

# BOOTSTRAPPING FOR PIECE-WISE CONVEX ENDMEMBER DISTRIBUTION DETECTION

Alina Zare\*, Paul Gader†, Tim Allgire\*, Dmitri Dranishnikov,† Ryan Close†

\*Electrical and Computer Engineering, University of Missouri

†Computer and Information Science and Engineering, University of Florida

## ABSTRACT

A hyperspectral endmember detection and spectral unmixing algorithm that finds multiple sets of endmember distributions is presented. If endmembers are represented as random vectors, then they can be characterized by a multivariate probability distribution. These distributions are referred to as endmember distributions. The proposed method combines the Piece-wise Convex Multiple Model Endmember Detection (PCOMMEND) algorithm, the Sparsity Promoting Iterated Constrained Endmembers (SPICE) algorithm, and the Competitive Agglomeration (CA) algorithm to estimate endmember distributions. The goal is to produce distributions that are suitable for inclusion into the Normal Compositional Model (NCM). PCOMMEND forms a fuzzy partition of the spectral pixels into a collection of fuzzy convex sets. Each convex set is defined by a set of endmembers and the linear mixing model. In this way, non-convex hyperspectral data are more, accurately characterized. The SPICE algorithm estimates the number of endmembers, the endmembers, and the abundances for each convex set. This process is repeated several times; each time a set of endmembers is produced. The collection of all such sets is merged into a single set of endmembers. This set is clustered using the CA algorithm, which estimates the number of endmembers by estimating the number of clusters and prototypes for each cluster in the single set of endmembers. These prototypes are taken to be the means of endmember distributions. The covariances are estimated by assigning each endmember to the closest prototype and estimating the covariance of that set. The resulting distributions are suitable for the NCM model. Results are shown for the PAVIA data set.

## 1. INTRODUCTION

The spectral signatures of the pure materials in a hyperspectral scene are often referred to as *endmembers* [1]. *Spectral unmixing* is the task of decomposing pixels from a hyperspectral image into their respective endmembers and *abundances*. Abundances are the proportions of every endmember in each pixel in a hyperspectral image. The standard model used to perform spectral unmixing is the *linear mixing model*. This

model states that every pixel is a convex combination of endmembers in the scene. This has been shown in the literature to hold in cases where the spectra of the endmember are mixed by the spatial resolution of the imaging sensor [1]. If the linear mixing model holds, the 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 shown in Equation (1),

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

where  $N$  is the number of pixels in the image,  $M$  is the number of endmembers,  $\epsilon_i$  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 Equation (2),

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

Under the standard linear mixing model, hyperspectral pixels are assumed to be convex combinations of all of the endmembers in a scene. However, consider scenes composed of multiple regions that do not share common materials. Pixels in each region are convex combinations of a subset of the endmembers in the scene (i.e., mixtures of the endmembers associated with the corresponding region). The set of all image spectra will consist of a union of all of the simplices defined by the multiple subsets of endmembers and the union on these simplices is unlikely to be convex. Several spectral unmixing methods that assume each pixel is a mixture of a small number of the endmembers from the full scene have been investigated. These include work in sparse unmixing methods [2–5], piece-wise convex unmixing methods [6–8], and methods such as [9, 10] where each pixel can be modeled with a different numbers and types of endmembers and [11] spatially-dependent endmembers.

The proposed method extends the use of the piece-wise convex unmixing model to incorporate NCM[12] and to estimate the number of endmembers needed for each convex set. Under NCM, endmembers are represented using Gaussian distributions rather than a single point in high dimen-

sional space. By using a full distribution, the spectral variability resulting from environmental conditions, the inherent variability in a material, and others is captured in the endmember distribution representation.

The algorithm presented here combines PCOMMEND [6], the SPICE [13] algorithm and the CA algorithm [14] to partition input data into convex regions to represent non-convex hyperspectral data sets, represent endmember spectral variability through the use of endmember distributions, and estimate the number of endmember distributions needed for each convex set.

In the following, Section 2 describes the proposed method, Section 3 presents results on the HySens Pavia data set, and Section 4 provides discussion and future work.

## 2. METHODOLOGY

The proposed method partitions the input hyperspectral data using PCOMMEND, estimates endmember distributions and the number of endmember distributions on each partition by repeatedly applying the SPICE algorithm and then clustering the set of estimated endmembers using the CA algorithm. Pseudo-code of the proposed method is shown below.

