Automated optimal channel selection for spectral imaging sensors

John Gruninger[*a], Robert Sundberg [a], Marsha Fox [a] and Robert Levine [a], William Mundkowsky [b], Michael S. Salisbury [c] and Alan H. Ratcliff [c]
Spectral Sciences, Inc. [a], Surface Optics Corp. [b], Air Force Research Laboratory/SNJ [c]

ABSTRACT

A method of optimizing the selection of spectral channels in a spectral/spatial remote sensor has been developed that is applicable to the design of multispectral, hyperspectral and ultraspectral resolution sensors. The approach is based on an end member analysis technique that has been refined to select the most information dense channels. The algorithm operates sequentially and at any step in the sequence, the channel selected is the most independent from all previously selected channels. After the channel selection process, highly correlated channels, which are contiguous to those selected, can be merged to form bands. This process increases the signal to noise for the new broader spectral bands. The resulting bands, potentially of unequal width and spacing, collect the most uncorrelated spectral information present in the data. The band selection provides a physical interpretation of the data and has applications in spectral feature selection and data compression.

Keywords: End-member, Hyperspectral, Multispectral, Ultraspectral, Remote Sensing, Spectral Imaging

1. INTRODUCTION

Hyperspectral data cubes, or scenes, contain a spectrum for each pixel which can consist of hundreds of channels. The pixel spectra are highly correlated and there is high correlation among the measured channels. It is desirable to compress this information into useful bands by eliminating or concatenating redundant channels. Commonly applied techniques involve eigenvector analysis of the scene covariance matrix or autocorrelation matrix and singular value decomposition. These have desirable optimum properties for data compression, but by blending linear combinations of all channels into the eigenvectors, a physical interpretation is more difficult and their utility for sensor design studies is more limited. Techniques which use matrix factorization as a step to find a reduced set of physical bands include an iterative method developed by Price[1,2] and Zhao-Liang[3]. More recently, a method was developed[4] in which QR factorization (Q is a matrix of orthonormal columns and R is upper triangular matrix) coupled with analysis of principal components is used to find independent bands. The approach in our work also leads to a factorization, but not an orthogonal factorization. Our approach most closely follows that of Bowles[5] who use a modified Gram Schmidt process with a pivoting strategy to generate end-members.

Hyperspectral scenes typically contain far fewer materials than either channels or pixels. Measurements from aircraft of satellites may result in spatial resolutions that lead to several pixels with more than one of the unique materials in view. Mixture models which use physical sets of component spectra, end-members, offer both interpretation and data reduction, although they are not as efficient as eigenvector techniques for compression. It is desirable to find both end-members and a reduced set of bands. In these physical models the end-member spectra and their abundances are positive, the end-members form the edges of a convex cone. Most methods that derive end-members from the scene utilize the mathematics of convex sets[6,7,8]. Keshava and others have recently reviewed several techniques[9]. The end-members extracted from the scene may be pure materials or if the "pure" material does not exist in the scene the end-members are the "most pure", the pixels with the highest abundance of a material. The number of materials present in the scene is typically assumed to be equal to the rank of the matrix formed from the hypercubes with pixel spectra forming columns and channel vectors forming rows. The maximum number of linearly independent bands is equal to the rank as well. This interpretation is complicated by variations in material spectra within a scene induced by the variations in illumination of the materials or in environmental interactions. The small variations in material spectra together with the high correlations among spectra of different materials and sensor noise combine to make rank determination challenging. There are usually several plausible models over a range of assumed rank, with the transition from variations in material spectra to noise hard to distinguish. There is a close connection between the "best conditioned" basis and the convex cone which contains the data matrix[10]. The condition of a basis can be measured

[*]john@spectral.com; 781 273-4770; fax 781 270-1161; http//www.spectral.com; Spectral Sciences, Inc., 99 South Bedford Street, Suite 7, Burlington, MA, USA 01803-5169

through the magnitude of the determinant of its metric matrix, the Gramian. The "best conditioned" basis is the basis with the largest Gramian. The Gramian is equal to the volume enclosed by the basis. This basis points to extreme points of the data. The basis that is extracted from the scene encloses the maximum volume and is a convex cone that contains the data.

