

PCE: Piecewise Convex Endmember Detection

Alina Zare, *Member, IEEE*, and Paul Gader, *Senior Member, IEEE*

Abstract—A new hyperspectral endmember detection method that represents endmembers as distributions, autonomously partitions the input data set into several convex regions, and simultaneously determines endmember distributions (EDs) and proportion values for each convex region is presented. Spectral unmixing methods that treat endmembers as distributions or hyperspectral images as piecewise convex data sets have not been previously developed. Piecewise convex endmember (PCE) detection can be viewed in two parts. The first part, the ED detection algorithm, estimates a distribution for each endmember rather than estimating a single spectrum. By using EDs, PCE can incorporate an endmember's inherent spectral variation and the variation due to changing environmental conditions. ED uses a new sparsity-promoting polynomial prior while estimating abundance values. The second part of PCE partitions the input hyperspectral data set into convex regions and estimates EDs and proportions for each of these regions. The number of convex regions is determined autonomously using the Dirichlet process. PCE is effective at handling highly mixed hyperspectral images where all of the pixels in the scene contain mixtures of multiple endmembers. Furthermore, each convex region found by PCE conforms to the convex geometry model for hyperspectral imagery. This model requires that the proportions associated with a pixel be nonnegative and sum to one. Algorithm results on hyperspectral data indicate that PCE produces endmembers that represent the true ground-truth classes of the input data set. The algorithm can also effectively represent endmembers as distributions, thus incorporating an endmember's spectral variability.

Index Terms—Convex geometry model, Dirichlet process, endmember, hyperspectral, spectral variability, unmixing.

I. INTRODUCTION

SPECTRAL signatures representing the constituent materials in an imaged scene are referred to as *endmembers* [1]. For example, in an image containing a grassy field and a lake, there may be one endmember spectrum corresponding to the grass, one endmember spectrum corresponding to the lake water, and mixed pixels that would contain spectra composed of some combination of the grass and water endmembers. *Spectral unmixing* is often performed to decompose mixed pixels into their respective endmembers and *abundances*. Abundances are

the proportions of the endmembers in each pixel in a hyperspectral image. Spectral unmixing relies on the definition of a mixing model. The standard mixing model is the *convex geometry model* (also known as the *linear mixing model*) which assumes that every pixel is a convex combination of endmembers in the scene [1]–[3]. Under the convex geometry model, endmembers are the spectra found at the corners of a convex region enclosing all the spectra in a hyperspectral scene. This model can be written as

$$\mathbf{x}_i = \sum_{k=1}^M p_{ik} \mathbf{e}_k + \epsilon_i, \quad i = 1, \dots, N \quad (1)$$

where N is the number of pixels in the image, M is the number of endmembers, ϵ_i is an error term, p_{ik} is the proportion of endmember k in pixel i , and \mathbf{e}_k is the k th endmember. The proportions of this model satisfy the constraints in the following equation:

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

The piecewise convex endmember (PCE) detection algorithm uses the Dirichlet process to determine the number of convex regions needed to describe an input hyperspectral image. For each convex region, PCE uses the endmember distribution (ED) detection algorithm to estimate a distribution for each endmember rather than estimating single spectra. Overall, this algorithm is a stochastic EM algorithm. Proportions that conform to the convex geometry model are estimated for every pixel in each convex region. The ED and Dirichlet process techniques utilize Bayesian machine learning approaches to learn EDs and partition the data set into convex regions while simultaneously estimating proportion values for each data point.

Several endmember detection and spectral unmixing algorithms are described in the literature. Many rely on the pixel purity assumption and search for endmembers within the data set [4]–[7]. By restricting the endmembers to be data points from the scene, these methods cannot find endmembers when pure pixels cannot be found in the image. Many methods have also been developed based on nonnegative matrix factorization [8]–[11], independent components analysis [12], [13], and others [14], [15]. 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 well-suited endmembers for nonconvex data sets. PCE partitions the input hyperspectral set into distinct contexts using the Dirichlet process and estimates a set of EDs for each context. Consider the data shown in Fig. 1. This figure displays the labeled Airborne Visible/Infrared

Manuscript received December 8, 2008; revised September 30, 2009 and December 7, 2009. Date of publication March 15, 2010; date of current version May 19, 2010. This work was supported in part by the U.S. Army Research Office and in part by the U.S. Army Research Laboratory and was accomplished under Cooperative Agreement DAAD19-02-2-0012. The views and conclusions contained in this work are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office, Army Research Laboratory, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

The authors are with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: azare@cise.ufl.edu; pgader@cise.ufl.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2010.2041062

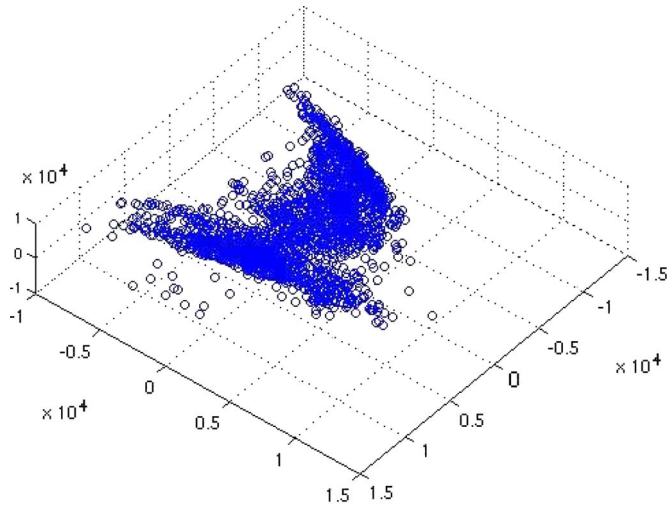


Fig. 1. Plot of labeled pixels from the AVIRIS Indian Pines hyperspectral data set after dimensionality reduction using PCA. Every 10th labeled pixel in the image is plotted.

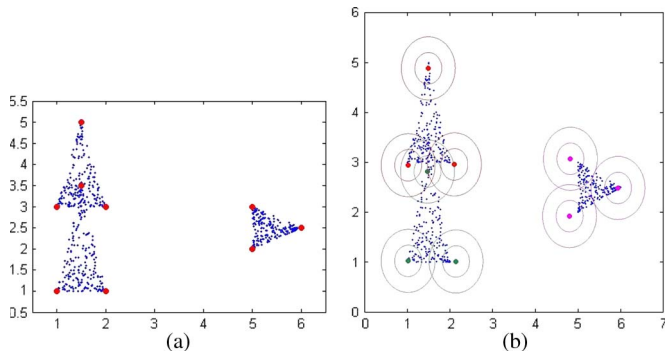


Fig. 2. (a) Two-dimensional data generated from three sets of endmembers. Small points correspond to the input data set. Large points correspond to the endmembers from which the data were generated. Each triangle of data points was generated from three of the endmembers. (b) Two-dimensional data results found using PCE. Small points correspond to the input data set. Large points correspond to the mean endmembers for each ED. Thin curves correspond to both the first and second standard deviation curves from each ED.

Imaging Spectrometer (AVIRIS) Indian Pines hyperspectral data set after using principal component analysis (PCA) for dimensionality reduction from 220 to 3 bands. A striking feature of this real hyperspectral data is that the data set is not convex. The PCE algorithm accounts for this and describes the data using a piecewise convex representation of the data.

For further illustration, consider the 2-D data and its PCE results in Fig. 2. The 2-D data were generated from three sets of endmembers. The piecewise convex representation was able to appropriately partition the data set into three convex regions and find three sets of endmembers. These endmembers cannot be found with methods that assume a single convex region.

The majority of endmember detection and spectral unmixing algorithms use a single spectrum to represent an endmember. PCE incorporates each endmember's spectral variability by representing each endmember as a full distribution rather than a single spectrum. Using a distribution to characterize a hyperspectral class has been previously discussed in [3]. However, in [3], the probability density models are used to treat each class as a cluster rather than an endmember in the convex geometry

model. PCE uses distributions to define endmembers during spectral unmixing and within the convex geometry model. Therefore, hyperspectral pixels are described using convex combinations of EDs. This representation incorporates spectral variability into the convex geometry model and can describe pixels as mixtures of multiple EDs.

PCE uses the Dirichlet process to find the number of convex regions for an input image and partitions all spectra in the image into these regions. Previous endmember detection methods using the Dirichlet distribution, Dirichlet process, and Monte Carlo methods have been developed. In [16], a method that uses the Dirichlet distribution as a prior distribution for the abundance values while estimating endmembers and performing spectral unmixing is presented. In this method, an expectation–maximization type algorithm is employed to iteratively estimate endmembers and abundances. This method, like most previous endmember detection algorithms, assumes the convex geometry model and defines a single convex region to describe the input image. In [17], a reversible jump Markov chain Monte Carlo method to perform unmixing given a set of library endmember spectra is presented. In this algorithm, given a set of known library spectra, abundance values and the subset of endmembers found in each pixel are sampled. Like PCE, this method encourages sparsity in the abundance vectors since each pixel uses only a subset of endmembers in the scene. However, in [17], the endmembers are assumed to be known. In contrast, EDs are estimated in PCE. The authors have previously published a method using the Dirichlet process to perform endmember detection and spectral unmixing [18]. In [18], a Dirichlet process is used to sample abundance values and determine the number of endmembers for a hyperspectral scene. However, in [18], single endmember spectra are used rather than EDs, and a single convex region is used to describe the image rather than the piecewise convex representation presented here.

