w$ is set to a scaled version of the mean of variances of the initial partitions.

## III. Results

The Sample PCUE algorithm was run on simulated data and the AVIRIS Indian Pines and Cuprite scenes. Results are shown on these data sets and discussed in the following sections.

### A. Simulated Data

The S-PCUE algorithm was run on simulated data generated from three sets of two endmember distributions. The purpose of this simulated data set was to show that the proposed method is able to recover interior endmembers that are not found using conventional single partition unmixing algorithms. Each data point in this simulated set was generated as a convex combination of two endmembers with a total of three set of endmember pairs. Simulated data generated in this fashion was selected to simulate the pairwise-mixing seen in hyperspectral imagery such as shown in [29]. Figure 3 shows the simulated data set. The results found using the parameters settings described in Section II-C are shown in Figure 2(b). As can be seen, the data are appropriately partitioned and correct endmembers are estimated. For comparison, the SPICE [30]

$$P(z_i = z_j|\mathbf{z}_{-i}, \mathbf{x}_i) = C\frac{n_{-i,j}}{\alpha + N - 1} f(\mathbf{x}_i|\mathbf{p}_i, \mathbf{E}_r, z_j = r) f(\mathbf{E}_r) f(\mathbf{p}_i|z_i = r), j = 1, \ldots, R, j \neq i \quad (12)$$

$$P(z_i = R + 1|\mathbf{z}_{-i}, \mathbf{x}_i) = C\frac{\frac{\alpha}{K}}{\alpha + N - 1} f(\mathbf{x}_i|\mathbf{p}_i^*, \mathbf{E}^*) f(\mathbf{E}^*) f(\mathbf{p}_i^*) \quad (13)$$

and VCA [2] algorithms were also run on this data set. Both methods did not find the endmember located at $(2, 3, 2)$.

The algorithm was run for 50,000 iterations and, from the 50,000 samples generated, the majority of samples partitioned the data into three convex regions, as shown in Figure 2(c). The result shown in Figure 2(b) was selected by, first, determining the number of convex sets found by the majority of the samples and, then, returning the result that had the largest likelihood and the appropriate number of convex sets. This method for selecting the final result was used in all of the runs of S-PCUE.

### *B. AVIRIS Indian Pines*

S-PCUE was applied to the June 1992 AVIRIS Indian Pines data set as well, shown in Figure 1. This data set, collected over an agricultural area in northern Indiana, was collected during very early growth stages. The data in this image is highly mixed resulting from the fields having very low crop cover and varying levels of residue from the previous year's crops. The S-PCUE algorithm was applied to a subset of the points selected from the image (every $10^{th}$ labeled pixel, resulting in 1031 pixels) for 50,000 iterations using the parameters settings described in Section II-C. As shown in the results in 4, the data are appropriately partitioned into three convex regions: (Partition 1) woods, grass and trees, (Partition 2) hay, corn and soybean, and (Partition 3) stone-steel towers. For example, the algorithm pulls out a distinct hay endmember, whose proportion map is shown in Figure 5(b). The soybean and corn fields, which are primarily soil and crop residue are shown in the second two endmembers in Figure 5(b).

For comparison, the SPICE and VCA algorithms were applied to the AVIRIS Indian Pines data as well. For SPICE, the parameters used were $\mu = 0.001$ and $\gamma = 1$. SPICE resulted in estimating 7 endmembers after initializing with 20 endmembers. The resulting proportion maps estimated by SPICE are shown in Figure 7. VCA was used to estimate 9 endmembers from the Indian Pines data set. VCA was set to estimate 9 endmembers since this was the number of endmembers found by the proposed S-PCUE algorithm. The proportion maps corresponding to the endmembers estimated by VCA are shown in Figure 8. Neither the SPICE nor VCA algorithms were able to estimate the strong hay endmember found by the S-PCUE algorithm. However, both SPICE and VCA were able to estimate a single stone-steel towers endmember which the S-PCUE assigned to a separate convex region (and, thus, three endmembers). As discussed below, future work will include estimating the number of endmembers needed per convex region. In the case of the convex region associated with stone-steel towers here, only a single endmember may be needed.

### *C. AVIRIS Cuprite*

S-PCUE was also applied to the AVIRIS Cuprite data set collected over Cuprite, NV. This area has little vegetation and visible mineral coverage. A mineral map of the area can be found in [31]. The S-PCUE algorithm was applied to a subset of the points selected from the image (every $101^{st}$ pixel, resulting in 3113 pixels) for 50,000 iterations using the parameters settings described in Section II-C. Figure 9 shows the 3-D plot of the data and the endmembers found following principal components analysis dimensionality reduction. Figure 10 displays the proportion maps associated with each of the estimated endmembers over the full data set.