---

### Algorithm 1 Bootstrapping for Endmember Distribution Estimation

---

- Set the number of convex regions,  $C$
  - Set the number of bootstrap repetitions,  $R$
  - Apply PCOMMEND to estimate the fuzzy partition of the input data into  $C$  convex sets
  - Assign each input data point to the partition to which it has the highest membership
  - FOR  $c \leftarrow 1 \dots C$ 
    - FOR  $r \leftarrow 1 \dots R$ 
      - \* Apply SPICE algorithm to  $X_c$ , the data points assigned to convex region  $c$ , to estimate endmembers,  $E_{c,r}$  and the number of endmembers
    - ENDFOR
    - Apply CA algorithm to  $\cup_{r=1}^R E_{c,r}$  to cluster estimated endmembers and estimate the number of endmember distributions
    - Compute mean and covariance of each endmember cluster to define the endmember distributions for partition  $c$
  - ENDFOR
- 

### 2.1. PCOMMEND

PCOMMEND is applied to partition the input hyperspectral data into convex regions. PCOMMEND estimates endmember sets and proportion values by minimizing the objective function using alternating optimization

$$J_P(\mathbf{E}, \mathbf{P}, \mathbf{U}) = \sum_{i=1}^C \left( \sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{E}_i \mathbf{p}_{ij})^T (\mathbf{x}_j - \mathbf{E}_i \mathbf{p}_{ij}) + \alpha \sum_{k=1}^{M-1} \sum_{j=k+1}^M (\mathbf{e}_{ik} - \mathbf{e}_{ij})^T (\mathbf{e}_{ik} - \mathbf{e}_{ij}) \right) \quad (3)$$

such that

$$p_{ik} \geq 0 \quad \forall k = 1, \dots, M; \quad \sum_{k=1}^M p_{ik} = 1 \quad (4)$$

$$u_{ik} \geq 0 \quad \forall k = 1, \dots, M; \quad \sum_{k=1}^M u_{ik} = 1 \quad (5)$$

where  $\mathbf{x}_j$  is a  $d \times 1$  column vector representing the  $j^{th}$  pixel,  $C$  is the number of endmember sets being estimated,  $\mathbf{p}_{ij}$  is the vector of proportion values associated with pixel  $j$  for the  $i^{th}$  endmember set,  $\mathbf{E}_i$  is a  $d \times M$  matrix such that each column of  $\mathbf{E}_i$ ,  $\mathbf{e}_{ik}$ , is the  $d \times 1$  vector representing the  $k^{th}$  endmember in the  $i^{th}$  endmember set, weight  $u_{ij}$  is the membership value of the  $j^{th}$  data point in the  $i^{th}$  endmember set,  $\alpha$  is a fixed parameter used to balance the two terms of the objective function, and  $m$  is a fixed parameter which controls the degree of sharing across endmember sets.

The first term of this objective computes the squared Euclidean distance between each input hyperspectral data point and their estimate using the estimated endmembers and proportion values. The second term is used to constrain the size of each simplex by minimizing the squared Euclidean distance between each pair of endmembers in an endmember set. The membership values,  $u_{ij}$ , in the first term indicate the degree to which pixel  $j$  is represented by the endmembers of convex set  $i$ .

After applying PCOMMEND, each data point is assigned to the convex region with the largest associated membership value. Although PCOMMEND partitions the data into convex regions, the algorithm assumes that all convex regions are represented with the same number of endmembers nor does it estimate the number of endmembers needed for each convex region. Therefore, after partitioning the data into convex regions, the SPICE algorithm is applied repeatedly to each convex region to estimate endmembers. These endmembers are then clustered using CA to determine endmember distributions and the number of endmember distributions needed for each convex region. This resulted in estimating 5 endmember distributions for each of the two partitions (for a total of 10 endmember distributions).