In the following, Section II presents the ED detection algorithm. Section III reviews the Dirichlet process mixture model. Section IV presents the PCE detection algorithm. Results are shown in Section V. Conclusions and a discussion on future work are given in Section VI.

II. ED: ED DETECTION

The ED detection algorithm has the unique property of representing endmembers as random vectors, thereby calculating endmember distributions rather than single spectra. EDs are found by assuming a model for each endmember and iteratively updating the proportion vectors for each pixel and the parameters of the EDs. ED was developed for use within PCE. However, since ED incorporates spectral variability when performing spectral unmixing and endmember determination, applications for ED may extend beyond use within the PCE algorithm.

Assuming the convex geometry model in (1), each input hyperspectral pixel is a convex combination of the endmembers. In the following, all EDs are assumed to be Gaussian with mean spectra \mathbf{e}_k and known covariance matrices \mathbf{V}_k . It follows that each pixel is a multivariate Gaussian random variable

whose distribution is defined by the convex combination of the endmembers' Gaussian distributions

$$f(\mathbf{x}_j|\mathbf{E}, \mathbf{p}_j) \propto \exp \left\{ -\frac{1}{2} \mathbf{R}^T \left(\sum_{k=1}^M p_{jk}^2 \mathbf{V}_k \right)^{-1} \mathbf{R} \right\} \quad (3)$$

where

$$\mathbf{R} = \mathbf{x}_j - \sum_{k=1}^M p_{jk} \mathbf{e}_k$$

\mathbf{E} is the matrix of mean spectra for all EDs, \mathbf{e}_k and \mathbf{V}_k are the mean spectrum and covariance matrix for the k th ED, M is the number of EDs being determined, and p_{jk} is the j th data point's proportion value for the k th endmember [19]. The joint likelihood for all the hyperspectral pixels is assumed to be the product of the individual likelihoods

$$f(\mathbf{X}|\mathbf{E}, \mathbf{P}) \propto \prod_{j=1}^N f(\mathbf{x}_j|\mathbf{E}, \mathbf{p}_j). \quad (4)$$

Each hyperspectral data point has a unique abundance vector. Although all the data points share the same set of EDs, their unique abundance vectors result in each data point having a unique Gaussian distribution. This is shown in (3) where the maximum likelihood value of the data point \mathbf{x}_j is $\mathbf{p}_j \mathbf{E}$.

In order to provide a tight fit around the input hyperspectral data set, the prior on the endmembers is defined using the sum of squared distances between the means of the EDs. This is similar to the prior on the endmembers used by the Iterated Constrained Endmembers (ICE) and Sparsity Promoting Iterated Constrained Endmembers (SPICE) algorithms for endmember detection [20], [21]

$$f(\mathbf{E}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\mathbf{S}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{4} \sum_{k,l} (\mathbf{e}_k - \mathbf{e}_l)^T \mathbf{S}^{-1} (\mathbf{e}_k - \mathbf{e}_l) \right\}. \quad (5)$$

Initially, the Dirichlet distribution was considered for the prior on the abundance values. However, since the Dirichlet distribution is not a conjugate prior to $f(\mathbf{x}_j|\mathbf{E}, \mathbf{p}_j)$, a simple update formula cannot be used. Instead, constrained nonlinear optimization is required when updating abundance values. As abundances approach zero (which is very desirable and common), the log of the Dirichlet distribution is very steep and approaches $-\infty$, causing instability when using nonlinear optimization techniques. Therefore, the polynomial prior in (6) was developed to be the prior on the abundance vectors

$$P(\mathbf{p}_j) = \frac{1}{Z} \left(\sum_{k=1}^M b_k + 1 - \sum_{k=1}^M b_k (p_{jk} - c_k)^2 \right) \quad (6)$$

where Z is a normalization constant given by

$$Z = \frac{\sqrt{M} \left(\sum_{k=1}^M b_k + 1 \right)}{(M-1)!} - \sqrt{M} \sum_{k=1}^M \frac{b_k}{(M-1)!} \left(\left(c_k - \frac{1}{M} \right)^2 + \frac{M-1}{(M+1)M^2} \right). \quad (7)$$

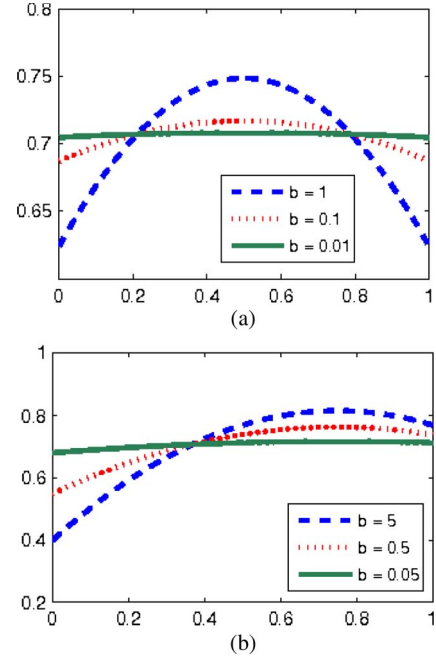


Fig. 3. Plots of ED's abundance prior for $M = 2$ and various \mathbf{c} and \mathbf{b} values. The x -axis is the first abundance value. The y -axis is the prior probability value for the abundance vector. (a) $\mathbf{c} = [0.5, 0.5]$. (b) $\mathbf{c} = [0.75, 0.25]$.

The \mathbf{p} and \mathbf{c} vectors are constrained to be nonnegative and sum to one

$$p_{jk} \geq 0 \quad \forall k = 1, \dots, M; \quad \sum_{k=1}^M p_{jk} = 1$$

$$c_k \geq 0 \quad \forall k = 1, \dots, M; \quad \sum_{k=1}^M c_k = 1.$$

The vector \mathbf{c} is the maximum likelihood value for \mathbf{p} . The b_k terms control the steepness of the prior. The polynomial prior prefers abundance vectors which are binary, i.e., vectors with a single abundance with value one and the rest with value zero. This is a result of the normalization constant Z .

The numerator of the polynomial prior is maximized when \mathbf{c} is equal to \mathbf{p} . The normalization constant in the denominator is minimized when \mathbf{c} is binary. Thus, when both the \mathbf{p} and \mathbf{c} vectors are binary, the polynomial prior is maximized. This property introduces sparsity within abundance vectors which, when combined with the flexibility achieved by representing endmembers by distributions, represents a major advance in automated determination of meaningful endmembers and abundances.

If several endmembers adequately describe a data point, the polynomial prior will place all weight on one endmember rather than spreading the abundance across endmembers, encouraging the method to use the minimum number of endmembers needed. Furthermore, many different points can be assigned abundance values of one with respect to a given endmember because of the variance of the ED. Examples of this prior for abundance vectors of length two are shown in Fig. 3. In addition, plots showing the abundance prior as a function of \mathbf{c} are shown in Fig. 4.

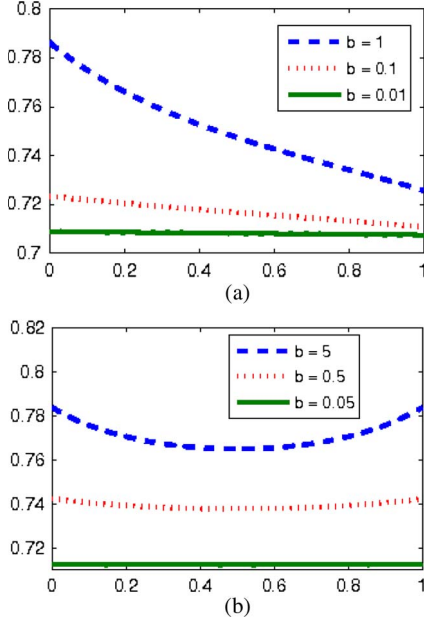


Fig. 4. Plots of ED's abundance prior for $M = 2$ and various \mathbf{p} and \mathbf{b} values. The x -axis is the first \mathbf{c} value. The y -axis is the prior probability value for \mathbf{c} . (a) $\mathbf{p} = [0.45, 0.55]$. (b) $\mathbf{p} = [0.5, 0.5]$.

The algorithm proceeds by iteratively maximizing $f(\mathbf{X}|\mathbf{E}, \mathbf{P})f(\mathbf{E})f(\mathbf{P})$, where $f(\mathbf{P})$ is the joint likelihood of all the abundance vectors. Given initial estimates of the EDs and \mathbf{c} in the polynomial prior, abundance vectors are updated by maximizing the log of $f(\mathbf{x}_j|\mathbf{E}, \mathbf{p}_j)P(\mathbf{p}_j)$ given by (3) and (6) with respect to \mathbf{p}_j for each data point. This is a constrained nonlinear optimization problem. In the current Matlab implementation, this is maximized using Matlab's `fmincon` function in the optimization toolbox. Following an update of the abundance vectors, $f(\mathbf{X}|\mathbf{E}, \mathbf{P})P(\mathbf{E})$ defined in (4) and (5) is maximized with respect to means of the EDs, \mathbf{e}_k for $k = 1, \dots, M$. This maximization is performed directly by taking the derivative of the log of the product and setting it equal to zero, as shown in (8). The following equation does not constrain \mathbf{e}_k to be positive:

$$\begin{aligned} \mathbf{e}_k^T = & \left(\sum_{j=1}^N \left(\mathbf{x}_j - \sum_{l \neq k} a_{jl} \mathbf{e}_l \right) \right)^T \\ & \times \left(\sum_{l \neq k} a_{jl}^2 \mathbf{V}_l \right)^{-1} a_{jk} + \sum_{l \neq k} \mathbf{e}_l^T \mathbf{S}^{-1} \\ & \times \left(\sum_{j=1}^N a_{jk}^2 \left(\sum_{l \neq k} a_{jl}^2 \mathbf{V}_l \right)^{-1} + (M-1) \mathbf{S}^{-1} \right)^{-1}. \quad (8) \end{aligned}$$

The third step of the iteration updates the \mathbf{c} vector in the abundance prior given the abundance vectors for all the data points. The third step is also a nonlinear optimization problem solved using Matlab's `fmincon` function.

