Sparsity Promoting Iterated Constrained Endmember Detection with Integrated Band Selection

Alina Zare, Paul Gader *Senior Member, IEEE*

***Abstract*— An extension of the Iterated Constrained Endmembers (ICE) that incorporates sparsity promoting priors to find the correct number of endmembers and simultaneously select informative spectral bands is presented. In addition to solving for endmembers and endmember fractional maps, this algorithm attempts to autonomously determine the number of endmembers required for a particular scene. The number of endmembers is found by adding a sparsity-promoting term to ICE's objective function. Additionally, hyperspectral band selection is performed by incorporating weights associated with each hyperspectral band. A sparsity promoting term for the band weights is added to the objective function to perform band selection.**

***Index Terms*—Sparsity Promotion, Endmember, Hyperspectral Imagery**

I. Introduction

AUTONOMOUS endmember detection is a difficult problem in hyperspectral imaging. Many endmember extraction algorithms have been formulated but the majority of these algorithms require the knowledge of the number of endmembers required for a scene. The problem of autonomously determining the number of required endmembers to a large extent has not been tackled.

We provide an extension of the Iterated Constrained Endmembers (ICE) Algorithm [1] that provides better estimates of the number of endmembers required for a dataset. This extension adds a sparsity-promoting term to the ICE objective function and is, therefore, referred to as SPICE. This added term encourages the pruning of unnecessary endmembers.

Band selection is also performed by incorporating band weights into the objective function. Band selection is performed simultaneously with endmember determination by adding a sparsity promoting term for the band weights into SPICE's objective function.

In Section II, we review the ICE algorithm and discuss the sparsity promoting extension for both endmember determination and band selection. It is assumed that the reader is familiar with the endmember detection and hyperspectral band selection problems. In Section III, we present results from artificial and real image data. Section IV is the conclusion.

II. ICE with Sparsity Promotion

A. Review of ICE Algorithm

The ICE Algorithm performs a least squares minimization of the residual sum of squares (RSS) based on the convex geometry model. The convex geometry model assumes that every pixel in a scene is a linear combination of the endmembers of the scene. The convex geometry model can be written as

$$\mathbf{X}_i = \sum_{k=1}^{M} p_{ik}\mathbf{E}_k + \varepsilon_i \qquad i = 1,\ldots,N \tag{1}$$

where N is the number of pixels in the image, M is the number of endmembers, ε_i is an error term, p_{ik} is the proportion of endmember k in pixel i, and $\mathbf{E}_k$ is the k^{th} endmember. The proportions satisfy the constraints

$$\sum_{k=1}^{M} p_{ik} = 1,\ p_{ik} \geq 0, k = 1,\ldots,M\ . \tag{2}$$

By minimizing the RSS, subjected to the constraints in (2), the error between the pixel spectra and the pixel estimate found by the ICE algorithm for the endmembers and their proportions is minimized.

$$RSS = \sum_{i=1}^{N}\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right)^T\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right) \tag{3}$$

As described in [1], the minimizer for RSS is not unique. Therefore, the ICE algorithm adds a sum of squared distances (SSD) term to the objective function.

$$SSD = \sum_{k=1}^{M-1}\sum_{l=k+1}^{M}\left(\mathbf{E}_k - \mathbf{E}_l\right)^T\left(\mathbf{E}_k - \mathbf{E}_l\right) = M(M-1)V \tag{4}$$

where V is the sum of variances (over the bands) of the simplex vertices. The second equality is shown in [1]. As done in [1], V is used in the objective function instead of $M(M-1)V$ in an effort to make this term independent of the number of endmembers, M. This term is related to the size of the area bounded by the endmembers. Therefore, by adding this term to the objective function, the algorithm finds endmembers that provide a tight fit around the data.

Therefore, the objective function used in the ICE algorithm is

$$RSS_{reg} = (1-\mu)\frac{RSS}{N} + \mu V \quad (5)$$

where μ is the regularization parameter that balances the RSS and SSD terms of the objective function.

