Blind recovery of biochemical markers of brain cancer in MRSI

Shuyan Du[a], Xiangling Mao[b], Dikoma Shungu [b] and Paul Sajda[a]

[a]Department of Biomedical Engineering, Columbia University, New York, NY USA
[b]Department of Radiology, Mount Sinai School of Medicine, New York, NY USA

ABSTRACT

We present an algorithm for blindly recovering constituent source spectra from magnetic resonance spectroscopic imaging (MRSI) of human brain. The algorithm is based on the non-negative matrix factorization (NMF) algorithm,[1,2] extending it to include a constraint on the positivity of the amplitudes of the recovered spectra and mixing matrices. This positivity constraint enables recovery of physically meaningful spectra even in the presence of noise that causes a significant number of the observation amplitudes to be negative. The algorithm, which we call constrained non-negative matrix factorization (cNMF), does not enforce independence or sparsity, though it recovers sparse sources quite well. It 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. We incorporate the cNMF algorithm into a hierarchical decomposition framework, showing that it can be used to recover tissue-specific spectra, e.g., spectra indicative of malignant tumor. We demonstrate the hierarchical procedure on ^{1}H long echo time (TE) brain absorption spectra and conclude that the computational efficiency of the cNMF algorithm makes it well-suited for use in diagnostic work-up.

Keywords: non-negative matrix factorization (NMF), blind source separation (BSS), magnetic resonance spectroscopic imaging (MRSI), hierarchical decomposition, brain cancer

1. INTRODUCTION

Magnetic Resonance Spectroscopic Imaging (MRSI) is an imaging technique whereby high resolution magnetic resonance spectra are acquired across a volume of tissue.[3] MRSI has emerged as a powerful biochemical imaging modality to be used as adjunct to conventional structural MRI. MRSI plays an important clinical role in the noninvasive diagnostic detection, identification, treatment and monitoring of various diseases, most notably neurological disorders. It is also important for obtaining additional diagnostic information in pathologies where clinical history as well as the results of MRI studies and other neurological examinations are inconclusive. Clinicians are more frequently referring patients for metabolic "work up" with MRSI. This increased demand for clinical MRSI scans, and its complementarity with structural MRI, have brought about the need for effective computer-assist tools that can provide integrated biochemical and morphological features of biological tissue and disease processes critical for diagnostic or prognostic assessment.

Brain tumor ^{1}H MR spectra are typically characterized by the signal intensity change of several important biochemicals.[4] N-acetyl-aspartate (NAA, single resonance peak at 2.02ppm) is thought to exist primarily in neurons. The reduction or absence of an NAA spectral peak is attributed to a low density of neuronal cells in brain tumors. Choline (CHO, single resonance peak at 3.22ppm) is typically elevated in tumors compared with normal brain tissue. This is thought to be due to accelerated membrane synthesis of rapidly dividing cancer cells. Creatine (CR, singlet at 3.04 and 3.9ppm) is often reduced in tumors, but the significance of decreased CR in terms of tumor metabolism is not clear. Lactic acid (LAC, doublet at 1.33ppm) is often observed in tumor spectra, partially due to the preference for aerobic glycolysis which is prominent in highly metabolic tumors.

In MRSI, each tissue type can be viewed as having a characteristic spectral profile or set of profiles corresponding to the chemical composition of the tissue. In tumors, for example, metabolites are heterogeneously

Direct correspondence to Paul Sajda, E-mail: ps629@columbia.edu

distributed and in a given voxel multiple metabolites and tissue types may be present.[5] The observed spectra are therefore a combination of different constituent spectra. The signal measured in MRSI 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 result 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 constituent) and N rows (one for each voxel). $\mathbf{X}$ and $\mathbf{S}$ have L columns (one for each resonance band).

Since we interpret $\mathbf{A}$ as concentrations, we can assume the matrix to be non-negative. In addition, since the constituent spectra, $\mathbf{S}$, represent amplitudes of resonances, for the case of long echo time (TE) imaging (which we consider below), 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 MRSI, however measurement noise may lead to violations of the positivity constraint of the observed spectra.

