Open access • Journal Article • DOI:10.1109/TSP.2011.2180718

Hyperspectral Image Unmixing Using a Multiresolution Sticky HDP — Source link

Roni Mittelman, Nicolas Dobigeon, Alfred O. Hero

Institutions: University of Michigan, University of Toulouse

Published on: 01 Apr 2012 - IEEE Transactions on Signal Processing (IEEE)

Topics: Mixture model, Gibbs sampling, Hierarchical Dirichlet process, Hidden Markov model and Markov random field

Hyperspectral Image Unmixing Using a Multiresolution Sticky HDP

Roni Mittelman, *Member, IEEE*, Nicolas Dobigeon, *Member, IEEE*, and Alfred O. Hero, III, *Fellow, IEEE*

***Abstract*—This paper is concerned with joint Bayesian endmember extraction and linear unmixing of hyperspectral images using a spatial prior on the abundance vectors. We propose a generative model for hyperspectral images in which the abundances are sampled from a Dirichlet distribution (DD) mixture model, whose parameters depend on a latent label process. The label process is then used to enforces a spatial prior which encourages adjacent pixels to have the same label. A Gibbs sampling framework is used to generate samples from the posterior distributions of the abundances and the parameters of the DD mixture model. The spatial prior that is used is a tree-structured *sticky hierarchical Dirichlet process* (SHDP) and, when used to determine the posterior endmember and abundance distributions, results in a new unmixing algorithm called spatially constrained unmixing (SCU). The directed Markov model facilitates the use of scale-recursive estimation algorithms, and is therefore more computationally efficient as compared to standard Markov random field (MRF) models. Furthermore, the proposed SCU algorithm estimates the number of regions in the image in an unsupervised fashion. The effectiveness of the proposed SCU algorithm is illustrated using synthetic and real data.**

***Index Terms*—Bayesian inference, hidden Markov trees, hyperspectral unmixing, image segmentation, spatially constrained unmixing, sticky hierarchical Dirichlet process.**

I. Introduction

HYPERSPECTRAL imaging provides a means of identifying natural and man-made materials from remotely sensed data [1], [2]. Typical hyperspectral imaging instruments acquire data in hundreds of different subbands for each spatial location in the image. Therefore each pixel is a sum of spectral responses of constituent materials in the pixel region, defined by the spatial resolution of the instrument.

Spectral unmixing [3] is the process by which the hyperspectral data is deconvolved under a linear mixing model (LMM). In the LMM the observed spectrum in each pixel is described as a linear combination of the spectra of several materials (*endmembers*) with associate proportions (*abundances*). A common solution to the unmixing problem is to use a two stage approach: endmember extraction followed by an inversion to compute the abundances. Two of the most popular endmember extraction algorithm are the N-FINDR [4] algorithm, and vertex component analysis (VCA) [5]. However these methods assume the existence of pure pixels in the observed image, i.e., they assume that for each of the materials there is at least one pixel where it is observed without being mixed with any of the other materials. This assumption may be a serious limitation in highly mixed scenes. There have been several approaches presented in the literature to address the pure pixel assumption. In [6], a convex optimization based unmixing algorithm which uses a criterion that does not require the pure pixel assumption was presented. However the observation model assumes noise-free measurements and therefore the algorithm may not be effective at low signal-to-noise ratio (SNR). A Bayesian linear unmixing (BLU) approach which estimates the endmembers and abundances simultaneously and avoids the pure pixel assumption was presented in [7]. BLU also uses priors on the model variables which ensure endmember nonnegativity, abundance nonnegativity and sum-to-one constraints, and was shown to outperform the N-FINDR and VCA based unmixing algorithms for highly mixed scenes. Therefore, BLU can be considered as a state-of-the-art algorithm.

A common assumption underlying the aforementioned unmixing algorithms is that the abundance vector for each pixel is independent of other pixels. When the spatial resolution (the size of the region that is represented by each pixel) is low it might be expected that neighboring pixels have different proportions of endmembers, however as the spatial resolution increases neighboring pixels are more likely to share similar spectral characteristics. Even low resolution images may have patches that are characterized by similar abundances, e.g., a large body of water or a vegetation field. Including spatial constraints within the unmixing process has been receiving growing attention in the literature, and has been demonstrated to improve unmixing performance. In [8], spatially constrained unmixing was considered, however, the abundance nonnegativity and sum-to-one constraints as well as the endmember nonnegativity constraint were not enforced. The algorithms in [9] and [10] use spatial constraints to perform endmember extraction, however they rely on the pure pixel assumption. In [11] a spatially constrained abundance estimation algorithm that uses Markov random fields (MRF) [12] and satisfies the abundance nonnegativity and sum-to-one constraints was presented, however the endmembers were estimated separately without including any spatial constraints.

The MRF prior has been used extensively in domains such as texture [13] and hyperspectral [14], [15] image segmentation. Although MRF based algorithms perform well they suffer

Manuscript received May 06, 2011; revised September 03, 2011 and September 03, 2011; accepted December 01, 2011. Date of publication December 21, 2011; date of current version March 06, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Visa Koivunen. This work was partially supported by the AFOSR by Grant FA9550-06-1-0324.

R. Mittelman and A. O. Hero III are with the Department of Electrical Engineering and Computer Science, University of Michigan, MI 48109 USA (e-mail:rmittelm@umich.edu; hero@umich.edu).

N. Dobigeon is with the University of Toulouse, IRIT/INP-ENSEEIHT/TéSA, BP 7122, 31071 Toulouse Cedex 7, France (e-mail:Nicolas.Dobigeon@enseeiht.fr).

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/TSP.2011.2180718

from several drawbacks. Inference in MRF is computationally expensive, and parameter estimation in the unsupervised setting is difficult [16]. Furthermore MRF estimation performance is highly sensitive to tuning parameters [17]. Although there exist methods such as the iterated conditional modes algorithm [18] that reduce the computational complexity related to inference in MRF, these methods usually only converge to a locally optimal solution, and limit the range of priors that may be employed in a Bayesian formulation. A common image processing alternative to MRF is the multiresolution Markov models defined on pyramidally organized trees [19], which allow for computationally efficient scale-recursive inference algorithms to be used, and can be constrained to enforce local smoothness by increasing the self-transition probabilities in the Markov model.

In this paper, we develop a spatially constrained unmixing (SCU) algorithm that simultaneously segments the image into disparate abundance regions and performs unmixing. The abundances within each region are modeled as samples from a Dirichlet distribution (DD) mixture model with different parameters, thus the nonnegativity and sum-to-one physical constraints are naturally satisfied. Specifically we use a mixture model with three components. The first two mixture components capture the abundance homogeneity within the region by setting the precision parameter of the DD mixture components to be relatively large, and the third component models the outliers using a DD whose parameters are all set to one (this is equivalent to a uniform distribution over the feasibility set that satisfies the nonnegativity and sum-to-one constraints). We avoid the need to define the number of disparate homogeneous abundance regions *a priori* by employing a *hierarchical Dirichlet process* (HDP) [20] type of prior. The standard HDP is a nonparametric prior in the sense that it allows the number of states in the Markov process to be learned from the data, and has been previously used for hidden Markov models and hidden Markov trees (HMT) [21], [22]. The multiresolution Markov model which we use differs from the hidden Markov models in [21] and [22] since the observations are only available at the bottommost level of the tree. To encourage the formation of spatially smooth regions we use the sticky HDP (SHDP) [23]. Our method has several advantages compared to the spatially constrained unmixing algorithms in [8]–[11]: (a) it is based on a directed multiresolution Markov model instead of a MRF and thus it allows the use of inference schemes which exhibit faster mixing rates; (b) the spatial dependencies are used to estimate both the abundances and the endmembers simultaneously, rather than just the abundances or endmembers; (c) it does not require the pure pixel assumption; (d) the number of regions that share the same abundances is inferred from the image in an unsupervised fashion. The SCU algorithm that we present here extends the work in [43] by modeling the abundance vectors in each cluster as samples from a DD mixture model with different parameters for each region, as opposed to fixed abundances that are shared by all pixels within the region. Our experimental results show that in low SNR the spatial constraints implemented by the SCU algorithm significantly improves the unmixing performance.

This paper is organized as follows. Section II presents background on the LMM for hyperspectral imaging, and the abundance model. Section III presents the multiresolution prior and background on the SHDP. Section IV presents the spatially constrained unmixing algorithm. Section V presents the experimental results, and Section VI concludes this paper.

II. Problem Formulation

A. Hyperspectral Imaging With the LMM

A hyperspectral image is composed of P pixels $\{\mathbf{y}_p\}_{p=1}^P$, where each $\mathbf{y}_p$ is a L-dimensional vector representing different spectral bands of the reflected electromagnetic field. In the LMM each pixel measurement $\mathbf{y}_p$ is a convex combination of R spectra vectors $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ called endmembers, corrupted by additive Gaussian noise $\mathbf{n}_p = [n_{p,1}, \ldots, n_{p,L}]^T$

$$\mathbf{y}_p = \sum_{r=1}^{R} \mathbf{m}_r a_{p,r} + \mathbf{n}_p \tag{1}$$

where $a_{p,r}$ denotes the proportion of the r^{th} material in the p^{th} pixel. The vector $\mathbf{a}_p = [a_{p,1}, \ldots, a_{p,R}]^T$ which is called the abundance, must satisfy nonnegativity and sum-to-one constraints

$$\begin{cases} a_{p,r} \geq 0, \quad \forall r = 1, \ldots, R \\ \sum_{r=1}^{R} a_{p,r} = 1 \end{cases}. \tag{2}$$

We denote by $\mathcal{A}$ the set of feasible abundances that satisfy the constraints (2). Similarly, since spectra are nonnegative, the endmember must satisfy

$$m_{r,l} \geq 0, \quad \forall r = 1, \ldots R, \quad l = 1, \ldots, L. \tag{3}$$

Concatenating the vectors (1) into a matrix we have the equivalent matrix version of (1)

$$\mathbf{Y} = \mathbf{MA} + \mathbf{N} \tag{4}$$

where

$$\begin{aligned} \mathbf{Y} &= [\mathbf{y}_1, \ldots, \mathbf{y}_P], & \mathbf{M} &= [\mathbf{m}_1, \ldots, \mathbf{m}_R] \\ \mathbf{A} &= [\mathbf{a}_1, \ldots, \mathbf{a}_P], & \mathbf{N} &= [\mathbf{n}_1, \ldots, \mathbf{n}_P]. \end{aligned} \tag{5}$$

The objective of hyperspectral unmixing is to estimate the matrices $\mathbf{M}$ and $\mathbf{A}$, given the observations $\mathbf{Y}$.

B. Dimensionality Reduction

Due to the sum-to-one and nonnegativity constraints the abundance vectors lie on a simplex; a subspace of codimension 1. We use the principal component analysis (PCA) approach that was proposed in [7] to accomplish dimensionality reduction. Let $\bar{\mathbf{y}}$ denote the mean vector of the columns of $\mathbf{Y}$, and let $\boldsymbol{\Upsilon}$ denote the covariance matrix estimated using the columns of $\mathbf{Y}$. Let $\mathbf{D}$ denote a diagonal matrix with the $\mathcal{K}$ largest eigenvalues of $\boldsymbol{\Upsilon}$ arranged along the diagonal, and similarly let $\mathbf{V}$ denote a matrix with columns that are the appropriate eigenvectors, then the PCA reduced endmember takes the form

$$\mathbf{t}_r = \mathbf{P}(\mathbf{m}_r - \bar{\mathbf{y}}) \tag{6}$$

where $\mathbf{P} = \mathbf{D}^{-\frac{1}{2}}\mathbf{V}^T$. Equivalently we have that

$$\mathbf{m}_r = \mathbf{U}\mathbf{t}_r + \bar{\mathbf{y}} \tag{7}$$

where $\mathbf{U} = \mathbf{V}\mathbf{D}^{\frac{1}{2}}$. The endmember matrix $\mathbf{M}$ can therefore be expressed using $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_R]$ as follows:

$$\mathbf{M} = \mathbf{U}\mathbf{T} + \bar{\mathbf{y}}1_R^T \tag{8}$$

where $1_R = [1, \ldots, 1]^T \in \mathbb{R}^R$.

The feasibility set for the endmember is expressed in terms of the PCA reduced representation as

$$\mathcal{T}_r = \left\{ \mathbf{t}_r \middle| \bar{y}_l + \sum_{k=1}^{\mathcal{K}} u_{l,k} t_{k,r} \geq 0, \quad l = 1, \ldots, L \right\} \tag{9}$$

where $\bar{\mathbf{y}} = [\bar{y}_1, \ldots, \bar{y}_L]^T$, $u_{l,k} = (\mathbf{U})_{l,k}$, and $t_{k,r}$ is the k^{th} entry of $\mathbf{t}_r$.

C. The Dirichlet Distribution

The Dirichlet probability density function (PDF) with parameter vector $\boldsymbol{\lambda}$ is [41]

$$p(\mathbf{a}) = \mathrm{Dir}(\boldsymbol{\lambda}) = \frac{\Gamma\left(\sum_{r=1}^{R} \lambda_r\right)}{\prod_{r=1}^{R} \Gamma(\lambda_r)} \prod_{r=1}^{R} a_r^{\lambda_r - 1} \tag{10}$$

where $\mathbf{a} \in \mathcal{A}$. An alternative representation of the Dirichlet parameter vector $\boldsymbol{\lambda}$ is given by

$$s = \sum_{r=1}^{R} \lambda_r \tag{11}$$

$$\mathbf{m} = \mathbb{E}[\mathbf{a}] = \frac{\boldsymbol{\lambda}}{s} \tag{12}$$

where $\mathbf{m}$ is the mean of the DD, and s is known as the precision parameter.

