RECOVERY OF CONSTITUENT SPECTRA IN 3D CHEMICAL SHIFT IMAGING USING NON-NEGATIVE MATRIX FACTORIZATION

Paul Sajda[*]*, Shuyan Du, Truman Brown*

Department of Biomedical Engineering
and Department of Radiology
Columbia University
New York, NY USA

Lucas Parra

Adaptive Image and
Signal Processing
Sarnoff Corporation
Princeton, NJ USA

Radka Stoyanova

Biostatistics,
Fox Chase Cancer Center
Philadelphia, PA USA

ABSTRACT

In this paper we describe a non-negative matrix factorization (NMF) for recovering constituent spectra in 3D chemical shift imaging (CSI). The method is based on the NMF algorithm of Lee and Seung [1], extending it to include a constraint on the minimum amplitude of the recovered spectra. This constrained NMF (cNMF) algorithm 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 ^{31}P human brain data are compared to Bayesian Spectral Decomposition (BSD) [2] which considers a full Bayesian treatment of the source recovery problem and requires computationally expensive Monte Carlo methods. The cNMF algorithm is shown to recover the same constituent spectra as BSD, however in about 10^{-4} less computational time.

1. INTRODUCTION

Chemical shift imaging (CSI) is an imaging modality whereby high resolution nuclear magnetic resonance (NMR) spectra are acquired across a volume of tissue [3]. *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 characterizations of tissue and/or a means for optimizing the signal-to-noise ratio (SNR). The more common atomic resonances include ^{1}H (protons)[1], ^{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 tissue and disease processes.

Each tissue type can be viewed as having a characteristic spectral profile corresponding to the chemical composition of the tissue. In a given voxel multiple tissue types may be present. The observed spectra is therefore a combination of different constituent spectra. The signal measured in CSI is the response to a coherent stimulation of the entire tissue. As a result the amplitudes of the different coherent resonators are additive and their spectral magnitudes are additive. The overall gain with which a tissue type contributes to this addition is proportional to its concentration in each voxel. As a results we can explain the observed spectra $\mathbf{X}$ as

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

where the columns in $\mathbf{A}$ represent the concentration, or abundance, of the constituent material and the rows in $\mathbf{S}$ their corresponding 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). $\mathbf{X}$ and $\mathbf{S}$ have L columns (one for each resonance frequency). Figure 1 shows an example of CSI data for ^{31}P spectroscopy. The data shows spectra for an axial slice of human brain. In this example the complete CSI acquisition was 8-by-8-by-8 voxels.

Since we interpret $\mathbf{A}$ as concentrations, we can assume the matrix to be non-negative. In addition, since the constituent spectra, $\mathbf{S}$, represent amplitudes of resonances, in theory the smallest resonance amplitude is zero, corresponding to the absence of resonance at a given frequency. The factorization of Equation 1 is therefore constrained by,

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

In CSI spatial inhomogeneity of the magnetic field may introduce an unknown phase shift which must be separately estimated for every voxel. Inhomogeneity within a voxel and errors in the estimation of that phase as well as measurement noise may lead to violations of the positivity constraint of the observed spectra.

[*]This research was supported by an NSF CAREER Award (BES-0133804) to P.S. and the DoD Multidisciplinary University Research Initiative (MURI) program adminstered by the Office of Naval Research (N00014-01-0625). Direct correspondence to ps629@columbia.edu.

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

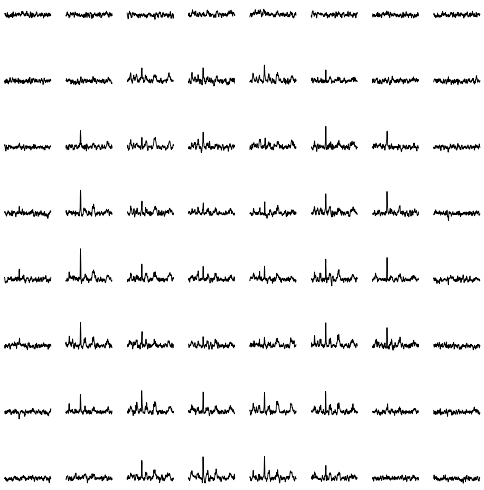


Fig. 1. Example ^{31}P CSI data from Ochs et al. [2]. Shown is an 8-by-8 voxel axial slice of spectra taken of the brain of a healthy subject. Spectra near the edges are almost exclusively noise. The complete dataset consists of 512 voxels with spectra of 369 points (bands).