We develop a fast algorithm which exploits only the non-negativity of $\mathbf{A}$ and $\mathbf{S}$ for blindly separating multi-voxel MRSI data. The algorithm is based on the non-negative matrix factorization (NMF) algorithm of Lee and Seung.[1,2] We further develop NMF, within a maximum likelihood framework, to include a function forcing negative amplitude spectral values in the recovered source spectra and abundance matrices to be zero. The method can be viewed as a subspace reduction whereby negative 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).

In this paper, we demonstrate how to apply cNMF hierarchically, with the hierarchical framework automatically removing residual water and lipids and recovering specific spectral signatures of tissue types level-by-level. The framework is very computationally efficient and can thus be used when a patient is in the MR scanner.

2. APPROACH

With a Gaussian noise assumption for MRSI data,[6] the recovery of $\mathbf{A}$ and $\mathbf{S}$ can be formulated as maximum likelihood estimation, which is equivalent to minimizing the negative log-likelihood,

$$\begin{aligned}
\mathbf{A}_{ML}, \mathbf{S}_{ML} &= \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmax}}\, p(\mathbf{X}|\mathbf{A},\mathbf{S}) \\
&= \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmax}}\, \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\|\mathbf{X}-\mathbf{AS}\|^2}{2\sigma^2}} \\
&= \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmin}}\, \|\mathbf{X}-\mathbf{AS}\|^2 \\
\text{subject to} &: \quad \mathbf{A} \geq 0,\ \mathbf{S} \geq 0.
\end{aligned} \tag{3}$$

where σ^2 is the variance of the Gaussian noise and $(\mathbf{AS})$ its mean.

The basic idea of Lee and Seung's NMF algorithm is to construct a gradient descent over the objective function of the negative log-likelihood, $F = \|\mathbf{X} - \mathbf{AS}\|$, and convert an additive update rule to a multiplicative rule. The gradients of F for $\mathbf{A}$ and $\mathbf{S}$ are given by,

$$\frac{\partial F}{\partial \mathbf{A}_{i,m}} = -2 * ((\mathbf{XS}^T)_{i,m} - (\mathbf{ASS}^T)_{i,m})$$
$$\frac{\partial F}{\partial \mathbf{S}_{m,\lambda}} = -2 * ((\mathbf{A}^T\mathbf{X})_{m,\lambda} - (\mathbf{A}^T\mathbf{AS})_{m,\lambda}) \quad (4)$$

where i, m and λ are in dices for the matrices of $\mathbf{A}$ an $\mathbf{S}$, as $i = 1, ..., N$, $m = 1, ..., M$, and $\lambda = 1, ..., L$.

Using the gradients, we can construct the additive update rules,

$$\mathbf{A}_{i,m} \leftarrow \mathbf{A}_{i,m} + \delta_{i,m}[(\mathbf{XS}^T)_{i,m} - (\mathbf{ASS}^T)_{i,m}]$$
$$\mathbf{S}_{m,\lambda} \leftarrow \mathbf{S}_{m,\lambda} + \eta_{m,\lambda}[(\mathbf{A}^T\mathbf{X})_{m,\lambda} - (\mathbf{A}^T\mathbf{AS})_{m,\lambda}] \quad (5)$$

By appropriately choosing the free parameters, δ and η, we obtain the multiplicative updating rules,

$$\mathbf{A}_{i,m} \leftarrow \mathbf{A}_{i,m}\frac{(\mathbf{XS}^T)_{i,m}}{(\mathbf{ASS}^T)_{i,m}}$$
$$\mathbf{S}_{m,\lambda} \leftarrow \mathbf{S}_{m,\lambda}\frac{(\mathbf{A}^T\mathbf{X})_{m,\lambda}}{(\mathbf{A}^T\mathbf{AS})_{m,\lambda}} \quad (6)$$

By formulating the updates as multiplicative rules in Equation 6 we 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 $\mathbf{S}$, non-negativity will not be guaranteed even if $\mathbf{A}$ and $\mathbf{S}$ are initialized as non-negative. This can lead to physically unrealistic solutions for the recovered spectra–i.e. spectra with negative amplitudes.

