

Analyzing Hyperspectral Data with Independent Component Analysis

Jessica Bayliss

Dept. of Computer Science, Univ. of Rochester
Rochester, NY 14627

J. Anthony Gualtieri

Applied Information Systems, Code 935, NASA/GSFC
Greenbelt, MD 20771

Robert F. Cromp

Applied Information Systems, Code 935, NASA/GSFC
Greenbelt, MD 20771

ABSTRACT

Hyperspectral image sensors provide images with a large number of contiguous spectral channels per pixel and enable information about different materials within a pixel to be obtained. The problem of spectrally unmixing materials may be viewed as a specific case of the blind source separation problem where data consists of mixed signals (in this case minerals) and the goal is to determine the contribution of each mineral to the mix without prior knowledge of the minerals in the mix. The technique of Independent Component Analysis (ICA) assumes that the spectral components are close to statistically independent and provides an unsupervised method for blind source separation. We introduce contextual ICA in the context of hyperspectral data analysis and apply the method to mineral data from synthetically mixed minerals and real image signatures.

Keywords: hyperspectral, ICA, spectral unmixing, Cuprite

1. INTRODUCTION

Hyperspectral image sensors provide images with a large number of contiguous spectral channels per pixel and enable information about different materials within a pixel to be obtained. Past approaches such as the spectral mixture modeling of Adams and Smith¹ spectrally unmix different materials at a sub-pixel level through assuming basic knowledge of the main materials of interest (called endmembers) and that the spectrum is mainly composed of these materials. These assumptions may be inaccurate due to incomplete knowledge of a scene, leading to incorrect results.^{21,10}

A more recent approach by Maselli, Pieri, and Conese avoids this problem by using specially chosen pixels from the scene as endmembers.¹³ After a cluster analysis of the scene, maximum likelihood fuzzy classification is used to pick out the pixels which are most representative of the typical spectral features of the clusters and these pixels are chosen as endmembers. After choosing the endmembers in this fashion, spectral mixture analysis is performed to yield classifications into dark vegetation, very bright soil, dark soil, bright soil, bright vegetation, and water. Unfortunately, the pixels chosen as endmembers may consist of mixtures themselves. Thus, it remains very difficult to determine the real endmembers in a scene without prior knowledge of the scene.

In order to address this issue, we note that the determination of endmembers may be viewed as a specific case of the *blind source separation* problem.¹¹ In this paradigm, the data consists of linearly mixed signals (in this case minerals) and the goal is to determine which minerals are being mixed without prior knowledge of what minerals are in a scene. A relatively new unsupervised algorithm, Independent Component Analysis (ICA) approximates the solution to this problem by assuming independence of the spectral information for different materials. This strong assumption gives ICA the ability to use all of the statistics in the mixed signal rather than just first order (the mean)

How Contextual ICA works

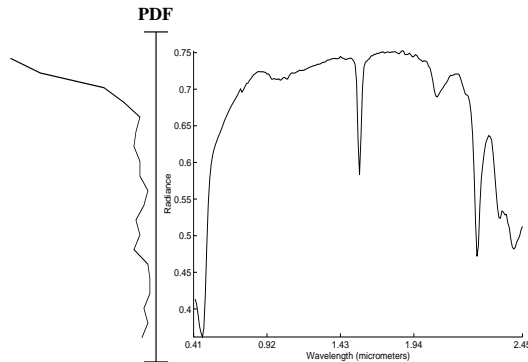


Figure 1. ICA is able to unmix signals by minimizing the difference between the true probability density function (pdf) of the mixed signals and the pdf of the model of the mixture as occurring from independent sources. In this way, signal separation becomes an optimization problem that can be handled by a neural network paradigm. Contextual ICA (cICA) uses previous wavelength information to condition the current sample in order to take advantage of the structure of the wavelength signal.

or second order (the covariance) statistics. Algorithms such as principal component analysis (PCA) use only second order statistics.

However it was found that plain ICA was unable to reliably perform the separation desired and this led us to use a generalization of ICA by Pearlmutter and Parra known as contextual ICA (cICA)^{17,18}. In the sequel we introduce this algorithm, and discuss the benefits and drawbacks to using this algorithm for spectrally unmixing minerals in an image scene. We have applied this approach to both synthetically mixed minerals and mixtures from a real image scene. We show that in both cases the minerals are correctly unmixed, although there is a degradation in performance from synthetically mixed minerals to minerals in a real scene.

2. INDEPENDENT COMPONENT ANALYSIS

As introduced by Comon,⁹ Independent Component Analysis (ICA) approximates the factor analysis of multiple linearly mixed source signals through assuming that the sources are statistically independent from one another. This relationship between the sources and mixed signals may be expressed as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (1)$$

where \mathbf{x} is a vector of n mixed sources, \mathbf{s} is a vector of independent sources, and \mathbf{A} is an $n \times n$ mixing matrix. Given this source/mixed signal relationship, the goal of blind source separation is to recover the vector of sources, \mathbf{s} , when given only the n mixed signals \mathbf{x} and the knowledge that these sources are independent from one another. Note that since the source signals are described by a statistical model and not by a deterministic model, we will not be able to recover either the magnitudes of the sources, nor the the order in which the signals were mixed. Thus we are not expecting that the $\mathbf{W} = \mathbf{A}^{-1}$, but only that

$$\mathbf{W}\mathbf{A} = \mathbf{P}\mathbf{C} \quad (2)$$

where \mathbf{P} is an arbitrary permutation matrix (all zeros except, for a 1 in every row and column), and where \mathbf{C} is an arbitrary scaling matrix (only non-zero elements on the diagonal).

