Nonnegative Matrix Factorization for Rapid Recovery of Constituent Spectra in Magnetic Resonance Chemical Shift Imaging of the Brain

Paul Sajda*, *Member, IEEE*, Shuyan Du, Truman R. Brown, Radka Stoyanova, Dikoma C. Shungu, Xiangling Mao, and Lucas C. Parra, *Member, IEEE*

Abstract—We present an algorithm for blindly recovering constituent source spectra from magnetic resonance (MR) chemical shift imaging (CSI) of the human brain. The algorithm, which we call constrained nonnegative matrix factorization (cNMF), does not enforce independence or sparsity, instead only requiring the source and mixing matrices to be nonnegative. It is based on the nonnegative matrix factorization (NMF) algorithm, extending it to include a constraint on the positivity of the amplitudes of the recovered spectra. This 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. We demonstrate and characterize the algorithm's performance using ^{31}P volumetric brain data, comparing the results with two different blind source separation methods: Bayesian spectral decomposition (BSD) and nonnegative sparse coding (NNSC). We then incorporate the cNMF algorithm into a hierarchical decomposition framework, showing that it can be used to recover tissue-specific spectra given a processing hierarchy that proceeds coarse-to-fine. We demonstrate the hierarchical procedure on ^{1}H brain data and conclude that the computational efficiency of the algorithm makes it well-suited for use in diagnostic work-up.

Index Terms—Blind source separation (BSS), chemical shift imaging (CSI), hierarchical decomposition, magnetic resonance (MR), magnetic resonance spectroscopy (MRS), nonnegative matrix factorization (NMF).

*P. Sajda is with the Laboratory of Intelligent Imaging and Neural Computing, Department of Biomedical Engineering, Columbia University, 351 Engineering Terrace Building, Mail Code 8904, 1210 Amsterdam Ave., New York, NY 10027 USA (e-mail: ps629@columbia.edu).

S. Du is with the Laboratory of Intelligent Imaging and Neural Computing, Department of Biomedical Engineering, Columbia University, New York, NY 10027 USA (e-mail: sd2012@columbia.edu).

T. R. Brown is with the Hatch Magnetic Resonance Research Laboratory, Neurological Institute, Department of Radiology and Biomedical Engineering, Columbia University, New York, NY 10032 USA (e-mail: trb11@columbia.edu).

R. Stoyanova is with the Department of Biostatistics, Fox Chase Cancer Center, Philadelphia, PA 19111 USA (e-mail: stoyan@fccc.edu).

D. C. Shungu and X. Mao are with the Citigroup Biomedical Imaging Center, Weill Medical College of Cornell University, New York, NY 10021 USA (e-mail: dcs7001@med.cornell.edu; xim2004@med.cornell.edu).

L. C. Parra is with the Department of Biomedical Engineering, The City College of New York, New York, NY 10031 USA (e-mail: parra@ccny.cuny.edu).

Digital Object Identifier 10.1109/TMI.2004.834626

I. INTRODUCTION

ACCURATE and highly specific diagnosis of intracranial tumors is critical for minimizing the morbidity and mortality of brain cancer. Imaging plays an important role in such diagnosis since there is a tradeoff between overaggressive resection of tumor and maintenance of critical brain function. Although biopsy is still considered the gold-standard for diagnosis and grading [1], it comes with significant risks including a 1.7% mortality rate [2] and 8% chance of other serious complications [3]. Structural magnetic resonance imaging (MRI) has played an important role as a noninvasive modality for determining tumor extent, grading, and monitoring treatment and tumor reoccurrence. The specificity of MRI for initial diagnosis, however, is highly variable (between 30% and 90% [3], [4]) and depends on tumor type. Magnetic resonance spectroscopy (MRS) is a comparatively new diagnostic method that was adopted into clinical practice approximately ten years after MRI. Automated sequence development has paved the way for increased MRS application in clinical radiology, with more reliable and user-friendly hardware, improved data analysis, and confidence of interpretation [5]–[7].

Chemical shift imaging (CSI) is an imaging modality whereby high-resolution MR spectra are acquired across a volume of tissue [8]. *In vivo* CSI can measure different biochemical markers by tuning to particular nuclear resonance frequencies, thus providing precise characterization of tissue and/or a means for optimizing the signal-to-noise ratio (SNR). The more common nuclei include ^{1}H (proton), ^{13}C (carbon), ^{19}F (fluorine), and ^{31}P (phosphorus). CSI allows for the noninvasive characterization and quantification of molecular markers with clinical utility for improving detection, identification, and treatment for a variety of diseases, most notably brain cancers [9]–[11].

Tumor ^{1}H MR spectra are characterized by the signal intensity change of several important brain tissue types and metabolites. Signal intensity of N-acetyl-aspartate (NAA, single resonance peak at 2.02 ppm) is reduced in tumors, with NAA thought to exist primarily in viable neurons [12]. Choline (CHO, singlet at 3.22 ppm) is typically elevated compared with normal brain tissue and is thought to be due to accelerated membrane synthesis of rapidly dividing cancer cells [13], [14]. Creatine (CR, singlet at 3.04 and 3.9 ppm) is often reduced in tumors, but the significance of decreased CR in terms of tumor metabolism is not clear. Lactic acid (LAC, doublet at

1.33 ppm) is often observed in tumor spectra [10], [15]–[17], which is partially due to the preference for anaerobic glycolysis [18] often prominent in highly metabolic tumors [5]. Combined with MRI, CSI can provide an integrated biochemical and morphological view of biological tissue and disease processes.

The interpretation of CSI data is quite challenging: a typical dataset consists of hundreds of spectra, typically having low SNR with peaks that are numerous and overlapping. Sophisticated and accurate approaches for quantitation of resonance peaks in a single spectrum (either the time or frequency domain) are widely available in the NMR community (for reviews of these methods refer to [19], [20]). However, when applied to CSI data, these methods produce a long list of peak-areas, which have to be additionally analyzed for successful interpretation. Moreover, the spectra on CSI datasets are closely related and this has created the need for approaches that can analyze an entire set simultaneously, taking advantage of the relationships among the spectra to improve the quality of the analysis. Approaches that are able to do this are particularly useful for spectra with low SNR as they utilize the collective power of the data. Here, we present such a method that identifies the shapes of the underlying constituent source spectra and their contribution to each voxel in the CSI data.

In CSI, 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 distributed and in a given voxel multiple metabolites and tissue types may be present [17], [21], [22]. The observed spectra are, 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 spectral amplitudes of the different coherent resonators 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, one can explain the observed spectra $\mathbf{X}$ as

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N} \qquad (1)$$

where the columns in $\mathbf{A}$ represent the concentration, or abundance, of the constituent materials 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). Fig. 1 shows an example of an axial slice taken from a three-dimensional (3-D) ^{31}P CSI acquisition of human head.

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

$$\mathbf{A} \geq 0 \text{ and } \mathbf{S} \geq 0. \qquad (2)$$

In CSI, however, noise may lead to violations of the positivity constraint of the observed absorption spectra. Thus, the observed spectra, $\mathbf{X}$, may have negative values.