We modify the algorithm by first adding an initialization step for the matrices, constructing a non-negative random $\mathbf{A}$ and estimating $\mathbf{S}$ by solving a non-negatively constrained least squares problem,

$$\underset{\mathbf{A},\mathbf{S}}{\text{argmin}} \|\mathbf{X} - \mathbf{AS}\|^2 \quad \text{subject to } \mathbf{S} \geq 0. \quad (7)$$

Before solving this least squares problem, we must define the dimensionality of our matrices, namely choosing M, the number of recovered sources. An important element of the algorithm is the fact that the factorization in Equation 1 includes an explicit subspace reduction from an N dimensional space into a constrained M dimensional space. Such a compression or "bottleneck" has been shown to be useful in having the subspace capture statistical regularities in the data[7].[8] In this paper principal component analysis (PCA) is used to determine M. 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 constrained 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 constrain the space of possible solutions by exploiting the fact that resonances and abundances are non-negative, and negative spectral and abundance magnitude values are disallowed since we assume that they are due to baseline noise. We enforce this constraint by introducing a threshold on $\mathbf{S}$,

$$S_{m,\lambda} = \begin{cases} S_{m,\lambda} & S_{m,\lambda} > 0 \\ \epsilon & S_{m,\lambda} \leq 0 \end{cases} \quad (8)$$

where ϵ is some very small positive value *. We treat $\mathbf{A}$ symmetrically, using the same positivity constraint as mentioned above, to ensure $\mathbf{A}$ remains non-negative, given the possibility of negative values in $\mathbf{X}$. To summarize, the procedure for updating $\mathbf{A}$ and $\mathbf{S}$ is,

*Note that the spectral amplitudes cannot be set exactly to zero given the update rules for $\mathbf{A}$ and $\mathbf{S}$. We therefore use $\epsilon = 2.2204 \times 10^{-16}$ (the value of floating point relative accuracy used by MATLAB 6.5)

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

We call this 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 edges of the cone, lie in the positive quadrant of the L dimensional points defined by the rows of the observations $\mathbf{X}$ which 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 negative values of the vectors in $\mathbf{S}$ 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.

We further extend the methodology by applying cNMF hierarchically, with the spectral recovery and subspace reduction constraining which observations are used in the next hierarchical level of recovery. A flow diagram, shown in Figure 1, illustrates the basic idea. The approach enables an automatic "drilling down" of the source space, ultimately increasing the specificity of the recovered spectra based on a natural, and physically meaningful, hierarchy (e.g. head, brain, tumor, malignancy, etc).

In the hierarchical decomposition, each level of analysis uses spectral images at the full spatial resolution, effectively integrating prior knowledge about spatial distributions of brain-like spectral signatures into the cNMF source spectra recovery. Principal component analysis (PCA) is used to preprocess the spectra, much in the same way as in Ochs et al.,[9] to align spectra, select voxels containing signals and estimate the number of sources.[10] cNMF is then applied to recover constituent spectra and the corresponding concentration spatial distributions at this level are analyzed. The spectrum corresponding to the largest abundances/concentrations in central parts of the spatial distribution map is taken as brain spectrum. The spatial distribution map corresponding to the selected brain spectrum is thresholded to automatically generate a mask for the next hierarchical level. This mask effectively selects voxels for the next level of processing. The masking can be seen as a form of data selection for improving the specificity of the separation, with voxels removed by the mask considered to be signals not related to brain tissue (e.g., water, lipids, etc).

The selected voxels in the mask are used in the next hierarchical level of cNMF source recovery. Again PCA is used to determine the number of constituent spectra at this level and cNMF is used to recover source spectra and their corresponding spatial distributions. Due to the masking of the previous level, the observations are more specific to brain. Again the brain-like spectrum is selected and a mask is generated from thresholding based on the corresponding spatial distribution map. This procedure continues to the next with the hierarchical level until the desired number of levels (and specificity) is reached.