Although the ED algorithm was developed for use within the PCE algorithm, applications of the ED algorithm may extend beyond this. This may occur since, using EDs, the spectral vari-

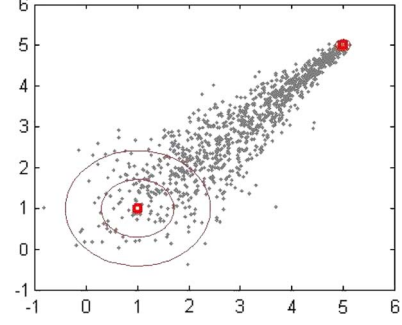


Fig. 5. Data points generated from linear combinations of two EDs. The ED centered at (5,5) has a diagonal covariance whose elements are all equal to 0.005. The ED centered at (1,1) has a diagonal covariance whose elements are all equal to 0.5. Data points are shown in light gray. Mean spectra and standard deviation curves for the EDs are shown with squares and thin curves, respectively. Both the first and second standard deviation curves are shown for each ED.

ation which occurs due to varying environmental conditions or inherent variability can be measured in controlled environments and then incorporated and utilized during endmember detection or spectral unmixing. For example, if endmember means and covariances are estimated from a spectral library, these can be held constant during the ED algorithm while spectral unmixing is performed, and if necessary, additional EDs are learned.

The use of the ED model can represent a wide variety of data. For example, the data points in Fig. 5 were generated using two EDs. The standard model using convex combinations of single endmember spectra would require three endmembers to represent the data while maintaining a small reconstruction error.

As can be seen in Fig. 1, some of the data appear to follow a linear combination of Gaussian distributions like that shown in Fig. 5.

III. OVERVIEW OF GIBBS SAMPLING FOR THE DIRICHLET PROCESS MIXTURE MODEL

The Dirichlet process mixture model applies a Dirichlet process prior to the mixing proportions of a mixing model, allowing for a countably infinite number of mixture components [18], [22]. Consider N data points $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, each of which is assumed to have been independently generated by some distribution $f_i(\cdot, \phi_i)$, where ϕ_i is the vector of parameters that defines the process generating observation \mathbf{x}_i . Under the Dirichlet process mixture model, ϕ_i is generated by some unknown distribution G [23]. Then, G is distributed according to the Dirichlet process, i.e., $\mathcal{D}(\alpha G_0)$, where G_0 is the base distribution and α is the concentration parameter [22]. Therefore, the complete model can be written as [22], [24]

$$\begin{aligned} \mathbf{x}_i & \sim f(\cdot|\phi_i) \\ \phi_i & \sim G \\ G & \sim \mathcal{D}(\alpha G_0). \end{aligned} \quad (9)$$

Under this model, the values $\phi_i, i = 1, \dots, N$, generated from G are members of a set of $M \leq N$ distinct values denoted as $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M\}$ corresponding to the parameters for each

mixture component. In other words, several data points can be generated from the same mixture component [23].

To simplify the model, G can be integrated out to express the prior of each ϕ_i in terms of the base distribution G_0 and all other parameter sets [22]–[25]

$$\phi_i | \phi_{-i} \sim \frac{1}{\alpha + N - 1} \sum_{j=1, i \neq j}^N \delta(\phi_j) + \frac{\alpha}{\alpha + N - 1} G_0 \quad (10)$$

where ϕ_{-i} is the set of component distributions for all data points other than i , N is the number of data points, $\delta(\phi_i)$ is the distribution over parameters with all weight concentrated at parameter set ϕ_i , and G_0 is the prior distribution for the component parameters [24].

As shown in [24], the likelihood of a data point, given the component parameters, can be combined with the probability of a class label given all other labels in (10). Then, the Gibbs sampler can be used to sample indicator variable values and component parameter values. Furthermore, if G_0 is a conjugate prior to the likelihood distributions (component distributions), then only the indicator variables of the observations need to be sampled. In this case, the conditional distributions are expressed as [22]

$$\begin{aligned} f(c_i = c_j \text{ for some } j \neq i | \mathbf{c}_{-i}, \mathbf{x}_i) \\ = C \frac{n_{-i,j}}{\alpha + N - 1} \int f(\mathbf{x}_i | \boldsymbol{\theta}) H_{-i,c_j}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ f(c_i \neq c_j \forall j \neq i | \mathbf{c}_{-i}, \mathbf{x}_i) \\ = C \frac{\alpha}{\alpha + N - 1} \int f(\mathbf{x}_i | \boldsymbol{\theta}) G_0(\boldsymbol{\theta}) d\boldsymbol{\theta} \end{aligned} \quad (11)$$

where C is a normalizing constant and H_{-i,c_j} is the posterior distribution of the component parameters given prior G_0 and current indicator values \mathbf{c}_{-i} [22]. Given that G_0 is a conjugate prior to the likelihood distributions, the integrals in (11) can be computed analytically. These integrals remove the need to include component parameters in the Markov chain which significantly reduces the search space for the Gibbs sampler [22].

IV. PCE: PCE DETECTION

In this section, a novel method for endmember detection using the Dirichlet process is presented. Existing endmember detection algorithms generally assume that all pixels in a hyperspectral image are convex combinations of a single set of endmembers. However, some hyperspectral images may be better represented using several sets of endmembers. This algorithm partitions the input hyperspectral data set into convex regions each with its own set of EDs. By using the Dirichlet process, the PCE detection algorithm learns the number of convex regions needed to represent an input hyperspectral image and estimates EDs and proportion values for each convex region.

This method differs from the Dirichlet process mixture model since each convex region is represented with a set of EDs for which each data point has a unique abundance vector. Thus, as previously shown in (3), each data point is a random variable with a unique distribution. Each data point having a

unique distribution contrasts with the Dirichlet Process Mixture Model (DPMM) approach where data points from each cluster are assumed to be identically distributed.

PCE performs Gibbs sampling with Dirichlet process priors to sample the partition to which each data point belongs. The probability of sampling a partition is computed using the likelihood of a data point belonging to a convex combination of the associated EDs

$$\begin{aligned} f(r_i = r_j \quad j \neq i | \mathbf{r}_{-i}, \mathbf{x}_i) \\ = C \frac{n_{-i,j}}{\alpha + N - 1} \int f(\mathbf{x}_i | \mathbf{p}_i^{r_j}, \mathbf{V}^{r_j}, \mathbf{E}^{r_j}) H_{-i,r_j} \\ \times (\mathbf{E}^{r_j}, \mathbf{V}^{r_j}, \mathbf{P}^{r_j}) d\mathbf{p}_i^{r_j} d\mathbf{E}^{r_j} \\ = C \frac{n_{-i,j}}{\alpha + N - 1} \mathcal{N} \\ \times (T(T + S)^{-1} \mathbf{p} \mathbf{E}^{r_j} \\ + S(T + S)^{-1} \mathbf{c} \mathbf{E}^{r_j}, S + T(T + S)^{-1} S) \\ f(r_i \neq r_j \quad \forall j \neq i | \mathbf{r}_{-i}, \mathbf{x}_i) \\ = C \frac{\alpha}{\alpha + N - 1} \int f(\mathbf{x}_i | \mathbf{E}^*) G_0(\mathbf{E}^*) d\mathbf{E}^* \\ = C \frac{\alpha}{\alpha + N - 1} \mathcal{N} \\ \times (\mathbf{V}_0(\mathbf{V}_0 + \mathbf{V})^{-1} \mathbf{x}_i \\ + \mathbf{V}(\mathbf{V}_0 + \mathbf{V})^{-1} \mu_0, (\mathbf{V}_0^{-1} + \mathbf{V}^{-1})^{-1} + \mathbf{V}) \end{aligned} \quad (12)$$

where r_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 r_j , N is the total number of data points, and α is the innovation parameter for the Dirichlet process. The matrices T and S correspond to $\sum_k c_k^2 V_k$ and $\sum_k p_{ik}^2 V_k$, respectively. The matrices \mathbf{V} and \mathbf{V}^{r_j} are the covariance matrices associated with new and existing EDs. In the current implementation of this algorithm, all covariance matrices for EDs are set to the same constant matrix value.

