Recovery of constituent spectra using non-negative matrix factorization

Paul Sajda[a], Shuyan Du[a] and Lucas Parra[b]

[a] Department of Biomedical Engineering, Columbia University, New York, NY USA
[b] Adaptive Image and Signal Processing, Sarnoff Corporation, Princeton, NJ USA

ABSTRACT

In this paper a constrained non-negative matrix factorization (cNMF) algorithm for recovering constituent spectra is described together with experiments demonstrating the broad utility of the approach. The algorithm is based on the NMF algorithm of Lee and Seung,[1,2] extending it to include a constraint on the minimum amplitude of the recovered spectra. This constraint enables the algorithm to deal with observations having negative values by assuming they arise from the noise distribution. The cNMF algorithm does not explicitly enforce independence or sparsity, instead only requiring the source and mixing matrices to be non-negative. The algorithm is very fast compared to other "blind" methods for recovering spectra.[3,4] cNMF can be viewed as a maximum likelihood approach for finding basis vectors in a bounded subspace. In this case the optimal basis vectors are the ones that envelope the observed data with a minimum deviation from the boundaries. Results for Raman spectral data, hyperspectral images, and ^{31}P human brain data are provided to illustrate the algorithm's performance.

Keywords: Non-negative Matrix Factorization (NMF), Blind Source Separation, Raman Spectroscopy, Hyperspectral Imaging (HSI), Chemical Shift Imaging (CSI), Nuclear Magnetic Resonance (NMR) Spectroscopy

1. INTRODUCTION

Spectroscopic imaging makes use of spectral variations resulting from the interaction of energy with a sample to quantify/characterize the composition of the sample. Spectra, which are the resultant plots of the detected energy intensity versus wavelength (or mass, momentum, frequency, etc.) provide information about atomic and molecular energy levels, molecular geometries, chemical bonds, interactions of molecules, and related processes. Spectra are often used to identify the components of a sample (qualitative analysis) and may also be used to measure the abundance/concentration in a sample (quantitative analysis). While structural images provide morphological information, spectroscopy provides complementary information about chemical composition.

There are a variety of spectroscopic imaging methods, which are often characterized by the nature of the energy sources used to generate the spectra.

Raman Spectroscopy: Raman spectra, which are based on the vibrational motions of molecules, provide a chemical fingerprint suitable for identification and discrimination of a wide range of materials.[5] For example, Raman spectroscopic techniques utilize laser irradiation to provide specific information about biological tissue, important for both disease classification and chemical and morphological analysis. The technique is applied widely in diagnosis, such as in blood analyte measurement, atherosclerosis quantification, and breast cancer detection.

Hyperspectral Imaging (HSI) : An imaging method often used for remote sensing, hyperspectral imaging utilizes the reflectance spectrum of ambient light to produce spectra across a large number of wavelengths/bands ($>$ 30 bands). Typical wavelength ranges are from 0.4 to 2.5 μm with spatial resolutions of 10-30 m. Each pixel in the resultant image represents a spectral response that is indicative of materials that are exposed on the Earth's surface. Absorption features in these spectra are determined by chemical composition and physical structure of the material. Hyperspectral data have been studied for a wide range of applications including environmental monitoring, mineral exploration, vegetation mapping, hazardous material remediation, and water quality control.

Chemical Shift Imaging (CSI): CSI is an imaging modality whereby high resolution nuclear magnetic resonance (NMR) spectra are acquired across a volume of tissue.[6] *In vivo* CSI allows for the non-invasive characterization and quantification of molecular markers with clinical utility for improving detection, identification, and treatment for a variety of diseases, most notably neurological disorders. CSI can be tuned to different biochemical markers through varying the atomic resonances, thus providing precise characterization of tissue and/or a means for optimizing the signal-noise-ratio (SNR). Common atomic resonances include ^{1}H (protons)*, ^{19}F (Fluorine), and ^{31}P (Phosphorus). Together with structural magnetic resonance imaging (MRI), CSI can provide an integrated biochemical and morphological view of biological tissues and disease processes.

1.1. Linear Mixture Model for Spectral Data

In the three spectroscopic imaging modalities described above, one often assumes that the observations/measurements represent a mixture of some underlying constituent materials and, to a first approximation, this is a linear mixture. Given observed spectra $\mathbf{X}$, we can therefore say that,

$$\mathbf{X} = \mathbf{AS} + \mathbf{N} \tag{1}$$

where the columns in $\mathbf{A}$ represent the concentration/abundance of the constituent materials and the rows in $\mathbf{S}$ the corresponding constituent spectra. $\mathbf{N}$ represents additive noise. The abundance matrix $\mathbf{A}$ has M columns (one for each material) and N rows (one for each voxel/pixel). $\mathbf{X}$ and $\mathbf{S}$ have L columns (one for each frequency/band). Since we interpret $\mathbf{A}$ as concentrations, we can assume the matrix to be non-negative. In addition, since the constituent spectra, $\mathbf{S}$, represents amplitudes of reflectance/resonances, in theory its smallest value should be zero, corresponding to the absence of reflectance/resonance at a given frequency. The factorization of Equation 1 is therefore constrained by,

$$\mathbf{A} \geq 0 \text{ and } \mathbf{S} \geq 0. \tag{2}$$

However, due to the presence of additive noise, the observations $\mathbf{X}$ may themselves be negative.