Following the approach of Pearlmutter and Parra,¹⁷ we would like to know the true probability distribution $P(\mathbf{x})$ from which the samples in our mixed signals have been drawn. Notice from Figure 1 that we are looking at the probability density function (pdf) of the minerals from the y axis rather than the x axis. This is because the minerals are *linearly* mixed in this direction. While we can not calculate this distribution, we can model it. Using a maximum-likelihood approach,¹⁴ the goal becomes to reduce the difference between our model of the probability

density function (pdf) and the actual pdf. In order to achieve this using ICA, we must find a vector of parameters \mathbf{w} that maximize the log-likelihood that a set of mixed signals \mathbf{x} could have arisen from a random process in which the sources are linearly mixed. As discussed in Olshausen,¹⁵ this is formally equivalent to minimizing the Kullback-Leibler distance (G) between the *actual* joint probability of the signals $P^*(\mathbf{x})$ and *our model* of the joint probability $P(\mathbf{x}; \mathbf{w})$,

$$\begin{aligned} G(P^*(\mathbf{x}), P(\mathbf{x}; \mathbf{w})) &= \int P^*(\mathbf{x}) \log \frac{P^*(\mathbf{x})}{P(\mathbf{x}; \mathbf{w})} d\mathbf{x} \\ &= H[P] - \int P^*(\mathbf{x}) \log P(\mathbf{x}; \mathbf{w}) d\mathbf{x}. \end{aligned} \quad (3)$$

This leaves us with the entropy of the *fixed* input distribution $P^*(\mathbf{x})$ minus the likelihood of $P^*(\mathbf{x})$ given $P(\mathbf{x}; \mathbf{w})$, and the \mathbf{w} which minimizes G maximizes the likelihood. As we do not have access to the actual Kullback-Leibler distance G , we may still obtain an unbiased estimate of it by taking a sample \mathbf{x} from $P^*(\mathbf{x})$,

$$G^*(P^*(\mathbf{x}), P(\mathbf{x}; \mathbf{w})) = H[P] - \log P(\mathbf{x}; \mathbf{w}). \quad (4)$$

In order to obtain a learning rule for the neural network architecture, we use stochastic gradient ascent,

$$\frac{dG^*}{d\mathbf{w}} = -\frac{d \log P(\mathbf{x}; \mathbf{w})}{d\mathbf{w}}. \quad (5)$$

Let \mathbf{W} be an $n \times n$ matrix, and let $\mathbf{x} = \mathbf{W}^{-1}\mathbf{s}$. \mathbf{s} is an n -dimensional vector of sources whose components s_j are drawn from n independent parameterized one-dimensional density functions $f_j(s_j; \mathbf{w}_j)$. The estimated density of \mathbf{x} is denoted $P(\mathbf{x}; \mathbf{w})$ where \mathbf{w} is a concatenation of the elements of \mathbf{W} with the parameters $\mathbf{w}_1, \dots, \mathbf{w}_n$ of the densities f_1, \dots, f_n . Expanding $\log P(\mathbf{x}; \mathbf{w})$ from above, we obtain

$$\log P(\mathbf{x}; \mathbf{w}) = \log |\mathbf{W}| + \sum_j \log f_j(s_j; \mathbf{w}_j). \quad (6)$$

This gives us the following formulas:

$$\frac{dG^*}{d\mathbf{W}} = -\mathbf{W}^{-T} - \left(\frac{f'_j(s_j; \mathbf{w}_j)}{f_j(s_j; \mathbf{w}_j)} \right)_j \mathbf{x}^T, \quad (7)$$

$$\frac{dG^*}{d\mathbf{w}} = -\frac{f'_j(s_j; \mathbf{w}_j)}{f_j(s_j; \mathbf{w}_j)}, \quad (8)$$

where $(\text{expr}(j))_j$ denotes the column vector whose elements are $\text{expr}(1), \dots, \text{expr}(n)$. In the Bell and Sejnowski algorithm (plain ICA),⁵ $f_j(s_j; \mathbf{w}_j)$ is taken to be the derivative of the logistic cumulative distribution function, $g(\cdot) = (1 + \exp(-\cdot))^{-1}$ and the parameter vector \mathbf{w}_j holds the j^{th} component of a vector of *bias terms* \mathbf{w}_0 , giving us

$$\frac{dG^*}{d\mathbf{W}} = -\mathbf{W}^{-T} + (\mathbf{1} - 2\mathbf{y})\mathbf{x}^T, \quad (9)$$

$$\frac{dG^*}{d\mathbf{w}} = \mathbf{1} - 2\mathbf{y}, \quad (10)$$

with $\mathbf{y} = g(\mathbf{W}\mathbf{x} + \mathbf{w}_0)$ and $\mathbf{1}$ is vector of 1's.