Conventional NMR spectra analysis often imposes a parametric or explicit model for $\mathbf{S} = \mathbf{S}(\theta)$, and considers one voxel at a time, $\mathbf{x}_i$ to invert the linear problem $\mathbf{x}_i = \mathbf{S}(\theta)\mathbf{a}_i$ (i.e. a constrained least-squares solution). More recently, there have been efforts to simultaneously exploit the statistical structure of multi-voxel spectra to solve Equation 1 as a blind source separation (BSS) problem. For example, Nuzillard et al. [4] use second order blind identification (SOBI) [5] 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. There are many cases where the constituent spectra can be highly correlated, and thus an orthogonality or independence assumption is incorrect.

Ochs et al [2] formulate Equation 1 within a Bayesian framework to simultaneously solve for $\mathbf{A}$ and $\mathbf{S}$. Using a Markov chain Monte Carlo (MCMC) procedure, they sample the posterior space of $p(\mathbf{A}, \mathbf{S}|\mathbf{X})$ subject to the likelihood $p(\mathbf{X}|\mathbf{S}, \mathbf{A})$ (the noise distribution) and priors $p(\mathbf{A})$, $p(\mathbf{S})$. These priors include positivity and sparseness in $\mathbf{A}$ and $\mathbf{S}$, however make no assumptions about orthogonality or independence. Their results have shown good separation for highly correlated constituent spectra. However the approach is computationally expensive, given the MCMC procedure.

In this paper we describe a fast algorithm which exploits only the non-negativity of $\mathbf{A}$ and $\mathbf{S}$ for blindly separating multi-voxel CSI data. The algorithm is based on the non-negative matrix factorization (NMF) algorithm of Lee and Seung [1][6]. 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 cNMF algorithm is four orders of magnitude faster than the Bayesian MCMC approach and converges to the same solution for real-mixture, multi-voxel CSI data. In the following sections we describe the cNMF algorithm and present results for ^{31}P data, a subset of which is shown in Figure 1.

2. CONSTRAINED NON-NEGATIVE MATRIX FACTORIZATION

We begin by reviewing the formulation of the NMF algorithm of Lee and Seung. 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 given the empirical noise distribution in Figure 2), one can formulate the problem as a maximum likelihood estimation, where we minimize the negative log-likelihood,

$$\begin{aligned} \mathbf{A}_{ML}, \mathbf{S}_{ML} &= \underset{\mathbf{A,S}}{\text{argmax}}\, p(\mathbf{X}|\mathbf{A}, \mathbf{S}) \\ &= \underset{\mathbf{A,S}}{\text{argmin}}\, \|\mathbf{X} - \mathbf{AS}\|^2 \\ \text{subject to}: \quad & \mathbf{A} \geq 0,\ \mathbf{S} \geq 0. \end{aligned} \tag{3}$$

Defining F as the negative log likelihood, the gradients for $\mathbf{A}$ and $\mathbf{S}$ are given by,

$$\begin{aligned} \frac{\partial F}{\partial A_{i,m}} &= -2 * \left((XS^T)_{i,m} - (ASS^T)_{i,m}\right) \\ \frac{\partial F}{\partial S_{m,\lambda}} &= -2 * \left((A^TX)_{m,\lambda} - (A^TAS)_{m,\lambda}\right) \end{aligned} \tag{4}$$