The endmembers found by S-PCUE was compared to spectra in the USGS Spectral library. The spectra with the smallest spectral angle distance to the estimated endmembers were identified. These spectra are shown in Figure 12. The average spectral angle distance between the estimated endmembers and identified spectra in the USGS spectral library was 0.037.

For comparison, the SPICE and VCA algorithms were run on the same AVIRIS Cuprite data. The VCA algorithm was set to estimate 9 endmembers since S-PCUE found 9 endmembers on this data. The SPICE parameters used were $\mu = 0.001$ and $\gamma = 1$. SPICE found 9 endmembers after pruning 11 of the 20 initial endmembers. Figures 14 and 16 display the endmembers found and their closest matching spectra in the USGS spectral library. The average spectral angle distance between the SPICE and VCA endmembers and spectral library spectra were 0.136 and 0.071, respectively. Thus, the proposed method estimated endmembers that have a smaller spectral angle distance between minerals matched in the USGS spectral library.

## IV. Conclusions and Future Work

The S-PCUE 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. Thus, the proposed method provides an endmember estimation and spectral unmixing algorithm given a non-linear representation of the data. Future work to further develop this method will include extending the algorithm to estimating covariance parameters for each endmember distribution and the number of endmember distributions per set. For example, consider the stone-steel towers partition shown in Figure 5(c), this partition may be best represented with one or two endmembers, however, the current algorithm is restricted to the same number of endmembers per convex region. Currently, the method also does not restrict the endmembers to be non-negative which is generally the case for hyperspectral imagery. Since, after applying dimensionality reduction, such as PCA or MNF, the resulting data often has negative components, this restriction

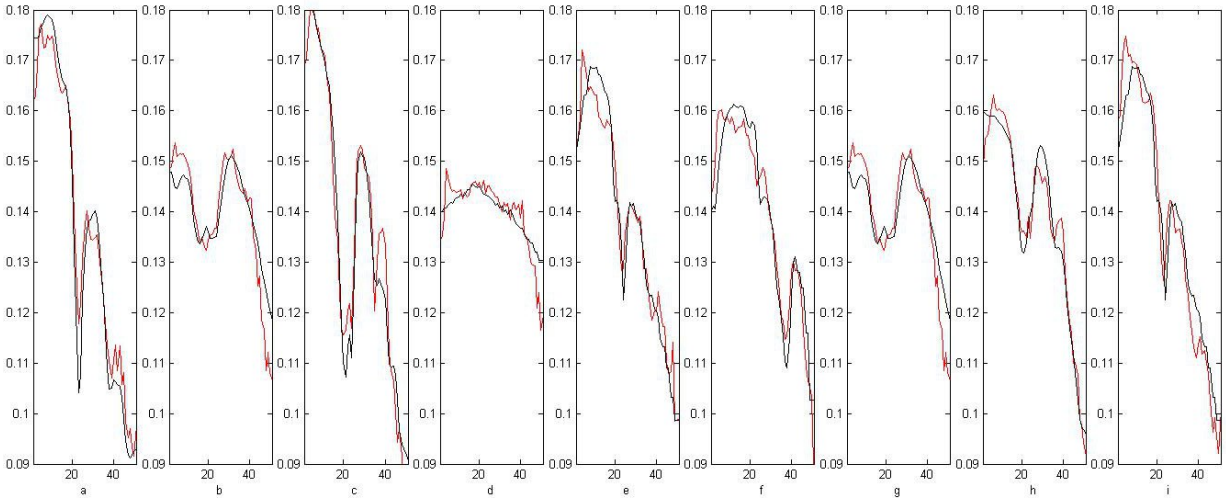



Fig. 11. Estimated endmembers (red) found using S-PCUE on the AVIRIS Cuprite data and USGS spectral library spectra (Black). The USGS spectra were selected by having the smallest spectral angle distance to the estimated endmembers. The USGS spectra selected for each endmember were: (a) muscovite-medhi-Al CU91-252d.26143 (b) buddingtonite cu93-260b.24428 (c) alunite na dickite_mv99-6-26b.23945 (d) hematite_gds27.9282 (e) hematite_coat_br93-25a.26778 (f) calc_mont_amx6.24507 (g) buddingtonite_cu93-260b.24428 (h) alunite_gds82.1019 (i) hematite_coat_br93-25a.26778

Fig. 12.