Unfortunately, this algorithm makes no use of the structure in the order of the wavelength information as it only makes use of cumulative histograms. If these histograms are Gaussian, then this algorithm will be unable to separate

the sources. This algorithm does work well on super-Gaussian signals (signals with positive kurtosis) and has been used for separating mixed speech signals.⁶ In order to obtain an idea of the kurtosis of minerals, the kurtosis of 498 minerals was calculated and the average was found to be 0.52, with quite a few minerals having a Gaussian or a negative kurtosis. This makes the signals impossible to unmix using regular ICA.

In order to avoid this problem, a generalization to ICA known as contextual ICA (cICA) is used^{17,18} Pearlmutter and Parra note that there is no restriction on the form of the distributions f_j in equations 7 and 8 and so derive a learning rule that is conditioned on nearby or other signal information as follows. In the case of learning mineral signatures, the *other information* used is the wavelength information at τ_{max} bands preceding the specific wavelength being sampled. The distributions that describe u_j are taken as a mixture of k logistic density functions. In particular (where t and τ are hyperspectral band indices),

$$f_j(u_j(t)|u_j(t-1), u_j(t-2), \dots, \text{other info}, \dots; \mathbf{w}_j) = \sum_k m_{jk} h((u_j(t) - \bar{u}_{jk})/\sigma_{jk})/\sigma_{jk}, \quad (11)$$

where σ_{jk} is a scale parameter for logistic density k of source j and is an element of \mathbf{w}_j , the mixing coefficients m_{jk} are elements of \mathbf{w}_j and are constrained by $\sum_k m_{jk} = 1$, and $h(t) = dg/dt$ is the logistic density function. The component means, \bar{u}_{jk} , are taken to be linear functions of the recent values of that source,

$$\bar{u}_{jk} = \sum_{\tau=1}^{\tau_{max}} a_{jk}(\tau) u_j(t - \tau) + b_{jk}, \quad (12)$$

where the coefficients $a_{jk}(\tau)$ and the bias b_{jk} are elements of \mathbf{w}_j . Now as part of the stochastic gradient descent on \mathbf{w}_j the quantities $m_{jk}, \sigma_{jk}, a_{jk}, b_{jk}$ must also be included in the gradient descent. The derivation of the learning rule for this form of cICA may be found in Pearlmutter and Parra's paper.¹⁷ We have taken $\tau_{max} = 12$ and $k = 19$, with initial values of $1/k, 1/k, 1$, and 0 for $\sigma_{jk}, m_{jk}, a_{jk}$, and b_{jk} , respectively. The learning rate was begun at 0.025 and then decreased by 0.89 after every 97 iterations. In matlab a typical run of 6600 updates on an alpha 2100 5/250 took 46 sec. The matlab code for the algorithm may be found on-line. *

In practice, the matrix inverse is very expensive to calculate and Amari, Cichocki, and Yang² have increased the computational efficiency of the algorithm by multiplying equation 7 by $W^T W$. Since this is a positive-definite matrix, it does not effect the stochastic gradient convergence criteria. This multiplication contains the benefits of eliminating the matrix inversion, making the algorithm scale-invariant to the true mixing matrix \mathbf{A} , and leads to the new learning rule

$$\frac{dG^*}{d\mathbf{W}} = \mathbf{W} + \left(\frac{f'_j(s_j; \mathbf{w}_j)}{f_j(s_j; \mathbf{w}_j)} \right)_j \mathbf{s}^T \mathbf{W}. \quad (13)$$

The learning rules tell where the formulas come from, but not necessarily the usefulness of the method or its drawbacks. The signal sources will most likely be complex and nonlinear. The reason that cICA may handle these sources is because they are mixed *linearly*. For problems such as speech separation (the cocktail party problem) this is a reasonable assumption, but there are many interesting problems that cICA can not handle because it only separates linear mixes. In these cases, nonlinear varieties of PCA that deal with higher order statistics as discussed in Karhunen et al¹² may be used more profitably.

One of the drawbacks of cICA is that unlike PCA, it can not be analytically calculated. Due to this limitation, the neural network algorithm described above may take a long time to converge and in some cases may diverge. For the extra time spent, the user of the cICA algorithm gains the ability to use the higher order statistics of the mixed signals and to achieve nonorthogonal basis functions from the network. PCA offers only orthogonal basis functions and uses only second order statistics from the signal.

Why emphasize nonorthogonality and higher order statistics? As discussed by Bell and Sejnowski,⁴ the higher order statistics in small image patches yield edges. In larger patches, the higher order statistics correspond to semi-local parts of an image or object. The transient attack or decay of a mineral signature over wavelength is another

*All matlab code and data used in this paper may be found on-line at <http://www.cs.rochester.edu/u/bayliss/spectral/spectral.html>