One of the primary goals of spectroscopic analysis is to recover the constituent spectra $\mathbf{S}$ and estimate the abundances $\mathbf{A}$. Such techniques for "unmixing" or "source recovery" can be classified as either supervised or unsupervised.

1.2. Supervised vs. Unsupervised Methods

Supervised spectral unmixing relies on prior knowledge about the reflectance/resonance patterns of candidate constituent materials (sometimes referred to as *endmembers*), or domain knowledge and a series of semi-automatic steps to find the constituent materials in a particular scene. Given knowledge about the endmembers one can find the abundances simply by solving a constrained least square problem,[4] for example imposing a parametric or explicit model for $\mathbf{S} = \mathbf{S}(\theta)$, and considering one pixel/voxel at a time, $\mathbf{x}_i$ to invert the linear problem $\mathbf{x}_i = \mathbf{S}(\theta)\mathbf{a}_i$.[7] Such supervised techniques require substantial user interaction to recover the endmembers and easily lead to error, as a pixel can be misinterpreted (e.g. a pixel is taken as one pure endmember while it is actually a mixture of several endmembers). Another type of supervised unmixing approach obtains endmembers directly from a database, which also can be problematic. For example, environmental changes, such as atmospheric absorption, can significantly affect the spectral profile and therefore the endmember extracted from a database may not be a good model for a variety of environmental contexts.

More recently, unsupervised methods have been used to try to identify endmembers and mixtures directly from the observed data $\mathbf{X}$. For example, there have been efforts to simultaneously exploit the statistical structure of multi-pixel/voxel spectra to solve Equation 1 as a blind source separation (BSS) problem. Parra et al.[4] developed a maximum a posteriori (MAP) method for estimating the abundances and spectra in HSI. Their MAP approach imposes constraints on the positivity of the mixing matrix and the sparsity, in the space of innovations, of the source matrix. However convergence is slow, with solutions heavily dependent upon meta-parameter choices used in the optimization.

*When protons are imaged the imaging modality is often referred to as magnetic resonance spectroscopy imaging (MRSI).

Nuzillard et al.[8] use second order blind identification (SOBI)[9] to separate ^{13}C spectra. Problematic with the approach is the assumption that the constituent spectra are orthogonal, which is required for the SOBI algorithm. While in many cases the constituent spectra can be highly correlated, an orthogonality assumption is incorrect. Similar to Parra et al., Ochs et al[3] formulate Equation 1 within a Bayesian framework to estimate $\mathbf{A}$ and $\mathbf{S}$ for CSI spectra. Their Bayesian Spectral Decomposition (BSD) uses a Markov chain Monte Carlo (MCMC) procedure to sample the posterior space of $p(\mathbf{A}, \mathbf{S}|\mathbf{X})$ subject to the likelihood $p(\mathbf{X}|\mathbf{S}, \mathbf{A})$ (i.e. the noise distribution) and priors $p(\mathbf{A})$, $p(\mathbf{S})$. These priors include positivity and sparseness on $\mathbf{A}$ and $\mathbf{S}$, however make no assumptions about orthogonality. Their results show good separation for highly correlated constituent spectra. However the approach is computationally expensive, given the MCMC procedure.

In this paper, a fast algorithm is proposed which exploits only the non-negativity of $\mathbf{A}$, $\mathbf{S}$ for blindly separating various types of spectroscopic data. The algorithm is based on the NMF algorithm by Lee and Seung.[1,2] We further develop the NMF approach, within a maximum likelihood framework, to include a function for forcing low amplitude spectral values in the recovered sources to be zero. The method can be viewed as a subspace reduction whereby small amplitudes of the constituent sources are forced to zero - i.e. forced to the edges of a polygonal conic subspace spanned by the constituent spectra. We term this algorithm constrained non-negative matrix factorization (cNMF). The algorithm bears some resemblance to the non-negative sparse coding (NNSC) method of Hoyer,[10] which also has as its basis the NMF algorithm. However NNSC imposes an explicit sparsity constraint while cNMF does not. Perhaps more importantly, cNMF is formulated to systematically deal with observations having negative values, a situation which often comes up in practice, yet other algorithms typically have difficulty with. In the following sections we describe the cNMF algorithm and present results for Raman Spectroscopy, Hyperspectral Imaging, and ^{31}P CSI data.

2. CONSTRAINED NON-NEGATIVE MATRIX FACTORIZATION

We begin by reviewing the NMF algorithm of Lee and Seung,[1] formulating it within the context of a noise model. The basic idea of the algorithm is to construct a gradient descent over an objective function that optimizes $\mathbf{A}$ and $\mathbf{S}$. For example, with $\mathbf{N}$ modeled as i.i.d. Gaussian noise (a reasonable assumption, for example, given the empirical noise distribution of the CSI data shown in Figure 4), one can formulate the problem as maximum likelihood estimation,

$$\begin{aligned} \mathbf{A}_{ML}, \mathbf{S}_{ML} &= \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmax}}\, p(\mathbf{X}|\mathbf{A}, \mathbf{S}) \\ &= \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmax}}\, \mathcal{N}(\mathbf{X} - \mathbf{AS}, \Sigma) \end{aligned} \tag{3}$$

where Σ is the covariance of the Gaussian distribution and $(\mathbf{AS})$ the mean.