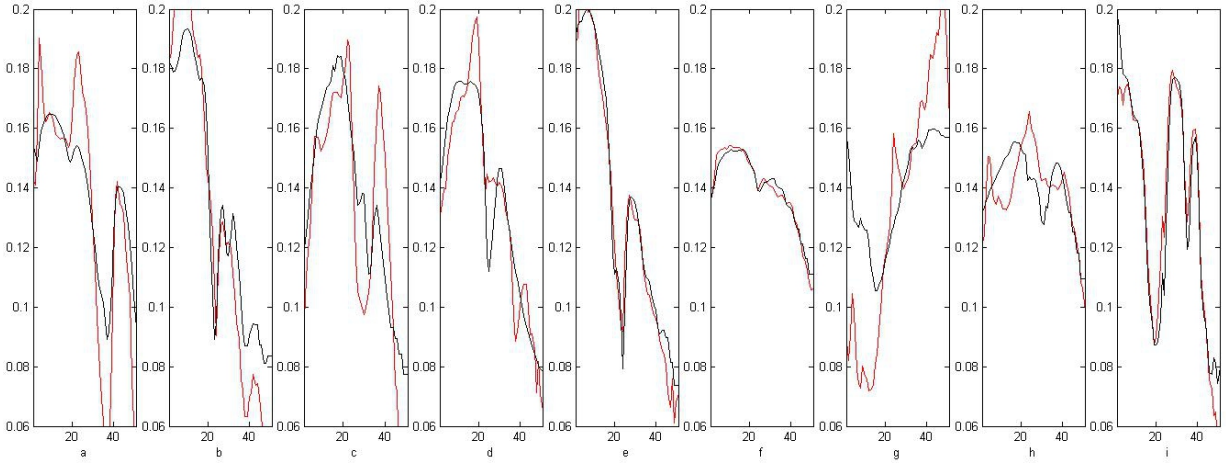



Fig. 13. Estimated endmembers (red) found using SPICE on the AVIRIS Cuprite data and USGS spectral library spectra (Black). The USGS spectra were selected by having the smallest spectral angle distance to the estimated endmembers. The USGS spectra selected for each endmember were: (a) calcite_gds304.4124 (b) chlorite_mixture_cu93-65a.24855 (c) nontronite_swa1.16168 (d) montmorillonite_stx1.14644 (e) halloysite_kaolinite_cm29.25328 (f) montmorillonite-Na_cu93-52.25995 (g) buddingtonite_gds85.3924 (h) jarosite_gds98.11375 (i) alunite_cu91-217g1.23924

Fig. 14.

was not included. However, future work can include exploring the use of truncated Gaussians for endmember proposal and prior distributions to restrict the endmembers to be non-negative. Also, incorporating spatial correlations is an area of future work [32].

The proposed method applies a fully stochastic MCMC algorithm to a high dimensional problem. The advantage of this is that the algorithm is guaranteed to converge to the joint distribution over the parameters. The implication is that one can truly expect to find globally optimal solutions by examining these distributions. Therefore, MCMC algorithms will almost certainly eventually find better solutions than algorithms that find local optima. Given these excellent convergence properties, it seems that MCMC algorithms should always be used. The difficulty is that the speed of the algorithm is slow compared to many existing unmixing algorithms in the literature. Furthermore, it is not obvious how to speed up such an algorithm since they appear to be inherently sequential. Generally, one cannot run Markov chains in parallel and the take the union of the results while maintaining the convergence properties because they are different chains. However, Gopal and Casella [33] have made a tremendous theoretical breakthrough in the area of parallelization of MCMC algorithms. Roughly speaking, they have devised and demonstrated a method for running Markov chains in parallel in a way that maintains the convergence properties. Thus, MCMC methods can now be implemented correctly on clusters of computers consisting of tens, hundreds, or thousands of processors. Currently their methods require some analysis of each specific problem. It seems clear that, once these methods are further