The variance of the DD takes the form [41]

$$\mathrm{Var}(a_i) = \frac{\mathbb{E}[a_i](1 - \mathbb{E}[a_i])}{s+1} \tag{13}$$

and for $i \neq j$ we have that

$$\mathrm{Cov}(a_i a_j) = \frac{\mathbb{E}[a_i]\mathbb{E}[a_j]}{s+1}. \tag{14}$$

This implies that the DD PDF becomes more peaked around its mean as the precision parameter increases.

D. The Spatially Constrained Abundance Model

The unmixing algorithm that we present in this paper segments the image into disparate regions. Let the different regions $\mathcal{R}_k$, $k = 1, \ldots, K$ denote disjoint sets composed of pixel indices, and let the indicator function of a pixel in region $\mathcal{R}_k$ be defined as

$$1_{\mathcal{R}_k}(p) = \begin{cases} 1 & \text{if } p \in \mathcal{R}_k \\ 0 & \text{otherwise} \end{cases}. \tag{15}$$

The class label associated with pixel p thus takes the form

$$z_p = \sum_{k=1}^{K} k 1_{\mathcal{R}_k}(p). \tag{16}$$

The abundance of the p^{th} pixel is generated using

$$\mathbf{a}_p | \boldsymbol{\lambda}_{z_p}^{(0,1)}, \boldsymbol{\upsilon} \sim \upsilon_0 \mathrm{Dir}(\boldsymbol{\lambda}_{z_p}^{(0)}) + \upsilon_1 \mathrm{Dir}(\boldsymbol{\lambda}_{z_p}^{(1)}) + \upsilon_2 \mathrm{Dir}(\boldsymbol{\lambda}_{z_p}^{(2)}) \tag{17}$$

where $\boldsymbol{\lambda}_k^{(2)} = 1_R$ for $k = 1 \ldots, K$, $\upsilon_0 + \upsilon_1 + \upsilon_2 = 1$, $\upsilon_0, \upsilon_1, \upsilon_2 \geq 0$, and we denoted $\boldsymbol{\upsilon} = [\upsilon_0 \;\; \upsilon_1 \;\; \upsilon_2]^T$, and $\boldsymbol{\lambda}_k^{(0,1)} = \{\boldsymbol{\lambda}_k^{(0)}, \boldsymbol{\lambda}_k^{(1)}\}$. The first two mixture components capture the spatial homogeneity of the abundances by setting the precision parameter to a relatively large value, whereas the third component accounts for the outliers by setting υ_2 to a relatively low value. The prior for the parameters $\boldsymbol{\lambda}_k^{(i)}$, $i = 0, 1$ follows:

$$\boldsymbol{\lambda}_k^{(i)} \sim s_0^{(i)} \bar{\mathbf{a}}_k \tag{18}$$

where $\bar{\mathbf{a}}_k \sim \mathrm{Dir}(1_R)$, $s_0^{(i)} \sim \mathrm{U}[l_i, u_i]$, and where $\mathrm{U}[l_i, u_i]$ denotes the uniform distribution on the interval $[l_i, u_i]$ and l_i, u_i are parameters that satisfy $1 < l_i < u_i$. By setting l_i to a relatively large value we can model the homogeneity of the abundances within a region. Furthermore we require $u_0 < l_1$ and $\upsilon_0 > \upsilon_1 \geq \upsilon_2$ in order for the likelihood function to be sufficiently peaky and facilitate the estimation of the labels $\{\mathbf{z}^{(\ell)}\}_{\ell=0}^{\mathcal{L}}$. We fixed the parameter values to $l_0 = 120$, $u_0 = 150$, $l_1 = 60$, $u_1 = 90$, $\upsilon_0 = 0.85$, $\upsilon_1 = 0.1$, $\upsilon_2 = 0.05$ throughout this work.

Let ψ_p denote a discrete random variable which determines which mixture component $\mathbf{a}_p$ was sampled from, then we can describe the generative model for the abundances using

$$\mathbf{a}_p | z_p, \psi_p \sim \mathrm{Dir}(\boldsymbol{\lambda}_{z_p}^{(\psi_p)}) \tag{19}$$

where the prior of ψ_p is a multinomial probability mass function (PMF) of the form

$$\psi_p \sim \upsilon_0 \delta_0(\psi_p) + \upsilon_1 \delta_1(\psi_p) + \upsilon_2 \delta_2(\psi_p). \tag{20}$$

We also denote the sets $\Psi = \{\psi_p\}_{p=1}^{P}$ and $\Psi_k^{(i)} = \{p | p \in \mathcal{R}_k, \psi_p = i\}$ for $k = 1, \ldots, K$, $i = 0, \ldots, 2$.

In the sequel the labels $z_p, \forall p = 1, \ldots, P$ will be denoted $z_p^{(1)}$ since they are going to be associated with the maximal resolution subset of a multiresolution tree in the SHDP representation described here.

E. Lower-Dimensional Abundance Representation

Since the abundances must satisfy the nonnegativity and sum-to-one constraints, they can be rewritten using the partial abundance vectors [7]

$$\mathbf{a}_p = \begin{bmatrix} \mathbf{c}_p \\ a_{p,R} \end{bmatrix} \quad \text{with} \quad \mathbf{c}_p = \begin{bmatrix} a_{p,1} \\ \vdots \\ a_{p,R-1} \end{bmatrix} \tag{21}$$

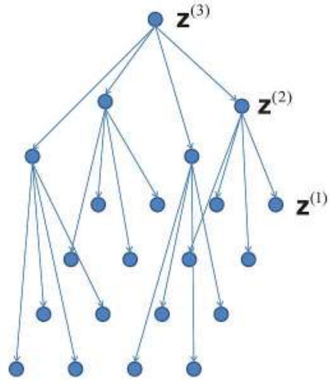


Fig. 1. A quadtree lattice.

and $a_{p,R} = 1 - \sum_{r=1}^{R-1} a_{p,r}$. The feasibility set is

$$\mathcal{S} = \left\{ \mathbf{c}_p \Big| \sum_{r=1}^{R-1} c_{p,r} \le 1, c_{p,1}, \ldots, c_{p,r} \ge 0 \right\}. \tag{22}$$

III. The Multiresolution Sticky HDP

The LMM and the abundance model described in the previous section provide a statistical model for the observations, conditioned on the labels $\{z_p\}_{p=1}^{P}$. To complete this model we also require a prior distribution for these labels, where in this paper we propose to use a multiresolution SHDP which can encourage the formation of spatially smooth label regions and determine the number of regions in an unsupervised fashion. Consider the quadtree lattice shown in Fig. 1, where the nodes are discrete random variables that take their values from the set $\{1, \ldots K\}$, where K denotes the number of class labels. We use the notation $z_p^{(\ell)}, \forall p = 1, \ldots, \frac{P}{4^{\ell-1}}$, to denote the labels at the ℓ^{th} level, where $\mathcal{L}$ denotes the number of levels in the quadtree. We also define the vector $\mathbf{z}^{(\ell)}$ containing all the labels at the ℓ^{th} level

$$\mathbf{z}^{(\ell)} = \left[z_1^{(\ell)}, \ldots, z_{\frac{P}{4^{\ell-1}}}^{(\ell)} \right]^T. \tag{23}$$

The labels $\{z_p^{(1)}\}_{p=1}^{P}$ at the bottommost level of the quadtree are associated with the appropriate pixels of the hyperspectral image. We assume here that the number of pixels in the image is equal to the number of leaves at the bottommost level of the quadtree, otherwise one can increase the size of the tree and prune all the branches that have no descendants that correspond to image pixels. Our prior for the labels assumes a Markovian relationship between the labels on the quadtree lattice. Specifically, let us define the likelihoods

$$\pi_{ki}^{(\ell)} = p\left(z_p^{(\ell)} = i | z_{\mathrm{Pa}(\ell,p)}^{(\ell+1)} = k \right) \tag{24}$$

where $\mathrm{Pa}(\ell, p)$ denotes the parent of node p at the ℓ^{th} level, then the joint probability mass function of all the labels takes the form

$$p\left(\mathbf{z}^{(1:\mathcal{L})} | \boldsymbol{\pi}^{(1:\mathcal{L}-1)} \right) = \prod_{p=1}^{\frac{P}{4^{\mathcal{L}-1}}} p\left(z_p^{(\mathcal{L})} \right) \prod_{\ell=1}^{\mathcal{L}-1} \prod_{p'=1}^{\frac{P}{4^{\ell-1}}} \pi^{(\ell)}_{z^{(\ell+1)}_{\mathrm{Pa}(\ell,p')} z^{(\ell)}_{p'}} \tag{25}$$

where $\mathbf{z}^{(1:\mathcal{L})} = \{\mathbf{z}^{(\ell)}\}_{\ell=1}^{\mathcal{L}}$, and $\boldsymbol{\pi}^{(1:\mathcal{L}-1)} = \{\boldsymbol{\pi}^{(\ell)}\}_{\ell=1}^{\mathcal{L}-1}$. We also define the vector consisted of the transition probabilities from class label k at the $(\ell+1)^{\text{th}}$ level

$$\boldsymbol{\pi}_k^{(\ell)} = [\pi_{k1}^{(\ell)}, \ldots, \pi_{kK}^{(\ell)}]^T. \tag{26}$$

The quadtree model can be used to enforce spatial smoothness by using a prior for $\boldsymbol{\pi}_k^{(\ell)}$ $\forall k = 1, \ldots, K, \ell = 1, \ldots, \mathcal{L}-1$ which encourages larger self-transition probabilities, i.e., $\pi_{kk}^{(\ell)} > \pi_{ki}^{(\ell)}$ $\forall i \neq k, \ell = 1, \ldots, \mathcal{L}-1$. Another issue is the choice of the number of class labels K. These type of problems are known as model order selection where common approaches such as the AIC [24] and BIC [25], optimize a criterion which advocates a compromise between the model fitting accuracy and the model complexity. The drawback is that the AIC and BIC require a scoring function to be computed for every considered number of parameters in order to choose the optimal model. Another approach is reversible jump Markov chain Monte Carlo samplers [26], [27] where moves between different parameter spaces are allowed. However such methods require accurate tuning of the jump proposals and are not computationally efficient [28]. Dirichlet processes (DP) provide a nonparametric prior for the number of components in a mixture model [20] and facilitate inference using Monte Carlo or variational Bayes methods [29], [30]. The HDP is an extension of the DP which allows for several models to share the same mixture components, and can be used to infer the state space in a Markov model. The SHDP augments the HDP by encouraging the formation of larger self-transition probabilities in the Markov model, thus it provides an elegant solution to all the requirements of our multiresolution prior. Next we provide an introduction to the DP, HDP, and SHDP in the context of the tree prior which is used in the unmixing algorithm presented in this paper.

A. Dirichlet Processes

The DP denoted by $\mathrm{DP}(\gamma, H)$ is a probability distribution on a measurable space [31], and can be represented as an infinite mixture model where each component is drawn from the base measure H, and the mixture weights are drawn from a stick-breaking process that is parameterized by positive real number γ [32]. Specifically, to sample from the DP one can sample the from the infinite mixture model

$$G_0(\boldsymbol{\theta}) = \sum_{k=1}^{\infty} \beta_k \delta(\boldsymbol{\theta}, \boldsymbol{\theta}_k), \quad \boldsymbol{\theta}_k \sim H \tag{27}$$

where $\{\beta_k\}_{k=1}^{\infty}$ is the stick-breaking process constructed as follows:

$$\nu_k \sim \mathrm{Beta}(1, \gamma), \qquad k = 1, 2, \ldots \tag{28}$$

$$\beta_k = \begin{cases} \nu_k, & k = 1 \\ \nu_k \prod_{\ell=1}^{k-1}(1 - \nu_\ell), & k = 2, 3, \ldots \end{cases} \tag{29}$$

where $\mathrm{Beta}(\cdot, \cdot)$ denotes the Beta distribution. The stick-breaking process is commonly denoted by $\mathrm{GEM}(\gamma)$, and can be interpreted as dividing a unit length stick into segments that represent the proportions β_k of the different mixture components $\boldsymbol{\theta}_k$. Observing (28) and (29), we note that setting γ such that ν_k is likely to be closer to one (i.e., smaller γ) leads to having fewer mixing components with nonnegligible weights and vice versa. Therefore the parameter γ expresses the prior belief on the effective number of mixture components. Equivalently, the generating process for a sample $\boldsymbol{\theta}'_i$ from $\mathrm{DP}(\gamma, H)$, can be represented using an indicator random variable ζ_i

$$\begin{aligned} \boldsymbol{\beta}|\gamma &\sim \mathrm{GEM}(\gamma), \quad \boldsymbol{\theta}_k \sim H, \quad k = 1, \ldots \\ \zeta_i|\boldsymbol{\beta} &\sim \mathrm{Mult}(\boldsymbol{\beta}), \quad \boldsymbol{\theta}'_i = \boldsymbol{\theta}_{\zeta_i} \end{aligned} \tag{30}$$

where $\mathrm{Mult}(\cdot)$ denotes a multinomial distribution.

Another interpretation of the DP is through the metaphorical *Chinese restaurant process* (CRP) representation [20] which follows from the *Pólya* sequence sampling scheme [33]. According to the CRP, a customer that is represented by the sample index i enters a restaurant with infinitely many tables each serving a dish $\boldsymbol{\theta}_k$. The customer can either sit at a new table where no one else is sitting, with probability that is proportional to γ, or sit at any other table with probability that is proportional to the number of other customers that are already sitting at that table. If the customer sits at a new table then he also chooses the dish served at that table by sampling the probability measure H, otherwise the customer selects the dish $\boldsymbol{\theta}_k$ that is served at the chosen table.