## 2.2. SPICE

The SPICE algorithm simultaneously performs spectral unmixing, estimates endmembers and estimates the number of endmembers for an input hyperspectral data set. This is done using alternating optimization to minimize the following objective function,

$$\begin{aligned}
 J_S(\mathbf{E}, \mathbf{P}) = & (1 - \mu) \sum_{j=1}^N (\mathbf{x}_j - \mathbf{E}_i \mathbf{p}_{ij})^T (\mathbf{x}_j - \mathbf{E}_i \mathbf{p}_{ij}) \\
 & + \mu \sum_{k=1}^{M-1} \sum_{j=k+1}^M (\mathbf{e}_{ik} - \mathbf{e}_{ij})^T (\mathbf{e}_{ik} - \mathbf{e}_{ij}) \\
 & + \sum_{k=1}^M \gamma_k \sum_{i=1}^N p_{ik} \quad (6)
 \end{aligned}$$

such that

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

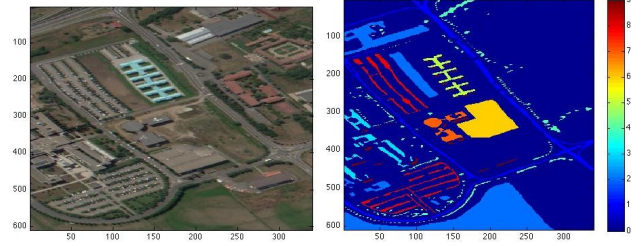
where  $\mu$  is a fixed parameter used to balance the two terms of the objective function and  $\gamma_k = \frac{\Gamma}{\sum_{i=1}^N p_{ik}}$  computed using the proportion values from the previous iteration and fixed parameter  $\Gamma$ . The two terms of this objective computes the squared Euclidean distance between each input hyperspectral data point and their estimate using the estimated endmembers and proportion values and the sum of squared differences between the estimated endmembers (similar to the first two terms in PCOMMEND). The third term of this objective is a sparsity promoting term on the proportion values associated with each endmember. This term is used to determine the number of needed endmembers for an input data set by driving the proportion values for unneeded endmembers to zero.

## 2.3. Competitive Agglomeration Clustering

Competitive Agglomeration is a fuzzy clustering algorithm that clusters input data and determines the number of clusters. CA is applied to the endmembers estimated by repeatedly applying SPICE to each convex region. The resulting clusters provide an estimate of the endmember distributions (by computing the mean and covariance of each of the endmember clusters) and the number of endmember distributions needed for each convex region.

CA is applied to cluster endmembers and estimates the number of endmember distributions by minimizing the following objective function using alternating optimization.

$$\begin{aligned}
 J_C(\mathbf{R}, \mathbf{V}) = & \sum_{m=1}^M \sum_{n=1}^N v_{mn} (\mathbf{e}_n - \mathbf{r}_m)^T (\mathbf{e}_n - \mathbf{r}_m) \\
 & + \beta \sum_{m=1}^M \sum_{n=1}^N v_{mn}^2 \quad (7)
 \end{aligned}$$



**Fig. 1.** (a) Image of the ROSIS Pavia University data set (using bands 56, 29, and 12 to for R, G, and B channels to create the color image) collected over Pavia, Italy. (b) Groundtruth for the ROSIS Pavia data set with the following categories (0) Unlabeled (1) Asphalt (2) Meadows (3) Gravel (4) Trees (5) Metal Sheets (6) Bare Soil (7) Bitumen (8) Self-blocking Bricks (9) Shadow

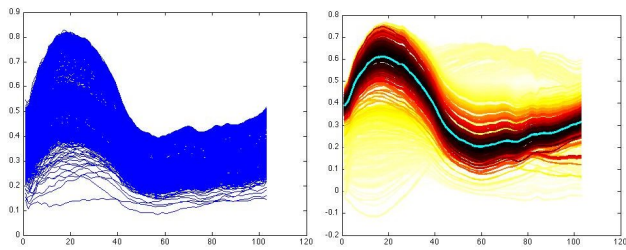
where  $\mathbf{r}_m$  is the mean of the  $m^{th}$  endmember distribution,  $\mathbf{e}_n$  is the  $n^{th}$  endmember estimated using the SPICE algorithm for the convex region under consideration,  $v_{nm}$  is the membership of the  $n^{th}$  endmember in the  $m^{th}$  endmember distribution, and  $\beta$  is a parameter used to balance the two terms of the objective function.

After clustering the estimated endmembers using CA, each endmember is assigned to the associated cluster with the highest membership value. The endmember distributions are defined by computing the mean and covariance of the endmembers associated with each cluster.