Maximizing the likelihood is equivalent to minimizing the negative log-likelihood, so Equation 3 can be written as,

$$\begin{aligned} \mathbf{A}_{ML}, \mathbf{S}_{ML} &= \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmin}}\, (-\log p(\mathbf{X}|\mathbf{A}, \mathbf{S})) \\ &= \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmin}}\, \|\mathbf{X} - \mathbf{AS}\|^2 \\ \text{subject to} \;:\quad & \mathbf{A} \geq 0,\ \mathbf{S} \geq 0. \end{aligned} \tag{4}$$

Defining the negative log-likelihood as $F = \|\mathbf{X} - \mathbf{AS}\|^2$, the gradients of F with respect to $\mathbf{A}$ and $\mathbf{S}$ are given by,

$$\begin{aligned} \frac{\partial F}{\partial \mathbf{A}_{i,m}} &= -2 * ((\mathbf{XS^T})_{i,m} - (\mathbf{ASS^T})_{i,m}) \\ \frac{\partial F}{\partial \mathbf{S}_{m,\lambda}} &= -2 * ((\mathbf{A^T X})_{m,\lambda} - (\mathbf{A^T AS})_{m,\lambda}). \end{aligned} \tag{5}$$

Using the gradients one can construct the additive update rules,

$$\begin{aligned}\mathbf{A}_{i,m} &\leftarrow \mathbf{A}_{i,m} + \delta_{i,m}[(\mathbf{X}\mathbf{S}^T)_{i,m} - (\mathbf{A}\mathbf{S}\mathbf{S}^T)_{i,m}] \\ \mathbf{S}_{m,\lambda} &\leftarrow \mathbf{S}_{m,\lambda} + \eta_{m,\lambda}[(\mathbf{A}^T\mathbf{X})_{m,\lambda} - (\mathbf{A}^T\mathbf{A}\mathbf{S})_{m,\lambda}].\end{aligned} \tag{6}$$

Note that there are two free parameters, which are the step sizes of the updates. Lee and Seung show that by choosing the step sizes as, $\delta_{i,m} = \frac{\mathbf{A}_{i,m}}{(\mathbf{A}\mathbf{S}\mathbf{S}^T)_{i,m}}$, $\eta_{m,\lambda} = \frac{\mathbf{S}_{m,\lambda}}{(\mathbf{A}^T\mathbf{A}\mathbf{S})_{m,\lambda}}$, the additive update rule can be formulated as a multiplicative update rule, with $\mathbf{X} = \mathbf{A}\mathbf{S}$ being a fixed point. In addition, they show that this optimization is equivalent to optimization over an auxiliary function, guaranteed to have the same minimum as Equation 4.[2] The multiplicative update rules for $\mathbf{A}$ and $\mathbf{S}$ therefore become,

$$\begin{aligned}\mathbf{A}_{i,m} &\leftarrow \mathbf{A}_{i,m}\frac{(\mathbf{X}\mathbf{S}^T)_{i,m}}{(\mathbf{A}\mathbf{S}\mathbf{S}^T)_{i,m}} \\ \mathbf{S}_{m,\lambda} &\leftarrow \mathbf{S}_{m,\lambda}\frac{(\mathbf{A}^T\mathbf{X})_{m,\lambda}}{(\mathbf{A}^T\mathbf{A}\mathbf{S})_{m,\lambda}}.\end{aligned} \tag{7}$$

Alternatively one might wish to assume a Poisson noise model, i.e., Poisson process with mean $(\mathbf{A}\mathbf{S})_{i,\lambda}$. In this case one would construct the likelihoods as,

$$\begin{aligned}p(\mathbf{X}_{i,\lambda}|\mathbf{A},\mathbf{S}) &= \frac{(\mathbf{A}\mathbf{S})_{i,\lambda}^{\mathbf{X}_{i,\lambda}}}{(\mathbf{X}_{i,\lambda})!}e^{-(\mathbf{A}\mathbf{S})_{i,\lambda}} \\ p(\mathbf{X}|\mathbf{A},\mathbf{S}) &= \prod_{i,\lambda} p(\mathbf{X}_{i,\lambda}|\mathbf{A},\mathbf{S}) \\ \log p(\mathbf{X}|\mathbf{A},\mathbf{S}) &= \log(\prod_{i,\lambda} p(\mathbf{X}_{i,\lambda}|\mathbf{A},\mathbf{S})) \\ &= \sum_{i,\lambda}\mathbf{X}_{i,\lambda}\log(\mathbf{A}\mathbf{S})_{i,\lambda} - \sum_{i,\lambda}(\mathbf{A}\mathbf{S})_{i,\lambda} - \sum_{i,\lambda}\log\mathbf{X}_{i,\lambda}!\end{aligned} \tag{8}$$

where $(\sum_{i,\lambda}\log\mathbf{X}_{\mathbf{i},\lambda}!)$ is a constant. To maximize the likelihood one constructs F as,

$$F = \sum_{i,\lambda}\mathbf{X}_{i,\lambda}\log(\mathbf{A}\mathbf{S})_{i,\lambda} - \sum_{i,\lambda}(\mathbf{A}\mathbf{S})_{i,\lambda} \tag{9}$$