---

[1]In this paper, we do not consider cases where spectra may have negative amplitudes, for example in the case of J-modulation.
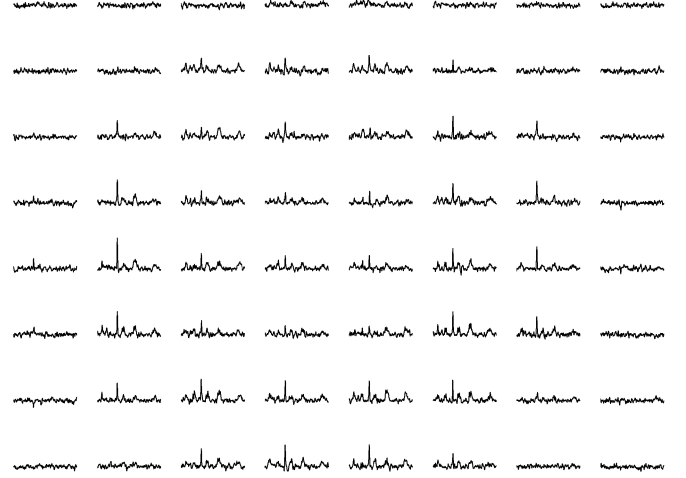

Fig. 1.   Example ^{31}P CSI data from Ochs *et al.* [27]. Shown is an 8×8 voxel axial slice of spectra taken of the brain from a healthy subject. Spectra near the edges are almost exclusively noise. The complete dataset consists of 512 voxels with spectra of 369 resonance bands. The chemical shift range is approximately 5 to -20 ppm.

Conventional MR spectra analysis often imposes an explicit or parametric model for $\mathbf{S} = \mathbf{S}(\theta)$, and considers one voxel at a time, $\mathbf{x}_i$, to invert the linear problem $\mathbf{x}_i = \mathbf{S}(\theta)\mathbf{a}_i$ with the constraint that $\mathbf{a}_i > 0$ (i.e., a constrained least-squares problem [23], [24]). Here, i is the index of matrices $\mathbf{X}$ and $\mathbf{A}$, representing the ith voxel, and $\mathbf{x}_i$ and $\mathbf{a}_i$ are, respectively, the observation vector and constituent abundance vector of the ith voxel. More recently, there have been efforts to simultaneously exploit the statistical structure of multivoxel spectra to solve (1) as a blind source separation (BSS) problem. For example, Nuzillard *et al.* [25] use second-order blind identification (SOBI) [26] to separate ^{13}C spectra. Ochs *et al.* [27] formulate (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, they 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 that exploits only the nonnegativity of $\mathbf{A}$ and $\mathbf{S}$ for blindly separating multivoxel CSI data. The algorithm is based on the nonnegative matrix factorization (NMF) algorithm of Lee and Seung [28], [29]. Due to noise, negative components may appear in the observations, and therefore we further develop the NMF approach, within a maximum likelihood framework, to include a positivity constraint forcing negative amplitude spectral values in the recovered sources and abundance distributions to be approximately zero. The method can be viewed as a subspace reduction with negative values of the constituent sources forced to zero, i.e., forced to the edges of a polygonal conic subspace spanned by the constituent spectra. We term this algorithm constrained nonnegative matrix factorization (cNMF). Although no sparsity constraint is applied, the algorithm can recover sparse spectra quite well. The cNMF algorithm is four orders of magnitude

faster than the Bayesian MCMC approach and converges to the same solution for absorption spectrum mixture, multivoxel CSI data. We further develop cNMF and show how it can be applied hierarchically, automatically removing residual water and lipids and recovering specific spectral signatures of tissue types using a coarse-to-fine decomposition strategy. Computationally, this framework is very efficient and can thus be used when a patient is in the MR scanner.

In the following, we describe the cNMF algorithm and present results for ^{31}P and ^{1}H multivoxel CSI datasets. Section II begins by formulating NMF within a maximum likelihood framework and describes how we have modified the NMF algorithm of Lee and Seung to introduce a constraint for ensuring positivity of the sources given noisy observations that may be negative. We also describe the role dimensionality reduction plays in recovering the sources. Section III describes how we incorporate cNMF into a hierarchical framework for systematically increasing the specificity of recovered spectral sources. Section IV then presents results, including a characterization of the algorithm's performance as a function of the SNR level, as well as examples of how the hierarchical framework is able to recover spectra indicative of malignant tumors.

II. CONSTRAINED NONNEGATIVE MATRIX FACTORIZATION

We begin by reviewing the formulation of the NMF algorithm of Lee and Seung [28]. The basic idea of the algorithm is to construct a gradient descent over an objective function that optimizes $\mathbf{A}$ and $\mathbf{S}$ and, by appropriately choosing gradient step sizes, to convert an additive update to a multiplicative one. For example, with $\mathbf{N}$ modeled as Gaussian noise (a reasonable assumption for CSI data), one can formulate the problem of recovering $\mathbf{A}$ and $\mathbf{S}$ as a maximum likelihood estimation

$$\mathbf{A}_{ML}, \mathbf{S}_{ML} = \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmax}}\, p(\mathbf{X}|\mathbf{A},\mathbf{S})$$
$$= \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmax}}\, \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\|\mathbf{X}-\mathbf{AS}\|^2}{2\sigma^2}}$$

$$\text{subject to: } \mathbf{A} \geq 0, \quad \mathbf{S} \geq 0 \quad (3)$$

where σ is the standard deviation of the Gaussian noise and $(\mathbf{AS})$ its mean.

Maximizing the likelihood is equivalent to minimizing the negative log-likelihood, and (3) can be written as

$$\mathbf{A}_{ML}, \mathbf{S}_{ML} = \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmin}}\, (-\log p(\mathbf{X}|\mathbf{A},\mathbf{S}))$$
$$= \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmin}}\, \|\mathbf{X}-\mathbf{AS}\|^2$$

$$\text{subject to: } \mathbf{A} \geq 0, \quad \mathbf{S} \geq 0. \quad (4)$$

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

$$\frac{\partial F}{\partial A_{i,m}} = -2 * \left((\mathbf{XS}^T)_{i,m} - (\mathbf{ASS}^T)_{i,m}\right)$$

$$\frac{\partial F}{\partial S_{m,\lambda}} = -2 * \left((\mathbf{A}^T\mathbf{X})_{m,\lambda} - (\mathbf{A}^T\mathbf{AS})_{m,\lambda}\right) \quad (5)$$

where i, m, and λ are indexes for the matrices of $\mathbf{A}$ and $\mathbf{S}$, as $i = 1, \ldots, N$, $m = 1, \ldots, M$, and $\lambda = 1, \ldots, L$. Using the gradients, we can construct the additive update rules

$$A_{i,m} \leftarrow A_{i,m} + \delta_{i,m} \left[(\mathbf{XS}^T)_{i,m} - (\mathbf{ASS}^T)_{i,m}\right]$$
$$S_{m,\lambda} \leftarrow S_{m,\lambda} + \eta_{m,\lambda} \left[(\mathbf{A}^T\mathbf{X})_{m,\lambda} - (\mathbf{A}^T\mathbf{AS})_{m,\lambda}\right]. \quad (6)$$

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 $\delta_{i,m} = A_{i,m}/(\mathbf{ASS}^T)_{i,m}$ and $\eta_{m,\lambda} = S_{m,\lambda}/(\mathbf{A}^T\mathbf{AS})_{m,\lambda}$, the additive update rule can be formulated as a multiplicative update rule, with $\mathbf{X} = \mathbf{AS}$ being a fixed point. The multiplicative update rules for $\mathbf{A}$ and $\mathbf{S}$, therefore, become

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

where Lee and Seung prove, using an auxiliary function, convergence of these update rules to a least a locally optimal maximization likelihood solution [29]. By formulating the updates as multiplicative rules in (7), we can ensure nonnegative $\mathbf{A}$ and $\mathbf{S}$, given both are initialized nonnegative and the observations, $\mathbf{X}$, are nonnegative. However, one problem is that, due to Gaussian noise in CSI data, the observations can have negative values. Since all observations are used in updating $\mathbf{A}$ and $\mathbf{S}$, nonnegativity will not be guaranteed even if they are initialized as nonnegative. 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 nonnegative random $\mathbf{A}$ and estimating $\mathbf{S}$ by solving a nonnegatively constrained least-squares problem

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

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 (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 [30], [31]. Except for the positivity constraints, the decomposition is completely arbitrary within that M-dimensional space. However, spectra and concentrations are nonnegative and so the M-dimensional degrees of freedom within that subspace are constrained by $M(N+L)$ linear boundary constraints (2). This is the portion of the space that corresponds to realistic solutions of the factorization. We disallow negative spectral magnitude values and instead assume that they are due to baseline noise. This is enforced by introducing a threshold constraint on $\mathbf{S}$

$$S_{m,\lambda} = \begin{cases} S_{m,\lambda} & S_{m,\lambda} > 0 \\ \epsilon & S_{m,\lambda} \leq 0 \end{cases} \quad (9)$$

where ϵ is some very small positive value.[2] We treat $\mathbf{A}$ symmetrically, using the same positivity constraint as mentioned above. This ensures $\mathbf{A}$ and $\mathbf{S}$ remain nonnegative, given the possibility of negative values in $\mathbf{X}$. To summarize, the procedure for updating $\mathbf{A}$ and $\mathbf{S}$ is as follows.