The prior distribution G_0 is Gaussian, where the mean μ_0 is set to the mean of the input data set and the covariance \mathbf{V}_0 is constant, i.e.,

$$G_0 = \mathcal{N} \left(\mu_0 = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_j, \mathbf{V}_0 \right). \quad (13)$$

The prior distribution combined with α , the innovation parameter in the Dirichlet process prior, dictates the probability of generating a new partition. The covariance matrix \mathbf{V}_0 is set to a large value to approximate a broad uniform prior over the data set.

Assuming that each ED is Gaussian with a known covariance matrix, the likelihood for an existing partition $f(\mathbf{x}_i | \mathbf{p}_i^{r_j}, \mathbf{V}^{r_j}, \mathbf{E}^{r_j})$ is determined by (3). The vector $\mathbf{p}_i^{r_j}$ contains the proportion values for the current data point in partition r_j . These proportion values are determined by maximizing the log

of $f(\mathbf{x}_j|\mathbf{E}, \mathbf{p}_j)P(\mathbf{p}_j)$ given by (3) and (6) with respect to \mathbf{p}_j given the endmembers of the partition \mathbf{E}^{rj} . The $f(\mathbf{x}_j|\mathbf{p}_j^{rj}, \mathbf{V}^{rj}, \mathbf{E}^{rj})$ value measures the ability of a set of endmembers to represent a data point by computing the distance between the data point and $\mathbf{p}_j^{rj} \mathbf{E}^{rj}$.

The distribution $H_{-i,r}(\mathbf{E}^r, \mathbf{V}^r, \mathbf{P}^r)$ is the prior distribution updated based on the data points assigned to the r th partition

$$H_{-i,r}(\mathbf{E}^r, \mathbf{P}^r) = \mathcal{N}\left(\mathbf{cE}^r, \sum_{k=1}^M c_k^2 \mathbf{V}_k^r\right) \quad (14)$$

where \mathbf{c} is the center from the polynomial abundance prior determined by maximizing (6) given \mathbf{P}^r and \mathbf{V}^r . By incorporating this updated prior, the likelihood depends not only on the distance to \mathbf{pE} but also to \mathbf{cE} . When the covariance matrices for EDs are equal, the updated prior depends on the distance to a point on the line segment connecting \mathbf{pE} and \mathbf{cE} , namely, $w_1 \mathbf{pE} + w_2 \mathbf{cE}$, where $w_1 = (\sum_{k=1}^M c_k^2) / (\sum_{k=1}^M (c_k^2 + p_{jk}^2))$ and $w_2 = (\sum_{k=1}^M p_{jk}^2) / (\sum_{k=1}^M (c_k^2 + p_{jk}^2))$.

As stated previously and shown in line 12 of the following pseudocode, in each iteration of the algorithm, a partition is sampled for the current data point. A partition is sampled by computing the likelihood of a data point belonging to each existing partition and the likelihood of a data point generating a new partition using (12). The unit interval is then divided into regions whose lengths are equal to each partition's normalized likelihood value. A random value from the unit interval is then generated. The corresponding partition whose region includes the generated random value is the partition that is sampled for the current data point.

After a partition is sampled, the parameters of the sampled partition are updated. This is done by updating the prior on the abundances [see (6)] with respect to \mathbf{c} for the given partition. After one or more iterations of the partition sampling scheme using the Dirichlet process, the EDs and all proportion values are updated using a designated number of iterations of the ED algorithm.

Several items in the following PCE pseudocode differ from the standard DPMM method. As stated in lines 10 and 13 of the pseudocode, a partition's parameters are updated when a data point is removed or added to the partition by updating the partition's \mathbf{c} vector in the polynomial abundance prior. In contrast, for the standard Gaussian DPMM method, the mean of the Gaussian cluster would be updated instead. Lines 16–18 of the pseudocode also differ from the standard DPMM method. After a set number of Gibbs sampling iterations in PCE, each partition's endmembers and proportion matrices are updated. In the standard DPMM, all values associated with each cluster are updated in each Gibbs iteration. PCE essentially performs a series of several Gibbs sampling runs each with a new set of endmembers. Overall, this algorithm is a stochastic EM method.

- 1: Initialize Partitions
- 2: **for** $r \leftarrow 1$ to $R_{initial}$ partitions **do**
- 3: Initialize \mathbf{E}^r and \mathbf{P}^r using ED
- 4: **end for**

- 5: **for** $k \leftarrow 1$ to number of total iterations **do**
- 6: **for** $i \leftarrow 1$ to number of Gibbs sampling iterations **do**
- 7: Randomly reorder data points in X
- 8: **for** $j \leftarrow 1$ to number of data points **do**
- 9: Remove \mathbf{x}_j from its current partition
- 10: Update the partition's \mathbf{c}
- 11: Compute Dirichlet process partition probabilities for \mathbf{x}_j using Equation (12).
- 12: Sample a partition for \mathbf{x}_j based on the Dirichlet process partition probabilities
- 13: Update new partition's \mathbf{c}
- 14: **end for**
- 15: **end for**
- 16: **for** $r \leftarrow 1$ to R_k partitions **do**
- 17: Update \mathbf{E}^r and \mathbf{P}^r using ED
- 18: **end for**
- 19: **end for**
- 20: $R_{final} = R_k$

V. EXPERIMENTAL RESULTS

The PCE algorithm was tested on real hyperspectral data. Results are shown and compared to results found using the ICE and SPICE endmember detection algorithms [20], [21]. The ICE algorithm is an iterative algorithm that alternates between solving for endmembers and abundances while holding the other constant. Endmembers and abundances are estimated by maximizing an objective function containing two terms, a squared error term between data points and their reconstruction using endmembers and abundances and a sum-of-squared distances term between the estimated endmembers to promote a tight fit around the data set. The SPICE algorithm is an extension of ICE which adds a sparsity promoting Laplacian prior to the objective function. This sparsity promoting prior is used to determine the number of endmembers while simultaneously estimating abundance and endmember spectra.

A. PCE Detection Results on the Aviris Indian Pines Data

PCE was tested on the labeled pixels of the June 1992 AVIRIS data set collected over the Indian Pines test site in an agricultural area of northern Indiana. The image has 145×145 pixels with 220 spectral bands and contains approximately two-third agricultural land and one-third forest and other elements. The soybean and corn crops in the image are in early growth stages and thus have only about a 5% crop cover [27], [28]. The remaining field area is soil covered with residue from the previous crop. The no till, min till, and clean till labels indicate the amount of previous crop residue remaining. No till corresponds to a large amount of residue, min till has a moderate amount, and clean till has a minimal amount of residue [28]. Fig. 6 shows band 10 (approximately $0.49 \mu\text{m}$) and the ground truth of the data set. Only 49% of the pixels in the image have ground-truth information [28].

Prior to running PCE, the data dimensionality was reduced from 220 bands to 6 dimensions using PCA. A total of 1037 pixels (every 10th labeled pixel) was selected from the data set and used in the PCE algorithm. Partitions on these data were

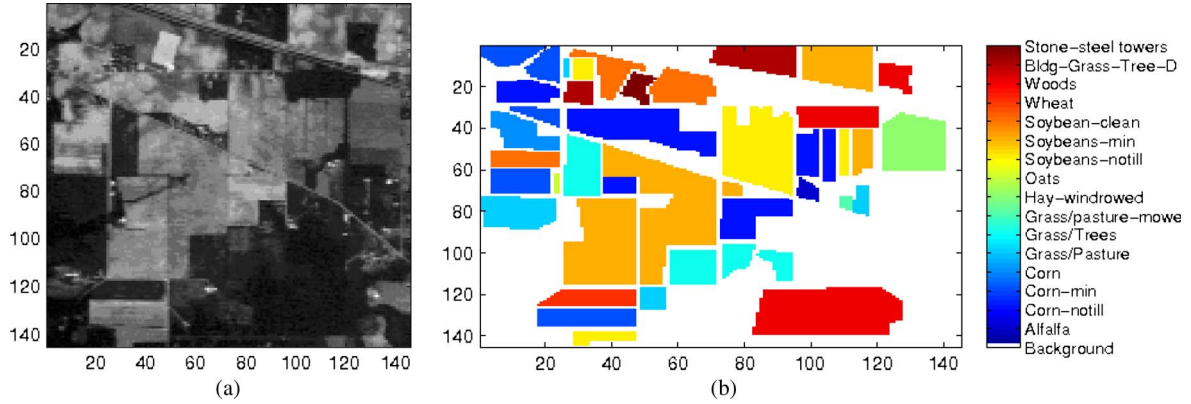


Fig. 6. (a) Band 10 ($\sim 0.5 \mu\text{m}$) of the AVIRIS Indian Pines data set. (b) Ground truth of the AVIRIS Indian Pines data set.

TABLE I
PARAMETER VALUES USED TO GENERATED PCE RESULTS. ALL COVARIANCE MATRICES
ARE DIAGONAL WITH ELEMENTS EQUAL TO THE VALUES SHOWN IN THE TABLE

Data Set	Data dimen.	Variance of data	Likelihood variance	α	ED likelihood variance	ED SSD variance	b_k
2D Data	2	2.16	0.005	2	0.010	1.000	0.001
PCA IP	6	0.05	0.005	1	0.005	0.001	0.001
Full Spectra Cuprite	51	0.12	0.005	1	0.005	0.005	0.001

initialized using the Kernel Global Fuzzy C-Means (KG-FCM) algorithm and the DPMM algorithm resulting in three partitions [29]. After initial partitions were found, endmembers for each partition were initialized using the ED algorithm. Each partition was restricted to three endmembers. The parameters used to generate results shown are listed in Table I.