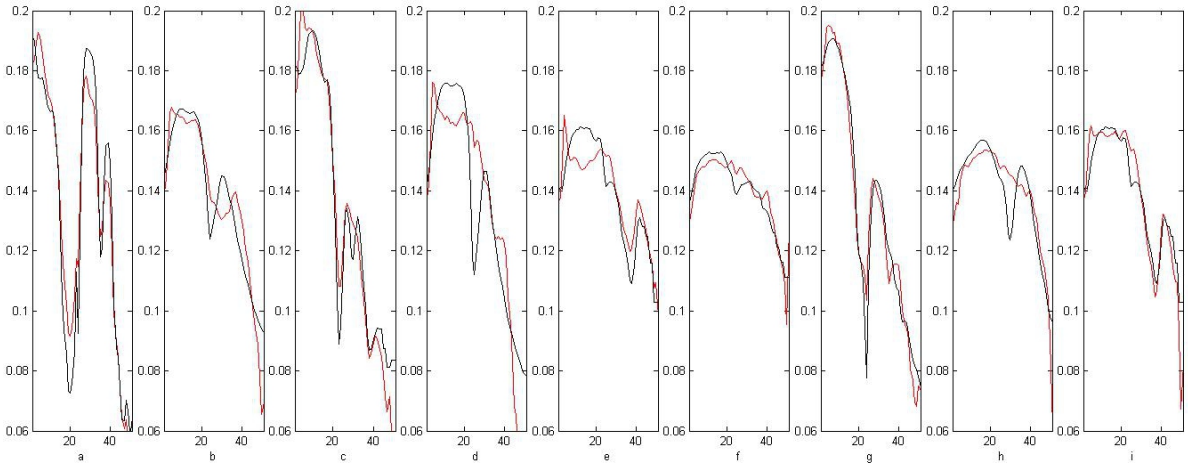



Fig. 15. Estimated endmembers (red) found using VCA on the AVIRIS Cuprite data and USGS spectral library spectra (Black). The USGS spectra were selected by having the smallest spectral angle distance to the estimated endmembers. The USGS spectra selected for each endmember were: (a) alunite_hs295.1425 (b)montmorillonite_cm20.14338 (c) chlorite_mixture_cu93-65a.24855 (d)montmorillonite_stx1.14644 (e) calc_mont_amx6.24507 (f) montmorillonite-Na_cu93-52.25995 (g) halloysite_nmnh106236.9052 (h) jarosite_sj1.11680 (i) calc_mont_amx6.24507

Fig. 16.

developed, MCMC methods will become extremely attractive and potentially the standard method for parameter estimation for difficult, high-dimensional problems such as hyperspectral image analysis. The work described here, together with the research of Dobigeon et al. [34]–[36] and others [22], [37]. represents progress in that direction. Hyperspectral unmixing is a very complex problem in general. In addition to piecewise convexity, there are a number of sources of nonlinearity that are only beginning to be considered for large-scale data sets. Complete analyses of hyperspectral data with multiple scattering, intimate mixtures, and multiple image components will certainly benefit from the type of research described in this paper.

## ACKNOWLEDGMENT

Research was partially 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.

## REFERENCES

[1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, 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. on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[3] 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.

[4] 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*, R. Green, Ed., vol. 1. Pasadena, CA: JPL Publ., 1995, pp. 23–26.

[5] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operators," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2025–2041, Sep. 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, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[8] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra Applications*, vol. 416, no. 1, pp. 321–331, Jul. 2005.

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

[10] A. Ambikapathi, T. Chan, W.-K. Ma, and C.-Y. Chi, "Chance-constrained robust minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 11, pp. 4194 –4209, nov. 2011.

[11] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via sparsity-constrained nonnegative matrix factorization," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 11, pp. 4282 –4297, nov. 2011.

[12] T.-M. Tu, "Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent components analysis approach," *Optical Engineering*, vol. 39, no. 4, pp. 897–906, 2000.

[13] J. Wang and C.-I. Chang, "Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2601–2616, Sep. 2006.

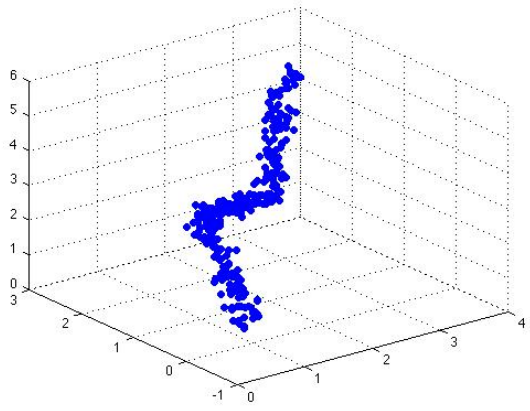
[14] W. Xia, X. Liu, B. Wang, and L. Zhang, "Independent component analysis for blind unmixing of hyperspectral imagery with additional constraints," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 6, pp. 2165 –2179, june 2011.

[15] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 32, no. 3, pp. 542–552, May 1994.

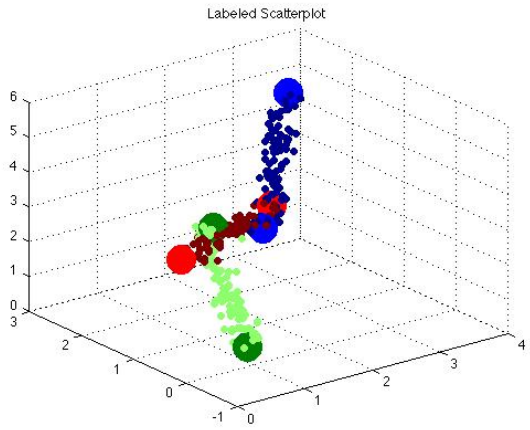
[16] G. X. Ritter, G. Urcid, and M. S. Schmalz, "Autonomous single-pass endmember approximation using lattice auto-associative memories," *Neurocomputing*, vol. 72, pp. 2101–2110, 2009.