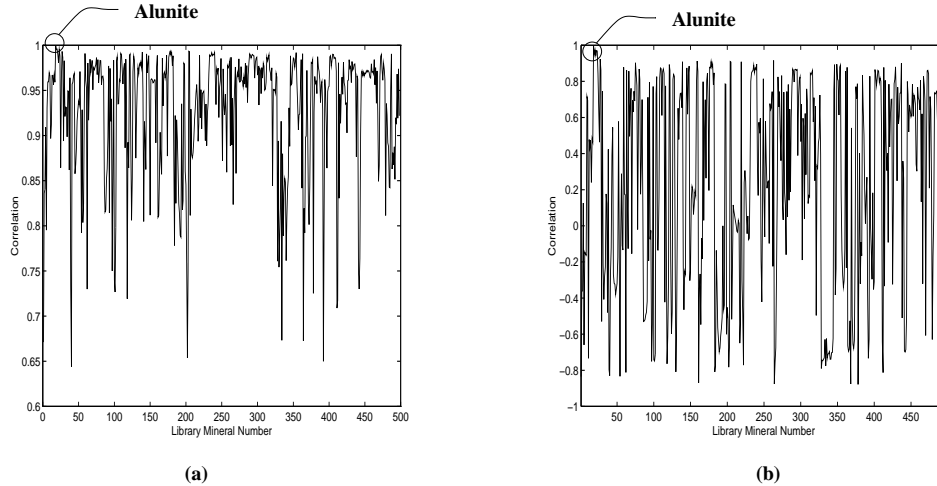


Figure 2. (a) The correlation of the 21st mineral in the spectral library (Alunite) with all the other minerals in the library. Note that most correlations are near 1. (b) The same correlation, but after the mean of each mineral has been removed. Whitening the data further reduces the correlation between minerals because removing the covariance removes statistical dependence.

example of higher order statistics. Nonorthogonal basis functions often characterize such information and are very important “features” of the signals of interest.

In order to avoid the extra time that cICA takes to converge on a solution, the mixed signal is often *whitened*. As first and second order statistics may dominate a signal, they are removed and the resulting *whitened* signal is used as input to the neural network. This amounts to changing the input vector \mathbf{x} to

$$\mathbf{x} \leftarrow 2(\text{cov}(\mathbf{x}))^{-\frac{1}{2}}(\mathbf{x} - \bar{\mathbf{x}}), \quad (14)$$

where $2(\text{cov}(\mathbf{x}))^{-\frac{1}{2}}$ is known as the whitening matrix.⁶

When whitening is performed, it is assumed that the higher order statistics give important information. This whitening also has the property of orthogonalizing the input data, making a network solution easier and faster to obtain. For on-line learning this adds the burden that all input signals must be transformed and all network outputs must be multiplied by the inverse of the whitening matrix before obtaining a final, possibly nonorthogonal answer.

3. APPROACH

Finding endmembers using cICA may be broken down into three different categories: preprocessing, cICA, and post-processing. Preprocessing the image data makes the unmixing problem “easier” to solve for the cICA algorithm. One way to accomplish this is to segment the data into similar regions before performing cICA on each section as similar regions are likely to have similar mixtures of endmembers. A small number of pixels from each region may then be chosen in order to find relevant endmembers in order to reduce the computational load of cICA.

The specific number of pixels chosen is dependent on a prior estimate of the number of endmembers in the area and thus introduces a bias in the approach. In practice, choosing a number of pixels greater than the actual number of endmembers is all right, but choosing a number of pixels less than the number of endmembers should be avoided as it yields incorrect results.

In order to use cICA for unmixing minerals, it is first necessary to make sure that the mineral data meets the basic assumptions used by cICA: the importance of higher order statistics, at least N pixels of data for N unmixed minerals, linear mixing, and independence. The transient attack or decay of a mineral signature over wavelength is an example of the higher order statistics in a mineral signature and appears important in recognizing the pattern of the waveform. As may be surmised above, there are many more pixels of spectral data than cICA may comfortably handle and so the second assumption is met. The third assumption may not hold, but is the central assumption for spectral unmixing. Unfortunately, as shown in Figure 2, the minerals are not even close to independent from each

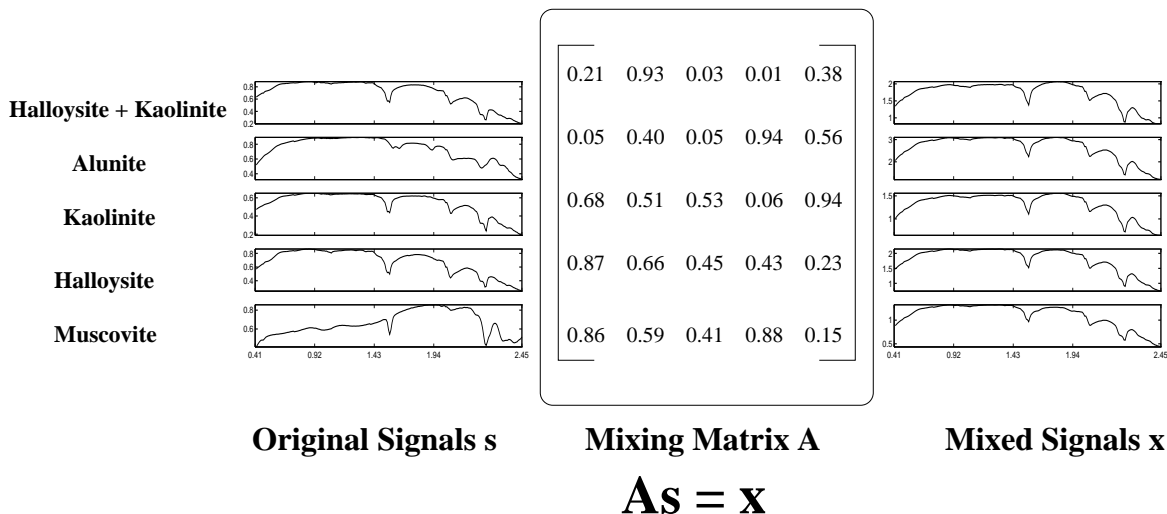


Figure 3. A synthetic mineral mixture is created by multiplying the original input signals s by a random mixing matrix A containing numbers between 0 and 1. Notice that one of the input minerals for this example is itself a mixture of Halloysite and Kaolinite. This means that there are four distinct minerals that need to be unmixed, but there are five channels of data. This was done in order to demonstrate what may happen when there are more channels of input than there are minerals to unmix, as may often be the case in real data where the precise number of endmembers is not known in advance.

other. While this sounds discouraging, there is a way to increase the independence for training the network: remove the lower order statistics. This may be accomplished by whitening the data and has the added benefit of lowering the time needed for convergence when running cICA.