3. EXPERIMENTAL RESULTS

We demonstrate spectral recovery results for two cases. Each case consists of 32-by-32 voxel axial view ^{1}H long echo time (TE) MRSI human brain data with each voxel having a spectrum of 1024 points. The ^{1}H MRSI data were recorded on a 1.5T GE Signa Horizon 5.x MR system using the multi-slice MRSI sequence of Duyn et al. with TE/TR=280/2300ms, field-of-view 240mm, 32x32 phase-encoding steps with circular k-space sampling, receiver bandwidth 1KHz and 256 sample points along the signal acquisition (time) domain. The raw data were

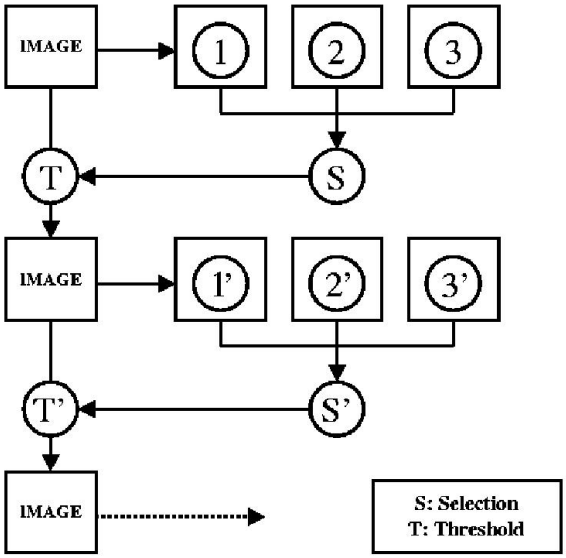

Figure 1. Hierarchical blind recovery. At the first level an image (data volume) having a given resolution is used as observations for recovering source spectra (e.g., 1, 2 and 3 shown in the figure). The source spectra are analyzed given prior information about the spectral signatures (i.e. spectra for brain versus muscle) and their spatial distributions (location of the head/brain in the image and its approximate shape). The result is the construction of a spatial mask, used for selecting those voxels which will be processed at the next level. At the second level, only those voxels passing through the mask are used as observations and cNMF is reapplied to recover new source spectra (e.g., 1',2' and 3'). This process is iterative and continues until the desired specificity is reached. The hierarchy is constructed using images having the full resolution at each level.

sorted by slice, zero-padded to 1024 sample points and then 3D Fourier transformed to yield 1024 frequency domain spectral voxels. Absorption spectra are used in our analysis. For both cases, a 2-level hierarchy is used. PCA analysis estimates 3-4 sources at the first level of the hierarchy. This is consistent with the head consisting of 3-4 primary tissue types: brain, lipids, and water. PCA selected spectra for signal containing voxels are input into cNMF at the first level of hierarchy with the four and three constituent sources recovered respectively shown in Figure 2(a) and 3(a). Analysis of both the spectra and spatial distributions clearly shows recovery of brain tissue, residual lipids and water. The results at the first level of hierarchical processing are then used to automatically generate a mask for the second level of the hierarchy– i.e. source recovery is constrained to include only those observations consistent with brain tissue. Recovered spectra and their corresponding spatial distributions for the second level of hierarchical processing are shown in Figure 2(b) and 3(b). From Figure 2(b) we see the three recovered spectra that are consistent with normal brain tissue, and malignant tumor tissue and residual lipid. The tumor spectrum (column 2) has spectral markers indicative of a high grade tumor, i.e., highly elevated CHO. From Figure 3(b) we see the recovered spectra that are consistent with normal brain tissue and malignant tumor tissue and residual water. The tumor spectrum (column 3) has markers for a low grade tumor, i.e., moderately elevated CHO and significantly decreased NAA. The spatial distribution maps of these constituent spectra are also shown, which can aid diagnosis when combined with structural MRI. Improvement in specificity and tumor delineation and extent are seen in both cases, and other datasets have yielded similar results. The cNMF requires only 20.6 seconds (2.3 GHz Intel Processor) to get the results shown in this paper.