and computes the gradients of F with respect to $\mathbf{A}$ and $\mathbf{S}$,

$$\begin{aligned}\frac{\partial F}{\partial \mathbf{A}_{i,m}} &= \sum_{\lambda}\mathbf{S}_{m,\lambda}\frac{\mathbf{X}_{i,\lambda}}{(\mathbf{A}\mathbf{S})_{i,\lambda}} - \sum_{\lambda}\mathbf{S}_{m,\lambda} \\ \frac{\partial F}{\partial \mathbf{S}_{m,\lambda}} &= \sum_{i}\mathbf{A}_{i,m}\frac{\mathbf{X}_{i,\lambda}}{(\mathbf{A}\mathbf{S})_{i,\lambda}} - \sum_{i}\mathbf{A}_{i,m}\end{aligned} \tag{10}$$

resulting in the additive updating rule,

$$\begin{aligned}\mathbf{A}_{i,m} &\leftarrow \mathbf{A}_{i,m} + \delta_{i,m}(\sum_{\lambda}\mathbf{S}_{m,\lambda}\frac{\mathbf{X}_{i,\lambda}}{(\mathbf{A}\mathbf{S})_{i,\lambda}} - \sum_{\lambda}\mathbf{S}_{m,\lambda}) \\ \mathbf{S}_{m,\lambda} &\leftarrow \mathbf{S}_{m,\lambda} + \eta_{m,\lambda}(\sum_{i}\mathbf{A}_{i,m}\frac{\mathbf{X}_{i,\lambda}}{(\mathbf{A}\mathbf{S})_{i,\lambda}} - \sum_{i}\mathbf{A}_{i,m}).\end{aligned} \tag{11}$$

By choosing $\delta_{i,m} = \frac{\mathbf{A}_{i,m}}{\sum_{\lambda}\mathbf{S}_{m,\lambda}}$, $\eta_{m,\lambda} = \frac{\mathbf{S}_{m,\lambda}}{\sum_{i}\mathbf{A}_{i,m}}$, one obtains the multiplicative update rule for the case of a Poisson noise model,

$$\mathbf{A}_{i,m} \leftarrow \mathbf{A}_{i,m} \frac{\sum_\lambda \mathbf{S}_{m,\lambda} \frac{\mathbf{X}_{i,\lambda}}{(\mathbf{AS})_{i,\lambda}}}{\sum_\lambda \mathbf{S}_{m,\lambda}}$$
$$\mathbf{S}_{m,\lambda} \leftarrow \mathbf{S}_{m,\lambda} \frac{\sum_i \mathbf{A}_{i,m} \frac{\mathbf{X}_{i,\lambda}}{(\mathbf{AS})_{i,\lambda}}}{\sum_i \mathbf{A}_{i,m}}. \tag{12}$$

By formulating the updates as multiplicative rules in Equation 7 and Equation 12 one can ensure non-negative $\mathbf{A}$ and $\mathbf{S}$, given both are initialized non-negative and the observations, $\mathbf{X}$, are non-negative. However one problem is that, due to noise, the observations can have negative values. Since all observations are used in updating $\mathbf{A}$ and $\mathbf{S}$, a possible solution will be recovered spectra with negative amplitudes, which is undesirable given our prior knowledge of physically meaningful spectra.

To mitigate the issue of negative values in the observations, it is instructive to view the factorization in Equation 1 as representing a subspace reduction from a L dimensional space into a constrained M dimensional space. Except for the positivity constraints, the decomposition is completely arbitrary within that M-dimensional space. However, spectra and concentrations are non-negative and so the M-dimensional degrees of freedom within that subspace are constraints by $M(N+L)$ linear boundary constraints (Equation 2). This is the portion of the space that corresponds to realistic solutions of the factorization. We wish to further constrain the space of possible solutions by disallowing small spectral magnitude values or negative values, and instead assume that they are due to baseline noise. We thus allow non-zero solutions within the linear subspace only when there is sufficient evidence to explain them above a minimum noise floor. We enforce this by introducing a threshold constraint on $\mathbf{S}$,

$$S_{i,j} = \begin{cases} S_{i,j} & S_{i,j} > \theta \\ \epsilon & S_{i,j} \leq \theta \end{cases} \tag{13}$$

where θ determines the noise floor which is strictly positive and ϵ is some very small positive value [†]. We treat $\mathbf{A}$ symmetrically, using the same threshold constraint as mentioned above. This ensures $\mathbf{A}$ remains non-negative, given the possibility of negative values in $\mathbf{X}$.

We initialize the matrices $\mathbf{A}$ and $\mathbf{S}$ by constructing a non-negative random $\mathbf{A}$ and estimating $\mathbf{S}$ using constrained least squares,

$$\underset{\mathbf{A},\mathbf{S}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{AS}\|^2 \text{ subject to } \mathbf{S} \geq 0. \tag{14}$$

To summarize, the procedure for updating $\mathbf{A}$ and $\mathbf{S}$ is,

1. Initialize: Choose dimensions of $\mathbf{A}$ and $\mathbf{S}$ and initialize with non-negative values (e.g. random $\mathbf{A}$ and constrained least-squares for $\mathbf{S}$).
2. Update $\mathbf{A}$
3. Force small values of $\mathbf{A}$ to be approximately zero.
4. Update $\mathbf{S}$
5. Force small values of $\mathbf{S}$ to be approximately zero.
6. Iterate (back to 2).

We call this the constrained non-negative matrix factorization (cNMF) algorithm.