This whitened data is then used as input to the cICA algorithm, which learns to unmix the minerals and was the subject of the last section. After running the unmixing algorithm, post-processing occurs in order to discover which minerals correspond to the unmixed output of cICA. In order to determine which materials are present in the region, the cICA output should be statistically correlated with a library of laboratory minerals. Note that the means of the minerals are too close to each other to give relevant information. Thus, it is necessary to subtract the means from all signals before correlating.

As discussed in the cICA section, this algorithm remains computationally expensive and so it is necessary to use a sparse set of representative pixel regions for training (rather than the whole image). However, after the basic materials in a section of an image have been determined, spectral mixture analysis (as proposed by Adams and Smith¹) may be used with each individual component as an endmember. This will yield classification of the entire image in a reasonable amount of time.

4. EXPERIMENTAL RESULTS

4.1. Synthetic Data Results

While a technique must be properly tested on real data to show its use in real applications, experiments using synthetic data may bring to light the full potential of an algorithm as well as its drawbacks under ideal conditions. The use of synthetic data also allows comparisons of different algorithms because ground truth is *created* in advance. The ideal performance may then be compared to results obtained from real images in order to highlight the differences between real and synthetic data and to show the differences in an algorithm's performance under varying conditions.

In order to make the transition from synthetic to actual data as smooth as possible, the minerals present in a specific location of the Cuprite, Nevada image were chosen from a spectral library for synthetic mixing. As the ground truth for the AVIRIS Cuprite data is from USGS images, the USGS AVIRIS mineral library was chosen.⁸ For the first experiment, five minerals from the chosen area of the Cuprite, Nevada image were synthetically mixed with a random mixing matrix. Because one of the minerals from the library was actually a mixture of the minerals in the Cuprite image (Halloysite + Kaolinite), this mineral was included in the synthetic mixture in order to see what