The ICE algorithm minimizes this objective function iteratively. First, given endmember estimates, the proportions for each pixel are estimated. For the first iteration of the algorithm, endmember estimates may be set to randomly chosen pixels from the image. This requires a least squares minimization of each term in (3). Since each of these terms is quadratic and subjected to the linear constraints in (2), the minimization is done using quadratic programming. After solving for the proportions, the endmembers are found using the current proportion estimates:

$$\mathbf{e}_j = \left\{ \mathbf{P}^T\mathbf{P} + \lambda\left(\mathbf{I}_M - \frac{\mathbf{1}\mathbf{1}^T}{M}\right)\right\}^{-1}\mathbf{P}^T x_j \quad (6)$$

where **P** is the N x M proportion matrix, $\mathbf{e}_j$ is the vector of endmember values in the j*th* band, $\mathbf{x}_j$ is the vector of all the pixel values in the j*th* band, $\mathbf{I}_M$ is the M x M identity matrix, **1** is the M-vector of ones and $\lambda = N\mu/\{(M-1)(1-\mu)\}$. This iterative procedure is continued until the value of RSS_{reg} is smaller than a tolerance value. Although the ICE algorithm is an excellent algorithm for finding endmembers when the number of endmembers is known, there is no automated mechanism in ICE to determine the correct number of endmembers. Our proposed extension uses sparsity promoting priors to alleviate this disadvantage.

B. Endmember Sparsity Promotion

To promote sparsity and prune unnecessary endmembers, the following sparsity promoting term is added to the ICE objective function.

$$SPT = \sum_{k=1}^{M}\gamma_k\sum_{i=1}^{N}|p_{ik}| = \sum_{k=1}^{M}\gamma_k\sum_{i=1}^{N}p_{ik} \quad (7)$$

where the last equality follows due to the constraints in (2). For this work, we take

$$\gamma_k = \frac{\Gamma}{\sum_{i=1}^{N} p_{ik}}. \quad (8)$$

Γ is a constant associated with the degree that the proportion values are driven to zero. The advantage of this expression for γ_k is that as the proportion values change during the minimization of the objective function, the weight associated with each endmember adjusts accordingly. If the sum of a particular endmember's proportion values becomes small, then the weight, γ_k, for that endmember becomes larger. This weight change accelerates the minimization of those proportion values.

As described in [2], the use of the sparsity promoting term in (7) can be seen as placing a sparsity promoting Laplacian prior on the parameters.

C. Band Selection

Band weights and a band sparsity promoting term are added to the objective function to perform band selection. The band sparsity promoting term is defined as a weighted sum of weights, with one term for each band. The weights in this sum are band sparsity variables:

$$BST = \sum_{j=1}^{d}\lambda_j|w_j| = \sum_{j=1}^{d}\lambda_j w_j \quad (9)$$

where

$$\lambda_j = \frac{\Lambda\sum_{k=1}^{M}\frac{1}{|\psi_k|}\sum_{x_i\in\psi_k}(\mathbf{X}_i-\mathbf{E}_k)^T(\mathbf{X}_i-\mathbf{E}_k)}{\sum_{k=1}^{M}(\mathbf{E}_k-\boldsymbol{\mu}_0)^T(\mathbf{E}_k-\boldsymbol{\mu}_0)}, \quad (10)$$

$$\psi_k = \left\{\mathbf{X}_i \middle| \arg\max_l p_{il} = k, p_{ik} \ge T\right\}, \quad (11)$$

Λ is a tunable parameter controlling the degree of sparsity among the band weights, and *T* is a constant threshold value. The band weights are subjected to the following constraints:

$$w_j \ge 0, j = 1,....d, \quad \sum_{j=1}^{d} w_j = d \quad (12)$$

where *d* is the number of bands. Note that, if a λ_j value is small, then the associated band weights can be large and still yield a small value in the objective function. Conversely, if a λ_j value is large, then the associated weight must be small. Hence, a large λ_j value for a particular band should lead to small weights for that band. Thus, the λ_j values are defined to depend on the ratio of the within-class to between class scatter. Each endmember has one class that consists of those points with high abundances with respect to the corresponding endmember. So, bands with small ratios separate the data well and are therefore encouraged to have large weights whereas bands with large ratios do not separate the data well and are encouraged to have small weights.