An intuitive understanding of cNMF via geometrical considerations can be developed. The manifold of possible solutions specified by Equation 1 and Equation 2 represent a M dimensional polygonal cone spanned by the M rows of $\mathbf{S}$. Positivity constraints on the spectra require that the row vectors of $\mathbf{S}$, representing the

[†]Note that the spectral amplitudes cannot be set to exactly zero given the update rules for $\mathbf{A}$ and $\mathbf{S}$

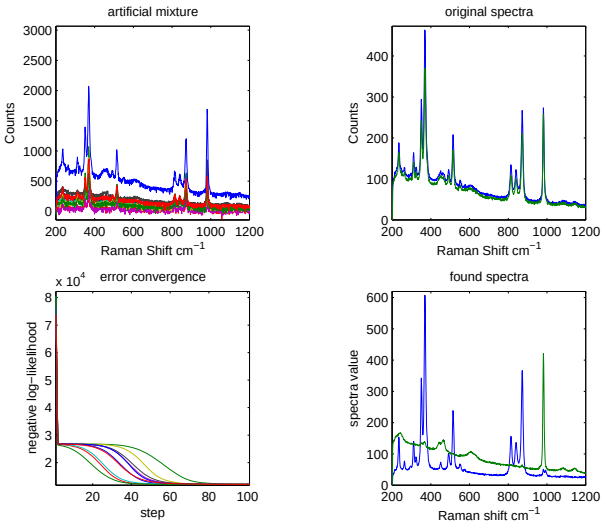


Figure 1. Results for Raman spectra using cNMF. (top left): Artificial mixture of Epsomite and Demantoid spectra with Gaussian noise, with only 10 "observations" shown; (top right): Initialization of **S** with two endmembers; (bottom left): Empirical convergence of negative log-likelihood in 10 independent runs; (bottom right): Recovered **S**.

edges of the cone, lie in the positive quadrant of the L dimensional points defined by the rows of the observations **X** must fall within that polygonal cone. The additive noise in the probabilistic model allows points to fall outside this cone with a certain likelihood. The aim of maximum likelihood is to find cone edge vectors that tightly envelope the observed L-points with the smallest possible deviation from the boundaries. By constraining small values of the vectors in **S** to be zero we force some polygon edges onto the boundaries (or edges) of the positive quadrant. This will possibly increase the noise required to explain points that fall outside the M-polygonal cone.

In theory this heuristic procedure is not guaranteed to converge monotonically because of the application of the threshold. However the procedure will not result in an unbounded cost since between threshold jumps, the cost function deceases monotonically. Thus the algorithm can be viewed as a gradient decent with occasional jumps that place the optimization in a new region of cost function space (i.e. optimization is locally monotonic). We see that, in practice, our algorithm always converges, with examples of the evolution of the negative log likelihood shown in Figures 1 and 4(c).

3. EXPERIMENTS

To demonstrate feasibility and performance of the cNMF algorithm we conducted experiments using various types of spectral data. These include artificial mixtures of Raman spectral data, real mixtures in hyperspectral aerial images and ^{31}P CSI data of human brain.

3.1. Raman Spectra: Artificial Mixtures

Raman spectral data were taken from a dataset provided by the Division of Geological and Planetary Sciences at CalTech. [‡] The spectra of Epsomite and Demantoid with bandwidth of 201 to 1200 Raman shift (cm^{-1}) are artificially mixed by a [256, 2] strictly positive random mixing matrix. To this mixture we add Gaussian noise so that 10% of the observation values are negativte. **X** therefore consists of 256 observations across 1000 bands. Since we have Gaussian noise, the multiplicative updating rule in Equation 7 is chosen and the updating procedure described previously is followed. Figure 1, shows the artificial mixture. An initial **S** , consisting of two non-negative spectra, is also shown. Note the strong overlap of the initialized spectra, and the recovered constituent spectra after approximately 60 iterations of the updating rule. Figure 1 also shows the negative

[‡]http://minerals.gps.caltech.edu/FILES/raman/Caltech_data/

log-likelihood for 10 independent initializations. Note in all cases there is monotonic convergence. We note that in some cases (not shown) convergence is not observed to be strictly monotonic, with the thresholding resulting in "jumps" in likelihood space. However in those cases we do see local convergence and we never observe the negative log-likelihood increasing unbounded.

A closer look at the recovered spectra in Figure 2(b) shows that they are nearly identical to the true spectra used in the artificial mixture (Figure 2(a)). The recovered mixing matrix $\mathbf{A}$ is also compared to the "true" abundance matrix used to construct the artificial mixture. Note that each [256,1] vector in the mixing matrix is reshaped into a [16,16] matrix in column sequence for display convenience.

We compare our results to those obtained using NNSC[10] (Figure 2(c)). Since NNSC requires all values in $\mathbf{X}$ to be non-negative, we modified the algorithm to including thresholding of negative observations. In NNSC, both $\mathbf{A}$ and $\mathbf{S}$ are randomly initialized to be strictly positive. We note that the recovered spectra include "flips" with some bands not as well-separated as cNMF. Though both algorithms are very fast, the cNMF algorithm converges to a more correct solution, while at the same time including no specific assumptions on the sparsity of the constituent spectra.