   1) Initialize: Choose dimensions of $\mathbf{A}$ and $\mathbf{S}$ (i.e., M) and initialize with nonnegative values (e.g., random $\mathbf{A}$ and constrained least-squares for $\mathbf{S}$).

   2) Update $\mathbf{A}$ [according to (7)].

   3) Force negative values of $\mathbf{A}$ to be approximately zero.

   4) Update $\mathbf{S}$ [according to (7)].

   5) Force negative values of $\mathbf{S}$ to be approximately zero.

   6) Iterate (back to 2).

We call this the constrained nonnegative matrix factorization (cNMF) algorithm.

An intuitive understanding of cNMF via geometrical considerations can be developed. The manifold of possible solutions specified by (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 hyper-quadrant of the L-dimensional space. Points defined by the rows of the observations $\mathbf{X}$ 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 approximately 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.

Based on Lee and Seung's convergence proof for NMF [29], Constraint (9) will not change the convergence characteristics of the cNMF algorithm, although the speed of convergence may be reduced. However, a least-squares initialization is useful for improving the speed of convergence. An empirical demonstration of convergence for cNMF is given in [32].

III. HIERARCHICAL RECOVERY OF SOURCE SPECTRA

A. Hierarchical cNMF Decomposition

The cNMF algorithm can be applied 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 Fig. 2, illustrates the basic idea. The approach enables a "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, proliferation/grade, etc.). We consider two types of hierarchical decomposition, single-resolution and multiresolution, described below.

---

[2]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)
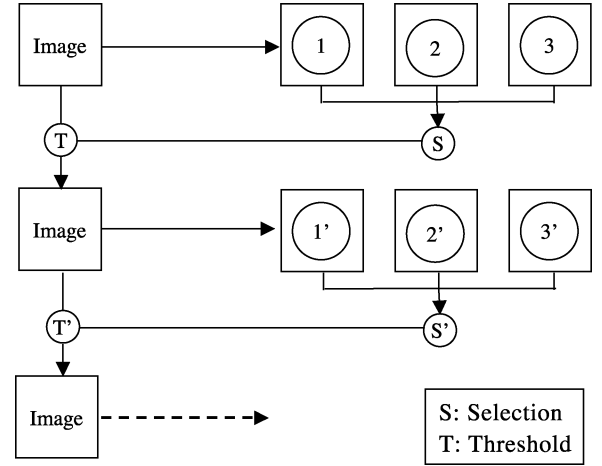

Fig. 2. Hierarchical blind recovery. At the first level, an image (data volume) having a given resolution is used as observations for recovering source spectra (e.g., 1, 2, and 3 shown in the figure). The source spectra are analyzed given prior information about the spectral signatures (e.g., spectra for brain versus muscle) and their spatial distributions (e.g., 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. Note that the hierarchy can be constructed using images having the full resolution at each level, or constructing the hierarchy on a pyramid, thus proceeding from coarse-to-fine resolution.

B. Single-Resolution Hierarchical Decomposition

In single-resolution hierarchical decomposition, each level of analysis uses spectral images at the same 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.* [27], to align spectra, select voxels containing signal, and estimate the number of sources (see [33] for details). cNMF is then applied to recover constituent spectra, and the corresponding spatial distributions at this level are analyzed. The spectrum corresponding to the largest abundance/concentration in central parts of the spatial distribution map is taken as brain. The spatial distribution map corresponding to the selected brain spectrum is thresholded to automatically generate a mask for the next hierarchical level. This mask selects voxels for the next level of processing. The masking can be seen as a form of active data selection [34] for improving the specificity of the separation, with voxels removed by the mask considered to be signals not containing significant 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 application of the mask in the previous level, the observations are more specific to brain. Again, the brain-like spectrum is selected and a mask is generated based on the corresponding spatial distribution map. This procedure continues until the desired number of levels (and specificity) is reached. The highest level in the hierarchy returns the recovered spectra and spatial distribution maps.
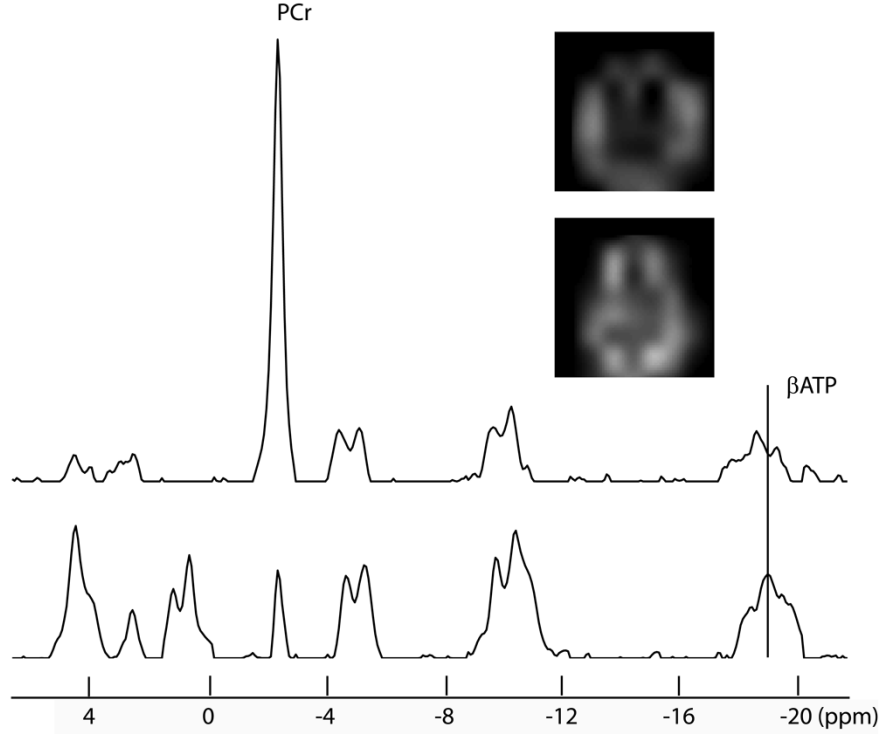
Fig. 3. ^{31}P spectral separation results using cNMF. (a) cNMF recovered spectra, (top) source 1 and (bottom) source 2. Source 1 shows sharp βATP peaks centered at -18.62 ppm with PCr at -2.52 ppm, indicative of muscle tissue. Source 2 shows βATP centered at -18.92 ppm with PCr at -2.52 ppm, which is indicative of brain tissue. Inset shows the spatial distribution of the recovered spectra for the fifth slice as axial relative concentrations of muscle and brain spectra. Note that these images are upsampled from 8×8 to 64×64 for visualization purposes.