Using the gradients we can construct the additive update rules,

$$\begin{aligned} A_{i,m} &\leftarrow A_{i,m} + \delta_{i,m}[(XS^T)_{i,m} - (ASS^T)_{i,m}] \\ S_{m,\lambda} &\leftarrow S_{m,\lambda} + \eta_{m,\lambda}[(A^TX)_{m,\lambda} - (A^TAS)_{m,\lambda}] \end{aligned} \tag{5}$$

Note that we have two free parameters, which are the step size of our updates. Lee and Seung show that by appropriately choosing the step size, δ and η ,

$$\begin{aligned} \delta_{i,m} &= \frac{A_{i,m}}{(ASS^T)_{i,m}} \\ \eta_{m,\lambda} &= \frac{S_{m,\lambda}}{(A^TAS)_{m,\lambda}} \end{aligned} \tag{6}$$

The additive update rule can be formulated as a multiplicative update rule, with $\mathbf{X} = \mathbf{AS}$ being a fixed point. In addition, they show that this optimization is equivalent to optimization over an auxiliary function, guaranteed to have

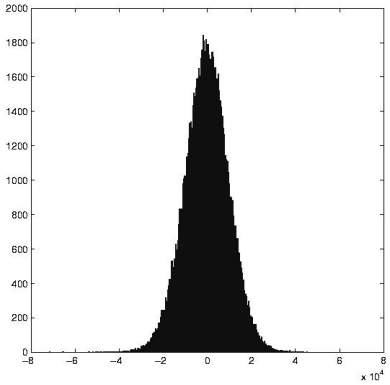


Fig. 2. Empirical noise estimate from ^{31}P CSI data. The noise distribution was computed using the 256 voxels that were determined, via PCA and visual inspection, to contain no signal.

the same minimum as Equation 3 [6]. The multiplicative update rules for $\mathbf{A}$ and $\mathbf{S}$ therefore become,

$$A_{i,m} \leftarrow A_{i,m} \frac{(XS^T)_{i,m}}{(ASS^T)_{i,m}}$$
$$S_{m,\lambda} \leftarrow S_{m,\lambda} \frac{(A^T X)_{m,\lambda}}{(A^T AS)_{m,\lambda}} \quad (7)$$

By formulating the updates as multiplicative rules we can ensure non-negative $\mathbf{A}$ and $\mathbf{S}$, given both are initialized non-negative. For our CSI experiments, initialization is done by first choosing the number of constituent spectra (sources) to recover by setting the dimensions of the two matrices $\mathbf{A}$ and $\mathbf{S}$. We then initialize the values of the matrices by constructing a non-negative random $\mathbf{A}$ and estimating $\mathbf{S}$ by solving a constrained least squares problem,

$$\underset{\mathbf{A},\mathbf{S}}{\text{argmin}} \ \|\mathbf{X} - \mathbf{AS}\|^2 \ \text{ subject to } \mathbf{S} \geq 0. \quad (8)$$

Equation 7 provides an updating rule for estimating $\mathbf{A}$ and $\mathbf{S}$ subject to the constraints of non-negativity. However one problem is that, due to noise, the observations, $\mathbf{X}$, can have negative values. Since the observations are used in updating $\mathbf{A}$ and $\mathbf{S}$, this means that a possible solution is recovered spectra with negative amplitudes. One option is to only utilize those values of $\mathbf{X}$ which are in fact non-negative. Though this solution yields correct results in the noise-free case, it does not recover the correct spectra in the case of real-mixtures (see results in Figure 3).