D. Band Selecting Sparsity Promoting Iterated Constrained Endmembers (B-SPICE)

Incorporating the band weights, adding the band sparsity term and endmember sparsity promoting term to the SPICE objective function yields

$$J = \eta\frac{RSS_B}{N} + \beta V_B + SPT + BST \quad (13)$$

where

$$RSS_B = \sum_{i=1}^{N}\left(\mathbf{W}\mathbf{X}_i - \sum_{k=1}^{M}p_{ik}\mathbf{W}\mathbf{E}_k\right)^T\left(\mathbf{W}\mathbf{X}_i - \sum_{k=1}^{M}p_{ik}\mathbf{W}\mathbf{E}_k\right),$$

$$SSD_B = \sum_{k=1}^{M-1}\sum_{l=k+1}^{M}(\mathbf{W}\mathbf{E}_k - \mathbf{W}\mathbf{E}_l)^T(\mathbf{W}\mathbf{E}_k - \mathbf{W}\mathbf{E}_l),$$

W is a diagonal matrix with the weights for each band lying along the diagonal, and η and β are the constant coefficient parameters for the RSS term and SSD terms.

In order to minimize this new objective function, the iterative procedure used in ICE can still be used. The endmembers are still found by solving (6) since neither SPT nor BST depend on the endmembers. When solving for the proportion values given endmember and band weight estimates, N quadratic programming steps, one for each data

point, can be employed to minimize (13) with respect to the constraints in (2).

Similarly, when solving for the band weights given the proportion and endmember estimates, (13) needs to be minimized using quadratic programming given the constraints in (12).

The minimization iterates between solving for the proportions, endmember and band weights. The band weights do not need to be updated each iteration. An update schedule for the band weights can be employed where they are updated every *n* iterations. By employing an update schedule, it allows the endmembers and proportions to settle initially before determining band weights. Furthermore, after the band weights have stabilized, they are no longer updated. This is done since (10) is computed using the current endmember and proportion values. After the band weights have stabilized, they are no longer updated to allow the objective function to minimize. This iterative minimization is continued until some stopping criteria are met. In the results shown in this paper, the stopping criteria was when the objective function value no longer changed significantly from iteration to iteration.

During the iterative minimization process, endmembers can be pruned as their proportion values drop below a pruning threshold. Every iteration of the minimization process, the maximum proportion values for every endmember can be calculated.

$$MAXP_k = \max_i \{p_{ik}\} \tag{14}$$

If the maximum proportion for an endmember drops below a threshold, then the endmember can be pruned from the endmember set.

III. Results

A. Toy Example

A toy 5-D example was initially used for testing the B-SPICE algorithm. Figure 1 shows the first two dimensions of the data set and the endmembers from which the data was generated. Three dimensions of Gaussian random noise were added to create unnecessary bands for removal using the band selection process. The data points were generated in the same fashion as the toy example in [1] with the addition of three unnecessary bands. Zero-mean, independent Gaussian random noise was added to the *x*- and *y*-coordinates of all the data generated.

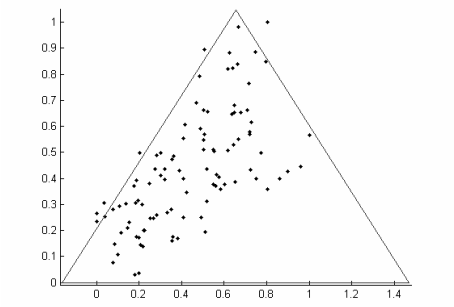

Figure 1: First two dimensions of Toy Data set were generated from the corners of the simplex shown above.

The results of three experiments comparing the ICE and B-SPICE algorithms are shown in Figure 2. The parameters for each algorithm, other than the sparsity promoting term, were held the same during the experiments: η was 100, β was set to 1, Λ was set to 0.01, and Γ was set to 1 for B-SPICE. The initial number of endmembers for all three runs was 20. Furthermore, the band weights were updated every fifth iteration, T was set to 0.54 and the endmember pruning threshold was set to 1e-10. The endmembers were initialized to the same values for each experiment comparing ICE and SPICE. These initial endmembers were chosen randomly from the data set.

As can be seen in the results shown, SPICE consistently determined that 3 endmembers was an appropriate number to represent the data set. ICE ended the algorithm with 11 endmembers.