C. Multiresolution Hierarchical Decomposition

In a multiresolution hierarchical decomposition, each level of analysis uses different spatial resolution spectral images for cNMF source recovery. PCA is again used to align the spectra and estimate, level-by-level, the number of source components. For example, when a three-level hierarchy model is used with the original spectral image having a resolution of $K \times K$, the next level is constructed to be $(K/2) \times (K/2)$ by averaging four neighboring spectra and subsampling in each dimension by two. The third level in the hierarchy is constructed from the second level using the same procedure, resulting in a $(K/4) \times (K/4)$ spectral image.

Starting with the lowest (coarsest) hierarchical level, all $(K/4) \times (K/4)$ spectra are input into cNMF, with the number of components having already been determined via PCA. Different from the single resolution case, PCA is not used to select signal voxels at the coarsest resolution; rather, all voxels at the coarsest resolution are used. As in the single resolution hierarchical case, recovered spectra are analyzed together with their spatial distributions to construct a mask consisting of brain-like spectra. This mask, constructed from the coarsest resolution, is up-sampled to $(K/2) \times (K/2)$ and applied to the $(K/2) \times (K/2)$ image in the hierarchy. Only voxels that pass through the mask are considered to contain brain tissue and used to estimate $\mathbf{A}$ and $\mathbf{S}$ at the next level. The process proceeds to the highest (finest) resolution.

Since the multiresolution hierarchical decomposition averages over voxels in computing the lower resolution spectral images, features may be filtered (i.e., blurred) which are specific to spectra of interest. Thus, some care must be taken to set the number of levels of the hierarchy (in practice, we have used three). However, the advantage of the multiresolution hierarchical decomposition is a significant reduction in computational cost (discussed in Appendix I).

IV. EXPERIMENTAL RESULTS

We first demonstrate the cNMF algorithm using ^{31}P CSI data of human brain, comparing results to two other decomposition methods, both in terms of fidelity of recovery and speed. We then demonstrate the hierarchical cNMF framework for ^{1}H CSI data, showing how the approach enables an increased specificity of tissue type to be recovered.

A. ^{31}P Chemical Shift Images of Human Brain

In this experiment, ^{31}P spectra of human head data from Ochs *et al.* [27] are used. This dataset comprises 256 ^{1}H decoupled ^{31}P spectra of typical peak SNR of approximately 6, which are selected by choosing axial slices with signals from 512 spectra ($8 \times 8 \times 8$ voxels) obtained by spatial and temporal fast Fourier transforms. Data were acquired from an entire human head using a quadrature dual tuned head coil on a 1.5-T Siemens Magnetom clinical imager/spectrometer operated with sequence details as follows: 250-μs radio-frequency pulse was followed by triangular 2.1-ms phase encode gradients applied in each of the three orientations and acquisition of 512 points (256 ms). CSI datasets were acquired in $8 \times 8 \times 8$ arrays (512 points in time) with six acquisitions per phase encoding step and $\mathrm{TR} = 1$ s, field-of-view 240 mm.
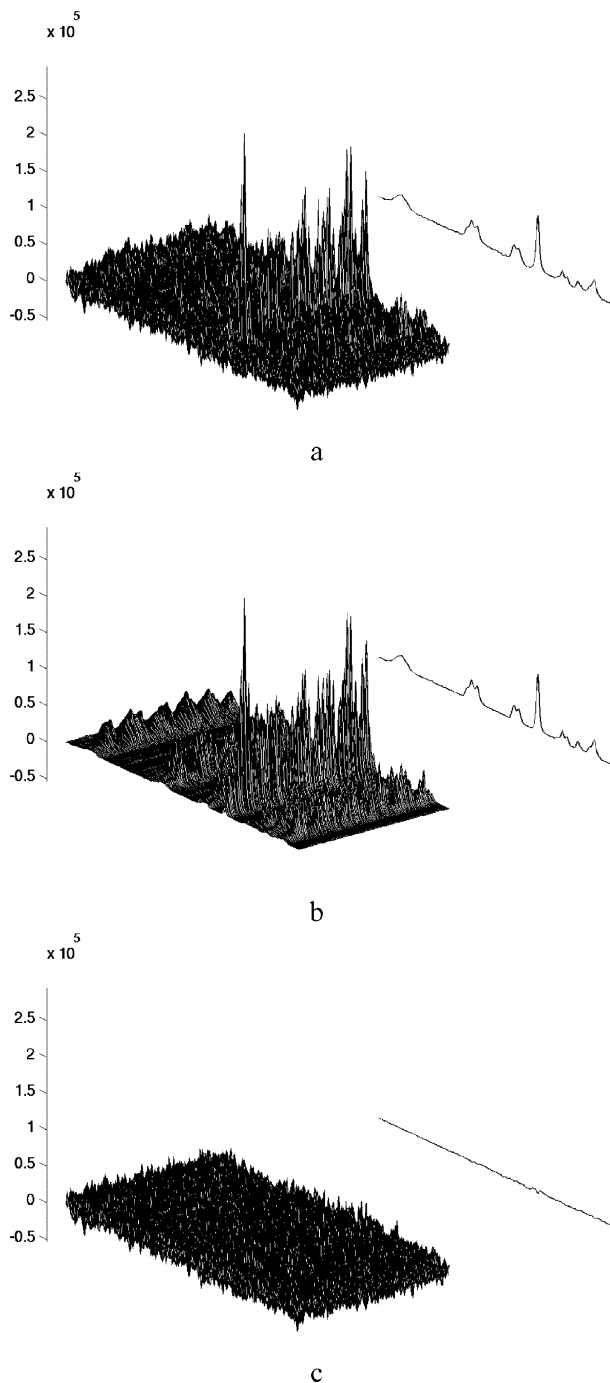
Fig. 4. Plots of full spectral dataset (observations), cNMF reconstructions, and residual for ^{31}P head data: each plot consists of 256 spectra having 369 resonance bands, each shown from left to right, together with an average across all the spectra (at the far right). (a) The observations, with low SNR and large variations apparent. (b) Reconstruction from the solution shown in Fig. 3. (c) Residuals. Compare to [27].

Fig. 5. ^{31}P spectral separation results using NNSC. Sparse control parameter $\lambda = 1$, learning rate $\mu = 1$. (a) Results of all observation $\mathbf{X}$ are input into the algorithm iteration (NNSC 1). (b) Results of only positive components of $\mathbf{X}$ are input into the algorithm (NNSC 2).