B. Hierarchical Dirichlet Processes

The HDP defines a set of probability measures G_j which are DPs that share the same base measure which is itself a DP. Let $G_0 \sim \mathrm{DP}(\gamma, H)$, then the HDP is obtained using

$$G_j \sim \mathrm{DP}(\alpha, G_0), \quad j = 1, \ldots, J \tag{31}$$

where J denotes the number of different groups that share the same base measure G_0, and α is a positive real number. The process of generating a sample $\boldsymbol{\theta}'_{ji}$ from $G_j(\boldsymbol{\theta})$ can be represented using the indicator variable notation

$$\begin{aligned} \boldsymbol{\beta}|\gamma &\sim \mathrm{GEM}(\gamma), \quad \boldsymbol{\theta}_k \sim H, \quad k = 1, \ldots \\ \boldsymbol{\pi}_j|\alpha, \boldsymbol{\beta} &\sim \mathrm{DP}(\alpha, \boldsymbol{\beta}) \\ \zeta_{ji}|\boldsymbol{\pi}_j &\sim \mathrm{Mult}(\boldsymbol{\pi}_j), \quad \boldsymbol{\theta}'_{ji} = \boldsymbol{\theta}_{\zeta_{ji}}. \end{aligned} \tag{32}$$

Similarly to the DP, the HDP can be interpreted using a representation that is analogous to the CRP and is known as the *Chinese restaurant franchise* (CRF). The CRF metaphor describes each of the processes $G_1, \ldots, G_J$ as restaurants that share the same global menu which offers dishes that are represented by the mixture components of G_0. A customer that enters the j^{th} restaurant can sit at an existing table with probability that is proportional to the number of other customers already sitting at that table, or at a new table with probability that is proportional to α. If the customer sits at a table which is already instantiated he chooses the same dish that is served at that table, otherwise he chooses a dish by randomly drawing from G_0. A dish which has already been served at any of the restaurants is sampled from G_0 with probability that is proportional to the number of all tables in the J restaurants that are serving that dish, and a new dish is sampled from H with probability that is proportional to γ.

Assuming a HDP prior, we can use $\boldsymbol{\pi}_j$ in (32) as the transition probabilities between the levels of the quadtree. The dishes correspond to samples from a base measure described by the PDF of $\boldsymbol{\lambda}_k^{(0,1)}$ which can be obtained from (18), customers correspond to label realizations at the nodes of the quadtree, and the restaurant that each customer is assigned to corresponds to the label of the parent node, where we use different restaurants for each level in the tree.

C. The Sticky HDP

An important property that is demonstrated by the HDP is that the base measure G_0 serves as the "average" measure of all the J restaurants [23]

$$\mathbb{E}[\boldsymbol{\pi}_j|\boldsymbol{\beta}] = \boldsymbol{\beta}. \tag{33}$$

The implication of (33) is that by using a HDP prior we assume that that the unique dish proportions are similar across different restaurants. This violates our requirement that the prior for the transition probabilities encourage larger self-transition likelihoods. The SHDP [23] modifies the restaurant specific dish likelihoods in (32) in the following manner:

$$\boldsymbol{\pi}_j|\alpha, \kappa, \boldsymbol{\beta} \sim \mathrm{DP}\left(\alpha + \kappa, \frac{\alpha\boldsymbol{\beta} + \kappa\boldsymbol{\delta}_j}{\alpha + \kappa}\right) \tag{34}$$

where κ is a nonnegative scalar, and $\boldsymbol{\delta}_j$ denotes a vector with 1 at the j^{th} entry and zeros in the rest. The "average" $\boldsymbol{\pi}_j$ now takes the form

$$\mathbb{E}[\boldsymbol{\pi}_j|\boldsymbol{\beta}] = \frac{\alpha}{\alpha + \kappa}\boldsymbol{\beta} + \frac{\kappa}{\alpha + \kappa}\boldsymbol{\delta}_j \tag{35}$$

and therefore the larger the parameter κ is, the more likely it becomes that the j^{th} dish, which is referred to as the *specialty dish*, would be selected at the j^{th} restaurant.

The metaphorical interpretation of the SHDP which is known as the *CRF with loyal customers*, is identical to the CRF with the sole difference that if a customer chooses to sit at an uninstantiated table at the j^{th} restaurant, then in order to sample a dish he flips a biased coin such that with probability proportional to κ he selects the j^{th} dish and with probability proportional to α the customer draws the dish from G_0. The table assignment of the i^{th} sample at the j^{th} restaurant thus follows:

$$p(t_{ji}|t_{j1}, \ldots, t_{ji-1}, \alpha, \kappa) \propto \sum_{t=1}^{T_j} \mathfrak{n}_{jt}\delta(t_{ji}, t) + (\alpha + \kappa)\delta(t_{ji}, T_j + 1) \tag{36}$$

where $\mathfrak{n}_{jt}$ denotes the number of customers sitting at the t^{th} table in the j^{th} restaurant, and T_j denotes the number of tables instantiated in the j^{th} restaurant. The dish assignment k_{jt} of a table t in the j^{th} restaurant is drawn using

$$\bar{k}_{jt}|\boldsymbol{\beta} \sim \mathrm{Mult}(\boldsymbol{\beta}) \tag{37}$$

$$w_{jt}|\alpha, \kappa \sim \mathrm{Ber}(\rho) \tag{38}$$

$$k_{jt}|\bar{k}_{jt}, w_{jt} = \begin{cases} \bar{k}_{jt}, & w_{jt} = 0 \\ j, & w_{jt} = 1 \end{cases} \tag{39}$$

where $\mathrm{Ber}(\cdot)$ denotes a Bernoulli distribution, $\rho = \frac{\kappa}{\alpha+\kappa}$, and w_{jt} is an auxiliary variable which is equal to zero or one, depending on whether the dish at the t^{th} table in the j^{th} restaurant

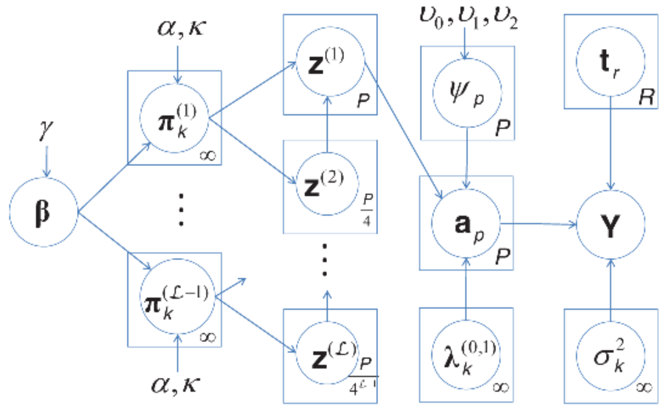


Fig. 2. Graphical model representation of the SCU algorithm, where $\boldsymbol{\lambda}_k^{(0,1)}$ is defined in (18), ψ_p is defined in (20), $\mathbf{a}_p$ is defined in (17), $\mathbf{t}_r$ is defined in (6), $\mathbf{z}^{(\ell)}$ is defined in (23), $\boldsymbol{\beta} \sim \mathrm{GEM}(\gamma)$, $\boldsymbol{\pi}_k^{(\ell)}$ is defined in (26), $\mathbf{Y}$ is the hyperspectral image, and σ^2 is the variance of the observation noise in (4).

was chosen by drawing from G_0, or by an override operation on the specialty dish.

1) Infinite Limit of a Finite Mixture Model: The DP and the derived HDP and SHDP can all be obtained as the limit of finite mixture models [34], [35]. Specifically for the SHDP the stick-breaking process $\boldsymbol{\beta}$ in (32) can be approximated as the K-dimensional DD

$$\boldsymbol{\beta}|\gamma \sim \mathrm{Dir}\left(\frac{\gamma}{K}, \ldots, \frac{\gamma}{K}\right) \tag{40}$$

and the restaurant specific dish probabilities in (34) can be approximated using the K-dimensional DD

$$\boldsymbol{\pi}_j|\alpha, \kappa, \boldsymbol{\beta} \sim \mathrm{Dir}(\alpha\beta_1, \ldots, \alpha\beta_{j-1}, \alpha\beta_j + \kappa, \alpha\beta_{j+1}, \ldots, \alpha\beta_K). \tag{41}$$

The above construction converges in distribution to the SHDP as $K \to \infty$.

2) Posterior Sampling in the SHDP: The approximation of the SHDP using finite mixture models leads to a simple form of the posterior distributions. By the conjugacy of the Dirichlet and multinomial distributions, it follows that:

$$\begin{aligned}\boldsymbol{\pi}_j|\mathfrak{n}, \alpha, \kappa, \boldsymbol{\beta} \sim \mathrm{Dir}(&\alpha\beta_1 + \mathfrak{n}_{j1}, \ldots, \alpha\beta_{j-1} + \mathfrak{n}_{j(j-1)} \\ &\alpha\beta_j + \mathfrak{n}_{jj} + \kappa, \alpha\beta_{j+1} + \mathfrak{n}_{j(j+1)}, \ldots, \alpha\beta_K + \mathfrak{n}_{jK})\end{aligned} \tag{42}$$

where $\mathfrak{n}_{jk} = \sum_{i=1}^{N_j} \delta(\boldsymbol{\theta}_k = \boldsymbol{\theta}'_{ji})$, which is also equivalent to the number of customers that are having the k^{th} dish in the j^{th} restaurant. The posterior for $\boldsymbol{\beta}$ takes the form

$$\boldsymbol{\beta}|\bar{\mathfrak{m}}, \gamma \sim \mathrm{Dir}\left(\frac{\gamma}{K} + \bar{\mathfrak{m}}_{\cdot 1}, \ldots, \frac{\gamma}{K} + \bar{\mathfrak{m}}_{\cdot K}\right) \tag{43}$$

where $\bar{\mathfrak{m}}_{\cdot k}$ denotes the number of tables that are serving the k^{th} dish all over the J restaurants, which were not instantiated by an override operation of the specialty dish.

The posterior for the number of tables serving the k^{th} dish in the j^{th} restaurant $\mathfrak{m}_{jk}$ in the SHDP takes the form [23]

$$\begin{aligned}&p(\mathfrak{m}_{jk} = \mathfrak{m}|\mathfrak{n}_{jk}, \boldsymbol{\beta}, \alpha) \\ &= \frac{\Gamma(\alpha\beta_k)}{\Gamma(\alpha\beta_k + \kappa\delta(k,j) + \mathfrak{n}_{jk})} s(\mathfrak{n}_{jk}, \mathfrak{m})(\alpha\beta_k + \kappa\delta(k,j))^{\mathfrak{m}}\end{aligned} \tag{44}$$

where $s(\mathfrak{n}, \mathfrak{m})$ are unsigned Stirling numbers of the first kind. Alternatively, it is possible to sample $\mathfrak{m}_{jk}$ by simulating table assignments from a CRP [23].

The posterior for the override auxiliary variables w_{jt} given in (38) was developed in [23]:

$$p(w_{jt}|k_{jt} = j, \boldsymbol{\beta}, \rho) \propto \begin{cases}\beta_j(1-\rho), & w_{jt} = 0 \\ \rho, & w_{jt} = 1\end{cases}. \tag{45}$$

The number of tables whose dish was selected by sampling from G_0 satisfies:

$$\bar{\mathfrak{m}}_{jk} = \begin{cases}\mathfrak{m}_{jk}, & j \neq k \\ \mathfrak{m}_{jj} - w_{j\cdot}, & j = k\end{cases} \tag{46}$$

where $w_{j\cdot} = \sum_{t=1}^{T_j} w_{jt}$. A sample from the posterior for $\bar{\mathfrak{m}}_{jk}$ can be obtained using Algorithm 1.

Algorithm 1: Posterior sampling of $\bar{\mathfrak{m}}_{jk}$

- For $j = 1, \ldots, J$, $k = 1, \ldots, K$
  1) Sample $\mathfrak{m}_{jk}$ using **(44)** or by simulating from **(36)**.
  2) For $t = 1, \ldots, \mathfrak{m}_{jk}$, sample w_{jt} from **(45)**.
  3) Compute $\bar{\mathfrak{m}}_{jk}$ using **(46)**.

IV. Spatially Constrained Hyperspectral Unmixing With a Pyramid Structured SHDP

In this section, we present the SCU algorithm where the graphical model representation is described in Fig. 2. We first describe each of the parameters' priors, and then present the posterior distributions that are used with the Gibbs sampling algorithm.

A. Parameter Priors

1) Label Transition Probabilities: The multiresolution Markov model described in Section III relies on the state transition probabilities $\boldsymbol{\pi}_k^{(\ell)}$, $\ell = 1, \ldots \mathcal{L} - 1$, $k = 1, \ldots, K$. The prior for these parameters is obtained from the SHDP with the finite mixture approximation perspective. Specifically, we have that $\boldsymbol{\beta} \sim \mathrm{GEM}(\gamma)$ is a stick-breaking process that is approximated using (40), and the prior for $\boldsymbol{\pi}_k^{(\ell)}$, $\ell = 1, \ldots, \mathcal{L} - 1$, $k = 1, \ldots, K$ is obtained similarly to (41) with $j = k$, regardless of the value of ℓ.

2) Abundances: As described in Section II-D the abundance of pixel p follows a DD mixture with a parameter vector which depends on the label $z_p^{(1)}$

$$f(\mathbf{A}|\{\boldsymbol{\lambda}_k^{(0,1)}\}_{k=1}^K, \mathbf{z}^{(1)}, \Psi) \propto \prod_{i=0}^{1} \prod_{p \in \Psi_k^{(i)}} \prod_{r=1}^{R} (a_{p,r})^{\lambda^{(i)}_{z_p^{(1)},r} - 1}. \tag{47}$$

where $\lambda^{(i)}_{z_p^{(1)},r}$ denotes the r^{th} element of the vector $\boldsymbol{\lambda}^{(i)}_{z_p^{(1)}}$.