3.2. Hyperspectral Aerial Images

In this experiment, ground truthed USGS hyperspectral data representing real mixtures are used. We select 10×10 blocks of pixels from three different regions[§] in the AVIRIS data of the Cuprite, Nevada mining district. We separate these 300 mixed spectra assuming 2 endmembers and Poisson noise, with the endmembers shown in Figure 3. After 4000 iterations, the algorithm converges, and recovered spectra are plot vs. sample spectra from the USGS data base library in Figure 3. Abundance maps for Kaolinite and Muscovite in these 3 regions are also shown in the figure. Regions 1 and 2 contain similar abundances for Kaolinite and Muscovite, while region 3 contains more Muscovite. These results are similar to those by Parra et al.[4]

Since hyperspectral data are smooth and not sparse in the original signal space, a sparsity assumption in the original signal space is not appropriate. Parra et al.'s[4] MAP algorithm imposes a sparsity constraint in the subspace of an auto-regressive innovation process, however our results indicate that this is not necessary. Only non-negativity is required. In addition, the cNMF algorithm is several orders of magnitude faster than the MAP approach.

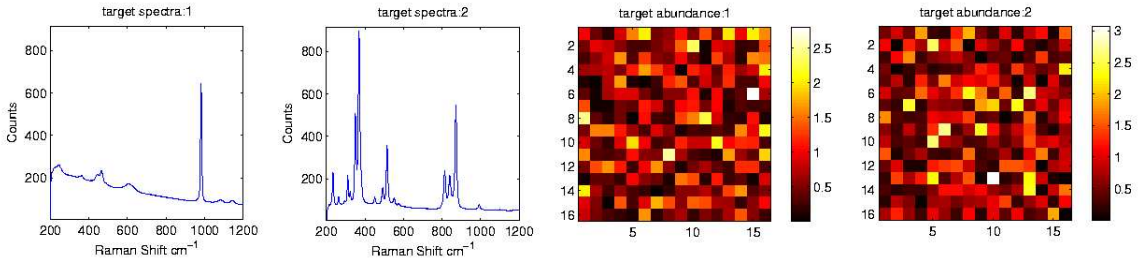
3.3. Chemical Shift Images of Human Brain

In this experiment, 512 ^{31}P spectra (8-by-8-by-8 voxels) of human brain data from Ochs et al.[3] were used, each spectra represented as a 369 point vector. Using the same preprocessing as in Ochs et al., we use principal component analysis (PCA) and visual inspection of the data to determine that 256 of the 512 voxels contain signal and that there are primarily two constituent sources. Thus we have M=2, N=256, L=369. Figure 4(a) shows an axial slice of the CSI data. Here we clearly see that spectra near the edges are almost exclusively noise. An empirical noise estimate, shown in Figure 4(b), indicates that a Gaussian noise model is reasonable. The recovered spectra are shown in Figure 5(a). It is clear that the recovered spectra are highly correlated and therefore any orthogonality assumption is inappropriate and would not lead to correct results. The results are consistent with the underlying biochemical characteristics of the tissue, with the top spectra being indicative of muscle tissue and the bottom of brain tissue. Figure 5(b) shows the 5th axial slice of the mixing matrix $\mathbf{A}$, which can be viewed as the relative concentration of the recovered spectra. In this case, we see the muscle spectra concentrated near the skull border while brain spectra largely internal to the muscle spectra. The recovered spectra are nearly identical to the BSD approach results in,[3] but the cNMF algorithm only takes 0.48 seconds to converge (2.3 GHz Intel Processor) while BSD takes 12,000 seconds (1.2 GHz Intel Processor).

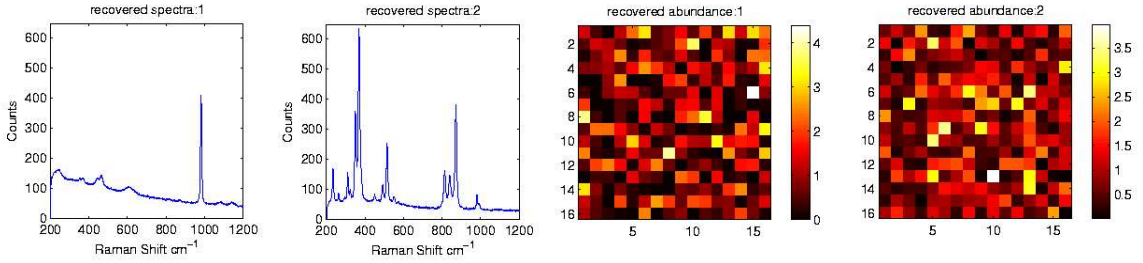
To demonstrate the importance of the noise model assumption[¶], we also generate results for the same set of CSI data except using the Poisson noise model formulation for the updates. Figures 6(a) and 6(b) are the

[§]Regions: (I) coordinate (265, 710) coordinate (275, 697) with Kaolinite and Muscovite 2, (III) coordinate (143, 661) with only Muscovite 2 (ftp://speclab.cr.usgs.gov/pub/cuprite/gregg.thesis.images/plate2.cuprite95.alpha.2um.image.wlocals.gif).