After preprocessing with PCA (described above), the matrices have dimensions: $M = 2$ source components, $N = 256$ voxels, and $L = 369$ resonance bands. After running cNMF for 100 iterations, the negative log-likelihood is minimized and the recovered spectra are shown in Fig. 3. It is clear that the recovered spectra are highly correlated and, therefore, any orthogonality assumption is inappropriate and would not lead to physically meaningful results. The results are consis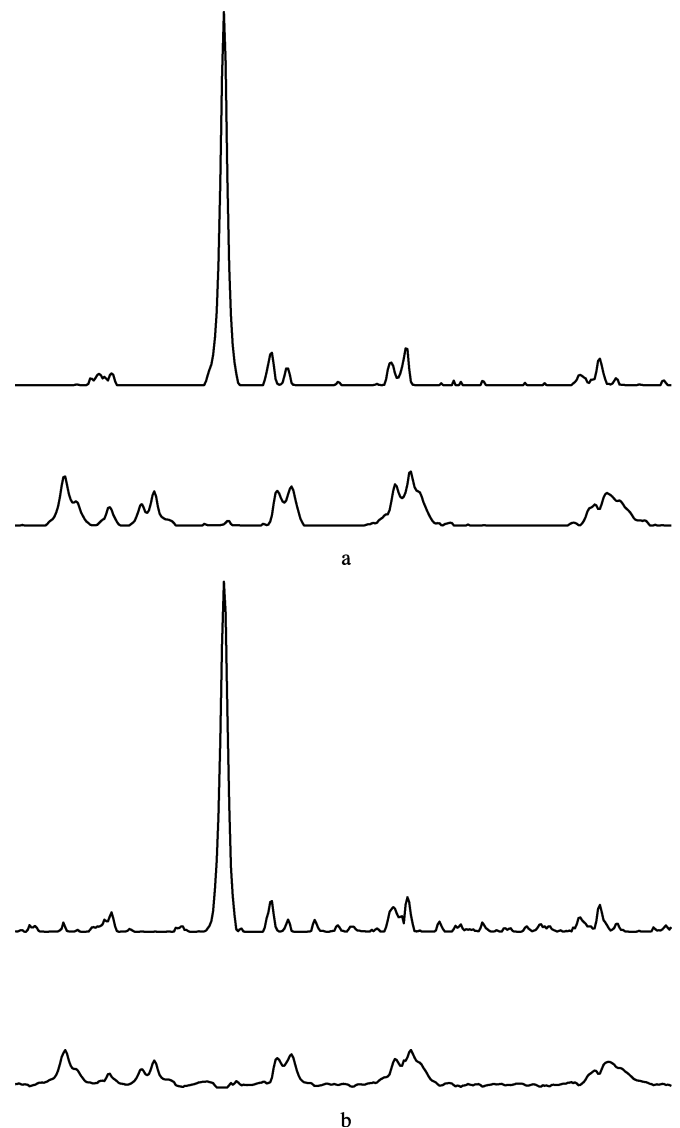tent with the underlying biochemical characteristics of the tissue, with the top spectrum being indicative of muscle and the bottom of brain tissue. The inset in Fig. 3 shows the fifth axial slice of the mixing matrix $\mathbf{A}$, which can be viewed as the relative concentration of the recovered spectra. In this case, we see the muscle spectrum is concentrated near the skull border while brain spectrum is largely internal to the muscle. The recovered spectra are nearly identical to those found using BSD; results are shown in [27]. However, the cNMF algorithm only requires 0.48 s to converge while BSD takes 12 000 s (2.3-GHz Intel Processor). Fig. 4 shows plots of the data, reconstruction from the cNMF algorithm, along with the residuals. The residual error is an order of magnitude smaller than the data and reconstruction.

We next compare cNMF to a similar decomposition algorithm. Nonnegative sparse coding (NNSC) [35] is a method for decomposing multivariate data into nonnegative sparse components which combines sparse coding and NMF. In linear sparse coding, the goal is to find a decomposition in which the hidden components are sparse, with probability densities that are highly
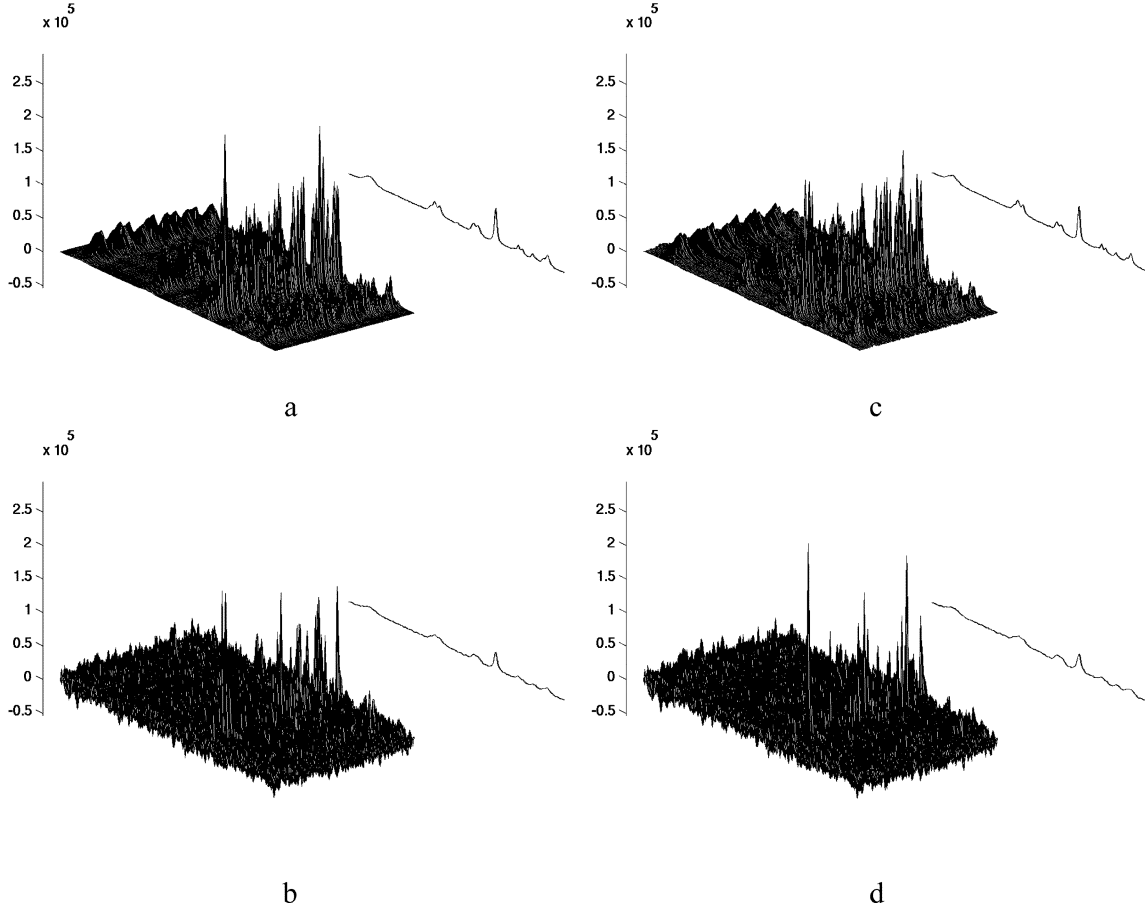
Fig. 6.   Plots of NNSC reconstruction and residuals for ^{31}P head data: (a) and (b) are reconstruction and residuals using all observations [spectra recovered in Fig. 5(a)] and (c) and (d) are the reconstruction and residuals when only using the nonnegative observations [spectra in Fig. 5(b)]. Note the larger residuals compared to cNMF.

peaked at zero and have heavy tails. This implies that the data can be well represented using only a few nonzero coefficients. Combining the goal of small reconstruction error with that of sparseness, one can arrive at the following objective function to be minimized:

$$C(\mathbf{A}, \mathbf{S}) = \frac{1}{2}\|\mathbf{X} - \mathbf{A}\mathbf{S}\|^2 + \lambda \sum_{i,j} f(\mathbf{S}_{ij}) \qquad (10)$$

where i and j are indexes of matrix $\mathbf{S}$. The tradeoff between sparseness and accurate reconstruction is controlled by the parameter λ, whereas the form of f defines how sparseness is measured. To achieve a sparse code, the form of f must be chosen correctly: a typical choice is $f(s) = |s|$, although similar functions that exhibit smoother behavior at zero can be chosen for numerical stability.

We used an existing implementation of NNSC [35][3] and compared results for this dataset. Since NNSC requires all values in $\mathbf{X}$ to be nonnegative, we evaluate the algorithm (1) using all observations (NNSC 1) and constraining negative recovered values to be approximately zero, results shown in Fig. 5(a) and (2) constraining all negative elements in the observation vectors to be zero (NNSC 2), results shown in Fig. 5(b). We note that the recovered spectra are not as well separated as cNMF, as evidenced by the large magnitude of the residuals shown in Fig. 6. Although both algorithms are very fast, cNMF converges to a more

[3]MATLAB code from http://www.cns.nyu.edu/~phoyer/.

physically meaningful solution, while at the same time not imposing an explicit assumption on the sparsity of the constituent spectra. Furthermore, NNSC is very sensitive to the choice of λ and μ; if not appropriately chosen, the algorithm will diverge or oscillate. cNMF has no such parameters to set.