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

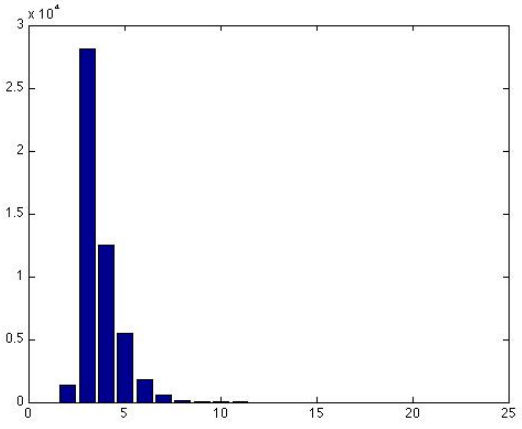
[18] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of



(a)

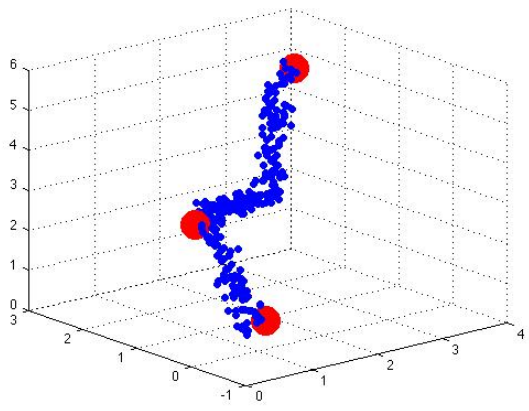



(b)

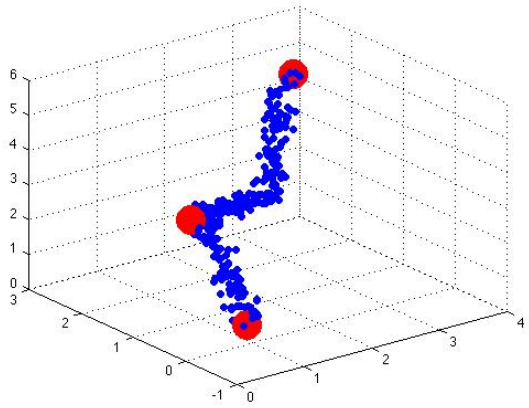



(c)

Fig. 2. (a) Scatter plot of simulated data (b) Results found using S-PCUE algorithm (c) Histogram of the number of convex sets over 50,000 samples



(a)



(b)

Fig. 3. (a) Results found using VCA algorithm on simulated data. (b) Results found using SPICE algorithm on simulated data.

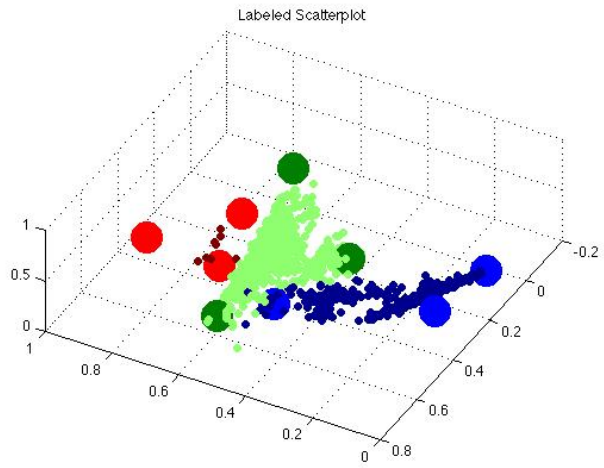


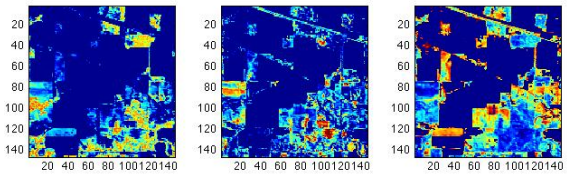

Fig. 4. Scatter plot of results found on the AVIRIS Indian Pines data set after PCA dimensionality reduction from 220 to 3 dimensions. The large points correspond to endmembers in the scene.

hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 6, pp. 2014 –2039, june 2011.

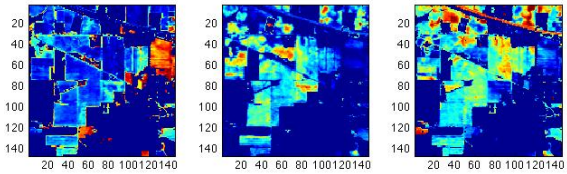
[19] M. T. Eismann and D. W. J. Stein, *"Stochastic mixture modeling," in Hyperspectral Data Exploitation: theory and applications*, C.-I. Chang, Ed. John Wiley & Sons, 2007, vol. 148.