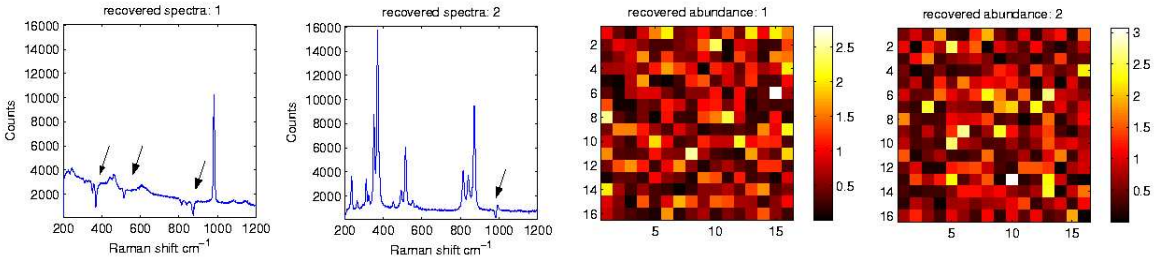
[¶]Note that Lee and Seung[1] remark that the choice of the cost function is not important. However our results indicate otherwise, especially when considering negative values in the observations. We believe this is clear if one formulates the cost function with respect to an underlying noise model.



(a)



(b)



(c)

Figure 2. (a) Target Raman spectra of Epsomite (1) and Demantoid (2), and corresponding target abundance distribution. (b) Recovered spectra of endmembers 1 and 2. In comparison, we see the recovered spectra are almost identical to the targets. Corresponding recovered material abundance maps are very close to the "true" abundance matrices. (c) Recovered Raman spectra by NNSC (arrows point to flipped peaks) and recovered abundance distribution of endmembers 1 and 2.

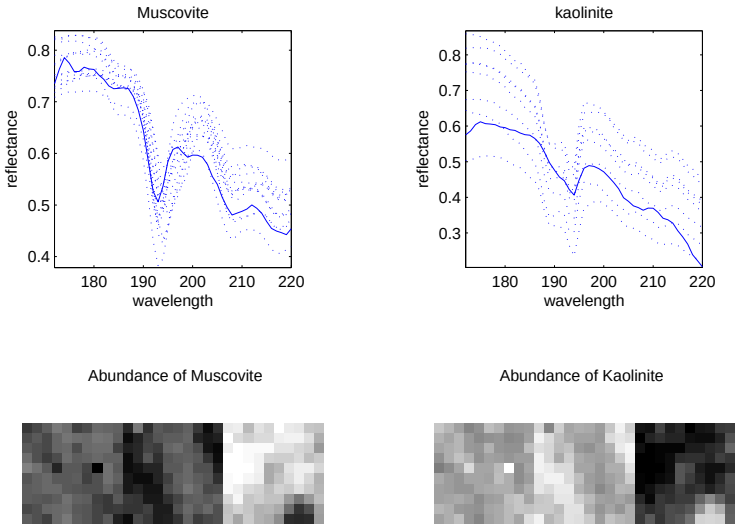
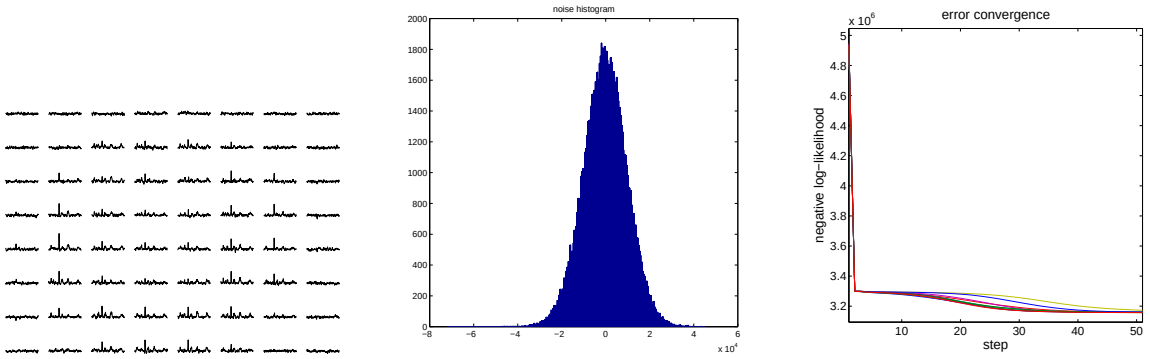


Figure 3. (Top left): Spectra of computed endmember (solid line) vs Muscovite sample spectra from the USGS data base library. Note that only part of the spectra is shown since the discriminating features are located between band 172 and 220; (top right): computed endmember (solid line) vs Kaolinite sample spectra from the USGS data base library; (bottom): Abundance of Kaolinite and Muscovite for three regions (lighter pixels represent higher abundance). Region 1 and 2 have similar abundances for Kaolinite and Muscovite, while region 3 contains more Muscovite.



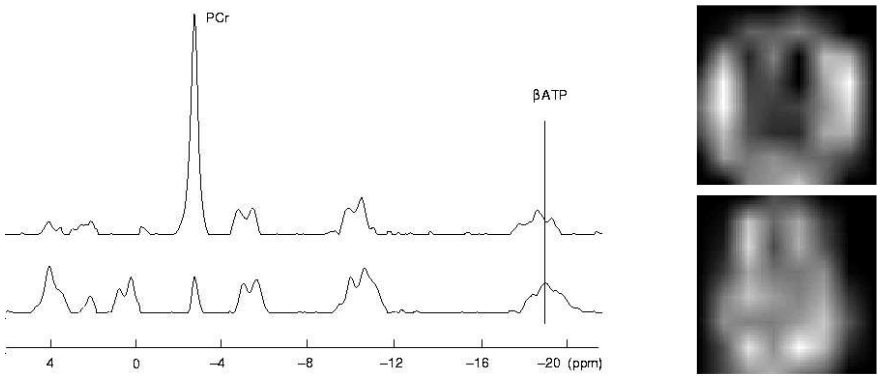
(a) (b) (c)