B. ^{31}P Artificial Mixture Data

To evaluate the performance of the cNMF algorithm as a function of SNR, we construct artificial mixtures from known target spectra with varying amounts of zero mean additive Gaussian noise. In real data, the relative SNR is approximately 6, and we therefore evaluate performance of the algorithm for relative SNR levels in the range 0.1–10. For each SNR level, we measure the root mean square (rms) error of the recovered spectra

$$E_{\mathrm{rms}} = \frac{1}{2}\|\mathbf{S}_1 - \hat{\mathbf{S}}_1\| + \frac{1}{2}\|\mathbf{S}_2 - \hat{\mathbf{S}}_2\| \qquad (11)$$

where $\mathbf{S}_1$ and $\mathbf{S}_2$ are the targets (known sources). We compute E_{rms} after initialization with quadratic programming $(E_{\mathrm{rms}}^{\mathrm{in}})$ and after 100 iterations of updating to convergence $(E_{\mathrm{rms}}^{\mathrm{conv}})$. We then compute the ratio of the two

$$E_{\mathrm{rms}}^{\mathrm{ratio}} = \frac{E_{\mathrm{rms}}^{\mathrm{conv}}}{E_{\mathrm{rms}}^{\mathrm{in}}}. \qquad (12)$$

This ratio measures how well the cNMF updating improves the separation over the initialization by quadratic programming.
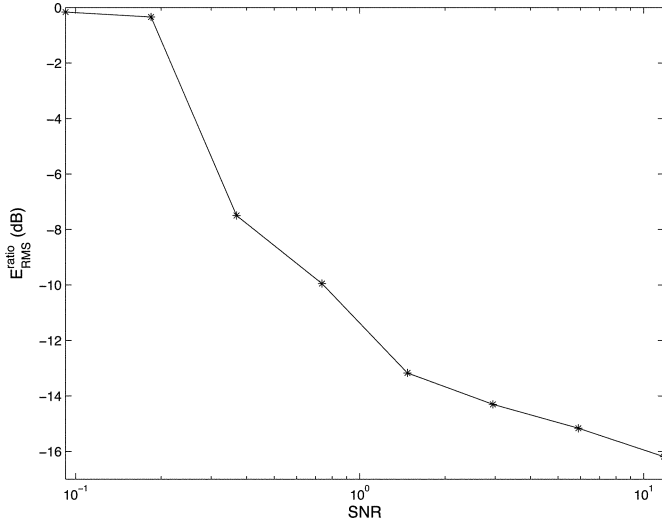
Fig. 7. Recovery error versus SNR level for cNMF. As the level of the SNR decreases, separation quality decreases, relative to the initialization provided by quadratic programming.
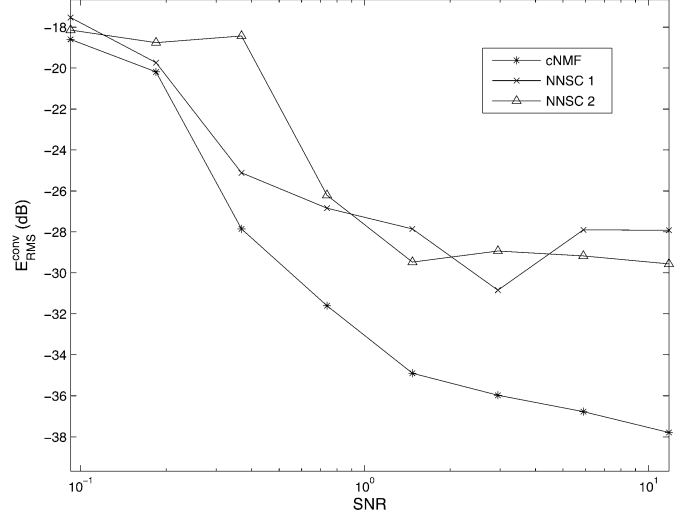

Fig. 8. Comparison of recovery error versus SNR level for cNMF, NNSC 1, and NNSC 2. Clear is that cNMF performance is better than NNSC for all tested SNR levels.

In addition $E_{\mathrm{rms}}^{\mathrm{conv}}$ demonstrates the overall error of the cNMF algorithm and can be used as a metric for comparing with other algorithms. Results for $E_{\mathrm{rms}}^{\mathrm{ratio}}$ are shown in Fig. 7. We see that the cNMF updates dramatically improve separation over a quadratic programming solution at all SNR levels ($E_{\mathrm{rms}}^{\mathrm{ratio}} < 1$), though as expected the effect is much less at low SNR levels.

We compare cNMF to NNSC (NNSC 1 and NNSC 2) in Fig. 8. Since cNMF and NNSC use different methods for initialization, we compare them using $E_{\mathrm{rms}}^{\mathrm{conv}}$. For all tested SNR levels, cNMF performance is better than NNSC. When the SNR is greater than 1, $E_{\mathrm{rms}}^{\mathrm{conv}}$ for NNSC is more than two times larger compared to cNMF, resulting in poor quality of separation. As SNR further decreases to 0.2, neither algorithm gives reasonable separation results.

C. Hierarchical Application on ^1H Brain Data

Results of applying the cNMF algorithm hierarchically for two cases of ^{1}H CSI human brain data are given below. 32×32 voxel axial view ^{1}H spectra of human brain are used, with each voxel having a spectrum of 1024 resonance bands. The ^{1}H CSI data were recorded on a 1.5-T GE Signa Horizon 5.x MR system using the multislice sequence of Duyn *et al.* [36] with $\mathrm{TE/TR} = 280/2300$ ms, field-of-view 240 mm, 32×32 phase-encoding steps with circular k-space sampling, receiver bandwidth 1 kHz and 256 sample points along the signal acquisition (time) domain. The raw data were sorted by slice, zero-padded to 1024 sample points, and then 3-D Fourier transformed to yield 1024 frequency-domain spectral voxels.

First, we show results for the single-resolution hierarchical decomposition. For both datasets, PCA analysis estimates three or four sources at the first level of the hierarchy. This is consistent with the head, at a coarse level, consisting of 3–4 primary tissue/biochemical types: brain, lipids, and water. Recovered sources for the first level of the hierarchy are shown in

Figs. 9(a) and 10(a). Analysis of both the spectra and spatial distributions clearly shows recovery of brain tissue, residual lipids, and water (in the sinuses). 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 Figs. 9(b) and 10(b). From Fig. 9(b), we see the three recovered spectra are consistent with normal brain tissue, high-grade tumor tissue, and residual lipids. The spatial distribution of these markers is also shown. Thus, the hierarchical approach improves the specificity of recovered sources. Similar improvement in specificity and tumor delineation and extent are seen in Fig. 10(b), where a low-grade tumor is well separated from normal brain tissue.

The multiresolution hierarchical decomposition is also evaluated for these data. A three-level hierarchy is used (32×32, 16×16, and 8×8). Recovered spectra and their corresponding spatial distributions are shown in Figs. 11 and 12. Comparing the results of the multiresolution hierarchical decomposition to the single-resolution hierarchical method, we see that at the highest level of the hierarchy they yield nearly identical results. However, the important distinguishing factor of the two hierarchical decompositions is that the multiresolution method is computationally less expensive (see Appendix I). Run-time for the two algorithms, including preprocessing, averaged 23.4 s for the single resolution hierarchy vs. 16.3 s for the multiresolution hierarchy (Pentium 2.3 GHz).