3) DD Parameters: As explained in Section II-D, we use $\boldsymbol{\lambda}_k^{(i)} \sim s_0^{(i)} \bar{\mathbf{a}}_k, i = 0, 1$ where the priors for $s_0^{(i)}$, and $\bar{\mathbf{a}}_k$ were described in Section II-D. Therefore $f(\boldsymbol{\lambda}_k^{(i)}) = f(s_{0,k}^{(i)}, \bar{\mathbf{a}}_k)$, $\forall k = 1, \ldots, K, i = 0, 1$. We assume that the parameters of the DD in every class are independent, therefore we have that

$$f(\{s_{0,k}^{(0)}, s_{0,k}^{(1)}, \bar{\mathbf{a}}_k\}_{k=1}^K) \propto \prod_{k=1}^{K} \prod_{i=0}^{1} 1_{[l_i, u_i]}(s_{0,k}^{(i)}) 1_{\mathcal{A}}(\bar{\mathbf{a}}_k). \tag{48}$$

Algorithm 2: The SCU algorithm

- *Initialization*: initialize $\{\mathbf{t}_{r,0}\}_{r=1}^{R}$, $\{\mathbf{a}_p\}_{p=1}^{P}$, $\{\boldsymbol{\lambda}_k^{(0,1)}\}_{k=1}^{K}$.
- *Iterations*: For $t = 1, \ldots$
  1) For $k = 1, \ldots, K$, sample σ_k^2 from **(72)**.
  2) Sample $\tilde{\mathbf{c}}^{(i)}_{z_p^{(1)}} \sim \mathcal{N}_S(\boldsymbol{\nu}_p, \boldsymbol{\Sigma}_p)$, $i = 1, \ldots, n_{mc}$ using **(61)**.
  3) For $\ell = 1 \ldots, \mathcal{L}$, $p = 1, \ldots, \frac{P}{4^{\ell-1}}$, compute the upward predictions $\varphi_p^{(\ell)}(z_p^{(\ell)})$ using **(58)** and **(60)**.
  4) For $\ell = 1 \ldots, \mathcal{L}$, $p = 1, \ldots, \frac{P}{4^{\ell-1}}$, sample the labels $z_{p,t}^{(\ell)}$ using **(56) and (57)**.
  5) Partition all the labels into dishes and restaurants as described in **Section IV-B.II**.
  6) Sample the number of tables serving the k^{th} dish at the j^{th} restaurant $\bar{\mathfrak{m}}_{jk}$, using Algorithm 1 with $J = (\mathcal{L} - 1)K$.
  7) Sample $\boldsymbol{\beta}$ from **(43)**.
  8) For $\ell = 1, \ldots, \mathcal{L} - 1$, $k = 1, \ldots, K$, sample $\boldsymbol{\pi}_k^{(\ell)}$ from **(62)**.
  9) For p $= 1, \ldots,$ P sample ψ_p from **(71)**.
  10) For $p \in \mathcal{R}_1$, sample $\mathbf{c}_p$ as described in **Section IV-B.III**.
  11) For $k = 1, \ldots, K$,
      — Sample $\bar{\mathbf{a}}_k$ using **(67)**.
      — For $i = 0, 1$, sample $s_{0,k}^{(i)}$ using **(65)**.
      — For $i = 0, 1$, set $\boldsymbol{\lambda}_k^{(i)} = s_{0,k}^{(i)} \bar{\mathbf{a}}_k$.
  12) For $r = 1, \ldots, R$, sample $\mathbf{t}_r$ from **(73)**.

where $\mathcal{A}$ denotes the abundance feasibility set (2).

4) The Indicator Variable ψ_p: The prior for ψ_p is given in (20), where as explained in Section II-D the parameters v_0, v_1, v_2 where chosen such that the prior promotes abundance PDFs which are peaky, and therefore facilitate the segmentation process.

5) Likelihood: We assume that the additive noise term in the LMM satisfies $\mathbf{n}_p \sim \mathcal{N}(0, \sigma_k^2 \mathbf{I})$ for all $p \in \mathcal{R}_k$, thus the likelihood of observing $\mathbf{y}_p$ takes the form

$$f(\mathbf{y}_p | z_p^{(1)}, \Theta) = \left(\frac{1}{2\pi\sigma^2_{z_p^{(1)}}} \right)^{\frac{L}{2}} \exp\left\{ -\frac{\|\mathbf{y}_p - \mathbf{M}\mathbf{a}_p\|^2}{2\sigma^2_{z_p^{(1)}}} \right\} \tag{49}$$

where $\Theta = \{\mathbf{M}, \mathbf{A}, \{\sigma_k^2\}_{k=1}^{K}\}$, and $\|\cdot\|$ denotes the standard Euclidean norm.

Since the noise vectors for each of the pixels $\mathbf{n}_p$, $p = 1, \ldots, P$ are assumed to be independent, the PDF of all the pixels takes the form

$$f(\mathbf{Y} | \mathbf{z}^{(1)}, \Theta) = \prod_{p=1}^{P} f(\mathrm{y}_p | z_p^{(1)}, \Theta). \tag{50}$$

6) Noise Variance Prior: The prior which we use for σ_k^2 is the conjugate prior

$$\mathcal{IG}(\nu_0, \gamma_0) \tag{51}$$

where $\mathcal{IG}(\cdot, \cdot)$ denotes an inverse-gamma distribution, and we used the parameter values $\nu_0 = 20$ and $\gamma_0 = 0.1$.

7) Projected Spectra Prior: Similarly to [7] we use a multivariate Gaussian that is truncated on the set $\mathcal{T}_r$ given in (9), as a prior for $\mathbf{t}_r$. The PDF therefore takes the form

$$f(\mathbf{t}_r | \mathbf{e}_r, \sigma_e^2 \mathbf{I}) \propto \exp\left\{ -\frac{\|\mathbf{t}_r - \mathbf{e}_r\|^2}{2\sigma_e^2} \right\} 1_{\mathcal{T}_r}(\mathbf{t}_r) \tag{52}$$

where the mean $\mathbf{e}_r$ is set using the endmembers found using VCA, and σ_e^2 is set to a large value (we used $\sigma_e^2 = 50$).

B. Gibbs Sampling and the Posterior Distributions

The estimation is performed using a Gibbs sampler [36] which generates a Monte Carlo approximation of the distribution of the random variables by generating samples from the posterior distributions iteratively, as outlined in Algorithm 2.

The posterior sampling schemes are described next in this section. Let $\{\tilde{\mathbf{x}}_t\}_{t=1}^{T_{mc}}$, denote the sequence generated by the Gibbs sampler for a random variable $\mathbf{x}$, then the minimum mean squared error (MMSE) estimate is approximated using:

$$\hat{\mathbf{x}} = \frac{1}{T_{mc} - T_{bi}} \sum_{t=T_{bi}+1}^{T_{mc}} \tilde{\mathbf{x}}_t \tag{53}$$

where T_{bi} denotes the number of burn-in iterations.

A byproduct of the unmixing algorithm is the segmentation that is given by the labels $\mathbf{z}_t^{(1)}$. Since the random vector $\mathbf{z}^{(1)}$ is discrete it can not be estimated like the abundances and endmembers using (53). One possible approach to segment the image is to select the $\mathbf{z}_t^{(1)}$ which maximize the posterior likelihood, however this approach tends to overfit the data [23]. The approach that we use in this work is known as the maximum of posterior marginals (MPM) [13], where the detected label for each pixel is that which occurs with the largest frequency over the sequence generated by the Gibbs sampler, i.e.

$$\hat{z}_p^{(1)} = \mathrm{argmax}_z \sum_{t=T_{bi}+1}^{T_{mc}} \delta(z_{p,t}^{(1)}, z). \tag{54}$$

1) Block Sampling the Labels' Posterior Distribution: The blocked sampler for the states of a HDP-HMT was presented in [22]. Here we present the particular case of the algorithm in [22] in which the observations are available at the leaf nodes alone, and which includes the SHDP extension of [23]. The approach is similar to the "upward-downward" procedure in hidden Markov trees [37]. The benefit of using a blocked sampler, as opposed to a direct assignment sampler which updates the label of a single node at a time, is that the mixing rate is improved significantly. A faster mixing rate translates into faster convergence.

The labels' posterior can be written as

$$p(\mathbf{z}^{(1:\mathcal{L})} | \boldsymbol{\pi}^{(1:\mathcal{L}-1)}, \mathbf{Y}, \Theta) = \prod_{p=1}^{\frac{P}{4^{\mathcal{L}-1}}} p(z_p^{(\mathcal{L})} | \boldsymbol{\pi}^{(1:\mathcal{L})}, \mathbf{Y}, \Theta)$$
$$\times \prod_{\ell=1}^{\mathcal{L}-1} \prod_{p=1}^{\frac{P}{4^{\ell-1}}} p(z_p^{(\ell)} | \mathbf{z}^{(\ell+1)}, \boldsymbol{\pi}^{(1:\mathcal{L})}, \mathbf{Y}, \Theta). \tag{55}$$

The interpretation of (55) is that given the appropriate conditional distributions, the block sampler is realized by sampling the labels at each level, going sequentially from the topmost level to the bottommost level. The conditional distributions in (55) admit the following expressions:

$$p(z_p^{(\mathcal{L})}|\boldsymbol{\pi}^{(1:\mathcal{L})}, \mathbf{Y}, \Theta) \propto p(z_p^{(\mathcal{L})})\varphi_p^{(\mathcal{L})}(z_p^{(\mathcal{L})}) \quad (56)$$

$$p(z_p^{(\ell)}|\mathbf{z}^{(\ell+1)}, \mathbf{Y}, \Theta) \propto \pi^{(\ell)}_{z^{(\ell+1)}_{\mathrm{Pa}(\ell,p)} z_p^{(\ell)}} \varphi_p^{(\ell)}(z_p^{(\ell)}) \quad (57)$$

where in this paper we used an equally likely distribution for the labels $z_p^{(\mathcal{L})}$. This choice ensures that the MCMC algorithm samples over the full depth of the tree. The upward predictions $\varphi_p^{(\ell)}(z_p^{(\ell)})$ are computed recursively using

$$\varphi_p^{(\ell)}(z_p^{(\ell)}) \propto \begin{cases} \prod_{p' \in \mathrm{c}(\ell,p)} \sum_{z_{p'}^{(\ell-1)}} \pi^{(\ell-1)}_{z_p^{(\ell)} z_{p'}^{(\ell-1)}} \varphi_{p'}^{(\ell-1)}(z_{p'}^{(\ell-1)}), & \ell > 1 \\ \int f(\mathrm{y}_p|\mathbf{a}_p, \Theta) 1_{\mathcal{A}}(\mathbf{a}_p) f(\mathbf{a}_p|\boldsymbol{\lambda}^{(0,1)}_{z_p^{(1)}}, \boldsymbol{v}) \mathrm{d}\mathbf{a}_p, & \ell = 1 \end{cases} \quad (58)$$

where $\mathrm{c}(\ell, p)$ denotes the set consisted of the children of node p at the ℓ^{th} level, and $f(\mathbf{a}_p|\boldsymbol{\lambda}^{(0,1)}_{z_p^{(1)}}, \boldsymbol{v})$ is obtained from (17). Equation (58) therefore constitutes the upward sweep in which the predictions $\varphi_p^{(\ell)}(z_p^{(\ell)})$ are calculated, whereas (56) implements a downward sweep in which the labels are sampled. Computing the integral in (58) is in general intractable, therefore we use Monte Carlo integration. Let c_0 denote the normalization constant of the truncated PDF $f(\mathbf{y}_p|\mathbf{a}_p, \Theta)1_{\mathcal{A}}(\mathrm{a}_p)$, then we first draw n_{mc} samples

$$\tilde{\mathbf{a}}_p^{(j)} \sim \frac{1}{c_0} f(\mathbf{y}_p|\mathbf{a}_p, \Theta) 1_{\mathcal{A}}(\mathbf{a}_p), \quad j = 1, \ldots, n_{mc} \quad (59)$$

and approximate the integral using

$$\int f(\mathbf{y}_p|\mathbf{a}_p, \Theta) 1_{\mathcal{A}}(\mathrm{a}_p) f(\mathbf{a}_p|\boldsymbol{\lambda}^{(0,1)}_{z_p^{(1)}}, \boldsymbol{v}) \mathrm{d}\mathbf{a}_p \approx \frac{c_0}{n_{mc}} \sum_{j=1}^{n_{mc}} f(\tilde{\mathbf{a}}_p^{(j)}|\boldsymbol{\lambda}^{(0,1)}_{z_p^{(1)}}, \boldsymbol{v}) \quad (60)$$

where in this paper we used $n_{mc} = 5$. We note that c_0 in (60) can be ignored for the purpose of approximating (58), and the sampling in (59) can be realized by sampling the partial abundance vector $\mathbf{c}_p$ from a Gaussian PDF that is truncated to the partial abundance feasibility set $\mathbf{c}_p^{(i)} \sim \mathcal{N}_S(\boldsymbol{\nu}_p, \boldsymbol{\Sigma}_p), \ i = 1, \ldots, n_{mc}$ where

$$\begin{cases} \boldsymbol{\Sigma}_p = \left[\frac{1}{\sigma^2_{z_p^{(1)}}} (\mathrm{M}_{-R} - \mathbf{m}_R 1_{R-1}^T)^T (\mathbf{M}_{-R} - \mathrm{m}_R 1_{R-1}^T) \right]^{-1} \\ \boldsymbol{\nu}_p = \boldsymbol{\Sigma}_p \left[\frac{1}{\sigma^2_{z_p^{(1)}}} (\mathbf{M}_{-R} - \mathbf{m}_R 1_{R-1}^T)^T (\mathrm{y}_p - \mathbf{m}_R) \right]. \end{cases} \quad (61)$$

We refer the reader to [7] for specific details regarding the implementation of efficient sampling from the truncated multivariate Gaussian distribution.

