Robust Endmember Detection using L_1 Norm Factorization

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

I. Introduction

THE widely-used convex geometry model for hyperspectral imagery assumes that spectra in a hyperspectral image are the convex combinations of the endmembers in the scene [1, 2],

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

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 (abundance) 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,

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

Given this model, spectral unmixing and endmember detection are the tasks of determining the endmembers and the proportions for every data point in the scene. Several endmember detection and spectral unmixing algorithms have been developed in the literature. However, the majority of these methods do not provide an autonomous way to estimate the number of endmembers and, thus, require the number of endmembers in advance. These methods include those based on Non-negative Matrix Factorization [3, 4], based on Indepenent Components Analysis [5, 6], and others [7, 8]. The number of endmembers is often unknown in advance.

Methods to estimate the number of endmembers from a data set have been developed as well. These methods include Virtual Dimensionality (VD), Transformed Gerschogorin Disk (TGD), the Noise-Adjusted TGD, and the Partitioned Noise-Adjusted Principal Components Analysis (PNAPCA) methods [9–11]. The VD method estimates the number of endmembers using the eigenvalues of the covariance and correlation matrices of the hyperspectral data set. The number of endmembers is set to the number of eigenvalues from the covariance and correlation matrices that differ based on some computed threshold. Due to the variances used when computing the thresholds, the VD method can be sensitive to noise in the data. The PNAPCA method relies on the use of the Maximum Noise Fraction (MNF) algorithm [12]. MNF simultaneously diagonalizes the data covariance matrix and whitens the noise covariance matrix for a data set. This requires an estimate of the noise covariance matrix in advance. PNAPCA partitions the noise-adjusted covariance matrix found by MNF and compares eigenvalues from the partitions to estimate the number of endmembers. However, an accurate estimate of the noise covariance matrix can be difficult to provide. The TGD method for estimating the number of endmembers uses the Gerschogorin disk theorem. This theorem provides a method of estimating the locations of eigenvalues for a matrix [13]. The TGD method defines a transformation that aims to cause the Gerschgorin disks associated with noise to have small radii and be located far from the disks associated with signal in the data. The Noise-Adjusted TGD applies this method to the noise-adjusted covariance matrix found using MNF. The TGD methods are also sensitive to noise.

The presented algorithm, the L_1-Endmembers algorithm, provides a noise and outlier robust method to simultaneously estimate endmember spectra, proportion values and the number of endmembers using an L_1 norm error term.

Both [14] and [15] use the L_1 norm and the Huber M-Estimator for matrix factorization. However, in hyperspectral endmember detection, the constraints on the abundances shown in Equation 2 are required to provide physically meaningful results. These constraints are not enforced in the matrix factorization methods presented in [14] and [15]. In addition, the L_1-Endmembers algorithm provides a sparsity-promoting term used to simultaneously estimate the number of endmembers needed for the image. This sparsity-promoting term is not used in the matrix factorization methods.

II. Endmember Detection using the Huber M-Estimator

The L_1-Endmembers algorithm estimates endmembers, proportions and the number of endmembers by iteratively optimizing an objective function containing three terms. The first term of the objective function computes the error between the input data points and their reconstruction using the estimated endmembers and proportion values.

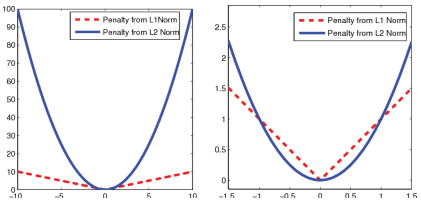
$$E_{L_1}(E, P) = \left\| X - EP^T \right\|_{L_1} \tag{3}$$

where $\|Z\|_{L_1} = \sum_i \sum_j |z_{ij}|$. This error term uses the L_1 norm when computing this cost function. In contrast, the L_2 norm is used in the SPICE and ICE algorithms [16, 17]. The use of the L_1 norm in this term provides resiliance against noise and outliers in the data. The robustness of the L_1 norm when compared to the L_2 is established [14, 18] and is illustrated in Figure 1(a). In the figure, the y-axis corresponds to the penalty for a given error using either the L_1 or L_2 norm. As you can see, the L_1 norm has a lower penalty value

A. Zare and P. Gader are with the Department of Computer & Information Science & Engineering, University of Florida, Gainesville, FL, 32611 USA e-mail: azare@cise.ufl.edu

Manuscript received December 11, 2009

for large errors providing a higher tolerance for outliers when compared to the L_2 norm. Figure 1 also illustrates the sparsity promoting properties of the L_1 norm over the L_2 norm. The L_2 norm prefers a series of many small errors instead of a small number of large errors.