V. DISCUSSION AND CONCLUSION

In this paper we present a fast algorithm for blindly recovering source spectra in CSI. 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
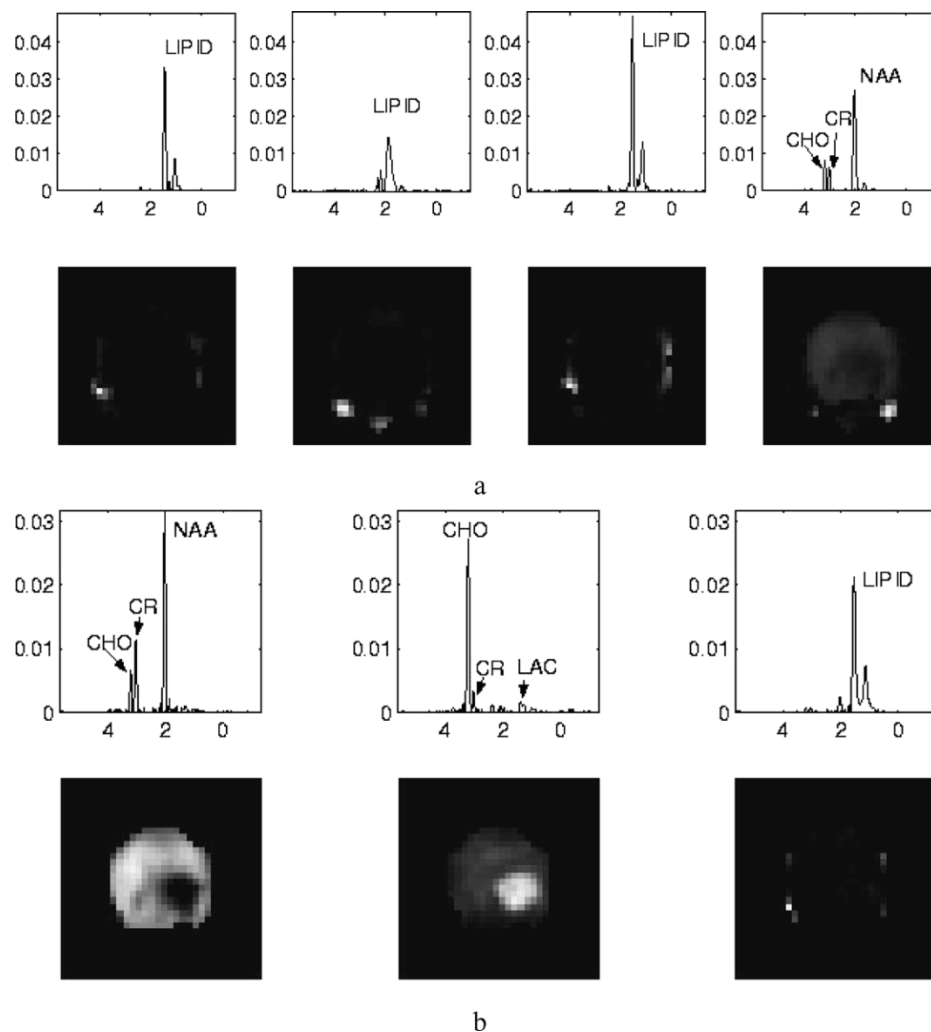
Fig. 9. Single-resolution hierarchy separation results of ^{1}H CSI human brain data case 1. (a) First-level separation results. First row is the recovered spectra, and second row is the spatial distribution of each recovered spectra. Shown in the spectra are biochemical markers for brain (fourth column) with peaks for CHO, CR, and NAA. Note that the corresponding spatial mapping of these markers maps to the region of the brain. Column 2 is a recovered spectrum for a misaligned lipid artifact. The fact that this is lipid is clear from the spatial distribution. Voxels having significant concentration of this spectrum are removed from subsequent analysis by the masking procedure. (b) Second-level separation results. Column 1 spectrum is indicative of normal brain tissue: low CHO, high CR, and high NAA; column 2 spectrum indicates high-grade malignant tumor tissue: highly elevated CHO, low CR, almost no NAA, and LAC peak appearance; column 3 spectrum indicates residual lipids.

due to noise. We term this the 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 nonnegativity constraints are used, with no explicit sparsity, independence, or orthogonality assumptions, though the algorithm recovers sparse sources quite well.

Experiments on ^{31}P and ^{1}H CSI data testify to the performance of the algorithm. For the ^{31}P CSI results, recovered spectra are clearly consistent with brain and muscle, the two dominant sources, with their physical interpretation further validated by the corresponding spatial distribution maps. Results for noisy artificial mixtures shows that cNMF recovers target spectra, for a variety of noise levels, and in all cases with greater fidelity than similar decomposition algorithms.

We also 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, proliferation/grade, etc.). The hierarchy can be cast within a multiresolution processing structure to reduce computational cost. The hierarchical decomposition algorithms are nearly automatic, except for threshold selection (i.e., the threshold on the eigenvalues of PCA for selecting the number of sources and the threshold for constructing the spatial mask) and selection of the number of levels in the hierarchy.

^{1}H MR has shown that, in tumors, metabolites are heterogeneously distributed and in a given voxel multiple metabolites and tissue types may be present. The observed spectra are therefore a combination of different constituent spectra, and are quite variable within and across subjects. This heterogeneity creates
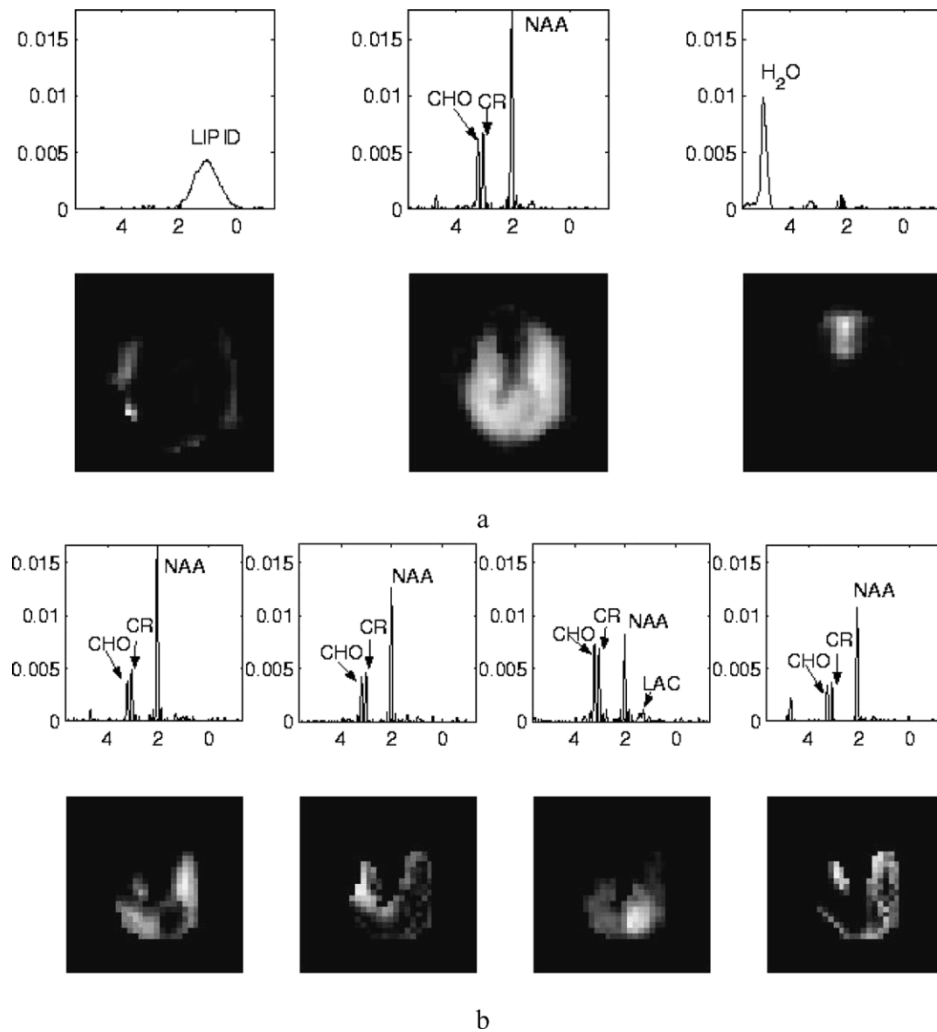
Fig. 10.   Single-resolution hierarchy separation results of ^{1}H CSI human brain data case 2. (a) First-level separation results. The layout of the figure is same as in Fig. 9. Shown in the second column spectrum are biochemical markers for brain. Also recovered are the sinuses (column 3) indicated by a water peak as well as a physically realistic spatial distribution for lipids (columns 1). (b) Second-level 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 (fourth column at 4.7 ppm); column 3 spectrum indicates low-grade cancerous tissue: moderately elevated CHO, significantly decreased NAA, and appearance of LAC peak.