Figure 4. Example ^{31}P CSI data from Ochs et al.[3] (a) 8-by-8 voxel axial slice of spectra taken of the brain of a healthy subject. Spectra near the edges are almost exclusively noise. The complete dataset consists of 512 voxels (8-by-8-by-8) with spectra having 369 points (chemical shifts in parts-per-million (ppm)). (b) Empirical noise estimate from ^{31}P CSI data. The noise distribution was computed using the 256 voxels that were determined, via PCA and visual inspection, to contain no signal. (c) Evolution of the negative log likelihood of cNMF, for 10 random initializations of $\mathbf{A}$ and $\mathbf{S}$. The figure is an empirical demonstration of algorithm convergence.

results. After 50 iterations, the algorithm converges, however the two spectra are not physically meaningful, as can also be seen in the corresponding abundance maps shown in Figure 6(b).

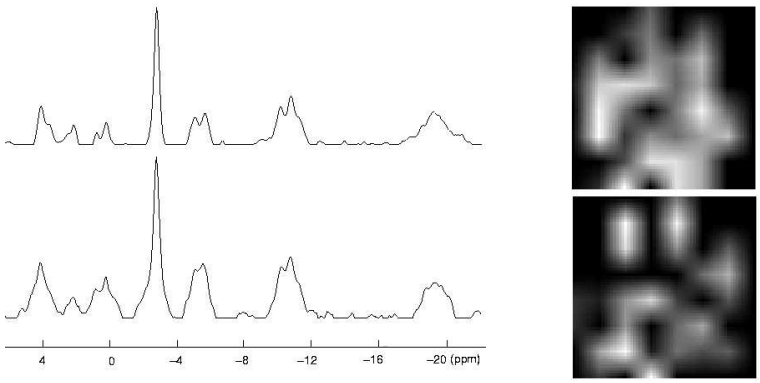
4. CONCLUSION

In this paper we have described an algorithm for recovering constituent (source) spectra in spectroscopic data, including Raman spectra, aerial hyperspectral imagery and CSI data. The algorithm, which we term cNMF, is



(a) (b)

Figure 5. Spectral separation results using cNMF. (a) cNMF recovered spectra, (top-source 1) muscle spectra and (bottom-source 2) brain spectra. (b) 5th axial slice of mixing matrix showing relative concentrations of muscle and brain spectra.



(a) (b)

Figure 6. Results using a Poisson noise model. (a) Recovered spectra. (b) 5th axial slice of mixing matrix.

based on an extension of the Lee and Seung NMF algorithm and can be viewed as a maximum likelihood approach for finding basis vectors in a subspace. The basis vectors are found such that they envelope the observed L-points with the smallest possible deviation from the boundaries. In cNMF only non-negativity constraints are imposed. In contrast to other algorithms, cNMF is formulated to deal with negative observations, which often occur with an underlying Gaussian noise model. We apply this algorithm to various types of spectroscopic data, which include both smooth and sparse signals, and report results which show that cNMF rapidly converges to correct and physically meaningful results. Such rapid convergence can be extremely useful, particularly if one is interested in real-time analysis during data acquisition.

5. ACKNOWLEDGMENTS

We would like to thank Truman Brown and Radka Stoyanova for the ^{31}P human brain data. This research was supported by an NSF CAREER Award (BES-0133804) to P.S. and the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research (N00014-01-0625).

REFERENCES

1. D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature* **401**, pp. 788–791, 1999.
2. D. Lee and H. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems 13*, pp. 556–562, MIT Press, 2001.
3. M. Ochs, R. Stoyanova, F. Arias-Mendoza, and T. Brown, "A new method for spectral decomposition using a bilinear Bayesian approach," *Journal of Magnetic Resonance* **137**, pp. 161–176, 1999.
4. L. Parra, C. Spence, A. Ziehe, K.-R. Mueller, and P. Sajda, "Unmixing hyperspectral data," in *Advances in Neural Information Processing Systems 13*, pp. 848–854, MIT Press, 2000.
5. B. Bulkin, "The Raman effect: An introduction," in *Analytical Raman Spectroscopy*, pp. 1–19, New York: John Wiley and Sons, Inc, 1991.
6. T. Brown, B. Kincaid, and K. Ugurbil, "NMR chemical shift imaging in three dimensions," *PNAS* **79**(11), pp. 3523–3526, 1982.
7. P. Sajda, S. Du, T. Brown, L. Parra, and R. Stoyanova, "Recovery of constituent spectra in 3D chemical shift imaging using non-negative matrix factorization," in *4th International Symposium on Independent Component Analysis and Blind Source Separation (ICA2003)*, pp. 71–76, 2003.
8. D. Nuzillard, S. Bourg, and J.-M. Nuzillard, "Model-free analysis of mixtures of NMR using blind source separation," *Journal of Magnetic Resonance* **133**, pp. 358–363, 1998.
9. A. Belouchrani, K. Abed-Meraim, J. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. SP* **45**, pp. 434–444, 1997.
10. P. O. Hoyer, "Non-negative sparse coding," in *Proc.IEEE Workshop on Neural Networks for Signal Processing*, pp. 557–565, 2002.