(a) (b)

Fig. 1. (a) Plot of the penalty values for the L_1 and L_2 norms for the interval of [-10, 10] and (b) for the interval of [-1.5, 1.5].

The Huber M-Estimator cost function approximates the L_1 norm and is shown in Equation 4.

$$\rho(t) = \begin{cases} \frac{1}{2}t^2 & \text{if } |t| \leq \gamma \\ \gamma|t| - \frac{1}{2}\gamma^2 & \text{if } |t| > \gamma \end{cases} . \tag{4}$$

where γ is a positive constant parameter. As $\gamma \rightarrow 0^+$, the Huber M-estimator becomes the L_1 norm [19]. Figure 2 shows a comparison of the Huber M-Estimator for many γ values in comparison to the L_1 norm.

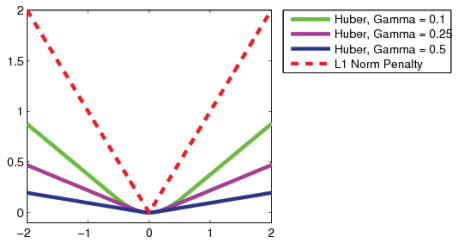


Fig. 2. Comparison of the Huber M-Estimator with Gamma Values of 0.5, 0.25 and 0.1 to the L1-Norm.

Using the Huber M-estimator, the first term of our L1-Endmembers objective function becomes

$$E_H(\mathbf{E}, \mathbf{P}) = \sum_{i=1}^{N} \rho\left(\mathbf{x}_i - \mathbf{p}_i\mathbf{E}\right) \tag{5}$$

Each term in this cost function, as shown in [14] and [15], can be rewritten as the quadratic programming problem shown in Equation 6 when solving for the proportion values.

$$\begin{aligned} \min_{\mathbf{p}\in R^M, \mathbf{z}\in R^D, \mathbf{t}\in R^D} & \frac{1}{2}\|\mathbf{z}\|_{L_2} + \gamma \mathbf{1}^T\mathbf{t} \\ s.t. & -\mathbf{t} \leq \mathbf{E}\mathbf{p}_i - \mathbf{x}_i - \mathbf{z} \leq \mathbf{t} \end{aligned} \tag{6}$$

where $\mathbf{x}_j$ is the i^{th} data point, $\mathbf{p}_i$ is the vector of proportion values for the i^{th} data point, D is the data dimensionality, M is the number of endmembers, and $\mathbf{1}$ is a vector of ones. A similar form is used when solving for endmembers.

The second term of our objective function promotes endmembers that provide a tight fit around the data.

$$E_V(\mathbf{E}) = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\left(\mathbf{e}_i - \mathbf{e}_j\right)^T\left(\mathbf{e}_i - \mathbf{e}_j\right) \tag{7}$$

where M is the number of endmembers.

The final term of our objective function uses the sparsity promoting properties of the Laplacian distribution to determine the number of endmembers. This term determines the number of endmembers by driving the proportions of unneeded endmembers to zero. Endmembers whose proportion values are driven to zero can be removed without any effect on the value of the cost function in Equation 5.

$$E_S(\mathbf{A}) = \sum_{k=1}^{M}\lambda_k\sum_{i=1}^{N}|p_{ik}| = \sum_{k=1}^{M}\lambda_k\sum_{i=1}^{N}p_{ik} \tag{8}$$

where $\lambda_k = \frac{\Lambda}{\sum_{i=1}^{N} p_{ik}}$ using the proportion values from the previous iteration. The second equality in Equation 8 comes from the non-negativity constraints on the proportion values in Equation 2. Both Equations 7 and 8 are also used in the SPICE objective function [16].

The complete objective function is $E = \alpha E_H + \beta E_V + E_S$ where α and β are coefficients determining the weight of the corresponding term. The Λ value used in E_S of the objective function controls the degree of sparsity used to determine the number of endmembers. In other words, Λ controls the scale value in the Laplacian prior on the abundance values. A larger value causes the Laplacian to have a sharper peak at zero.