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 exploiting the fact resonances typically occur in only a small fraction of bands, i.e. the spectra are zero in a large fraction of bands. We will disallow small spectral magnitude 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 $\mathbf{S}$,

$$S_{i,j} = \begin{cases} S_{i,j} & S_{i,j} > \theta \\ \epsilon & S_{i,j} \leq \theta \end{cases} \quad (9)$$

where θ determines the noise floor and ϵ is some very small value. [2]

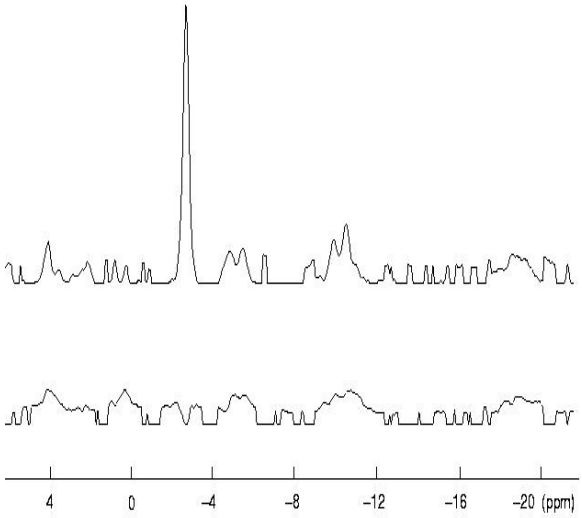


Fig. 3. Spectral separation results enforcing non-negative factorization by only using non-negative $\mathbf{X}$. Note that neither spectra is physically meaningful, though the top spectra has a mixture of features that might be indicative of muscle or brain spectra. Truth for the dataset was established by visual inspection and prior knowledge of the tissue distributions (see [2] for details). Compare to Figures 5 and 6

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. Update $\mathbf{S}$
4. Force small values of $\mathbf{S}$ to be approximately zero.

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

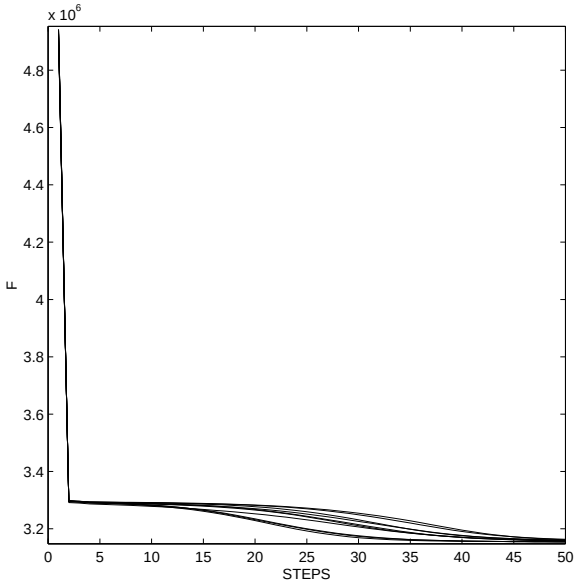


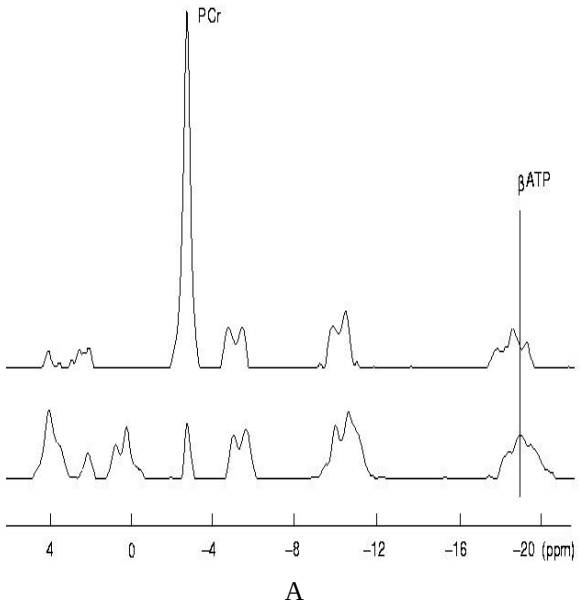
Fig. 4. Evolution of the negative log likelihood (F) of cNMF, for ten random initializations of $\mathbf{A}$ and $\mathbf{S}$. The figure is an empirical demonstration of convergence.