In order to compute abundance maps for the entire image, every data point was unmixed using each partition's set of endmembers, and the likelihood under each partition was computed. Every data point was then assigned to partition with the largest likelihood value. In addition, all endmembers whose maximum proportion value was less than 0.05 were removed. Following these steps, 13 clusters were found with a total of 14 endmembers. Fig. 7 shows the abundance maps for endmembers associated with more than 20 pixels.

For comparison, SPICE was run on the labeled Indian Pines data set. A total of 1037 normalized pixels (every 10th labeled pixel) was selected from the image and used to determine the endmembers. The initial number of endmembers, μ , and Γ were set to 20, 0.01, and 0.1, respectively. Six endmembers were found for the labeled pixels of the Indian Pines scene. The abundance maps are shown in Fig. 8. The endmembers roughly correspond to the following classes: (A) grass/pasture and woods; (B) hay-windrowed, alfalfa, and grass/pasture-mowed; (C) and (E) correspond to corn and soybean; (D) stone-steel towers; and (F) grass/trees, wheat, and woods.

The distribution of abundance values among endmembers in each ground-truth class found by SPICE is shown in Fig. 9. These distributions were computed by summing all the abundance values associated with an endmember in each ground-truth class. Each distribution was normalized by dividing by the number of points in the corresponding ground-truth class

where G_k is the set of pixels in ground-truth class k , N_k is the number of points in ground-truth class k , a_{il} is the i th data points' abundance value for the l th endmember, and h_{lk} is the k th value corresponding to the l th endmember.

For comparison with the SPICE results in Fig. 9, the normalized distributions of the abundance values across each endmember found by PCE were computed using (15). The distributions found are shown in Fig. 10. When comparing the SPICE and PCE distributions, the PCE results for each ground-truth class are significantly more concentrated than the SPICE results. This fact can be measured by computing Shannon's entropy for the normalized distribution associated with each ground-truth class [30]. A smaller entropy value indicates that a fewer number of endmembers are being used to describe each ground-truth class and that the endmembers are better representatives of the ground-truth classes. The sum of the Shannon entropies for SPICE's distributions of abundance values comes to 19.0. In contrast, the sum of the Shannon entropies for PCE's distributions of abundances is significantly lower at 9.4. This indicates that PCE produces endmembers which better represent the ground-truth classes.

Consider the wheat ground-truth class in the SPICE and PCE results. The SPICE abundance map associated with the most amount of wheat is shown in Fig. 8(f), and the corresponding distribution of abundance values is shown in Fig. 9(m). By examining the abundance map, it can be seen that many pixels other than wheat have nonzero abundance values associated with wheat's SPICE endmember. In contrast, very few pixels outside of the wheat ground-truth class share wheat's endmember. This is shown in the PCE abundance map in Fig. 7(i). Furthermore, by examining the SPICE distribution of abundance values for wheat, only about 60% of the wheat pixels' abundance values are associated with that endmember, whereas 100% of wheat's abundance values are placed with the associated endmember found using PCE.

$$h_{lk} = \frac{\sum_{i: \mathbf{x}_i \in G_k} a_{il}}{N_k} \quad (15)$$

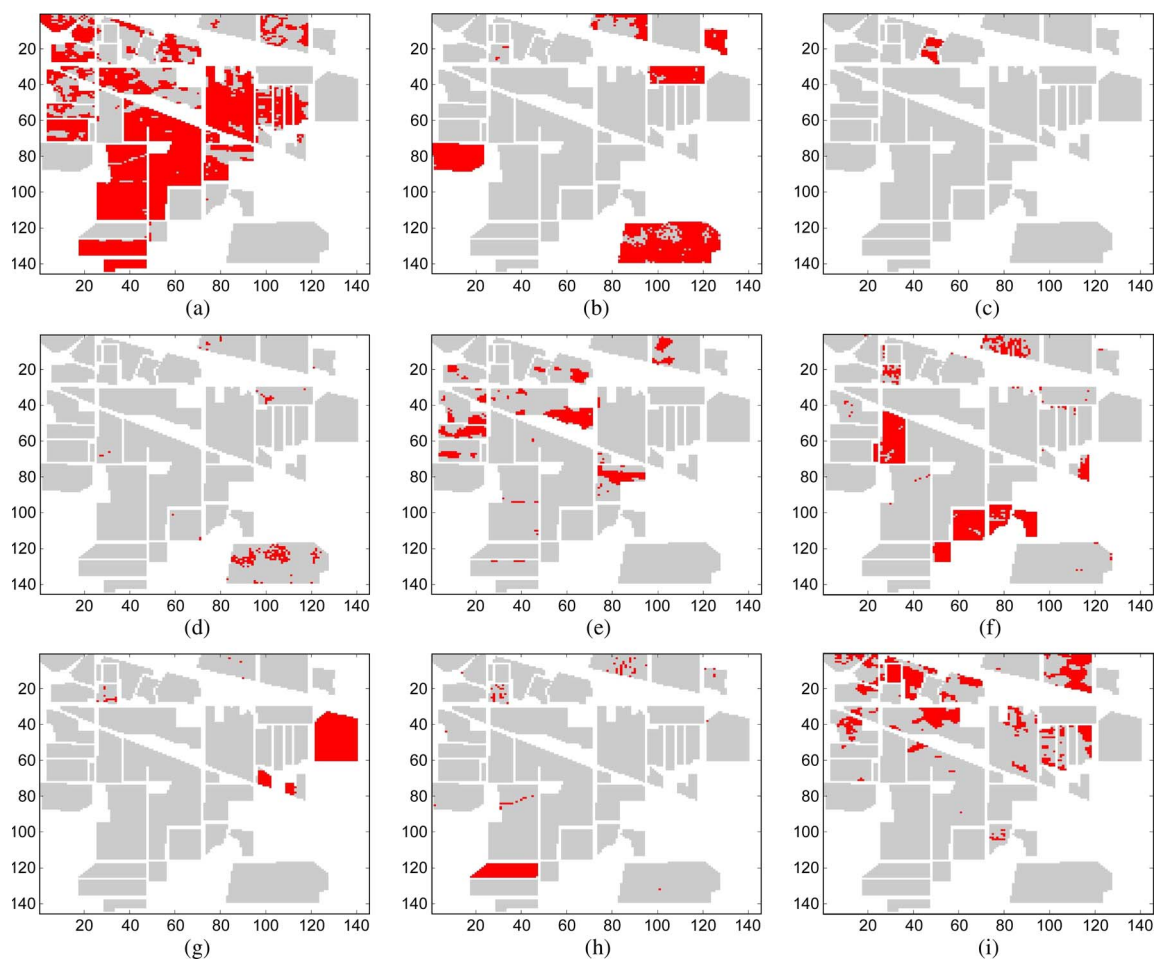


Fig. 7. Abundance maps found using PCE on labeled PCA-reduced AVIRIS Indian Pines data. Only abundance maps with greater than 20 associated pixels are shown. Pixels in white are unlabeled. Pixels in gray indicate pixels from another convex partition. Remaining pixels have abundance values of 1.

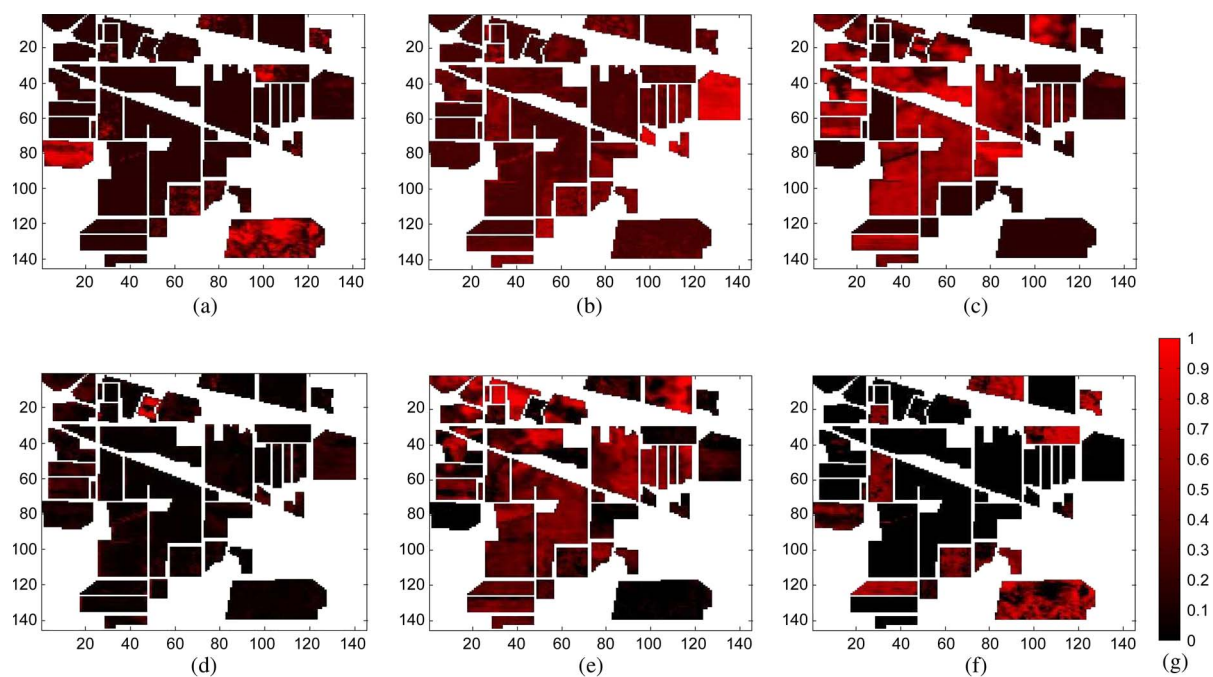


Fig. 8. Abundance maps generated by SPICE on the labeled AVIRIS Indian Pines data set. Pixels in white correspond to unlabeled pixels. Remaining pixels range from black (abundance value of zero) to abundance values of one. The scale for all the images is shown in (g).