Another criteria for conditioning is the condition number (largest eigenvalue of the metric matrix divided by the smallest). The basis with minimum condition number is similar to the basis with maximum Gramian. In finding the convex hull of the data of a given dimension, one searches an n-dimensional scatter plot for extreme points. One either finds a simplex of minimum volume that circumscribes the data[6], or a simplex of maximum volume that can be inscribed within the data[8]. Ifarraguerri and Chang[11] have an alternate approach that finds an optimum from outside the data points. Their approach like Gruninger[10] is capable of finding pure material spectra even when the material only appears within mixed pixels in the data. Minimizing or maximizing simplex volumes or Gramian determinants leads to the combinatorial problem of finding an optimum set of n-spectra or channels from a much larger set. Below we discuss a technique for finding end-members that is less computational than the techniques discussed above.

2. METHOD

Given the difficulty of determining the rank, we seek an approach that avoids rank determination as a preprocessing step. We also avoid the combinatorial problem. The approach taken to end-member selection and band selection is sequential and finds the most variable end-members or channels in order. Our approach is patterned on matrix algebraic techniques of finding a row or column basis for the matrix. The bases generated however, are convex cones.

A convex cone of a set of vectors $\{\mathbf{v}_k\}$ is the set of all vectors $\{\mathbf{V_p}\}$ where $V_p = \sum_k v_k a_{k,p}$ and the coefficients, $a_{k,p} \geq 0$. A convex hull has the added condition that the coefficients sum to one, $\sum_k a_{k,p} = 1$.

As the dimension of the number of bases increases, the errors in approximating the matrix by the larger convex cone decreases monotonically. We find a sequence of convex cones, which form approximating bases of rank k, with k ranging from 1 to n. This avoids the combinatorial problem of looking for the best pair or trio etc. It is sub-optimal in that the search is only for the best spectra or channel to add to the existing ones. The method allows the rank to be estimated after the sequence has been found. Additional criteria can be used to determine the number of end-members based on the spectral features found in the end-members and other properties of the mixture model. If the vectors of the convex cone of dimension n are "pure" materials and the scene contains only mixtures of these materials, then the coefficients should sum to one and the abundance matrix will be sparse.

We start with a hyperspectral data array where each column vector is a pixel spectrum and each row vector contains the radiances or apparent reflectances of the all of the pixels in the channel. We seek a column or row basis depending on whether end-members are sought or channel selection is sought respectively. Since scenes typically have many more pixels than channels, if a channel basis is sought, it is expedient to work with the autocorrelation matrix or a set of its eigenvectors rather than the full data matrix. Alternatively, the number of pixels or channels can be reduced via a quick clustering.

An additional preprocessing choice concerns normalization. Normalization of pixel spectra will suppress variations in brightness induced by illumination variations. Normalization may be appropriate for end-member selection, particularly when working with radiance data rather than with pseudo-reflectance data. For the selection of spectral channels that have significant signal the vectors are not normalized. This avoids the use of noise dominated channels.

Following preprocessing an extreme point is taken as the first vector. This extreme point can be selected automatically as the brightest, the darkest, the least like the average or some other vector selected by the user. This vector becomes an end-member or selected channel. For channel selection the brightest channel is usually selected, insuring maximum signal. This is the channel with the longest channel vector.

After a vector is selected, all remaining vectors have this vector removed from it by convex projection. The approach is patterned after a sequential approach to Gram-Schmidt orthogonalization, known as the Modified Gram-Schmidt method (MGS). MGS is preferred over the standard GS because of its numerically stability[12]. A convex projection rather than an orthogonal projection is used to remove a selected basis vector. The traditional orthogonal projection, O_V, which removes a selected vector, V, from one of the remaining vectors , v_k, is

$$O_V v_k = \left[1 - \frac{VV^T}{V^T V} \right] v_k = v_k - V c_k \ , \tag{1}$$

where c_k is the projection coefficient which can be positive or negative. The convex projection, P_V, yields $P_V v_k = v_k - V c_k$ for positive projection coefficient, and $P_V v_k = v_k$ for a negative projection coefficient.