2) Posterior Sampling of the State Transition Probabilities: The labels $\{\mathbf{z}^{(\ell)}\}_{\ell=1}^{\mathcal{L}}$ effectively partition the data into restaurants and dishes under the CRF with loyal customers metaphor. For instance assume that $z_p^{(\ell)} = i$ and $z^{(\ell+1)}_{\mathrm{Pa}(\ell,p)} = k$, then as discussed in Section III-B we interpret this as the i^{th} dish being served at the j^{th} restaurant where $j = (\ell - 1)K + k$. The posterior for $\boldsymbol{\pi}_k^{(\ell)}$ is therefore obtained similarly to (42) using

$$\begin{aligned} \boldsymbol{\pi}_k^{(\ell)}|\mathfrak{n}, \alpha, \kappa, \boldsymbol{\beta} \sim \mathrm{Dir}(&\alpha\beta_1 + \mathfrak{n}_{j1}, \ldots, \alpha\beta_{k-1} + \mathfrak{n}_{j(k-1)}, \\ &\alpha\beta_k + \mathfrak{n}_{ji} + \kappa, \alpha\beta_{k+1} + \mathfrak{n}_{j(k+1)}, \ldots, \alpha\beta_K + \mathfrak{n}_{jK}) \end{aligned} \quad (62)$$

where $j = (\ell - 1)K + k$. The posterior for β is similarly given in (43).

3) Sampling From the Abundance Posterior: The abundance posterior at pixel p takes the form

$$f(\mathbf{c}_p|\mathbf{T}, \mathbf{y}_p, \boldsymbol{\lambda}^{(\psi_p)}_{z_p^{(1)}}, \sigma^2_{z_p^{(1)}}) \propto \mathcal{N}(\mathbf{c}_p; \boldsymbol{\nu}_p, \boldsymbol{\Sigma}_p) \times \prod_{r=1}^{R-1} (c_{p,r})^{(\lambda^{(\psi_p)}_{z_p^{(1)},r} - 1)} (1 - 1_{R-1}^T \mathrm{c}_p)^{\lambda^{(\psi_p)}_{z_p^{(1)},R} - 1} 1_{\mathcal{S}}(\mathbf{c}_p) \quad (63)$$

with $\boldsymbol{\nu}_p$ and $\boldsymbol{\Sigma}_p$ given in (61).

The posterior for every element of $\mathbf{c}_p$ is

$$f(c_{p,r}|\mathbf{c}_{p,-r}, \mathbf{T}, \mathbf{y}_p, \boldsymbol{\lambda}^{(\psi_p)}_{z_p^{(1)}}, \sigma^2) \propto \mathcal{N}(c_{p,r}; \mu_r, \sigma_r) \times (c_{p,r})^{(\lambda^{(\psi_p)}_{z_p^{(1)},r} - 1)} (1 - 1_{R-2}^T \mathbf{c}_{p,-r} - c_{p,r})^{\lambda^{\psi_p}_{z_p^{(1)},R} - 1} 1_{\mathcal{S}_r}(c_{p,r}) \quad (64)$$

where $\mathbf{c}_{p,-r}$ denotes the length $R - 2$ vector obtained by excluding $c_{p,r}$ from $\mathbf{c}_p$, $\mathcal{S}_r = [0, 1 - 1_{R-2}^T \mathbf{c}_{p,-r}]$, and μ_r, σ_r are the mean and variance of the Gaussian conditional PDF of $c_{p,r}$ given $\mathbf{c}_{p,-r}$ which can be obtained from (61) using [42, p. 324].

Sampling from the posterior for $c_{p,r}$ proceeds by evaluating (64) on a linearly spaced points in the interval $\mathcal{S}_r$, and sampling from the obtained PMF.

4) Sampling From the Posterior for $\boldsymbol{\lambda}_k^{(0,1)}$: Instead of sampling from the posterior of $\boldsymbol{\lambda}_k^{(i)}$ $i = 0, 1$ directly, we sample from the posterior of $s_{0,k}^{(i)}$ and $\bar{\mathbf{a}}_k$ and set $\boldsymbol{\lambda}_k^{(i)} = s_{0,k}^{(i)} \bar{\mathbf{a}}_k$.

The posterior for $s_{0,k}^{(i)}$ takes the form

$$f(s_{0,k}^{(i)}|\mathbf{z}^{(1)}, \mathbf{A}, \bar{\mathbf{a}}_k, \Psi_k^{(i)}) \propto \frac{(\Gamma(s_{0,k}^{(i)}))^{|\Psi_k^{(i)}|} \prod_{r=1}^R g_{r,i}^{s_{0,k}^{(i)} \bar{a}_{k,r}}}{\left(\prod_{r=1}^R \Gamma(s_{0,k}^{(i)} \bar{a}_{k,r}) \right)^{|\Psi_k^{(i)}|}} 1_{[l,u]}(s_{0,k}^{(i)}) \quad (65)$$

where $g_{r,i} = \prod_{p \in \Psi_k^{(i)}} a_{p,r}$, and $|\Psi_k^{(i)}|$ denotes the cardinality of the set $\Psi_k^{(i)}$. We sample from the posterior for $s_{0,k}^{(i)}$ by evaluating (65) on a linearly spaced points in the interval $[l_i, u_i]$, and sampling from the obtained PMF.

Since the DD is in the exponential family it is easy to show that the posterior for $\bar{\mathbf{a}}_k$ is also in the exponential family, however it does not have the form of any standard PDF. We therefore propose a different approach to approximately sample from the posterior of $\bar{\mathbf{a}}_k$. Let $\mathbf{a}_p \sim \mathrm{Dir}(s_{0,k}^{(i)} \bar{\mathbf{a}}_k)$ for $p \in \Psi_k^{(i)}$, $i = 0, 1$, then using (12) and the strong law of large numbers we have that

$$\hat{\mathbf{a}}_k = \frac{1}{|\Psi_k^{(0)}| + |\Psi_k^{(1)}|} \sum_{p \in \Psi_k^{(0)} \cup \Psi_k^{(1)}} \mathbf{a}_p \xrightarrow{a.s.} \bar{\mathbf{a}}_k \quad (66)$$

as $|\Psi_k^{(0)} \cup \Psi_k^{(1)}| \to \infty$. Therefore assuming that the number of samples is large the distribution of the sample mean approx-

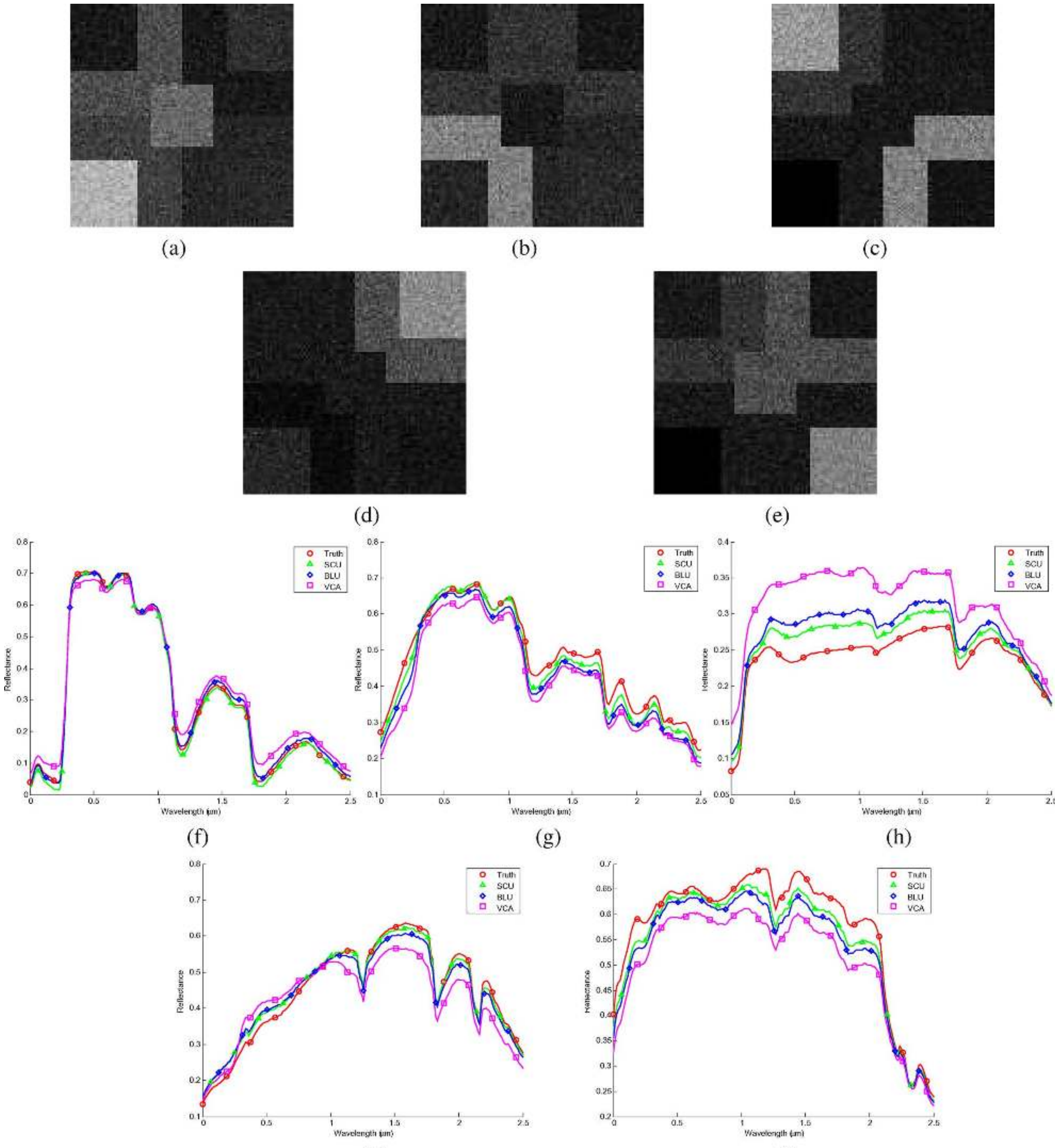


Fig. 3. The ground truth abundance maps (a)-(e), and the true endmembers and the endmembers estimated using the VCA, SCU and BLU algorithms for SNR of 15 db (f)-(j). (a) endm. 1 abundance map. (b) endm. 2 abundance map. (c) endm. 3 abundance map. (d) endm. 4 abundance map. (e) endm. 5 abundance map. (f) endm. 1 spectra. (g) endm. 2 spectra. (h) endm. 3 spectra. (i) endm. 4 spectra. (j) endm. 5 spectra.

imates the distribution of the posterior. Using the central limit theorem to approximate the PDF of the sample mean we approximate the posterior using

$$f(\bar{\mathbf{a}}_k|\mathbf{A}, s_{0,k}^{(0)}, s_{0,k}^{(1)}, \Psi) \cong \mathcal{N}\left(\hat{\mathbf{a}}_k, \bar{\mathbf{\Sigma}}_k\right) \tag{67}$$

where

$$\bar{\mathbf{\Sigma}}_k = \frac{1}{(|\Psi_k^{(0)}| + |\Psi_k^{(1)}|)^2}\left(\frac{|\Psi_k^{(0)}|}{s_{0,k}^{(0)}+1} + \frac{|\Psi_k^{(1)}|}{s_{0,k}^{(1)}+1}\right)\mathbf{\Sigma}_{\bar{\mathbf{a}}_k} \tag{68}$$

and where $\boldsymbol{\Sigma}_{\bar{\mathbf{a}}_k}$ is a symmetric matrix with the diagonal elements $\bar{a}_{k,i}(1-\bar{a}_{k,j})$ and off diagonal elements $\bar{a}_{k,i}\bar{a}_{k,j}$ for $i,j=1,\dots,R$, which follows from (13) and (14). Since $\bar{\mathbf{a}}_k$ is unavailable we replace it with the sample mean estimate $\hat{\mathbf{a}}_k$. It can be seen that the approximate posterior distribution (67) converges to a Dirac delta function as the number of samples becomes larger. In order to enforce the nonnegativity and sum-to-one constraints for $\bar{\mathbf{a}}_k$ we replace the Gaussian PDF with a DD with the same mean and covariance matrix. Therefore we can sample approximately from the posterior of $\bar{\mathbf{a}}_k$ using

$$\bar{\mathbf{a}}_k|\mathbf{A},\Psi,s_{0,k}^{(0)},s_{0,k}^{(1)}\sim \mathrm{Dir}(\hat{\mathbf{a}}_k s_k^*) \tag{69}$$

where $s_k^* = \frac{(|\Psi_k^{(0)}|+|\Psi_k^{(1)}|)^2(s_{0,k}^{(0)}+1)(s_{0,k}^{(1)}+1)}{|\Psi_k^{(0)}|(s_{0,k}^{(1)}+1)+|\Psi_k^{(1)}|(s_{0,k}^{(0)}+1)}-1$.

5) Sampling the Posterior for ψ_p: The posterior for ψ_p is of the form

$$p(\psi_p=i|\boldsymbol{\lambda}_{z_p^{(1)}}^{(i)})\propto \upsilon_i\int f(\mathrm{y}_p|\mathbf{a}_p,\Theta)1_{\mathcal{A}}(\mathbf{a}_p)\mathrm{Dir}(\mathbf{a}_p;\boldsymbol{\lambda}_{z_p^{(1)}}^{(i)})\mathrm{d}\mathbf{a}_p \tag{70}$$

where $\mathrm{Dir}(\mathbf{a};\boldsymbol{\lambda})$ denotes the PDF of a DD for the random vector a that is parameterized by $\boldsymbol{\lambda}$. Similarly to (60) we use Monte Carlo integration to approximate the integral. We first draw n_{mc} samples from (59), and then use the approximation

$$p(\psi_p=i|\boldsymbol{\lambda}_{z_p^{(1)}}^{(i)})\approx \upsilon_i\frac{c_0}{n_{mc}}\sum_{j=1}^{n_{mc}}\mathrm{Dir}(\tilde{\mathbf{a}}_p^{(j)};\boldsymbol{\lambda}_{z_p^{(1)}}^{(i)}) \tag{71}$$

where c_0 is a normalization constant which can be ignored. Therefore it is straightforward to sample ψ_p by drawing from the normalized PMF (71). In this paper, we used $n_{mc}=5$.