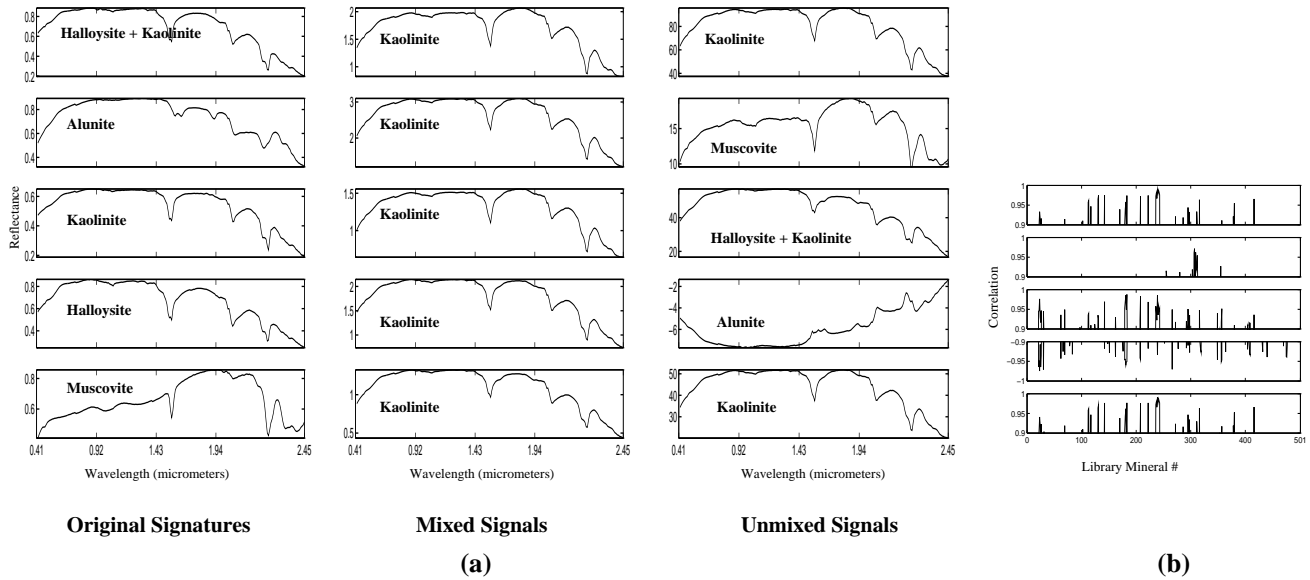


Figure 4. (a) Experiment 1 starts out with a synthetic mixture of library minerals (Halloysite CM13, Halloysite+Kaolinite CM29, Kaolinite CM3, Muscovite IL107, and Alunite GDS82 Na82). These minerals are then linearly mixed to yield five mixed channels. In order to show that the minerals need to be unmixed to yield the proper results, the highest correlation of each mixed channel to the library minerals is shown. The goal of cICA is to unmix these channels and the highest correlations with the unmixed signatures are (in order) Kaolinite KL502 (pxy1), Muscovite GDS114 Marshall, Halloysite+Kaolinite CM29, Alunite GDS82 Na82, and Kaolinite KL502 (pxy1). All of the original minerals are present in the unmixed results, but the Halloysite is only present in the Halloysite+Kaolinite mixture, while the Kaolinite is present on three of the resulting channels. Notice that some of the resulting channels have been permuted and are scaled as a result of applying the cICA algorithm for unmixing. (b) The correlations of the unmixed minerals in the first experiment with the 498 library minerals. These correlations demonstrate the need for care when interpreting results. Certain groups of minerals correlate very highly even after removing the mean of the signal and so more than one mineral may correlate above 0.9. The reason that negative correlations are shown for some of the channels is because these channels were permuted or “flipped” and thus correlate negatively with most minerals.

happened if there were only four minerals, but five channels worth of data as could very likely happen in situations where the real number of endmembers is not known in advance. This experimental setup is shown in Figure 3.

The results of unmixing using cICA are shown in Figure 4. All bands were used in order to unmix the mineral signatures. In order to determine the closest match between the library minerals and those obtained through unmixing, the unmixed results were correlated with the 498 minerals in the library after removing the means of all minerals and the highest correlation was chosen as the best match. The correlations are shown in Figure 4 .

4.2. AVIRIS Image Results

In order to test the method on real data, we obtained from R. N. Clark of the USGS,⁸ an AVIRIS image of reflectance data together with ground truth mineral classification maps for Cuprite, Nevada, taken in 1995. The ground truth is shown in Figure 5. Note that while the AVIRIS instrument measures radiance it must be converted to a surface reflectance by using physical models of atmospheric absorption and scattering, and knowledge of the incident solar radiation⁷ before it can be used in our method. The Cuprite area has been heavily studied.^{19,20} Since cICA claims to be able to unmix minerals, we wanted a section of the image where the ground truth contained a mixture of different minerals. We chose an 11×11 segment of the image which consisted of a mixture of Alunite, Kaolinite, and Muscovite.

Five pixels were chosen from different parts of the 11×11 image segment. Adjacent pixels were not chosen because in order to unmix the minerals, the mixtures must be different in order to make unmixing possible. This is not necessarily the case with adjacent pixels. Water absorption and other noisy bands (bands 107-113, 150-166, and 221-224) were removed from the data since they could affect results.

Cuprite, Nevada

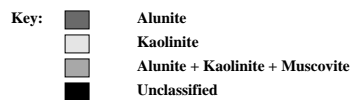
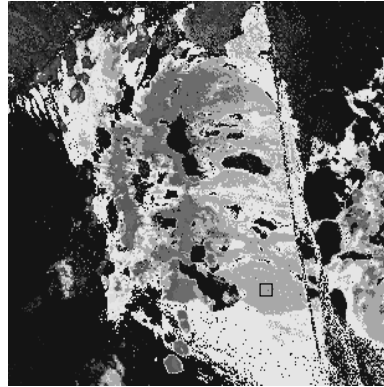


Figure 5. The ground truth map for the area of the Cuprite, Nevada site that was used. The black box in the picture shows the specific area where pixels were chosen to unmix the Alunite + Kaolinite + Muscovite mixtures. This particular mixture also may contain Halloysite and/or Dickite as these minerals are included in the Kaolinite classification.

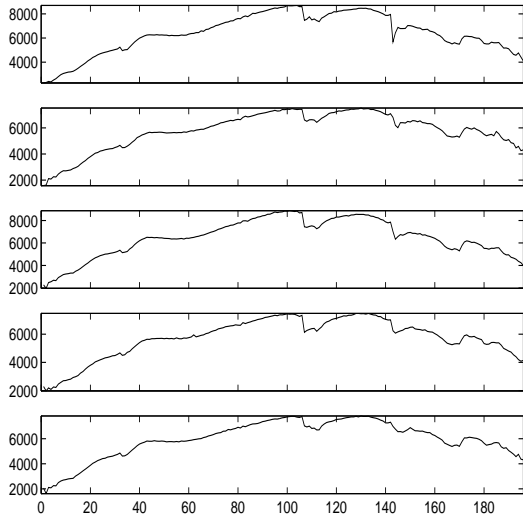
An image of the resulting 5 signals is shown in Figure 6. The first thing to notice about these mixtures is that the lower bands do not look like a mixture of Alunite, Muscovite, and Kaolinite. A synthetic mixture of the proper minerals is shown next to the real data in order to demonstrate this effect. It is not known why there is such a large difference between the lower bands of the library minerals and the real mixtures. It was noted that the bands between $1.55\mu\text{m}$ and $2.45\mu\text{m}$ are often the only bands used for the recognition of minerals.¹⁹ It was decided to use these bands in the Cuprite data, but because correlation between the unmixed minerals and the library minerals must be performed it was decided to include the bands between approximately $1.15\mu\text{m}$ and $1.55\mu\text{m}$ as these bands did not seem to have the problem of the lower bands. Thus, only AVIRIS bands above 77 were used.