Whenever constraints are active in the convex projection technique non-orthogonality is introduced, however our goal is to obtain a sequence of convex cones based on the selection of edge vectors and not an orthogonal basis. If convex projection constraints are active, then the vector(s) which are not orthogonalized to the newest selected end-member or channel are outside of the current cone in the sequence. The convex projection does not shorten these vectors, making them prime candidates to be selected in subsequent steps.

Our selection strategy is to select as the next member of the basis the vector that has the largest component outside the convex cone of the current members. The length of the component of v_k that is outside of the convex cone is given by $\left| v_k^0 \right| \sin(\theta)$ where $\left| v_k^0 \right|$ is the vectors original length and θ is the angle that the vector makes with the convex cone of the selected members. If the vectors are initially normalized our selection strategy selects the vector which make the maximum angle with the convex cone of the selected members. When the vectors are not normalized, then there is a length scaling. The length scaling weights signal to noise in the channel selection process. More general weighting of the channels can be directly implemented by pre-multiplying the data array with a weighting matrix. The weighting matrix could be a simple diagonal matrix leading to a scaling or the inverse of a covariance or scatter matrix from discriminant analysis leading to a selection of independent bands with optimal discriminating power. Our selection strategy has an exact analog in an orthogonalization scheme known as Augmented Modified Gram-Schmidt[13]. See Bowles[5] for the application of this orthogonalization technique to end-member determination. In the orthogonalization techniques, the vector selected has the largest component orthogonal to the subspace of the already chosen vectors. When selected, this vector is removed from the remaining vectors by orthogonal projection. Our technique and the orthogonalization technique will select the same vectors until a constraint becomes active at which time the orthogonal projection shortens the vector component which is outside of the cone while the convex projection does not. Thereafter, the unselected vectors have different lengths in the two strategies, and different vectors are chosen.

The result of our process is a sequence of convex cones that approximate the database with increasing accuracy. The magnitude of the components of vectors lying outside of the convex cone, the length of the residual vectors, decrease monotonically and asymptotically approach zero. The rank or dimension of the convex cone can be selected based on a plot of the root mean square residual as a function of dimension using the same criteria for determining rank from SVD. For example, if there is a break in the curve of the root mean square error versus the number of vectors, this could be taken as an indication of the appropriate dimension. For this technique the root mean square error curve will always lie above that of a SVD, since the SVD is the most efficient expansion.

Since a sequence of cones are available, one can use other criteria to estimate the rank. Most of the spectral differences among the end-members or channels will be contained in the first few vectors chosen, and those chosen later will have small variations from those already chosen. Direct inspection of the sequence of vectors will also aid in determining a useful working dimension to assume. For end-members, the abundance matrix should be sparse. Mixed pixels will typically have only a few materials with non-zero abundance. The investigation of abundance matrices for sparseness may help in the dimension determination. The residual scene should be relatively structure free spatially and should be primarily instrument noise. The convex cone coefficients for 'true' end-member reflectances should sum to unity, if atmospheric correction has removed all illumination effects. However, for materials in shadow, spectral shape variations are not currently removed by atmospheric compensation algorithms, although such an effort is underway[14].

Once channel selection has been performed, a broadening into bands can be performed. Our approach is to look at channels that are adjacent to the selected channels and merge those that are highly correlated together into a broad band. We have thresholded on cross correlation coefficient and used convex cone analysis to determine correlation. In the cone analysis, a channel will be highly correlated to one of the chosen channels if all but one of its expansion coefficients are close to zero. The one non-zero expansion coefficient for the highly correlated channel is large (since we do not normalize, the coefficients do not sum to one). The latter approach is more effective for the high correlation that is found among all the channels.