6) Sampling From the Posterior for σ_k^2: The posterior for σ_k^2 is an inverse Gamma distribution,

$$\sigma_k^2|\mathbf{A},\mathbf{T},\mathbf{Y},\mathrm{z}^{(1)}\sim\mathcal{IG}\left(\nu_0+\frac{PL}{2},\gamma_0+\frac{1}{2}\sum_{p\in\mathcal{R}_k}\|\mathbf{y}_p-\mathbf{M}\mathrm{a}_p\|^2\right). \tag{72}$$

7) Sampling the Posterior for $\mathbf{T}$: The posterior for $\mathbf{t}_r$, $r=1,\dots,R$ is also multivariate Gaussian truncated to the feasibility set $\mathcal{T}_r$. Let $\mathbf{T}_{-r}$ denote the matrix $\mathbf{T}$ with the r^{th} column removed, then we have that

$$\mathbf{t}_r|\mathbf{T}_{-r},\mathbf{A},\mathbf{Y},\ \{\sigma_k^2\}_{k=1}^K\sim\mathcal{N}_{\mathcal{T}_r}(\boldsymbol{\tau}_r,\boldsymbol{\Lambda}_r) \tag{73}$$

where

$$\begin{cases}\boldsymbol{\Lambda}_r=\left[\sum\limits_{p=1}^{P}\frac{1}{\sigma^2_{z_p^{(1)}}}a_{p,r}^2\mathrm{U}^T\mathbf{U}+\frac{1}{\sigma_e^2}\mathbf{I}_{\mathcal{K}}\right]^{-1}\\ \boldsymbol{\tau}_r=\boldsymbol{\Lambda}_r\left[\sum\limits_{p=1}^{P}\frac{1}{\sigma^2_{z_p^{(1)}}}a_{p,r}\mathbf{U}^T\boldsymbol{\epsilon}_{p,r}+\frac{1}{\sigma_e^2}\mathbf{e}_r\right]\end{cases} \tag{74}$$

with

$$\boldsymbol{\epsilon}_{p,r}=\mathbf{y}_p-a_{p,r}\bar{\mathbf{y}}-\sum_{j\neq r}a_{p,j}\mathbf{m}_j. \tag{75}$$

C. Computational Complexity

The main additional complexity incurred by the use of the SHDP is due to the computation of the upward predictions in (58). For $\ell>1$ the complexity of (58) is $\mathcal{O}(K^2P)$, however since the transition probabilities $\pi_{ki}^{(\ell)}$ are very sparse it can effectively be reduced to $\mathcal{O}(KP)$ without any noticeable effect on the performance. The complexity of the proposed SCU algorithm is therefore dominated by the Monte Carlo approximation used in (58) for $\ell=1$, which involves sampling from a truncated multivariate Gaussian. Let $\mathcal{O}_{\mathbf{a}}$ denote the complexity of sampling from a truncated multivariate Gaussian, then the complexity of SCU is approximately $(n_{mc}+1)P\mathcal{O}_{\mathbf{a}}$ whereas the complexity of BLU is approximately $P\mathcal{O}_{\mathbf{a}}$ since sampling the abundances entails the largest computational cost. Therefore the SCU runs about $n_{mc}+1$ times slower than the BLU.

TABLE I
ABUNDANCE MEANS FOR EACH OF THE REGIONS OF THE SYNTHETIC HYPERSPECTRAL IMAGE

Endm.	Region #								
	1	2	3	4	5	6	7	8	9
#1	0.3	0.25	0.15	0.1	0.1	0.2	0.2	0.7	0.4
#2	0.2	0.5	0.15	0.2	0.1	0.2	0.1	0.15	0.1
#3	0.2	0.1	0.5	0.1	0.6	0	0.1	0	0.1
#4	0.1	0.05	0.1	0.3	0.1	0	0.6	0.15	0.1
#5	0.2	0.1	0.1	0.3	0.1	0.6	0	0	0.3

TABLE II
THE SSE $\times 10^{-1}$ AND SAD $\times 10^{-1}$ OF THE ESTIMATED ENDMEMBERS FOR DIFFERENT SNR

SNR	Endm.	SSE			SAD		
		VCA	BLU	SCU	VCA	BLU	SCU
10db	1	1.91	0.46	**0.25**	0.78	0.35	**0.25**
	2	1.94	4.11	**1.17**	**0.2**	0.64	0.34
	3	7	1.6	**0.94**	0.91	**0.41**	0.43
	4	5.65	1.17	**0.93**	1.07	0.59	**0.46**
	5	20.19	**3.05**	3.2	0.85	**0.29**	0.33
15db	1	2.26	0.28	**0.26**	0.84	0.28	**0.26**
	2	7.38	3.46	**1.22**	0.7	0.57	**0.42**
	3	12.55	2.49	**1.04**	0.89	0.5	**0.38**
	4	5.62	1.15	**0.54**	0.98	0.49	**0.35**
	5	9.53	3.37	**1.67**	0.38	0.31	**0.24**
20db	1	2.36	0.3	**0.16**	0.86	0.3	**0.21**
	2	1.77	1.45	**0.04**	0.24	0.33	**0.08**
	3	6.2	3.03	**1.03**	1.06	0.56	**0.35**
	4	5.66	1.35	**0.38**	0.99	0.53	**0.29**
	5	9.32	3	**1.87**	0.39	**0.26**	0.28
25db	1	2.89	**0.53**	0.73	0.96	**0.4**	0.48
	2	1.75	0.71	**0.04**	0.25	0.21	**0.06**
	3	6.85	3.39	**0.75**	0.74	0.56	**0.35**
	4	5.69	1.41	**0.26**	1	0.53	**0.23**
	5	9.21	3.05	**1.34**	0.4	0.27	**0.23**

V. EXPERIMENTAL RESULTS

A. Simulations With Synthetic Data

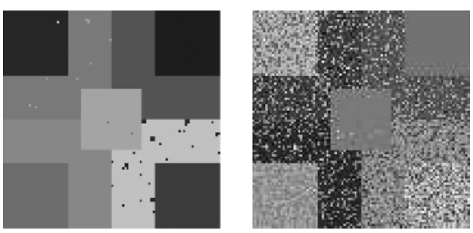
We generated a 100×100 synthetic hyperspectral image with 5 endmembers by simulating model (1) with $L=200$, where the endmember spectra were taken from [38], and the abundances were sampled from a DD with precision parameter set to 60 and the means that are given in Table I. The synthetic abundance maps are shown in Fig. 3(a)–(e).

The ground truth and estimated endmembers for the 15 db scenario are shown in Fig. 3(f)–(j), where the SNR was defined as follows:

$$\mathrm{SNR}=10\log_{10}\frac{\sum\limits_{l=1}^{L}\sum\limits_{p=1}^{P}(x_{lp}-\bar{x})^2}{LP\sigma^2} \tag{76}$$

TABLE III
THE SSE OF THE ESTIMATED ABUNDANCES FOR DIFFERENT SNR

SNR	Endm. #	Abundance SSE		
		VCA	BLU	SCU
10db	1	52.62	43.15	**23.84**
	2	96.5	54.71	**19.08**
	3	74.85	15.46	**8.25**
	4	105.1	24.96	**15.34**
	5	227.5	37.93	**30.54**
15db	1	77.46	37.28	**16.7**
	2	87.94	47.05	**16.56**
	3	71.22	20.66	**8.68**
	4	31.22	18.32	**7.01**
	5	49.44	35.31	**26.2**
20db	1	41.1	17.44	**6.19**
	2	38.2	27.04	**19.59**
	3	66.36	24.54	**8.85**
	4	42.81	16.81	**4.46**
	5	42.98	27.3	**23.27**
25db	1	27.21	9.52	**5.84**
	2	34.19	**17.51**	27.31
	3	54.37	26.66	**6.44**
	4	48.91	15.19	**3.74**
	5	45.89	23.15	**14.47**



(a) (b)

Fig. 4. The segmentation obtained for SNR 15 db using the proposed SCU sticky HDP algorithm (a), and nonsticky SCU (b). (a) SCU. (b) Nonsticky SCU.

Fig. 5. A color visible band image corresponding to the hyperspectral image of Cuprite, NV. The region of interest is marked with the black frame.

where $\mathbf{X} = \mathbf{MA}$, $x_{lp} = (\mathbf{X})_{lp}$, and $\bar{x} = \frac{1}{LP}\sum_{l=1}^{L}\sum_{p=1}^{P} x_{lp}$. The parameters that we used were $T_{bi} = 400 \text{ iterations}$, $T_{mc} = 800 \text{ iterations}$, and we used a truncation order of $K = 60$ to approximate the DPs. The number of levels in the quadtree

TABLE IV
THE MEAN, STANDARD DEVIATION, MINIMUM AND MAXIMUM, SSE $\times 10^{-1}$ OBTAINED USING 20 DIFFERENT INITIALIZATIONS FOR THE VCA, BLU AND PROPOSED SCU ALGORITHM FOR DIFFERENT SNRs

SNR	Material	VCA	BLU	SCU
10db	Kaolin #1	1.32	1.23	0.4
	Kaolin #2	0.97	0.6	0.6
	Alunite	1.08	1.03	0.7
	Montmorillonite	0.98	0.4	0.3
	Sphene	1.61	1.28	1
20db	Kaolin #1	1	0.28	0.23
	Kaolin #2	1.26	1.29	0.51
	Alunite	0.81	0.63	0.54
	Montmorillonite	1.08	0.33	0.29
	Sphene	1.74	1.53	0.99
30db	Kaolin #1	1.33	0.16	0.18
	Kaolin #2	0.99	0.69	0.36
	Alunite	1.19	0.53	0.51
	Montmorillonite	0.36	0.3	0.27
	Sphene	1.56	1.03	0.8

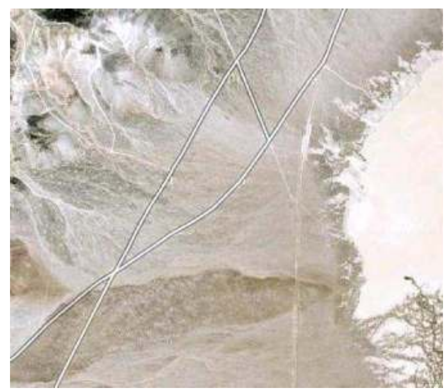


Fig. 6. A satellite image of the region of interest obtained from Google Maps. The white lines represent roads.

was set to the maximum possible levels, where as described in Section III for a 100×100 image we first extend the image to size 128×128 and prune all the branches that have no descendants that correspond to image pixels. The parameters κ, α, and γ were estimated using the method described in [23]. However, we observed that the performance is not sensitive to the exact values of these parameters. The DD parameters were initialized by using the k-means algorithm to cluster the abundances (estimated using VCA) into K classes, such that $\bar{\mathbf{a}}_k$, $k = 1, \ldots, K$ were set to the K centers, and the precision parameters were set to $s_{0,k}^{(0)} = 60$, $s_{0,k}^{(1)} = 120$. In this paper, we assume that the number of endmembers is known, however, in practice the number can be estimated using model selection methods such as [40]. It can be seen in Fig. 3(f)–(j) that the spectra that was estimated using the proposed SCU algorithm is generally closer to the true endmembers compared to the endmembers extracted using the VCA abd BLU algorithms. Table II compares the sum of squared errors (SSE) and the spectral angle distance (SAD) for the VCA, BLU, and SCU for different SNR, where the SAD is defined as

$$\mathrm{SAD}_r = \frac{\langle \hat{\mathbf{m}}_r, \mathbf{m}_r \rangle}{\|\hat{\mathbf{m}}_r\| \|\mathbf{m}_r\|}. \tag{77}$$

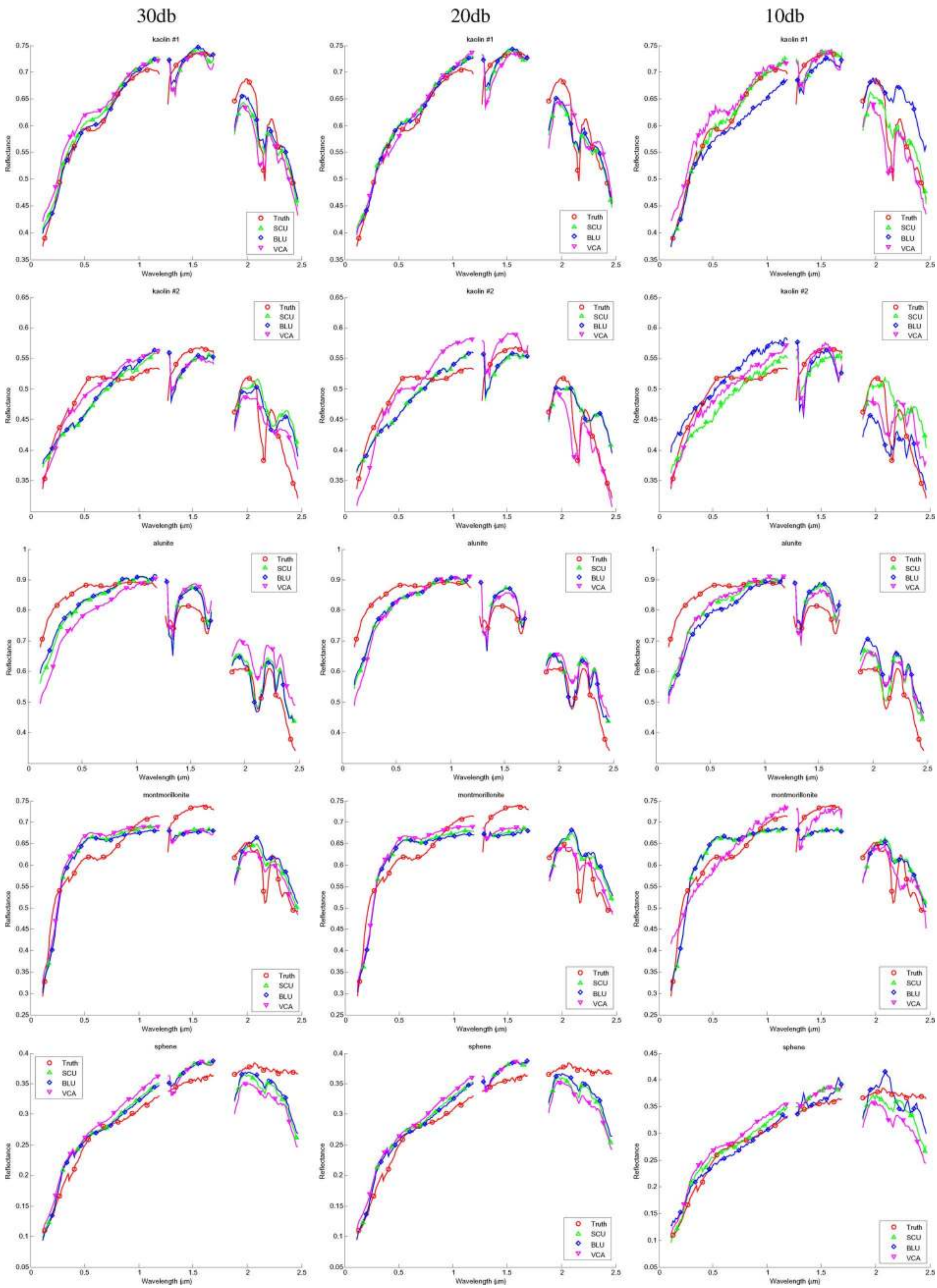


Fig. 7. The five estimated endmembers (Kaolin $\#1$, Kaolin $\#2$, Alunite $\#3$, Montmorillonite $\#4$, Sphene) for the proposed SCU algorithm as compared to ground truth and the VCA and BLU algorithms. SCU is competitive with the other methods at all SNRs.