The L_1-Endmembers algorithm iteratively updates the proportions, endmembers, and number of endmembers. During an proportion update, the quadratic programming problem is shown in Equation 9.

$$\begin{aligned} \min_{\mathbf{p}\in R^M, \mathbf{z}\in R^D, \mathbf{t}\in R^D} & \frac{1}{2}\|\mathbf{z}\|_{L_2} + \gamma \mathbf{1}^T\mathbf{t} + E_S \\ s.t. & -\mathbf{t} \leq \mathbf{E}\mathbf{p}_i - \mathbf{x}_i - \mathbf{z} \leq \mathbf{t} \\ & p_{ik} \geq 0 \quad \forall i, k \\ & \sum_{k=1}^{M} p_{ik} = 1. \end{aligned} \tag{9}$$

Similarly, the quadratic programming problem for solving for the endmembers during an endmember update is shown in Equation 10.

$$\begin{aligned} \min_{\mathbf{e}\in R^M, \mathbf{z}\in R^N, \mathbf{t}\in R^N} & \frac{1}{2}\|\mathbf{z}\|_{L_2} + \gamma \mathbf{1}^T\mathbf{t} + \beta E_V \\ s.t. & -\mathbf{t} \leq \mathbf{P}\mathbf{e}^j - \mathbf{x}^j - \mathbf{z} \leq \mathbf{t} \end{aligned} \tag{10}$$

where $\mathbf{x}^j$ is the vector of values from the j^{th} band across all data points, $\mathbf{e}^j$ is the vector containing the value from the j^{th} band over all endmembers, N is the number of data points, M is the number of endmembers, and $\mathbf{1}$ is a vector of ones.

After an iteration of updating endmembers and proportions, the proportions associated with each endmember can be examined. If the maximum proportion value falls below some threshold, the endmember can be pruned. During experimentation, the number of endmembers was found to be insensitive to this threshold value.

III. Experimental Results

L_1-Endmembers was run on both simulated and real hyperspectral imagery. These experiments examine the stability

accuracy of the algorithm. Comparisons to the SPICE algorithm that estimates endmembers, proportions and the number of endmembers using an L_2 error term is conducted. Also, the number of endmembers found by L_1-Endmembers is compared to the number found by the VD method.

A. Two-Dimensional Data with Increasing Noise Levels

In order to test the stability and accuracy of the algorithm, L_1-Endmembers was run on two-dimensional data with increasing levels of noise. The advantage of using this data set is that the true endmember values and number of endmembers are known. The data set was generated from three endmembers whose values are [0,0], [0,3] and [1,2]. Both L_1-Endmembers and SPICE were run 50 times for each noise level with random initializations. Zero-mean Gaussian random noise was added to each data point. The level of noise was increased by increasing the variance on the noise Gaussian. The parameters used for each run of the L_1-Endmembers algorithm were $\alpha = 1$, $\beta = 0.1$, $\Lambda = 0.5$, initial number of endmembers was 20, and the endmember pruning threshold was set to 1×10^{-9}. The parameters used for each run of the SPICE algorithm were $\mu = 5 \times 10^{-4}$, $\gamma = 0.5$, endmember prune threshold = 1×10^{-9} and the initial number of endmembers was set to 20 where the notation for these parameters comes from [16]. For both algorithms, the parameters were adjusted until an accurate result was found on the data without any noise; then, they were held constant for all remaining runs.

The number of runs with the correct number of endmembers found is shown in Table I.

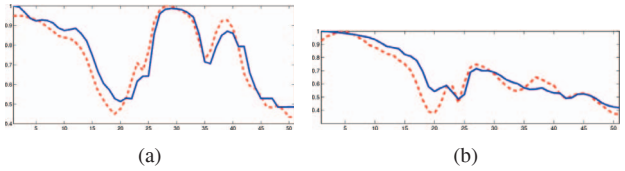
TABLE I
NUMBER OF RESULTS WITH THE CORRECT NUMBER OF ENDMEMBERS FOUND (OUT OF A TOTAL OF 50 RUNS)

Experiment	L_1-Endmembers	SPICE
2-D, No Noise	50	49
2-D, 0.1 Noise	50	43
2-D, 0.25 Noise	47	21

B. AVIRIS Cuprite Data Experiment

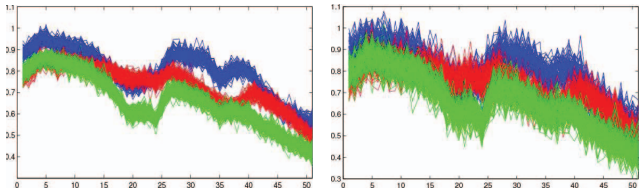
The L_1-Endmembers algorithm was also run on "Scene 4" of the AVIRIS Cuprite data set [20]. This data contains 51 contiguous spectral bands in the range of 1978 to 2477 nm. Like [16], L_1-Endmembers was run on a subset of pixels from the image using candidate points selected with the pixel purity index (PPI) [21]. PPI was run for 10,000 random projections and points within a distance of two from the boundary of the projection received increased purity indices. The 1,011 pixels with the highest PPI were used as the candidate points. Several of the points had a tie in PPI value. The parameters used for the L_1-Endmembers algorithm on the Cuprite data were the same as those used for the two-dimensional data. Several of these endmembers match the spectral shapes of materials known to be in the Cuprite Scene. A direct comparison of some of these materials with spectra from the USGS spectral library are shown in Figure 3 [22].