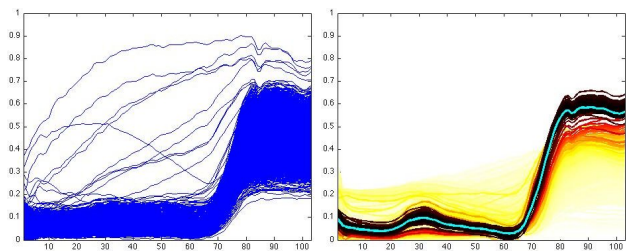
## 3. EXPERIMENTAL RESULTS

The proposed method was applied to the HySens collected by the Reflective Optics System Imaging Spectrometer (ROSIS) over Pavia University. These data were collected on July 8, 2002 over an urban area of Pavia in northern Italy by the ROSIS spectrometer. The image contains  $610 \times 340$  pixels with 103 bands. The image contains both natural and urban regions as shown in Figure 1. PCOMMEND was applied to this data to partition the data into 2 convex regions with the following parameter settings,  $C = 2, m = 2, M = 3, \alpha = 0.001$ . After partitioning the data with PCOMMEND, SPICE was applied to each convex region 250 times with the following parameter settings,  $\mu = 0.01, \Gamma = 20$ . Then, CA was applied to the endmembers found by repeatedly applying SPICE.

For evaluation, the endmember distributions were compared to the distribution of spectra assigned to each groundtruth class. Examples are shown in the following. Figure 2 shows a comparison of the pixels assigned to the Metal Sheet groundtruth class and the corresponding endmember distribution identified by the proposed approach. Figure 3 shows a comparison of the pixels assigned to the Trees groundtruth class and the corresponding endmember distribution identified by the proposed approach.



**Fig. 2.** (a) Spectra associated with ground truth class Metal Sheets (b) Endmember distribution estimated using the proposed approach matching the Metal Sheets ground truth class. The spectra are colored based on their membership to the endmember distribution as estimated by the CA algorithm.



**Fig. 3.** (a) Spectra associated with ground truth class Trees (b) Endmember distribution estimated using the proposed approach matching the Trees ground truth class.

#### 4. SUMMARY AND FUTURE WORK

In summary, the proposed algorithm applied PCOMMEND, SPICE, and CA to partition input data into convex sets and estimate endmember distributions, and the number of endmember distributions needed for each convex set. The proposed approach is can unmix non-convex data and estimate the spectral variability associated with each endmember,

## References

- [1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, pp. 44–57, 2002.
- [2] John B. Greer, "Sparse demixing," in *Proceedings of the SPIE, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVI*, Sylvia S. Shen and Paul E. Lewis, Eds., Orlando, Florida, May 2010, vol. 7695.
- [3] Zuyuan Yang, Guoxu Zhou, Shengli Xie, Shuxue Ding, Jun-Mei Yang, and Jun Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *Image Processing, IEEE Transactions on*, vol. 20, no. 4, pp. 1112–1125, april 2011.
- [4] M.-D. Iordache, J.M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 6, pp. 2014–2039, june 2011.
- [5] Yuntao Qian, Sen Jia, Jun Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via sparsity-constrained non-negative matrix factorization," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 11, pp. 4282–4297, nov. 2011.
- [6] A. Zare and P. Gader, "Piece-wise convex spatial-spectral unmixing of hyperspectral imagery using possibilistic and fuzzy clustering," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, Taipei, Taiwan, June 2011.
- [7] A. Zare and P. Gader, "PCE: Piece-wise convex endmember detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2620–2632, 2010.
- [8] A. Zare and P. Gader, "An investigation of likelihoods and priors for bayesian endmember estimation," *Proceedings of the 30th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, July 2010.
- [9] P.E. Dennison and D. A. Roberts, "Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE," *Remote Sensing of Environment*, vol. 87, pp. 123–135, 2003.
- [10] D.M. Rogge, B. Rivard, Jinkai Zhang, and Jilu Feng, "Iterative spectral unmixing for optimizing per-pixel endmember sets," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, no. 12, pp. 3725–3736, dec. 2006.
- [11] K. Canham, A. Schlamm, A. Ziemann, B. Basener, and D. Messinger, "Spatially adaptive hyperspectral unmixing," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 11, pp. 4248–4262, nov. 2011.
- [12] D. Stein, "Application of the normal compositional model to the analysis of hyperspectral imagery," in *Advances in Techniques for Analysis of Remotely Sensed Data, 2003 IEEE Workshop on*, oct. 2003, pp. 44–51.
- [13] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection for hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 3, pp. 446–450, July 2007.
- [14] Hichem Frigui and Raghu Krishnapuram, "Clustering by competitive agglomeration," *Pattern Recognition*, vol. 30, no. 7, pp. 1109–1119, 1997.