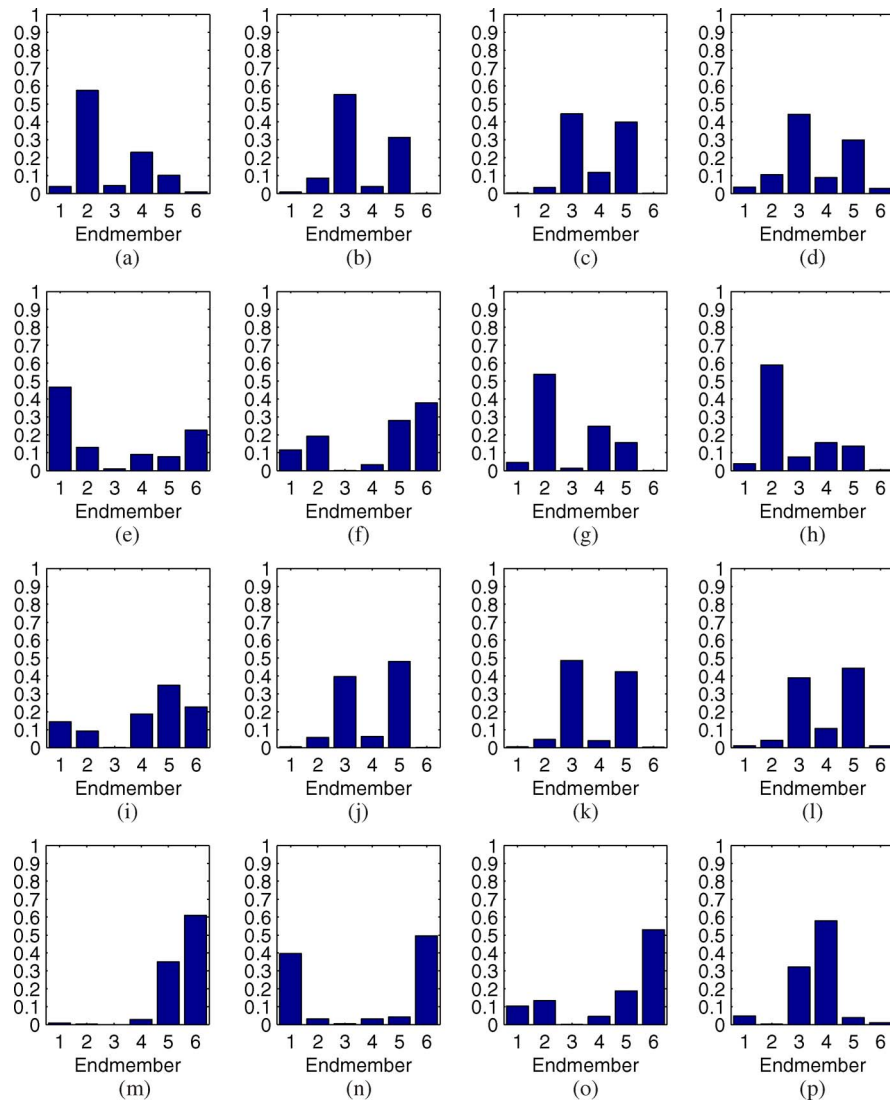


Fig. 9. Distribution of abundance values over the endmembers found by SPICE on labeled AVIRIS Indian Pines data in each ground-truth class. Distributions were computed according to (15). The sum of these distributions' Shannon entropy values is 19.0. The distributions correspond to the following ground-truth classes: (a) alfalfa, (b) corn-notill, (c) corn-min, (d) corn, (e) grass/pasture, (f) grass/trees, (g) grass/pasture-mowed, (h) hay-windrowed, (i) oats, (j) soybeans-notill, (k) soybeans-min, (l) soybean-clean, (m) wheat, (n) woods, (o) building-grass-trees-drive, and (p) stone-steel towers.

For the stone-steel towers ground-truth class, more than 70% of the pixels assigned to a single endmember using PCE and that endmember is not associated with any other ground-truth classes. The SPICE endmember most associated with the stone-steel towers ground-truth class is also used by every other ground-truth class.

The hay-windrowed [Fig. 10(h)], grass/pasture-mowed [Fig. 10(g)], and alfalfa [Fig. 10(a)] PCE distributions of abundance values show that they are associated with the same endmember. This can also be seen in the abundance map in Fig. 7(h). The corresponding SPICE distributions of abundance values for hay-windrowed, grass/pasture-mowed, and alfalfa in Fig. 9(a), (h), and (g) show that the three ground-truth classes have similar distribution shapes and share the same endmembers. However, the abundances found by SPICE are spread among three endmembers, whereas PCE placed their full weight with one endmember.

Soybean and corn ground-truth classes constitute a large majority of the Indian Pines scene. In the SPICE results,

abundance values associated with the soybean and corn classes are spread over all of the six endmembers found. In contrast, the PCE endmember results places nearly all soybean and corn abundances with the 2nd, 6th, and 10th endmembers.

Another indication that PCE is producing representative endmembers is found with the building/grass/trees/drive ground-truth class. This class is composed of a variety of material types. Interestingly, this is clearly shown in the class' PCE distribution of abundance values [Fig. 10(o)]. The abundance values for the class are spread across many endmembers.

In order to verify that the difference in the results between PCE and SPICE is not due to different data dimensionality and a different number of endmembers, the ICE algorithm was run on the same AVIRIS PCA-reduced Indian Pines data set discussed in this section. The ICE algorithm was employed rather than SPICE since the number of endmembers can be set to the same number found by PCE. The ICE algorithm was restricted to 14 endmembers, and the μ parameter was set to 0.01.