The band weights found by the three runs of the B-SPICE algorithm were averaged and found to be [1.86, 3.14, 0, 0, 0] for bands 1 to 5, respectively. This shows that the band sparsity term was effective at correctly selecting the useful bands.

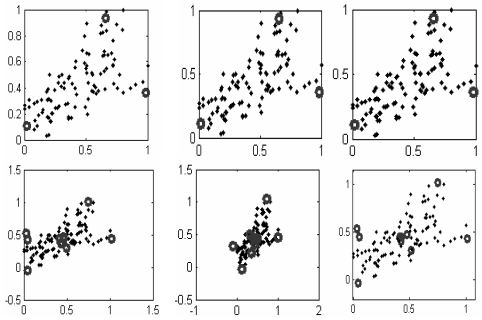

Figure 2: Comparison of B-SPICE (top) and ICE with pruning (bottom). Initial number of endmembers was 20. B-SPICE found 3 endmembers. ICE found 11 endmembers.

B. Indian Pines Results

B-SPICE was also run on the June 1992 AVIRIS data set collected over the Indian Pines Test site. The image has 145 x 145 pixels with 220 spectral bands and contains approximately 2/3 agricultural land and 1/3 forest and other elements. The soybean and corn crops in this scene are in early growth stages. Therefore, these regions are primarily bare soil and residues from previous crops [8].

As in [1], B-SPICE was run on a subset of pixels from the image using "candidate points" selected using the Pixel Purity Index (PPI) [6]. The candidate points in our experiments were chosen from 20,000 random projections. Points within a distance of 3 from the boundary of the projection received increased purity indices. The 1077 pixels with the highest PPI were used as the candidate points. In [1] , 1000 pixels were used during the experiments on the real image sets. We chose a PPI threshold that allowed us to have as close to 1000 pixels as possible (many pixels have the same PPI).

Before running B-SPICE, the image pixels were normalized. SPICE was initialized to 20 endmembers, η was set to 100, β was set to 1, Λ was set to 0.0, and Γ was set to 0.5. The initial number of endmembers for all three runs was 20.

Furthermore, the band weights were updated every fifth iteration, T was set to 0.54 and the endmember pruning threshold was set to 1e-10. The 8 endmembers found for this dataset using B-SPICE are shown in Figure 3. The resulting abundance maps are shown in Figure 5. B-SPICE pruned unnecessary endmembers and provided interpretable results that compare to results found by others on this dataset [7,8,9].

In Figure 5, the images were found to correspond to the following: (a) corresponds to woods, grass and pasture, (b) and (d) correspond to corn and soybean, (c) is hay-windrowed, (e) is grass and background, (f) corresponds to steel towers and roads, (g) is wheat and background, and (h) is trees and grass.

The band weight results for the Indian Pines Data set is shown in Figure 4. When compared to the found endmembers, in Figure 3, it can be seen that many of the water absorption bands have been removed, as would be expected. The results reduced the number of bands from 220 to 126.

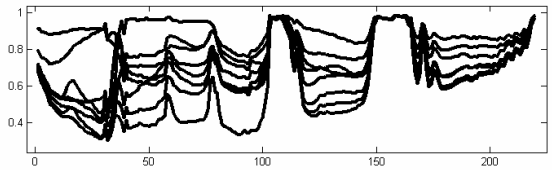

Figure 3: The 8 endmembers found using B-SPICE for the Indian Pines dataset.

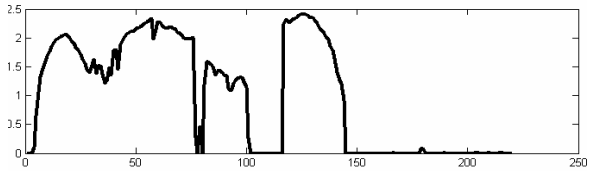

Figure 4: The band weights found using B-SPICE for the Indian Pines dataset.

IV. Conclusion

The B-SPICE algorithm extends the ICE algorithm with the addition of a sparsity promoting term. This term encourages the pruning of excess endmembers by penalizing the objective function when a large number of endmembers are being used. The sparsity promoting term drives the set of proportions associated unnecessary endmembers to zero at which point that endmember can be pruned from the set of endmembers representing the data. Similarly, the band sparsity term drive unnecessary band to zero.