[20] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model: Application to hyperspectral imagery," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1403–1413, June 2010.

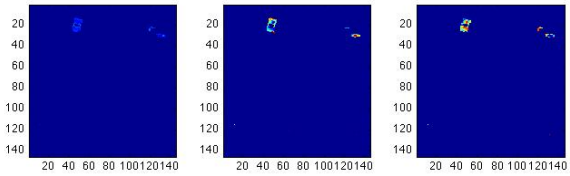
[21] A. Zare, "Hyperspectral endmember detection and band selection using bayesian methods," Ph.D. dissertation, University of Florida, 2009.



(a) Proportion maps corresponding to endmembers found for partition 1. This partition corresponds primarily to woods, grass, and trees.



(b) Proportion maps corresponding to endmembers found for partition 2. This partition corresponds primarily to hay (endmember 1), corn and soybean (endmembers 2 & 3).



(c) Proportion maps corresponding to endmembers found for partition 3. This partition corresponds primarity to stone-steel towers.

Fig. 5. Proportion maps found after unmixing using the endmembers estimated from the S-PCUE algorithm.

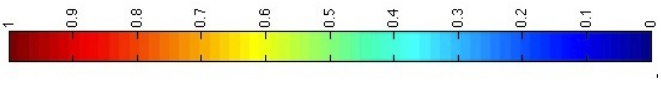
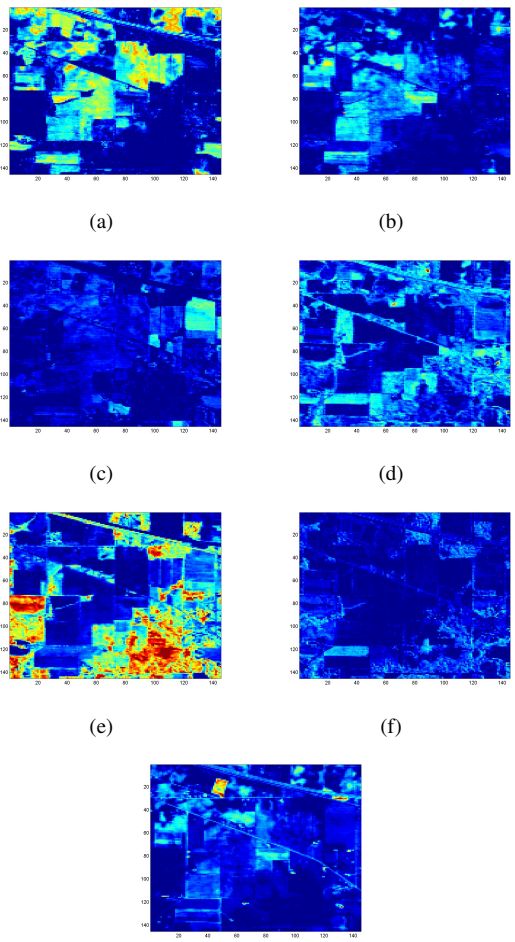


Fig. 6. Colorbar corresponding to all resulting proportion maps shown for the AVIRIS Indian Pines data. Proportion values of 0 correspond to dark blue and proportion values of 1 correspond to dark red.



(a) (b) (c) (d) (e) (f) (g)

Fig. 7. Proportion maps found after unmixing using the endmembers estimated from the SPICE algorithm.


[22] A. Zare and P. Gader, "Pce: Piece-wise convex endmember detection," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2620–2632, 2010.

[23] ——, "An investigation of likelihoods and priors for bayesian endmember estimation," *Proceedings of the 30th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, July 2010.

[24] AVIRIS, "Free standard data products." 2004, (2004, Sep) Jet Propulsion Laboratory, California Institute of Technology, Pasedena, CA. URL http://aviris.jpl.nasa.gov/html/aviris.freedata.html.

[25] S. B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1360–1367, July 2001.

[26] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 26, pp. 65–73, Jan. 1988.

[27] E. Lehmann and G. Casella, *Theory of Point Estimation, Second Edition*. Springer-Verlag, 1998.

[28] R. M. Neal, "Markov chain sampling methods for Dirichlet process mixture models," University of Toronto, Toronto, ON, Canada, Tech. Rep. 9815, Sep. 1998.

[29] M. Berman, A. Phatak, R. Lagerstrom, and B. R. Wood, "ICE: A new method for the multivariate curve resolution of hyperspectral images." *Journal of Chemometrics*, vol. 23, pp. 101–116, 2009.

[30] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection for hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 3, pp. 446–450, July 2007.

[31] G. Swayze, R. Clark, S. Sutley, and A. Gallagher, "Ground-truthing aviris mineral mapping at cuprite, nevada," *Summaries 3rd Annu. JPL Airborne Geosciences Workshop*, vol. 1, pp. 47–49, 1992.