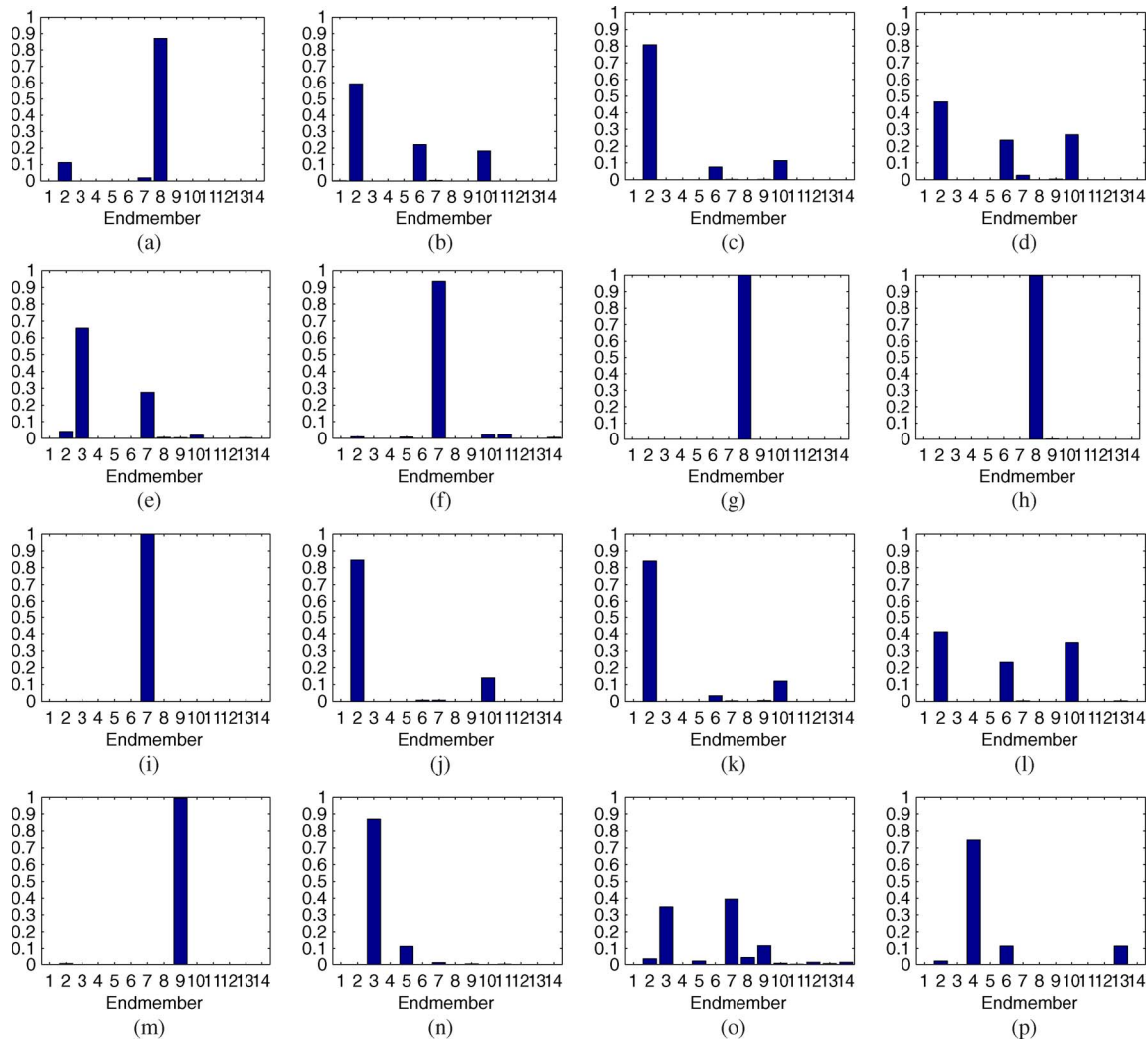


Fig. 10. Distribution of abundance values over the PCE endmember results on labeled PCA-reduced AVIRIS Indian Pines data in each ground-truth class. Distributions were computed according to (15). The sum of the distributions' Shannon entropy values is 9.4. The histograms correspond to the following ground-truth classes: (a) alfalfa, (b) corn-notill, (c) corn-min, (d) corn, (e) grass/pasture, (f) grass/trees, (g) grass/pasture-mowed, (h) hay-windrowed, (i) oats, (j) soybeans-min, (k) soybeans-clean, (l) soybean-clean, (m) wheat, (n) woods, (o) building-grass-trees-drive, and (p) stone-steel towers.

The sum of the Shannon entropies of ICE's distribution of abundance value from these results is 29.2. In comparison, PCE's value was 9.4. Therefore, although ICE was restricted to the same number of endmembers found using PCE, ICE did not produce endmembers that represent the ground-truth classes as well as PCE.

B. Results on Aviris Cuprite Data

PCE was also run on the AVIRIS Cuprite "Scene 4" data set collected over Cuprite, NV. These data contain 51 spectral bands in the range of 1978–2477 nm. PCE was applied to this data set to examine the quality of endmembers found by comparing the PCE results to spectra from the U.S. Geological Survey (USGS) Spectral Library with materials known to be found in the Cuprite scene [31] and results obtained using the vertex component analysis (VCA) [6] algorithm. As in [20], PCE was run on a subset of pixels from the image using "candidate points" selected using the pixel purity index (PPI) [5]. The candidate points in our experiments were chosen from 10 000 random projections where the PPIs of points within

a distance of two from the boundary of the projection were incremented. The 1011 pixels with the highest PPI were used as the candidate points. A PPI threshold that allowed as close to 1000 pixels as possible (many pixels have the same PPI) was chosen. This is the same data set used in [21]. Three partitions on this data set were initialized using the KG-FCM algorithm. After initial partitions were found, the endmembers for each partition were initialized using the ED algorithm. Each partition was restricted to four endmembers. The parameters used to generate results shown are listed in Table I.

PCE found ten partitions with 14 EDs. EDs whose abundance values summed to less than five over the entire data set were removed. In order to determine the quality of endmembers, they were compared to 100 USGS spectra of materials known to be found in the AVIRIS Cuprite scene: Alunite, Buddingtonite, Calcite, Chalcedony, Desert Varnish, Kaolinite, Montmorillonite, Muscovite, Nontronite, and Sphene. Fig. 11 shows the comparison of the endmembers found to the closest matching 2006 USGS Spectral Library spectra. VCA was also run on this data set with the same number of endmembers found by PCE. Fig. 12 shows the comparison of the VCA

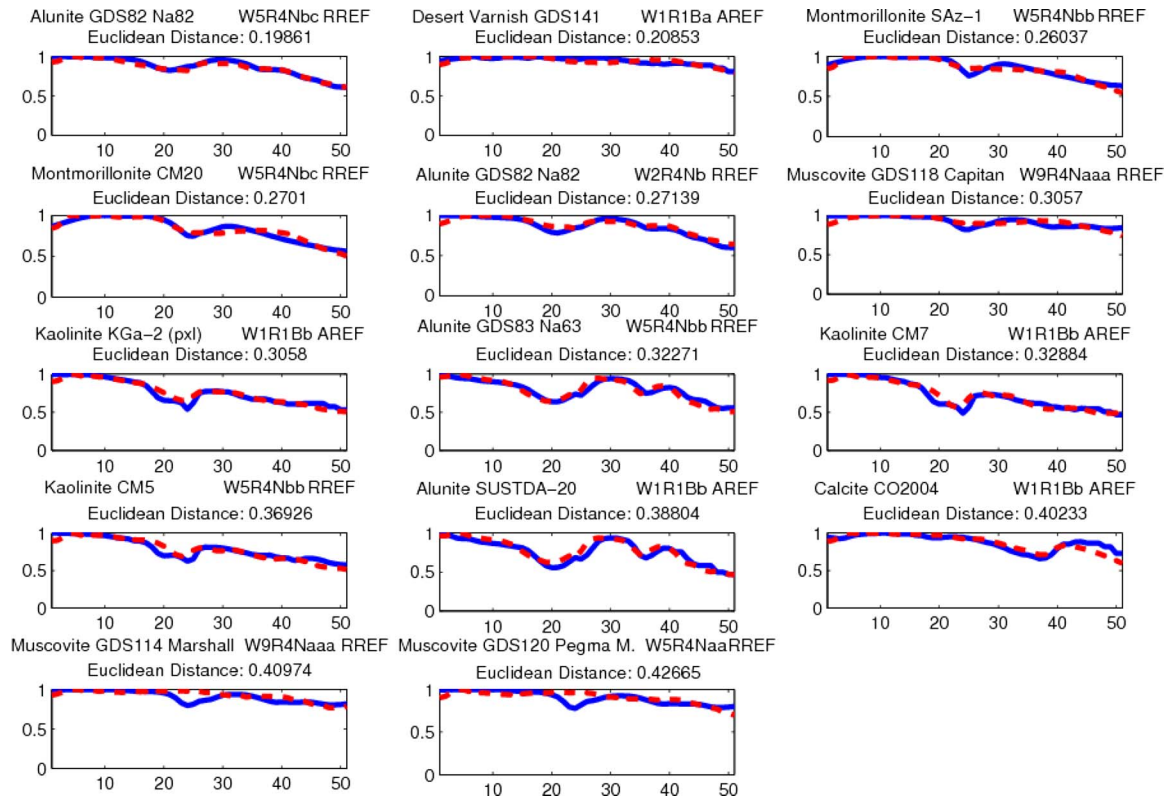


Fig. 11. Comparison of the mean of the EDs found using PCE on the AVIRIS Cuprite data set to USGS spectra. The estimated spectra are shown with dashed lines. The USGS spectra are shown with solid lines.

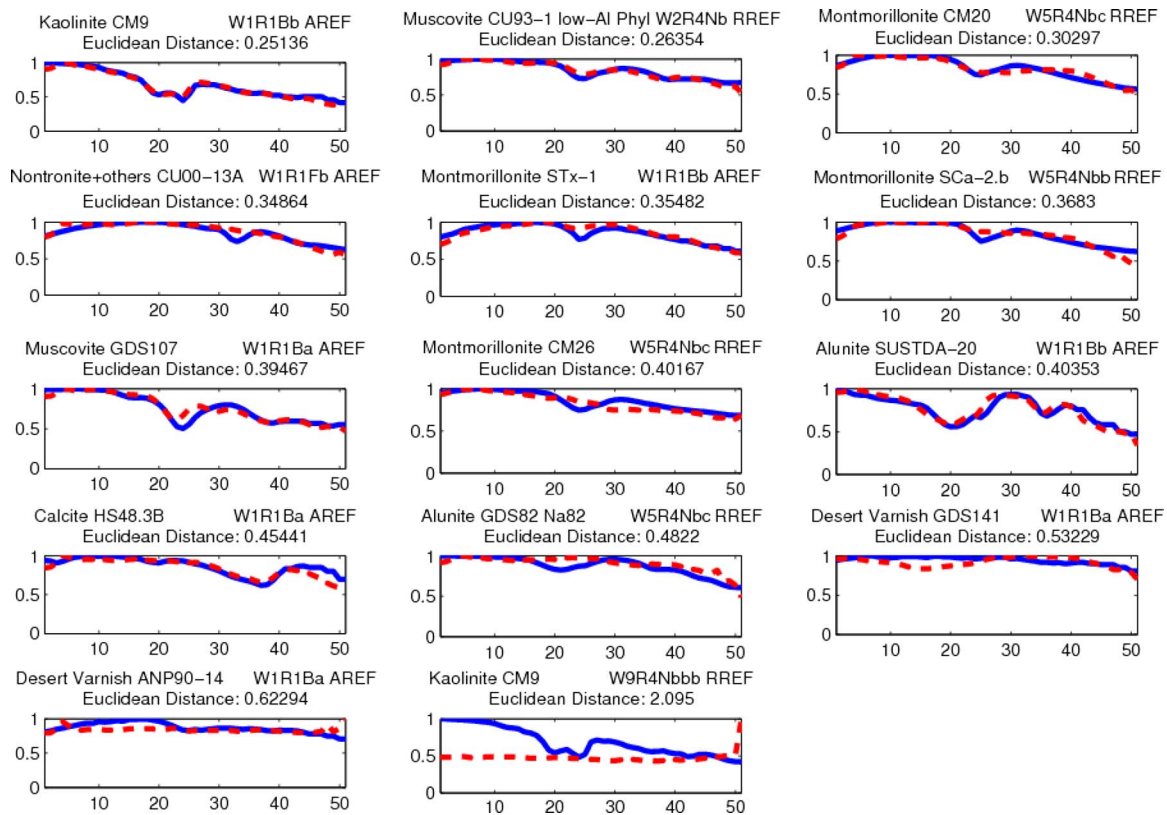


Fig. 12. Comparison of the mean of the EDs found using VCA on the AVIRIS Cuprite data set to USGS spectra. The estimated spectra are shown with dashed lines. The USGS spectra are shown with solid lines.