5. Iterate (back to 2).

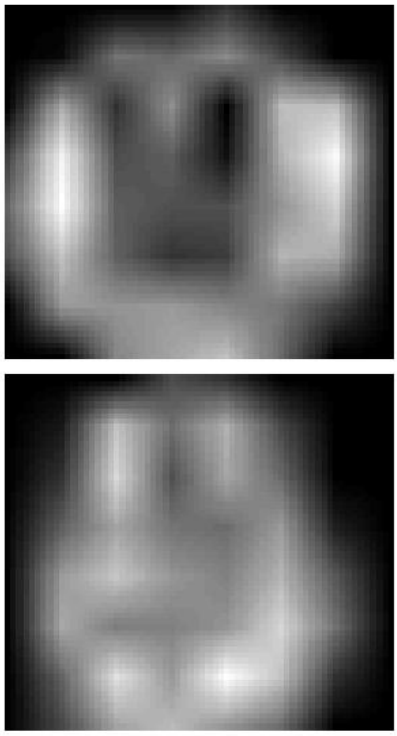
We call this the constrained non-negative matrix factorization (cNMF) algorithm.

We can develop an intuitive understanding of cNMF via geometrical considerations. The manifold of possible solutions specified by Equations 1 and 2 represent an 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 edges of the cone, lie in the positive quadrant of the L dimensional space of the observations. Positivity on $\mathbf{A}$ and the linear relation of Equation 1 state that the L dimensional points defined by the rows of the observations $\mathbf{X}$ 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 on the boundaries. By constraining small values of the vectors $\mathbf{S}$ 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. However, in practice our algorithm always converges. An example of the evolution of the negative log likelihood is shown in Figure 4. We hypothesize that the many degrees of freedom in $\mathbf{A}, \mathbf{S}$, and $\mathbf{N}$ allow alternative solutions with an equivalent goodness-of-fit and hence our heuristic does not alter convergence. More importantly, however, our solutions are more realistic in that small observed spectral magnitudes are explained as noise rather