The results of unmixing the minerals are shown in Figure 7. In order to decrease the possibility of false correlations, the average of the real mixed mineral channels was used in order to rule out incorrect correlations due to the fact that we do not use the average of the data in our method. The assumption is that the channel the unmixed mineral is contributing the most to is not going to have a reflectance average a lot more/less than that unmixed mineral. The unmixed signals contain the minerals associated with the ground truth minerals plus a contribution from Montmorillonite and a Kaolinite/Smectite combination. It is unknown whether there is a small amount of these minerals in that particular area of the image. It is possible that because less bands were used in the correlation, there are some false positives even after taking into account the average information in the signature.

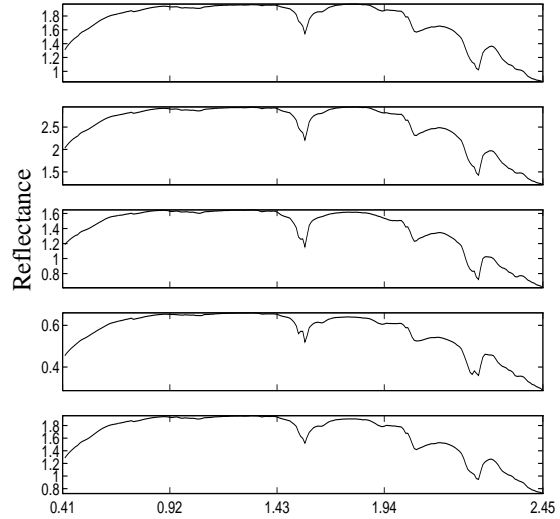
While the unmixed signals agree with the information about ground truth, the correlations give rise to doubts about trusting the results too much. The rise in minerals that correlate highly with the unmixed data signals is probably due to less bands in the AVIRIS data being used in the correlation. While whitening the data emphasizes the higher order statistics, it contains the problem of emphasizing the noise along with the higher order statistics. It is interesting to note that most of the noise in the resulting bands concentrates on the first channel.

5. FUTURE WORK

We have shown the ability of cICA to unmix synthetic and real data and have discussed the drawbacks of the approach. In the future these drawbacks should be addressed. Since the whitening process enhances all higher order statistics including noise, it becomes necessary to look for a way to reduce the noise before whitening the data. We propose that the data should be low pass filtered before whitening in order to decrease noise while keeping the statistics of interest. This approach has been followed by Olshausen and Field on two-dimensional image data¹⁶ and was introduced by Attick and Redlich.³

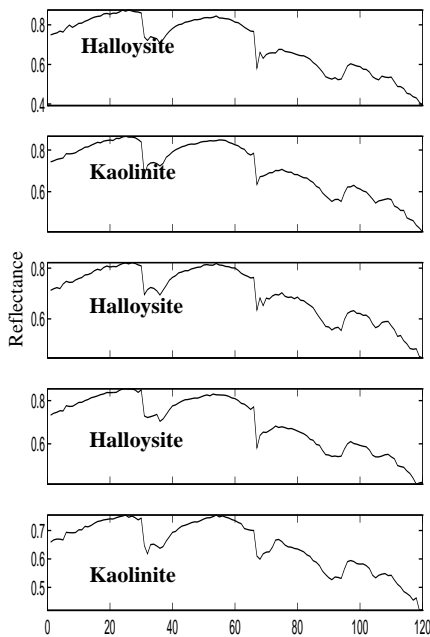


Real Data Mixtures

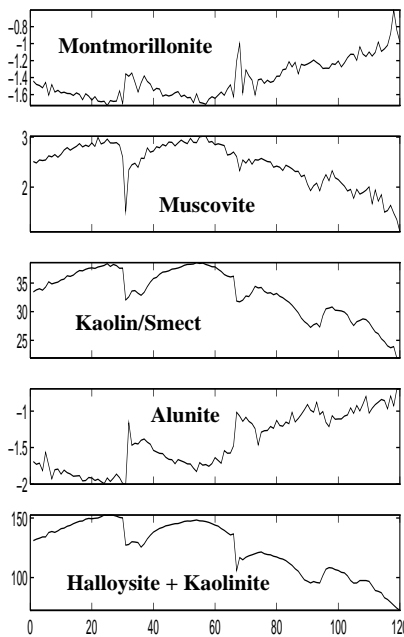


Synthetic Data Mixtures

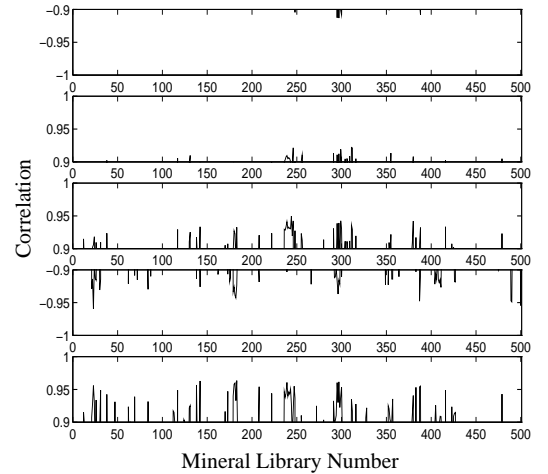
Figure 6. The real data is very different from a simple synthetic mixture containing Alunite, Muscovite, Kaolinite, Dickite, and Halloysite. The Dickite and Halloysite are added to the synthetic mixture as the ground truth lumps these minerals in with the Kaolinite, but even adding these minerals does not make the lower bands of the real data look like the lower bands of the synthetic data. Note that the real data mixture is shown with noisy bands removed and is scaled reflectance data.