endmembers to the closest matching 2006 USGS Spectral Library spectra. The endmembers found were compared to the USGS Spectral Library spectra using the Euclidean distance. The title above each plot in Figs. 11 and 12 list the name of the matching USGS library spectra and the Euclidean distance between the estimated endmembers and the spectral library spectra. The mean Euclidean distance between the estimated and spectral library spectra using PCE was 0.32, whereas for VCA, it was 0.53, indicating that the PCE algorithm found endmembers that better match the spectra from the USGS Spectral Library. Furthermore, a right-sided t-test in Matlab showed that the differences between the PCE distances and VCA distances were significant with a 96% confidence interval.

These Cuprite results were found with 100 iterations of the PCE algorithm. Each iteration consists of sampling a partition for each hyperspectral pixel and optimizing the endmembers and abundances for each partition using the ED method. Each iteration took 5.3 min. However, similar results were obtained in 25 iterations with the mean Euclidean distance for PCE at 0.33 and the mean Euclidean distance for VCA at 0.53. Future work will include determining the number of iterations required for PCE to obtain the endmember estimates as well as methods to speed up each iteration of the PCE algorithm.

VI. CONCLUSION AND FUTURE WORK

The PCE detection algorithm used the Dirichlet process to learn the number of convex regions needed to describe an input hyperspectral scene. For each convex region, an individual set of EDs and proportion values were determined. In contrast, previous endmember detection algorithms applied the same set of endmembers to every data point in a scene. By using PCE, different portions of an input hyperspectral scene can be represented using a separate set of endmembers. This results in better suited endmembers for all of the various regions in an input image.

In PCE, the ED detection algorithm was used to estimate EDs to incorporate spectral variability into the endmember detection model. Previously, endmember detection algorithms constrained endmembers to be single spectral vectors. By utilizing EDs, several pixels of the same material with some spectral variation can all be identified as having a full abundance for the same endmember.

The results in Fig. 10 show that PCE produces endmembers that are significantly better suited to a data set when compared to previous methods. This is shown with the fact that the abundance values for each ground-truth class are very compact and concentrated to a small set of endmembers.

During development and testing of the methods, several interesting areas for future research were uncovered. Currently, the ED algorithm assumes a constant and known covariance matrix for each ED. Investigations into methods of learning appropriate covariance matrices for each ED can be done. By learning covariance matrices, endmember distributions can be further tailored to the input data set. In addition, PCE utilizes optimization schedules and many parameter values. Studies on methods to determine the appropriate optimization schedules

and parameter values with regard to the input data set can be conducted.

PCE currently assigns each data point to a single partition. Investigations into methods of allowing data points to have partial membership in several partitions can be conducted. By allowing partial memberships, overlapping clusters are likely to be found using PCE.

Finally, since the current algorithm uses a sequential Gibbs sampler, the running time can be long. The Gibbs sampler requires several iterations for a "burn-in" period. Following the "burn-in" period, the Gibbs sampler samples values driven by the given likelihood until the distribution of partitions for the data points is learned. Future work includes investigating methods to shorten the running time needed for PCE. One potential option is through the use of variational methods which would approximate the solution with a much shorter running time than needed by the Gibbs sampler.

REFERENCES

- [1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [2] J. M. P. Nascimento and J. M. Bioucas-Dias, "Does independent component analysis play a role in unmixing hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 175–187, Jan. 2005.
- [3] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 79–116, 2003.
- [4] M. E. Winter, "Fast autonomous spectral end-member determination in hyperspectral data," in *Proc. 13th Int. Conf. Appl. Geologic Remote Sens.*, Vancouver, BC, Canada, 1999, pp. 337–344.
- [5] J. Boardman, F. Kruse, and R. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Proc. Summaries 5th Annu. JPL Airborne Geosci. Workshop*, R. Green, Ed., Pasadena, CA, 1995, vol. 1, pp. 23–26.
- [6] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
- [7] A. Plaza, P. Martinez, R. Perez, and J. Plazas, "Spatial/spectral end-member extraction by multidimensional morphological operators," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2025–2041, Sep. 2002.
- [8] D. Lee and H. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems*, vol. 13, Cambridge, MA: MIT Press, 2000, pp. 556–562.
- [9] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.
- [10] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra Appl.*, vol. 416, no. 1, pp. 321–331, Jul. 2005.
- [11] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, Jan. 2009.
- [12] T.-M. Tu, "Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent components analysis approach," *Opt. Eng.*, vol. 39, no. 4, pp. 897–906, Apr. 2000.
- [13] J. Wang and C.-I. Chang, "Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2601–2616, Sep. 2006.
- [14] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.
- [15] G. X. Ritter, G. Urcid, and M. S. Schmalz, "Autonomous single-pass endmember approximation using lattice auto-associative memories," *Neurocomputing*, vol. 72, no. 10–12, pp. 2101–2110, Jun. 2009.
- [16] J. M. P. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing algorithm via dependent component analysis," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Barcelona, Spain, Jul. 2007, pp. 4033–4036.
- [17] N. Dobigeon and J. Tourneret, "Library-based linear unmixing for hyperspectral imagery via reversible jump MCMC sampling," in *Proc. IEEE Aerosp. Conf.*, Mar. 2009, pp. 1–6.

- [18] A. Zare and P. Gader, "Endmember detection using the Dirichlet process," in *Proc. IEEE 19th Int. Conf. Pattern Recog.*, Dec. 2008, pp. 1–4.
- [19] D. D. Wackerly, W. Mendenhall, and R. L. Scheaffer, *Mathematical Statistics With Applications*, 5th ed. Belmont, CA: Wadsworth, 1996.
- [20] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "ICE: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.
- [21] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 446–450, Jul. 2007.
- [22] S. Jain and R. M. Neal, "A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model," Univ. Toronto, Toronto, ON, Canada, Tech. Rep. 2003, Jul. 2000.
- [23] M. West, P. Muller, and M. D. Escobar, "Hierarchical priors and mixture models with application in regression and density estimation," in *Aspects of Uncertainty*, P. R. Freeman and A. F. M. Smith, Eds. Hoboken, NJ: Wiley, 1994, pp. 363–386.
- [24] R. M. Neal, "Markov chain sampling methods for Dirichlet process mixture models," Univ. Toronto, Toronto, ON, Canada, Tech. Rep. 9815, Sep. 1998.
- [25] C. Rasmussen, "The infinite Gaussian mixture model," in *Advances in Neural Information Processing Systems*, vol. 12, S. A. Solla, T. K. Leen, and K. R. Muller, Eds. Cambridge, MA: MIT Press, 2000, pp. 554–560.
- [26] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2001.
- [27] M. Grana and M. J. Gallego, "Associative morphological memories for spectral unmixing," in *Proc. Eur. Symp. Artif. Neural Netw.*, Apr. 2003, pp. 481–486.
- [28] S. B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1360–1367, Jul. 2001.
- [29] G. Heo and P. Gader, "KG-FCM: Kernel-based global fuzzy c-means clustering algorithm," Dept. Comput. Inf. Sci. Eng., Univ. Florida, Gainesville, FL, 2008.
- [30] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006.
- [31] R. N. Clark, A. Swayze, R. Wise, K. E. Livo, T. M. Heofen, R. F. Kokaly, and S. J. Sutley, USGS Digital Spectral Library (splib05a): Spectroscopy Lab, U.S. Geological Survey, Jan. 2004. [Online]. Available: <http://speclab.cr.usgs.gov/spectral-lib.html>



Alina Zare (M'07) received the Ph.D. degree from the University of Florida, Gainesville, in 2008.

She is currently with the Department of Computer and Information Science and Engineering, University of Florida. Her research interests include sparsity promotion, machine learning, Bayesian methods, image analysis, and pattern recognition.



Paul Gader (SM'99) received the Ph.D. degree in mathematics from the University of Florida, Gainesville, in 1986.

He has worked as a Senior Research Scientist with Honeywells Systems and Research Center, a Research Engineer and Manager with the Environmental Research Institute of Michigan, and a Faculty Member with the University of Wisconsin, Missouri, and University of Florida, where he is currently a Professor of computer and information science and engineering. He has led teams involved in real-time handwritten address recognition systems for the U.S. Postal Service and teams that devised and tested several real-time algorithms in the field for mine detection. He has over 165 technical publications in the areas of image and signal processing, applied mathematics, and pattern recognition, including over 55 refereed journal articles. His research interests include land mine detection, handwriting recognition, mathematical morphology, machine learning, and fuzzy set and Choquet integration.