4. CONCLUSION

In this paper, we present a fast algorithm for blindly recovering source spectra in MRSI. The algorithm is an extension of Lee and Seung's NMF, where we include an explicit positivity constraint in the updating equations to recover physically meaningful spectra when there are negative observations due to noise. We term this the

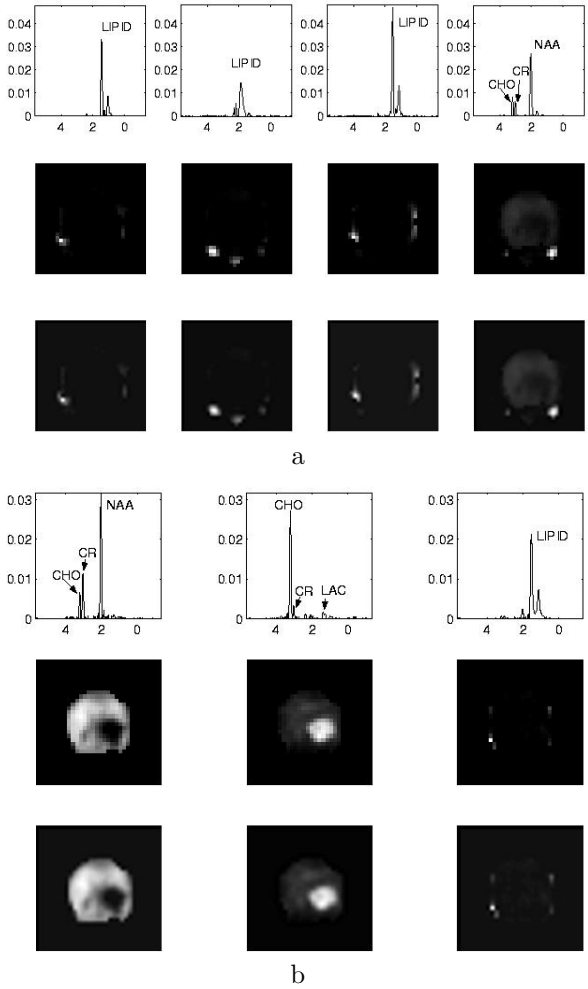

Figure 2. (a) First level of hierarchy, separation results. First row is the recovered spectra, second row is the spatial distribution of each recovered spectra and the third row is an up-sampled version of the spatial distribution (for visualization). Clear in the spectra are biochemical markers for brain (4th column) with peaks for CHO, CR and NAA. Note that the corresponding spatial mapping of these markers maps to the region of the brain. (b) Second level of hierarchy separation results. Column 1 spectrum is indicative of normal brain tissue: low CHO, high CR and high NAA; Column 2 spectrum indicates malignant tumor tissue: high CHO, low CR, low NAA and LAC peak appearance; Column 3 spectrum indicates residual lipid.