[32] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. on Geoscience and Remote Sensing*, 2010, submitted. [Online]. Available: http://arxiv.org/abs/1002.1059.

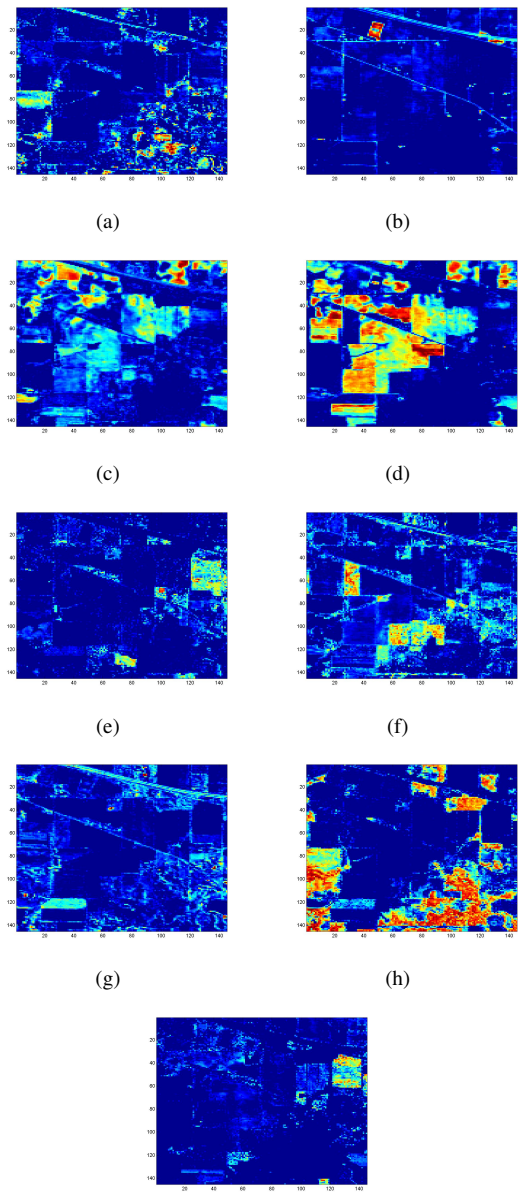
[33] V. Gopal, "Techniques of parallelization in markov chain monte carlo methods," Ph.D. dissertation, University of Florida, 2011.

[34] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. Hero, "Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery," *Signal Processing, IEEE Transactions on*, vol. 57, no. 11, pp. 4355 –4368, nov. 2009.

[35] A. Halimi, Y. Altmann, N. Dobigeon, and J. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 11, pp. 4153 –4162, nov. 2011.

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

[37] V. Mazet, S. Faisan, A. Masson, M.-A. Gaveau, and L. Poisson, "Unsupervised joint bayesian decomposition of a sequence of photoelectron spectra," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on*, june 2011, pp. 1 –4.



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 8. Proportion maps found after unmixing using the endmembers estimated from the VCA algorithm.

**Alina Zare** (S'07 - M'08) received the Ph.D. degree from the Department of Computer and Information Science and Engineering from the University of Florida, Gainesville, in December 2008. She is currently an Assistant Professor in the Electrical and Computer Engineering Department, University of Missouri, Columbia. Her research interests include machine learning, Bayesian methods, sparsity promotion, image analysis, pattern recognition, hyperspectral image analysis, and human geography.