Mixed Data Signals



Unmixed Signals



Correlations

Figure 7. The results of unmixing on real data show a decrease in performance, although the minerals are still able to be unmixed. All mixtures and results are labelled with their highest correlation from the mineral library. The performance decrease is most noticed in the correlations between the unmixed minerals and the library minerals. On four of the channels, the unmixed minerals correlate higher with more of the library minerals probably because less bands are being used in the correlation. Much of the remaining noise in the original bands found its way into the first channel of the unmixed minerals, causing a perhaps incorrect correlation with Montmorillonite.

As an iterative algorithm, cICA remains computationally intensive and only a sparse set of image data may be used in order to obtain endmembers. Another kind of analysis method for blind source separation might be less computationally intensive without losing the quality of the results. Furthermore, while preliminary results look promising, further analysis must be done in order to test the algorithm under a variety of conditions.

ACKNOWLEDGEMENTS

This research was facilitated in part by stipend support to Jessica Bayliss from the National Physical Science Consortium. Opportunities to engage in and complete the research reported in this article were provided by the NASA Goddard Space Flight Center.

REFERENCES

1. J.B Adams and M.O Smith. Spectral mixture modeling: A new analysis of rock and soil types at the viking lander 1 site. *J. of Geophysical Research*, 91(B8):8098–8112, 1986.
2. S. Amari, A. Cichocki, and H.H. Yang. A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems*, 8, 1996.
3. J.J. Atick and A.N. Redlich. Convergent algorithm for sensory receptive field development. *Neural Computation*, 5:45–60, 1993.
4. A.J. Bell and T.J. Sejnowski. The 'independent components' of natural scenes are edge filters. *To appear in Vision Research*.
5. A.J. Bell and T.J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1004–1034, 1995.
6. A.J. Bell and T.J. Sejnowski. Learning the higher-order structure of a natural sound,. *Network: Computation in Neural Systems*, 7, 1996.
7. Center for Study of Earth from Space (CSES), University of Colorado, Boulder, CO. *ATmospheric RE-Moval Program (ATREM)*, version 3.0 edition, July 1997. This software is available for anonymous ftp at [cses.colorado.edu:/pub/atrem](http://cses.colorado.edu/pub/atrem).
8. R. N .Clark. A comprehensive web site of hyperspectral imaging spectroscopy, spectral libraries and many results for Cuprite, Nevada is available at <http://speclab.cr.usgs.gov/>.
9. P. Comon. Independent component analysis: A new concept. *Signal Processing*, 36:287–314, 1994.
10. D.E. Sabol Jr., J.B Adams, and M.O. Smith. Predicting the spectral detectability of surface materials using spectral mixture analysis. *Proc. of the IEEE Internat'l Geoscience Remote Sensing Symp.*, 2:967–970, 1990.
11. C. Jutten and H. Herault. Blind separation of sources. part i. an adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 23:1–10, 1991.
12. J. Karhunen, E. Oja, L. Wang, R. Vigario, and J. Joutsensalo. A class of neural networks for independent component analysis. *IEEE Trans. on Neural Networks*, 8, May 1997.
13. F. Maselli, M. Pieri, and C. Conese. Automatic identification of end-members for the spectral decomposition of remotely sensed scenes. *Remote Sensing for Geography, Geology, Land Planning, and Cultural Heritage (SPIE)*, 2960:104–109, 1996.
14. J.M. Mendel and C.S. Burrus. *Maximum-likelihood deconvolution: a journey into model-based signal processing*. Springer-Verlag, 1990.
15. B.A. Olshausen. Learning linear, sparse, factorial codes. *MIT AI-memo*, 1580, 1996.
16. B.A. Olshausen and D.J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.
17. B.A. Pearlmutter and L.C. Parra. A context-sensitive generalization of ica. *Proc. ICONIP '96*, 1996. Available at <ftp://ftp.cnl.salk.edu/pub/bap/iconip-96-cica.ps.gz>.
18. B.A. Pearlmutter and L.C. Parra. Maximum likelihood blind source separation: A context-sensitive generalization of ica. In M.C. Mozer, M. I. Jordan, and T Petsche, editors, *Advances in Neural Information Processing Systems 9*, pages 613–619. MIT Press, 1997.
19. R.G. Resmini, M.E. Kappus, and W.S. Aldrich. Use of hyperspectral digital imagery collection experiment (hydice) sensor data for quantitative mineral mapping at cuprite, nevada. *Proc. of the 11th Thematic Conference on Geologic Remote Sensing*, 1:48–65, 1996.

20. H. Shipman and J.B. Adams. Detectability of minerals on desert alluvial fans using reflectance spectra. *J. of Geophysical Research*, 92(B10):10391–10402, 1987.
21. M.O Smith, D. Roberts, J. Hill, W. Mehl, B. Hosgood, J. Verdebout, G. Schmuck, D. Koechler, and J.B. Adams. A new approach to quantifying abundances of materials in multispectral images. *Internat'l Geoscience and Remote Sensing Symp.*, 4:2372–2374, 1994.