3. APPLICATION

We applied the technique to an AVIRIS Stennis reflectance cubes, see Figure 1. The atmospheric compensation was performed using the SSI Atmospheric Compensation Code (ACC). The algorithm was applied to the image to find end-members. In these examples, we chose to not normalize the pixel vectors and to select the brightest pixel as the starting vector. Locations in the scene where the first twenty end-members were selected are indicated with arrows in Figure 1. The adjacent numbers indicate the order in which the end-members were selected. The relative errors in approximating the scene pixels monotonically decrease as the dimensions of the cones formed by the algorithm increases. A plot of the maximum relative error as a function of dimension or number of end-members is shown in Figure 2. The relative error is the fraction of the length of a pixel that is outside of the cone. The maximum is the largest such error out of the entire data cube. The vector that has this maximum error would be the next vector selected to add to the existing cone. In this example, there is a rapid decrease in relative error with dimension up to six end-members, then a more gradual decrease up to twelve end-members the an extremely gradual decrease as the error asymptotically approach zero. This curve can be used to select the rank or dimension and an appropriate convex cone for an application. By selecting a cone of dimension twelve, for example, the end-member that would be added to make a 13 dimensional cone is the pixel with the largest relative error at dimension 12. All other pixels in the scene will have smaller residuals than this vector. A plot of this pixel and the convex cone representation of it is shown in Figure 3. It is interesting to note that this pixel spectrum is very similar to but less bright, (80% of the intensity), than the third end-member selected. The convex coefficients indicated that the 13th end-member modeled reflectance came from a 79% contribution of the third end-member and a 6% contribution from the sixth end-member. Also shown in Figure 3 are the spectra of the third end-member and the 20th end-member. The only significant contribution to the reflectance of spectrum came from end-member three with 56% contribution. All three of the pixels contain vegetation. The apparent reflectance varies by a factor of two. This variation may be caused by illumination differences between pixels of similar material, with the less bright pixels being more in shadow, or by differences in the plant materials themselves.

The algorithm was also applied to find channels. We chose to not normalize the channel vectors. The brightest channel vector was chosen as the first vector. The first eight channels selected are displayed in Figure 4 along with the first eight end-members. By not normalizing, the channels with significant signal are chosen and their locations are such that the eight materials can be identified. It is difficult to form broader bands based on standard cross correlation analysis since all of the bands are very highly correlated. To associate other channels to the selected channels we use convex analysis. This can be performed using a non-negative least squares algorithm[15]. The convex coefficients for each of the channels is shown in Figure 5. The technique leads to well localized bands of highly correlated contiguous channels. Band widths can be selected by placing a threshold on the magnitude of the convex coefficients. As with the end-members, the algorithm selected the brightest channels form the bands. By not normalizing the channels initially, the convex coefficients indicate the level of brightness in each channel. For channels which have more than one significant contribution for the selected channels, the coefficients indicate the relative strengths of the selected channels.

Figure 1. An AVIRIS Stennis apparent reflectance cube. The atmospheric compensation was performed using the SSI Atmospheric Compensation Code (ACC). Locations in the scene where the first 20 end-members were selected are indicated in white.