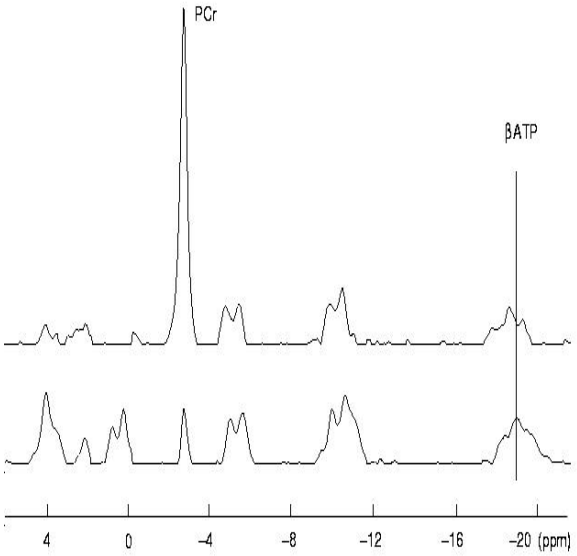


A

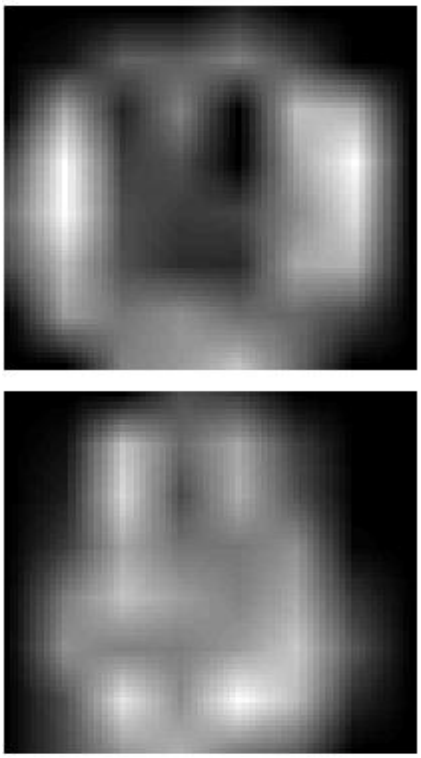


B

Fig. 5. Spectral separation results using BSD. (A) BSD recovered spectra, (top) source 1 and (bottom) source 2. Source 1 shows sharp βATP lines centered a -18.62 ppm with PCr at -2.52 ppm. This is indicative of muscle tissue. Source 2 shows βATP centered at -18.92 ppm with PCr at -2.52 ppm. This indicates brain tissue. (B) Fifth axial slice of mixing matrix showing relative concentrations of muscle and brain spectra. Note that these images are constructed by interpolating the 8-by-8 abundance values to 64-by-64.



A



B

Fig. 6. Spectral separation results using cNMF. (A) cNMF recovered spectra, (top-source 1) muscle spectra and (bottom-source 2) brain spectra. (B) Fifth axial slice of mixing matrix showing relative concentrations of muscle and brain spectra. Compare to Figure 5

than small spectral resonances.

3. RESULTS ON ^{31}P CSI BRAIN DATA

To demonstrate performance of the cNMF algorithm we compare results to the MCMC Bayesian Spectral Decomposition (BSD) method reported in Ochs et al. [2]. Such Bayesian formulations of source separation have been shown to be very flexible, in that they enable explicit definitions on the forms of the likelihood and priors.

We use the same preprocessing as described in [2]. 512 ^{31}P spectra (8-by-8-by-8 voxels) of human brain data were collected, with each spectra represented as a 369 point vector. Using principle component analysis (PCA) and visual inspection of the data, it was determined that 256 of the voxels contained signal and that there were primarily two constituent sources. Thus, the dimensionality of the problem is $M = 2, N = 256, L = 369$.

The recovered spectra using BSD is shown in Figure 5(A). One can see that the recovered spectra are highly correlated and therefore an orthogonality or independence assumption is inappropriate and would not lead to the 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 an axial slice (the fifth 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 with brain spectra largely internal to the muscle signal. The solution using BSD required 12000 seconds (1.2 GHz Intel processor).

Results using the cNMF procedure are shown in Figure 6. Note that the recovered spectra and relative concentrations distributions are nearly identical to the BSD approach. However the cNMF algorithm required only 0.48 seconds to converge (2.3 GHz Intel Processor).

4. DISCUSSION AND CONCLUSIONS

In this paper we have described an approach for recovering constituent (source) spectra in multi-voxel CSI. The approach, which we term cNMF, is 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 boundaries. We apply this algorithm to ^{31}P brain data used by Ochs et al. and report results that show cNMF is able to recover the same source spectra as BSD however in 10^{-4} the time. Both the cNMF and BSD impose non-negativity constraints, however BDS also imposes a sparsity constraint through priors on $\mathbf{A}$ and $\mathbf{S}$. Neither algorithm forces orthogonality or independence, as with many BSS algorithms [5, 7] which would lead to incorrect results in the case of CSI spectra. Finally, the speed of the cNMF algorithm enables near real-time analysis of

CSI data, potentially enabling a more directed diagnostic work-up.

5. REFERENCES

[1] D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[2] M.F. Ochs, R.S. Stoyanova, F. Arias-Mendoza, and T.R. Brown, "A new method for spectral decomposition using a bilinear Bayesian approach," *Journal of Magnetic Resonance*, vol. 137, pp. 161–176, 1999.

[3] T.R. Brown, B.M. Kincaid, and K. Ugurbil, "NMR chemical shift imaging in three dimensions," *PNAS*, vol. 79, no. 11, pp. 3523–3526, 1982.

[4] D. Nuzillard, S. Bourg, and J-M Nuzillard, "Model-free analysis of mixtures of nmr using blind source separation," *Journal of Magnetic Resonance*, vol. 133, pp. 358–363, 1998.

[5] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. SP*, vol. 45, pp. 434–444, 1997.

[6] D.D. Lee and H.S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems 13*. 2001, pp. 556–562, MIT Press.

[7] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.