Although results suggest that the B-SPICE algorithm removes the need to know the number of endmembers needed for a scene in advance, there are still a number of parameters that need to be set, e.g., the constant coefficients. It was observed that the coefficient for the RSS term needs to be several orders of magnitude larger than the other coefficients. Future work can include investigation of methods to automatically set these parameters.

Acknowledgment

Research was sponsored by the U. S. Army Research Office and U. S. Army Research Laboratory and was accomplished under Cooperative Agreement Number DAAD19-02-2-0012. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office, Army Research Laboratory, or the U. S. Government. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

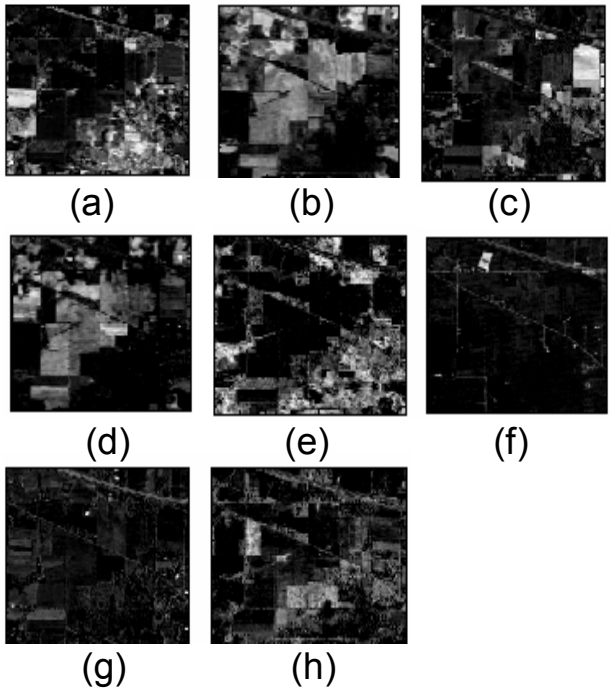

Figure 5: Abundance maps generated by SPICE on the Indian Pines data set.

References

[1] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Donne and J. F. Huntington, "ICE: A Statistical Approach to Identifying Endmembers in Hyperspectral Images," *IEEE Trans. On Geoscience and Remote Sensing*, vol. 42, Oct. 2004, pp. 2085–2095.

[2] A. Zare and P. Gader, "Sparsity Promoting Iterated Constrained Endmember Detection for Hyperspectral Imagery," *IEEE Geoscience and Remote Sensing Letters*, To Appear.

[3] P. Williams, "Bayesian Regularization and Pruning Using a Laplace Prior," *Neural Computation*, vol. 7, pp. 117-143, 1995.

[4] M. A. T. Figueiredo, "Adaptive Sparseness for Supervised Learning," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 25, Sept. 2003, pp. 1150–1159.

[5] "AVIRIS Free Standard Data Products," [Online]. September 2, 2004, [cited February 2007] Available: http://aviris.jpl.nasa.gov /aviris.freedata.html.

[6] J. Boardmann, F. Kruse, and R. Green, "Mapping target signatures via partial unmixing of AVIRIS data," *Summaries of the 5th Annu. JPL Airborne Geoscience Workshop*, vol. 1, AVIRIS Workshop, R. Green, Ed., Pasadena, CA, 1995, JPL Publ. 95-1, pp 23-26.

[7] L. Miao, H. Qi, and H. Szu, "Unsupervised Decomposition of Mixed Pixels Using the Maximum Entropy Principle," *Proceedings of the 18th International Conference On Pattern Recognition*, Vol. 1, 2006, pp 1067-1070.

[8] M. Grana, M.J. Gallego, "Associative Morphological Memories for Spectral Unmixing," *European Symposium on Artificial Neural Networks*, April 2003, pp 481-486.

[9] M. Grana, P. Sussner, G. Ritter, "Associative Morphological Memories for Endmember Detection in Spectral Unmixing," *The 12th IEEE International Conference on Fuzzy Systems*, Vol. 2, May 2003, pp. 1285-1290.