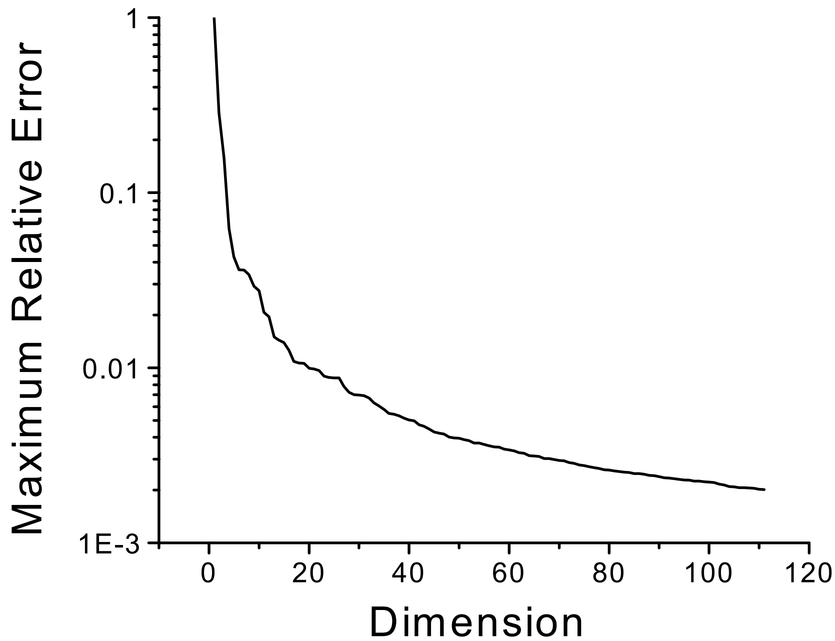

Figure 2. The maximum relative error as a function of dimension or number of end-members. The relative error is the fraction of the length of a pixel that is outside of the cone. The maximum is the largest such error out of the entire data cube.

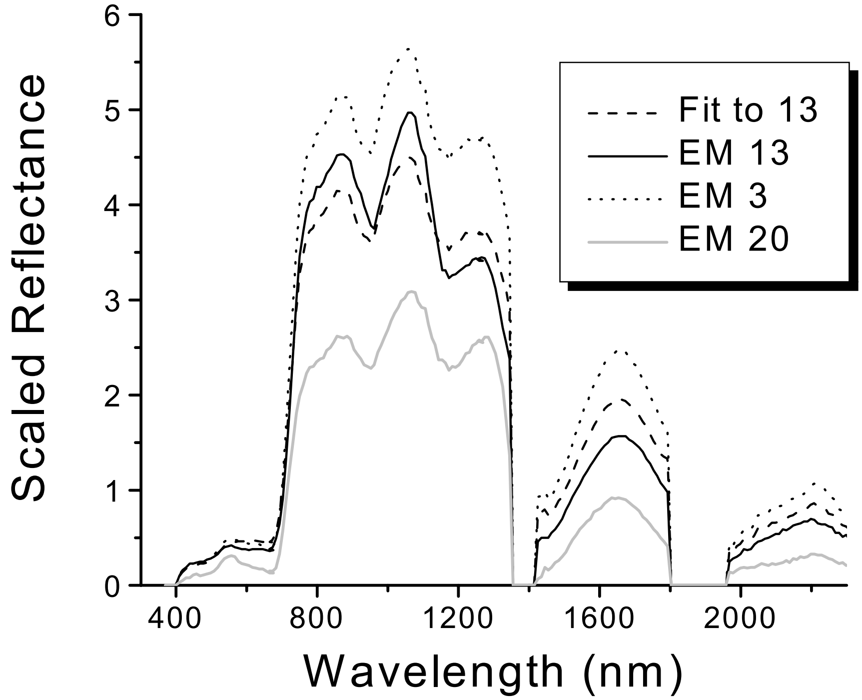

Figure 3. Convex cone representation to pixel spectrum for end-member 13 (dashed). Also shown are spectra for end-members 13, 3, and 20.