It can be seen that the SCU performs comparably or better in all cases. The abundance SSE using the three methods is shown in Table III, where it can be verified that the SCU obtains lower SSE for almost all of the cases compared to the VCA and BLU.

TABLE V
THE MEAN, STANDARD DEVIATION, MINIMUM AND MAXIMUM, SAD $\times 10^{-1}$ OBTAINED USING 20 DIFFERENT INITIALIZATIONS FOR THE VCA, BLU, AND PROPOSED SCU ALGORITHM FOR DIFFERENT SNRS

SNR	Material	SSE											
		Mean			Standard deviation			Best			Worst		
		VCA	BLU	SCU	VCA	BLU	SCU	VCA	BLU	SCU	VCA	BLU	SCU
10db	Kaolin #1	1.34	1.23	**1.02**	0.5	0.44	**0.34**	**0.41**	0.77	0.71	**1.89**	2.63	1.9
	Kaolin #2	**1.97**	3.17	2.45	0.3	0.35	**0.29**	**1.42**	2.55	1.9	**2.8**	3.86	3.34
	Alunite	11.87	9.97	**8.42**	3.98	1.81	**1.45**	9.28	6.69	**5.19**	25	14.34	**10.54**
	Mont.	**2.59**	3.62	2.98	0.4	0.64	**0.36**	**1.88**	2.94	2.53	**3.59**	5.88	3.92
	Sphene	3	**0.96**	1.23	0.84	**0.25**	0.26	2.05	**0.56**	0.72	5.25	**1.59**	1.82
15db	Kaolin #1	1.43	**1.01**	1.1	**0.27**	0.31	0.29	**0.45**	0.65	0.79	**1.67**	1.8	1.91
	Kaolin #2	**1.87**	2.52	2.51	**0.29**	0.58	0.52	**1.4**	1.57	2.11	**2.3**	4.37	4.58
	Alunite	10.48	6.94	**6.88**	2.08	**1.13**	1.16	9.4	**3.93**	4.38	14.65	9.38	**9.22**
	Mont.	**2.43**	3.7	2.98	**0.29**	1.31	0.41	**1.75**	2.93	2.51	**3.04**	9.24	4.16
	Sphene	3.22	**0.97**	1.5	1.49	**0.19**	0.31	2.34	**0.61**	0.9	7.78	**1.53**	2.32
20db	Kaolin #1	**1.35**	1.4	**1.35**	**0.12**	0.79	0.53	1	**0.6**	0.76	**1.57**	3.28	2.84
	Kaolin #2	**1.72**	2.91	2.42	**0.36**	1.09	0.64	1	**0.6**	0.76	**1.57**	3.28	2.84
	Alunite	10.1	6.2	**5.85**	**1.53**	2.52	**1.53**	9.48	**2.68**	2.84	14.69	12.7	**9.14**
	Mont.	**2.24**	4.39	3.38	**0.31**	1.07	0.78	**1.76**	3.07	2.53	**2.98**	6.99	5.42
	Sphene	4.06	**1.36**	1.52	1.99	0.57	**0.31**	2.54	**0.64**	0.82	9.81	3.57	**1.92**
25db	Kaolin #1	1.38	**1.12**	**1.12**	**0.04**	0.68	0.31	1.32	**0.59**	0.7	**1.47**	3.1	1.72
	Kaolin #2	**1.64**	2.48	2.55	**0.18**	0.94	0.81	**1.32**	1.85	1.98	**2.07**	4.85	5.05
	Alunite	10.7	**5.83**	7.1	2.3	2.03	**1.72**	9.54	**2.84**	4.83	16.79	**12.38**	13.82
	Mont.	**2.55**	4.47	3.54	**0.27**	2.19	1.44	**1.77**	2.89	2.6	2.95	11.84	8.43
	Sphene	2.93	**1.44**	1.69	1.59	0.57	**0.32**	2.39	**0.58**	0.75	9.87	3.68	**2.11**
30db	Kaolin #1	**1.29**	1.75	1.4	**0.15**	0.93	0.64	0.69	**0.61**	0.71	**1.14**	3.79	2.87
	Kaolin #2	**1.6**	2.95	2.75	**0.27**	0.98	0.79	**1.28**	1.83	1.55	**2.22**	4.7	4.39
	Alunite	10.52	**6.35**	7.26	2.09	2.85	**2.01**	9.61	**2.83**	3.39	16.75	14.52	**12.13**
	Mont.	**2.72**	5.54	3.79	**0.76**	2.58	1.61	**1.78**	3.14	2.65	**5.45**	13.35	9.03
	Sphene	3.83	1.51	**1.5**	2.34	0.43	**0.38**	2.41	0.81	**0.65**	9.96	2.86	**2.17**

We did not observe significant differences in performance for this simulated example under different initializations and therefore we do not report multiple random start statistics.

Fig. 4 shows the segmented images obtained using the proposed SCU algorithm when using the SHDP and the standard HDP. The SHDP identified 9 classes with very few misclassified pixels whereas the HDP identified 38. Therefore, the SHDP more accurately identified the underlying ground truth segmentation which had 9 classes. This demonstrates the significance of the stickiness property for segmentation purposes.

B. Simulations With Real AVIRIS Data

In this section, we test the new approach using the AVIRIS data of Cuprite, NV, [39] which has been used previously to demonstrate the performance of hyperspectral imaging algorithms [5], [7]. A color image synthesized from the hyperspectral image is shown in Fig. 5, where we used a 80×80 pixels region of interest which is marked with a black frame, to evaluate the performance of the proposed SCU algorithm. Fig. 6 also shows a satellite image of the region of interest obtained from Google Maps, where the roads present in the image are marked by the white lines. The ground truth for the endmembers in this dataset is available at [38]. The parameter values and initialization method that were used here were identical to those that were used for the synthetic image simulations, where we used $T_{bi} = 200 \text{ iterations}$, and $T_{mc} = 400 \text{ iterations}$, and the number of endmembers was set to 5. We ran the VCA, BLU, and SCU for 20 different times, where for each run we used the same endmember initialization obtained from the VCA algorithm for the BLU and SCU algorithms. The SNR of the image as estimated by the VCA algorithm is about 30 db, therefore to illustrate the benefits of the SCU algorithm in low SNR scenarios we also evaluated the performance when adding Gaussian noise to the hyperspectral image. Tables IV and V show the mean, standard deviation, worst, and best SSE and SAD of the endmembers over the 20 runs, for SNRs of 10, 20, and 30 db. It can be seen that VCA estimates the Kaolin #1, Kailin #2, and Montomorillonite endmembers quite well, which is most likely due to the existence of pure pixels in these materials for the scene under study. VCAÕs estimate of the Alunite and Sphene endmembers is much worse, probably due to the lack of such pure pixels. Comparing the BLU and SCU it can be seen that on average they perform comparably, however for the SCU the standard deviation and worst case SSE is generally better than for the BLU. This shows that the SCU is more robust to the initialization of the endmembers in (52) which is obtained here using the VCA.

Figs. 7 and 8 show the estimated endmembers and the abundance maps, respectively, from one of the 20 runs for different SNR. It can be seen in Fig. 7 that the endmembers estimated using the BLU and SCU are generally comparable. Fig. 8 demonstrates that the abundance maps obtained using the SCU degrade far less as the SNR decreases compared to the abundances estimated using the BLU. Although we only show the results of one of the 20 runs, the abundance maps of the other runs look very similar to those shown in Fig. 8 thus it is representative of all our simulations.

There is no available ground truth for the abundances, however we argue that since the roads that are present in the image and can be seen in Fig. 6 are man-made landmarks, the ground

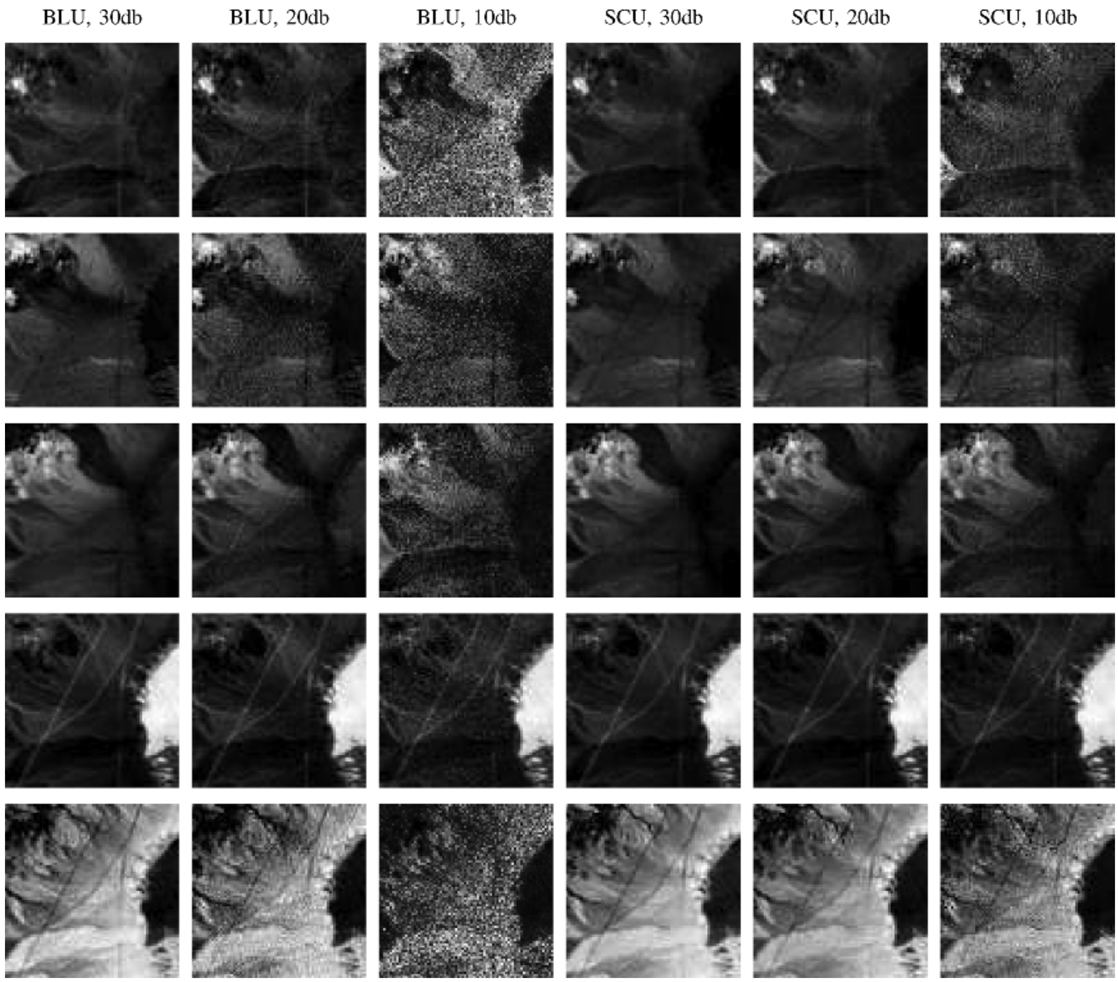


Fig. 8. The estimated abundances. The algorithm and SNR for each column is written at the top row. Each row describe the abundances of the same material (from top to bottom: Kaolin #1, Kaolin #2, Alunite, Montmorillonite, Sphene).