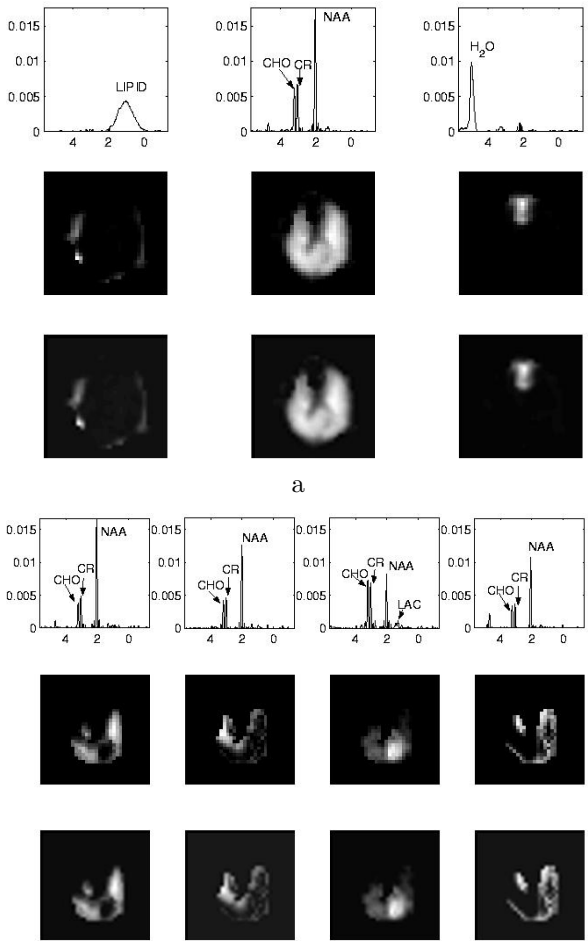

Figure 3. (a) First level of hierarchy, separation results. First row is the recovered spectra, second row is the spatial distribution of each recovered spectra and the third row is an up-sampled version of the spatial distribution. Clear in the spectra are biochemical markers for brain (2nd column) with peaks for CHO, CR and NAA. Note that the corresponding spatial distribution of these markers maps to the region of the brain. Also recovered are the sinuses (column 3) indicated by a water peak as well as a physically realistic spatial distribution for lipids (columns 1). (b) Second level of hierarchy separation results. Column 1, 2, 4 spectra are indicative of normal brain tissue: low CHO, high CR and high NAA. Note the small amount of residual water (4th column at 4.7 ppm); Column 3 spectrum indicates cancerous tissue: high CHO, low NAA and appearance of LAC peak.

constrained NMF algorithm (cNMF), which converges very rapidly and results in stable solutions for multiple initial conditions. cNMF 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 used, with no explicit sparsity, independence or orthogonality assumptions, though the algorithm recovers sparse sources quite well. We demonstrate how cNMF can be incorporated into hierarchical decomposition framework, enabling a "drilling down" into source space, with increasing specificity of the recovered spectra based on a natural and physically meaningful hierarchy (e.g., head, brain, tumor, etc) on ^{1}H brain data and report results that show cNMF is able to rapidly recover appropriate source spectra and differentiate between high and low grade tumors. The speed of cNMF algorithm enables rapid analysis of MRSI data, potentially enabling a more directed diagnostic work-up. The hierarchical decomposition algorithm is nearly automatic, except for threshold selection (the threshold on eigenvalues of PCA for selecting number of sources and the threshold for constructing the spatial mask) and selection of the number of levels in the hierarchy.

ACKNOWLEDGMENTS

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). Direct correspondence to ps629@columbia.edu.

REFERENCES

1. D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature* **401**, pp. 788–791, 1999.
2. D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems 13*, pp. 556–562, MIT Press, 2001.
3. T. Brown, B. Kincaid, and K. Ugurbil, "NMR chemical shift imaging in three dimensions," *PNAS* **79**(11), pp. 3523–3526, 1982.
4. F. A. Howe and K. S. Opstad, "^{1}H MR spectroscopy of brain tumours and masses," *NMR in Biomedicine* **16**, pp. 123–131, 2003.
5. S. Furuya, S. Naruse, M. Ide, H. Morishita, O. Kizu, S. Ueda, and T. Maeda, "Evaluation of metabolic heterogeneity in brain tumors using ^{1}H-chemical shift imaging method," *NMR in Biomedicine* **10**, pp. 25–30, 1997.
6. P. Sajda, S. Du, T. R. Brown, L. C. 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.
7. N. Tishby, F. Pereira, and W. Bialek, "The information bottleneck method," in *Proc. of 37th Annual Allerton Conference on Communication, Control and Computing*, pp. 368–377, 1999.
8. N. Friedman, O. Mosenzon, N. Slonim, and N. Tishby, "Multivariate information bottleneck," in *Proc. of 17th Conf. on Uncertainty in Artificial Intelligence (UAI)*, pp. 152–161, 2001.
9. 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* **137**, pp. 161–176, 1999.
10. R. Stoyanova and T. R. Brown, "NMR spectral quantitation by principal component analysis: III. a generalized procedure for determination of lineshape variations," *J. Magn. Reson.* **154**, pp. 163–175, 2002.