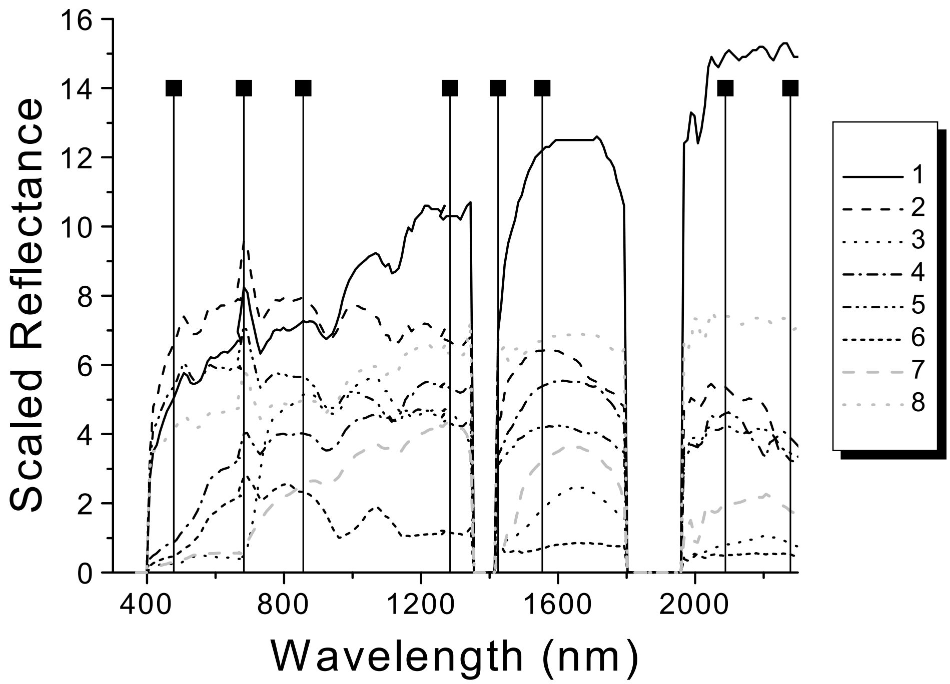

Figure 4. The first eight channels selected and the first eight end-members.

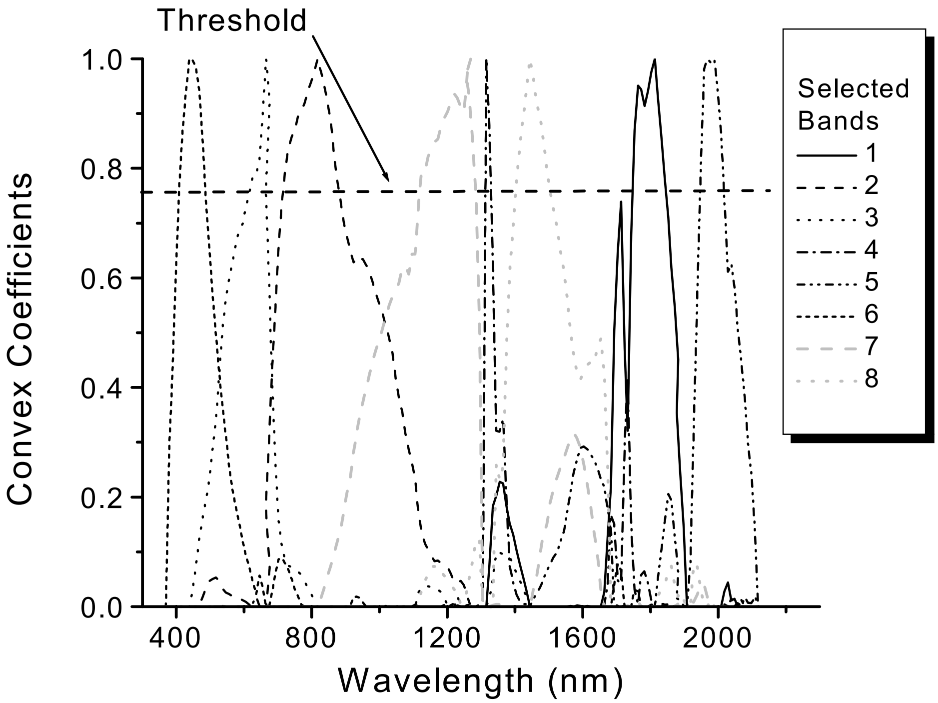

Figure 5. The convex coefficients for the AVIRIS channels for the eight selected channels. Band widths can be determined by setting a threshold on magnitude of convex coefficient.