Since the true number of endmembers for the full Cuprite data scene is not known, three hand-selected endmembers were chosen from the scene that correspond to alunite, kaolinite and calcite spectra. Two data sets were generated from these endmembers with differing levels of noise, these data sets are shown in Figure 4. L_1-Endmembers and the Virtual Dimensionality algorithms were applied to each data set. When running the VD algorithm, the NSP threshold method was used [9]. This thresholding method was selected because it produced the most accurate results on this data set when compared to the other VD thresholding methods. On the data set with a lower noise levels, both L_1-Endmembers and VD estimated the correct number of endmembers at three. However, with the increased noise level in Figure 4(b), the VD algorithm incorrectly estimated two endmembers when the L_1-Endmembers was able to maintain the correct estimate of three.



(a) (b)

Fig. 3. (a) Comparison of endmember found using L_1-Endmembers and an Alunite spectra from the USGS spectral library. (b) Comparison of endmembers found and a Kaolinite spectra from the USGS spectral library. Dashed lines are the endmembers found by L_1-Endmembers. Solid lines show spectra from the USGS spectral library.



(a) (b)

Fig. 4. (a) Data set generated from three AVIRIS Cuprite endmembers using zero-mean Gaussian noise with a variance of 0.025. Endmembers correspond to alunite, kaolinite and calcite. (b) using Zero-mean Gaussian noise with a variance of 0.05.

IV. CONCLUSIONS AND FUTURE WORK

The results from L_1-Endmembers display that the algorithm was extremely stable in the number of endmembers when compared to the SPICE algorithm and the Virtual Dimensionality methods for estimating the number of endmembers. Furthermore, the results shown for this algorithm were generated with the same parameter set for all of the data sets, from two-dimensional data to 51-dimensional real hyperspectral data. This indicates L_1-Endmembers lack of sensitivity to parameter value settings which is very desirable.

The L_1-Endmembers algorithm requires several quadratic programming steps per iteration: one for each band during the endmember update and one for each data point during the abundance vector updates. These quadratic programming steps can be used directly in large scale quadratic programming software packages such as CPLEX and, therefore, take advantage

of any running time reductions the software packages provide. However, investigations will be conducted into whether the specific form of this algorithm, particularly with respect to the constraints on the abundance values, can be used to reduce the running time.

REFERENCES

[1] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 79–116, 2003.

[2] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, pp. 44–57, 2002.

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

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

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

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

[7] A. Plaza, P. Martinez, R. Perez, and J. Plazas, "Spatial/spectral endmember extraction by multidimensional morphological operators," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2025–2041, Sep. 2002.

[8] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[9] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608–619, Mar. 2004.

[10] T.-M. Tu, P. S. Huang, and P.-Y. Chen, "Blind separation of spectral signatures in hyperspectral imagery," in *Proceedings of the IEEE: Vision, Image and Signal Processing*, vol. 148, no. 4, Aug. 2001, pp. 217–226.

[11] H.-T. Wu, J.-F. Yang, and F.-K. Chen, "Source number estimators using transformed Gerschgorin radii," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, pp. 1325–1333, Jun. 1995.

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

[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[14] Q. Ke and T. Kanade, "Robust l1 norm factorization in the prsence of outliers and missing data by alternative convex programming," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, Jun. 2005, p. 739746.

[15] O. L. Mangasarian and D. R. Musicant, "Robust linear and support vector regression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 9, pp. 950–955, 2003.

[16] 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.

[17] 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 Transactions on Geoscience and Remote Sensing*, vol. 42, pp. 2085–2095, Oct. 2004.

[18] N. Kwak, "Principal component analysis based on l1-norm maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 9, pp. 1672–1680, Sept. 2008.

[19] W. Li and J. Swetits, "The linear l1 estimator and the huber m-estimator," *SIAM J. Optimization*, vol. 8, pp. 45–475, 1998.

[20] (2004, Sep) AVIRIS free standard data products. Jet Propulsion Laboratory, California Institute of Technology, Pasedena, CA. URL http://aviris.jpl.nasa.gov/html/aviris.freedata.html.

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

[22] R. N. Clark, A. Swayze, R. Wise, K. E. Livo, T. M. Heofen, R. F. Kokaly, and S. J. Sutley. (2004, Jan.) USGS digital spectral library (splib05a). Spectroscopy Lab, U.S. Geological Surey. [Online]. Available: http://speclab.cr.usgs.gov/spectral-lib.html