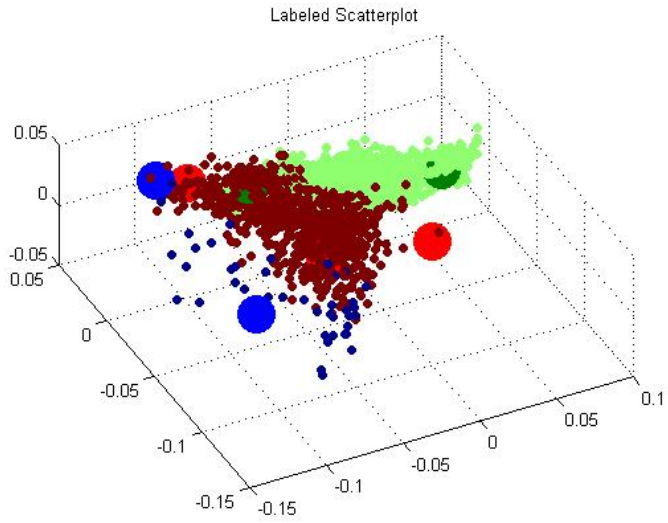
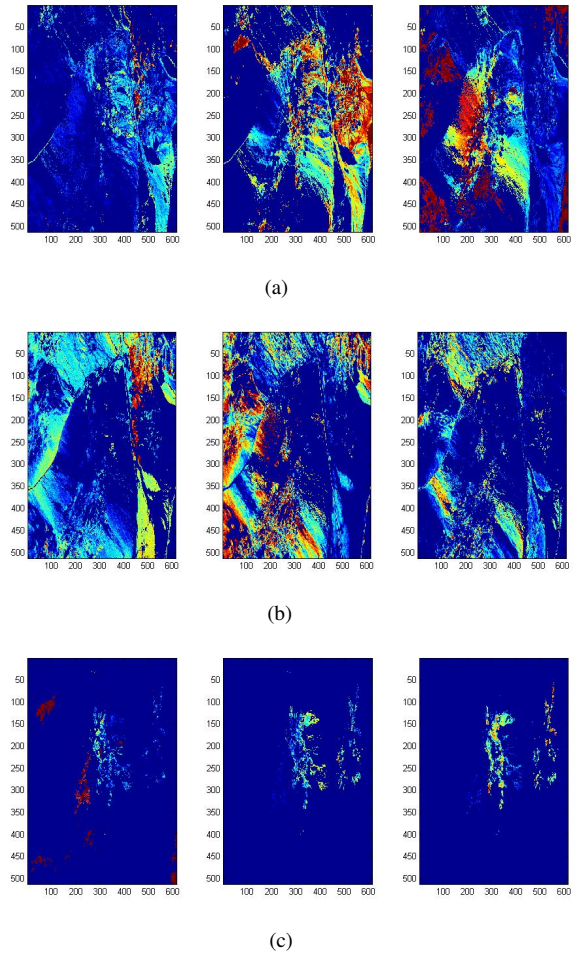



Fig. 9. Scatter plot of results found on the AVIRIS Cuprite data set after PCA dimensionality reduction from 51 to 3 dimensions. The large points correspond to endmembers in the scene.



(a) (b) (c)

Fig. 10. Proportion maps found on the AVIRIS Cuprite data set after unmixing using the endmembers estimated from the S-PCUE algorithm.

**Paul Gader** Paul Gader (M86-SM9-F11) received his Ph.D. in Mathematics for image processing related research in 1986 from the University of Florida. He has worked as Senior Research Scientist at Honeywell, as a Research Engineer and Manager at the Environmental Research Institute of Michigan, and as a faculty member at the University of Wisconsin - Oshkosh, the University of Missouri - Columbia, and the University of Florida, where he is currently a Professor and the Interim Chair of Computer and Information Science and Engineering.

He performed his first research in image processing in 1984 working on algorithms for detection of bridges in Forward Looking Infra-Red imagery as a summer student fellow at Eglin AFB. His dissertation research focused on algebraic methods for parallel image processing. He has since worked on a wide variety of theoretical and applied research problems including fast computing with linear algebra, mathematical morphology, fuzzy sets, Bayesian methods, handwriting recognition, automatic target recognition, bio-medical image analysis, landmine detection, human geography, and hyperspectral and LiDAR image analysis projects. Professor Gader has published hundreds of refereed journal and conference papers, and became a Fellow of the IEEE for his work in landmine detection.

**George Casella** George Casella was a Distinguished Professor in the Department of Statistics, and Distinguished Member of the Genetics Institute, at the University of Florida. He was active in many aspects of statistics, having contributed to theoretical statistics in the areas of decision theory and statistical confidence, to environmental statistics, and, more recently, in statistical genomics, landmine detection, and political science methodology. He also maintained active research interests in the theory and application of Monte Carlo and other computationally intensive methods. He was a Fellow of the American Statistical Association, the Institute of Mathematical Statistics, and an Elected Fellow of the International Statistics Institute. He has been listed as an ISI Highly Cited researcher, and has recently been elected Foreign Member of the Spanish Royal Academy of Sciences.

In other capacities, Casella also served as Theory and Methods Editor of the Journal of the American Statistical Association, 1996-1999, Executive Editor of Statistical Science, 2002-2004, and Joint Editor the Journal of the Royal Statistical Society, Series B. He has served on the Board on Mathematical Sciences of the National Research Council, 1999- 2003, and many other committees of the American Statistical Association and the Institute of Mathematical Statistics. Casella has authored seven textbooks: Variance Components, 1992, with S. R. Searle and C. E. McCulloch; Theory of Point Estimation, Second Edition, 1998, with Erich Lehmann, Statistical Inference Second Edition, 2001, with Roger Berger; Monte Carlo Statistical Methods, Second Edition, 2004, with Christian Robert; Statistical Analysis of Quantitative Traits, 2007, with C. X. Ma and R. Wu, Statistical Design, 2008, and An Introduction to Monte Carlo Methods with R, with Christian Robert.