4. CONCLUSIONS

The goal of this effort is to find a efficient method of reducing the complexity of hyperspectral data, while maintaining a physical representation of scene materials and important bands. Two challenging steps in such data reduction are the combinatorial problem of selecting a small set of n-pixels or channels from much larger sets to find a solution to an optimization problem and the determination of the appropriate size, n, of the small set. Our approach avoids the combinatorial problem by solving a sub-optimal problem of finding the best set sequentially. The appropriate size or dimension the convex set does not need to be determined prior to application of the method. The properties of the sequence of convex cones and convex analysis of the data can be utilized in the estimate of the appropriate size or dimension, n, to assume. The advantages of a sensor with 'designer' channel configurations are significant. By selecting only optimal channels for detection, the time required to collect a frame of data can be significantly reduced. Data transfer rates and processing overhead of satellite based systems could be substantially condensed if the number of channels can be reduced by half or more. Finally costs associated with production of high resolution sensors would be minimized.

ACKNOWLEDGEMENTS

This work was supported in part by AFRL/SN under Contracts F33615-00-C-1600 and F33615-99-C-1469. One of us (J.G.) also acknowledges generous support by Spectral Sciences, Inc. for development of the basic algorithms.

REFERENCES

1. J. C. Price, "Information Content of IRIS Spectra," *J. Geophys. Res.,* **80**, 1930-1936, 1975.
2. J. C. Price, "Band Selection Procedures for Multispectral Scanners", *Applied Optics*, **33**, 3281-3288, 1994.
3. L. Zhao-Liang, F. Becker, M. P. Stoll, Z. Wan and Y. Zhang, "Channel Selection for Soil Spectrum Reconstruction in 8-13µm Region." *J. Geophys. Res.,* **104**, 22271-22284, 1999.
4. M. Vélez-Reyes, L.O. Jiménez, D. M. Linares, and H. T. Velázquez, "Comparison of Matrix Factorization Algorithms for Band Selection in Hyperspectral Imagery." *SPIE*, **4049**, 288-297, 2000.
5. J. Bowles, M. Daniel, J. Grossmann, J. Antoniades, M. Baumback, and P. Palmadesso, SPIE Vol 3438,148-156, 1998.

6. M. D. Craig, "Minimum-Volume Transformations for Remotely Sensed Data," IEEE Trans. *Geoscience and Remote Sensing,* **32**, 99-109, 1994.
7. J. W. Boardman, "Analysis, Understanding and Visualization of Hyperspectral Data as Convex Sets in n-Space," *SPIE*, **2480**,14-22, 1995.
8. M. E. Winter, "N-FINDR: An Algorithm for Fast Autonomous Spectral End-Member Determination in Hyperspectral Data", *SPIE*, **3753**, 266-275, 1999.
9. N. Keshava, , J. Kerekes, D. Manolakis, and G. Shaw, "An Algorithm Taxonomy for Hyperspectral Unmixing," *SPIE,* **4049**,42, 2000.
10. J. Gruninger, and R. L. Sundberg., "Bounds on Component Spectra of Multispectral Images", *SPIE*, **3071**, 36-46, 1997.
11. A. Ifarraguerri, and C-I. Chang, "Multispectral and Hyperspectral Image Analysis with Convex Cones.", IEEE Trans. *Geoscience and Remote Sensing,* **37**, 756-770, 1999.
12. J. R. Rice, "Experiments on Gram-Schmidt Orthogonalization," *Math. Comp.,* **20**, 325-328, 1966.
13. G. Dahlquist, and Åke Björck , "Numerical Methods," translated by Ned Anderson, 201-204.
14. S. M. Adler-Golden, R. Y. Levine, M. W. Matthew, S. C. Richtsmeier, L. S. Bernstein, J. Gruninger, G. Felde, M. Hoke, G. Anderson and A. Ratkowski, "Shadow-Insensitive Material Detection/Classification with Atmospherically Corrected Hyperspectral Imagery," Summaries of the AVIRIS Earth Science and Applications Workshop, to be published, 2001.
15. C. L. Lawson, and R. J. Hanson., "Solving Least Squares Problems", Prentice Hall Englewood Cliffs New Jersey, 1974.