TABLE VI
THE VARIANCE $\times 10^{-2}$ OF THE ROAD PIXEL ABUNDANCES FOR VCA, BLU, AND PROPOSED SCU ALGORITHM FOR DIFFERENT SNRS

SNR	Material	SAD											
		Mean			Standard deviation			Best			Worst		
		VCA	BLU	SCU	VCA	BLU	SCU	VCA	BLU	SCU	VCA	BLU	SCU
10db	Kaolin #1	0.46	0.41	**0.37**	0.13	0.1	**0.07**	**0.23**	0.32	0.3	**0.61**	0.75	0.64
	Kaolin #2	**0.59**	0.81	0.72	0.07	0.06	**0.05**	**0.5**	0.62	0.55	**0.74**	0.91	0.85
	Alunite	1.02	0.91	**0.86**	0.15	**0.08**	**0.08**	0.91	0.77	**0.68**	1.5	1.11	**0.97**
	Mont.	**0.58**	0.69	0.62	0.04	0.05	**0.03**	**0.5**	0.62	0.58	**0.69**	0.88	0.72
	Sphene	1.26	**0.71**	0.8	0.16	0.09	**0.08**	1.05	**0.54**	0.62	1.68	**0.92**	0.98
15db	Kaolin #1	0.49	**0.37**	0.38	0.07	0.07	**0.06**	**0.24**	0.29	0.32	**0.57**	0.62	0.6
	Kaolin #2	**0.56**	0.72	0.72	0.08	**0.06**	**0.06**	**0.43**	0.58	0.65	**0.67**	0.89	0.99
	Alunite	0.96	**0.78**	**0.78**	0.09	**0.06**	0.07	0.92	**0.59**	0.62	1.15	0.92	**0.91**
	Mont.	**0.56**	0.69	0.62	0.03	0.1	**0.04**	**0.48**	0.62	0.57	**0.63**	1.1	0.74
	Sphene	1.29	**0.71**	0.89	0.26	**0.07**	0.09	1.12	**0.57**	0.69	2.05	**0.9**	1.11
20db	Kaolin #1	0.48	0.46	**0.42**	0.05	0.16	**0.09**	0.42	**0.28**	0.32	**0.55**	0.81	0.64
	Kaolin #2	**0.53**	0.72	0.71	**0.05**	0.16	0.09	**0.43**	0.48	0.59	**0.66**	1.03	0.95
	Alunite	0.95	0.73	**0.72**	**0.06**	0.13	**0.09**	0.92	**0.49**	0.5	1.15	1.07	**0.9**
	Mont.	**0.56**	0.75	0.66	**0.03**	0.09	0.07	**0.48**	0.63	0.58	**0.62**	0.96	0.84
	Sphene	1.44	**0.84**	0.89	0.33	0.14	**0.09**	1.16	**0.58**	0.66	2.3	1.35	**1.01**
25db	Kaolin #1	0.5	**0.38**	**0.4**	**0.05**	0.12	0.09	0.43	**0.28**	0.3	**0.54**	0.75	0.61
	Kaolin #2	**0.51**	0.71	0.7	0.07	0.11	**0.03**	**0.43**	0.63	0.65	**0.66**	1.02	0.82
	Alunite	0.97	**0.71**	0.79	0.09	0.11	**0.08**	0.92	**0.5**	0.66	1.23	**1.05**	1.11
	Mont.	**0.58**	0.75	0.67	**0.03**	0.15	0.11	**0.48**	0.62	0.58	**0.62**	1.25	1.05
	Sphene	1.23	**0.86**	0.94	0.24	0.15	**0.1**	1.13	**0.56**	0.63	2.31	1.4	**1.06**
30db	Kaolin #1	0.5	0.53	**0.47**	**0.05**	0.2	0.16	0.3	**0.28**	0.3	**0.53**	0.9	0.78
	Kaolin #2	**0.48**	0.7	0.69	0.07	**0.06**	0.07	**0.43**	0.6	0.49	**0.67**	0.86	0.85
	Alunite	0.97	**0.74**	0.8	**0.08**	0.15	0.11	0.93	**0.5**	0.55	1.23	1.14	**1.04**
	Mont.	**0.59**	0.83	0.69	**0.07**	0.17	0.13	**0.48**	0.64	0.59	**0.85**	1.32	1.09
	Sphene	1.38	**0.89**	**0.89**	0.38	**0.12**	**0.12**	1.14	0.66	**0.59**	2.32	1.24	**1.08**

truth should demonstrate the property that the abundances along the roads are more similar to each other. Table V shows the variance of the variance of the road pixels abundances, where it can be seen that the variance of road pixels abundances is lower when using the SCU compared to the VCA and BLU. This suggests that the SCU estimates the abundances more accurately compared to the other algorithms.

VI. Conclusion

We presented a Bayesian algorithm, called the spatially constrained unmixing (SCU) algorithm, which makes use of a spatial prior to unmix hyperspectral imagery. The spatial prior is enforced using a multiresolution Markov model that uses a sticky hierarchical Dirichlet process (SHDP) to determine the number of appropriate segments in the image, where the abundances are sampled from Dirichlet distribution (DD) mixture models with different parameters. We take the spatial homogeneity of the abundances into accounted by including DD mixture components with large precision parameters, whereas the outliers are modeled using a mixture component that corresponds to a uniform distribubution over the feasibility set which satisfies the nonnegativity and sum-to-one constraints. Large regions with similar abundances are most likely to be found in high resolution hyperspectral imagery, thus our proposed SCU approach is expected to be most beneficial in such images. However it is also useful in low resolution images that contain some large regions with similar abundances, e.g., a large body of water or a vegetation field. The experimental results with synthetic and real data demonstrate that our proposed SCU algorithm has improved endmember and abundance estimation performance, particularly in low SNR regimes.

References

[1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[2] J. A. Richards, "Analysis of remotely sensed data: The formative decades and the future," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 422–432, Mar. 2005.

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

[4] M. Winter, "Fast autonomous spectral end-member determination in hyperspectral data," in *Proc. 13th Int. Conf. on Appl. Geologic Remote Sens.*, Vancouver, British Columbia, Apr. 1999, vol. 2, pp. 337–344.

[5] J. M. Nascimento and J. M. Bioucas-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.

[6] T. H. Chan, C. Y. Chi, Y. M. Hunag, and W. K. Ma, "A convex analysis based minimum volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.

[7] 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 Process.*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.

[8] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2025–2041, Sep. 2002.

[9] S. Jia and Y. Qian, "Spectral and spatial complexity-based hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3867–3879, Dec. 2007.

[10] G. Martin and A. Plaza, "Spatial-spectral preprocessing for volume-based endmember extraction algorithms using unsupervised clustering," in *Proc. IEEE GRSS Workshop on Hyperspectral Image and Signal Process: Evolution in Remote Sens. (WHISPERS)*, Jun. 2010.

[11] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4239–4247, Nov. 2011.

[12] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Royal Stat. Soc. Ser. B*, vol. 36, no. 2, pp. 192–236, 1974.

[13] M. L. Comer and E. J. Delp, "Segmentation of textured images using a multiresolution Gaussian autoregressive model," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 408–420, Mar. 1999.

[14] G. Rellier, X. Descombes, F. Falzon, and J. M. Bioucas-Dias, "Texture feature analysis using Gauss Markov model in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1543–1551, Jul. 2004.

[15] R. Neher and A. Srivastava, "A Bayesian MRF framework for labeling terrain using hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1543–1551, Jul. 2004.

[16] C. A. Bouman, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. Image Process.*, vol. 3, no. 2, pp. 162–177, Mar. 1994.

[17] N. Bali and A. M. Djafari, "Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 217–225, Feb. 2008.

[18] J. Besag, "On the statatistical analysis of dirty pictures," *J. Royal Stat. Soc. Ser. B*, vol. 48, no. 3, pp. 259–302, 1984.

[19] A. S. Willsky, "Multiresolution Markov models for signal and image processing," *IEEE Proc.*, vol. 90, no. 8, pp. 1396–1458, Aug. 2002.

[20] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1566–1581, Dec. 2006.

[21] J. J. Kivinen, E. B. Sudderth, and M. I. Jordan, "Image denoising with nonparametric hidden Markov trees," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2007.

[22] J. J. Kivinen, E. B. Sudderth, and M. I. Jordan, "Learning multiscale representations of natural scenes using Dirichlet processes," in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, 2007.

[23] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "The sticky HDP-HMM: Bayesian nonparametric hidden Markov models with persistent states," MIT, Cambridge, MA, MIT LEADS Tech. Rep. P-2777, 2009.

[24] A. Hirotugu, "A new look at the statistical model identification," *IEEE Trans. Autom. Contr.*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[25] S. Gideon, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.

[26] P. J. Green, "Reversible jump MCMC computation and Bayesian determination," *Biometrika*, vol. 82, pp. 711–732, 1995.

[27] C. Andrieu, P. M. Djuric, and A. Doucet, "Model Selection by MCMC Computation," *Signal Process.*, vol. 81, no. 1, pp. 19–37, Jan. 2001.

[28] F. Bartolucci, L. Scaccia, and A. Mira, "Efficient Bayes factor estimation from the reversible jump output," *Biometrika*, vol. 92, no. 1, pp. 41–52, 2006.

[29] D. M. Blei and M. I. Jordan, "Variational inference for Dirichlet process mixtures," *Bayesian Anal.*, vol. 1, no. 1, pp. 121–144, 2005.

[30] J. Paisley and L. Carin, "Hidden Markov Mmdels with stick-breaking priors," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3905–3917, Oct. 2009.

[31] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *Ann. Statistics*, vol. 1, no. 2, pp. 209–230, Mar. 1973.

[32] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica*, vol. 4, pp. 639–650, 1994.

[33] D. Blackwell and J. B. Macqueen, "Ferguson distributions via Pólya urn schemes," *Ann. Statist.*, vol. 1, no. 2, pp. 353–355, 1973.

[34] H. Ishwaran and M. Zarepour, "Markov chain Monte Carlo approximate Dirichlet and beta two-parameter process hierarchical models," *Biometrika*, vol. 87, no. 2, pp. 371–390, 2000.

[35] H. Ishwaran and M. Zarepour, "Exact and approximate sum-representations for the Dirichlet process," *Canadian J. Statist.*, vol. 30, no. 2, pp. 269–283, Jun. 2002.

[36] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer, 2004.

[37] M. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet – based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[38] R. N. Clark, G. A. Swayze, R. Wise, E. Livo, T. Hoefen, R. Kokaly, and S. J. Sutley, USGS Digital Spectral Library Splib06a, U.S. Geological Survey, 2007, vol. 231, Digital Data Series [Online]. Available: http://speclab.cr.usgs.gov/spectral.lib06

[39] AVIRIS Free Data, Jet Propulsion Lab (JPL), Calif. Inst. Technol., Pasadena, CA, 2006 [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.freedata.html

[40] N. Dobigeon, J. Y. Tourneret, and C. I. Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, Jul. 2008.

[41] T. P. Minka, Estimating a Dirichlet Distribution [Online]. Available: http://research.microsoft.com/en-us/um/people/minka/papers/dirichlet/minka-dirichlet.pdf

[42] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[43] R. Mittelman and A. O. Hero, "Hyperspectral image segmentation and unmixing using hidden Markov trees," in *Proc. IEEE Conf. on Image Process. (ICIP)*, Hong Kong, Sep. 2010.

Roni Mittelman (S'08–M'09) received the B.Sc. and M.Sc. (*cum laude*) degrees in electrical engineering from the Technion— Israel Institute of Technology, Haifa, and the Ph.D. degree in electrical engineering from Northeastern University, Boston, MA, in 2002, 2006, and 2009, respectively.

Since 2009, he has been a postdoctoral research fellow with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. His research interests include statistical signal processing, machine learning, and computer vision.

Nicolas Dobigeon (S'05–M'08) was born in Angoulême, France, in 1981. He received the Eng. degree in electrical engineering from ENSEEIHT, Toulouse, France, and the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse, both in 2004, and the Ph.D. degree in signal processing also from the National Polytechnic Institute of Toulouse in 2007.

From 2007 to 2008, he was a Postdoctoral Research Associate with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. Since 2008, he has been an Assistant Professor with the National Polytechnic Institute of Toulouse (ENSEEIHT – University of Toulouse), within the Signal and Communication Group of the IRIT Laboratory. His research interests are centered around statistical signal and image processing with a particular interest to Bayesian inference and Markov chain Monte Carlo (MCMC) methods.

Alfred O. Hero III (S'79–M'84–SM'96–F'98) received the B.S. (*summa cum laude*) from Boston University, Boston, MA, in 1980 and the Ph.D. degree from Princeton University, Princeton, NJ, in 1984, both in electrical engineering.

Since 1984, he has been with the University of Michigan, Ann Arbor, where he is the R. Jamison and Betty Williams Professor of Engineering. His primary appointment is with the Department of Electrical Engineering and Computer Science and he also has appointments, by courtesy, with the Department of Biomedical Engineering and the Department of Statistics. In 2008, he was awarded the Digiteo Chaire d'Excellence, sponsored by Digiteo Research Park in Paris, located at the Ecole Superieure d'Electricite, Gif-sur-Yvette, France. He has held other visiting positions at LIDS Massachusetts Institute of Technology (2006), Boston University (2006), I3S University of Nice, Sophia-Antipolis, France (2001), Ecole Normale Supérieure de Lyon (1999), Ecole Nationale Supérieure des Télécommunications, Paris (1999), Lucent Bell Laboratories (1999), Scientific Research Labs of the Ford Motor Company, Dearborn, Michigan (1993), Ecole Nationale Superieure des Techniques Avancees (ENSTA), Ecole Superieure d'Electricite, Paris (1990), and M.I.T. Lincoln Laboratory (1987–1989). His recent research interests have been in detection, classification, pattern analysis, and adaptive sampling for spatio-temporal data. Of particular interest are applications to network security, multimodal sensing and tracking, biomedical imaging, and genomic signal processing.

Dr. Hero was awarded a University of Michigan Distinguished Faculty Achievement Award in 2011. He has been plenary and keynote speaker at major workshops and conferences. He has received several best paper awards including: a IEEE Signal Processing Society Best Paper Award (1998), the Best Original Paper Award from the *Journal of Flow Cytometry* (2008), and the Best Magazine Paper Award from the IEEE Signal Processing Society (2010). He received the IEEE Signal Processing Society Meritorious Service Award (1998), the IEEE Third Millenium Medal (2000), and the IEEE Signal Processing Society Distinguished Lecturership (2002). He was President of the IEEE Signal Processing Society (2006–2007). He was on the Board of Directors of IEEE (2009–2011) where he served as Director Division IX (Signals and Applications).