two immediate practical issues: 1) which spectrum should be selected for diagnosis given intrasubject variability and 2) what is a clear marker of disease given intersubject variability? In terms of the first issue, often the spectra of voxels within an MRI contrast-enhancing region are considered. However, this may not always be appropriate as some tumor types are nonenhancing. In addition, the center of the enhancing region in large tumors may be necrotic, while the proliferative boundary may be distributed across multiple mixed voxels. Thus, it would be beneficial to have a separate spectrum for each constituent tissue type rather than for each voxel.

With regard to the second issue, conventional spectral analysis estimates the relative concentrations of different metabolites by measuring the intensity of the resonance lines associated with each metabolite [37]. Relative metabolite concentrations are then used as a diagnostic index for disease. For example, the Choline-NAA index (CNI) measures how far the concentration ratio of CHO to NAA differs from the values of a normal

brain. Many such relations can be defined including concentrations of other metabolites such as Lactate, Glycine, Glutamate and Glutamine, Lipids, myo-Inositol, and more. However, given heterogeneity of tumor spectra, it is difficult to define a strict relation between metabolite concentrations and pathology [5].

The proposed method of cNMF has the potential to address both these problems. By representing each spectrum as a linear combination of constituent spectra one obtains a set of spectra that are the same across different voxels, i.e., the variability due to volume effect is removed. This should help overcome the difficulties of diagnostic classification. In addition, one obtains spectral images, which quantify the relative abundance of metabolites or tissue type contributing to observed spectrum in each voxel.

The goal of this work has been to demonstrate the potential clinical diagnostic value of the proposed method on brain tumors using routine long echo time (TE) spectra. Future research will consider short TE spectra, which have been shown
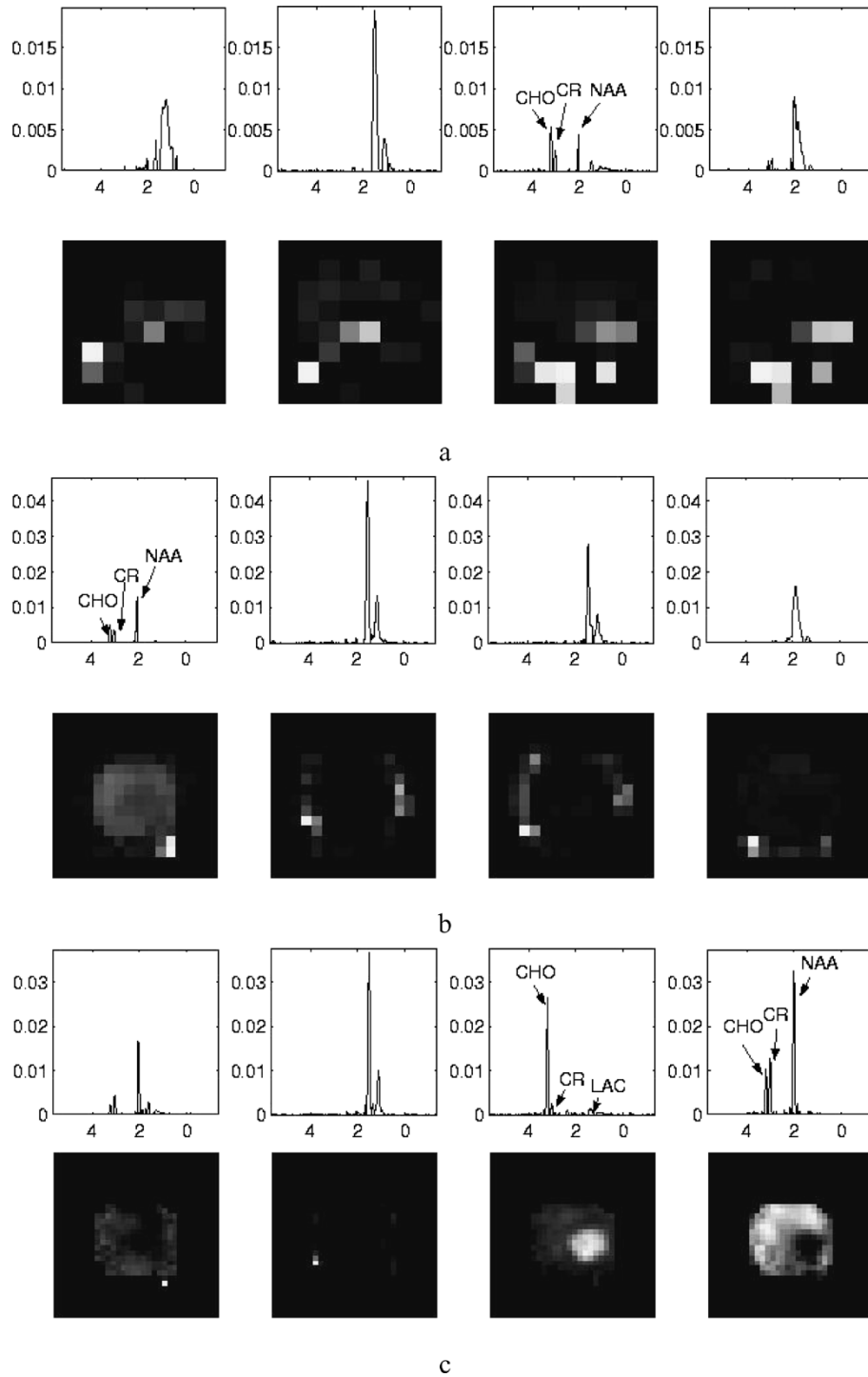
Fig. 11. Multiresolution hierarchy separation results of ^{1}H CSI human brain data case 1. (a) First-level separation result, resolution 8 × 8. The layout of the figure is the same as in Fig. 9. (b) Second-level separation results, resolution 16 × 16. (c) Third-level separation results, resolution 32 × 32. As in the single resolution case, the multiresolution hierarchical decomposition recovers spectra indicative of a high-grade malignant brain tumor (column 3).

to provide much richer diagnostic and prognostic information in small research studies. Current clinical practice does not consider short TE spectra as their analysis is complicated by severe spectral overlap and the variability of spectral profiles. However, we believe that the proposed spectral separation technique holds promise for short TE acquisitions, which in fact are more strongly constrained by nonnegativity of the resonance spectra.

APPENDIX I
COMPUTATIONAL COST OF THE HIERARCHICAL DECOMPOSITION

Given an observation matrix, $\mathbf{X}$, of dimension $[N, L]$, concentration mixing matrix $\mathbf{A}$ of dimension $[N, M]$, and source component matrix $\mathbf{S}$ of dimension $[M, L]$, we compute the computational cost of the hierarchical decomposition.
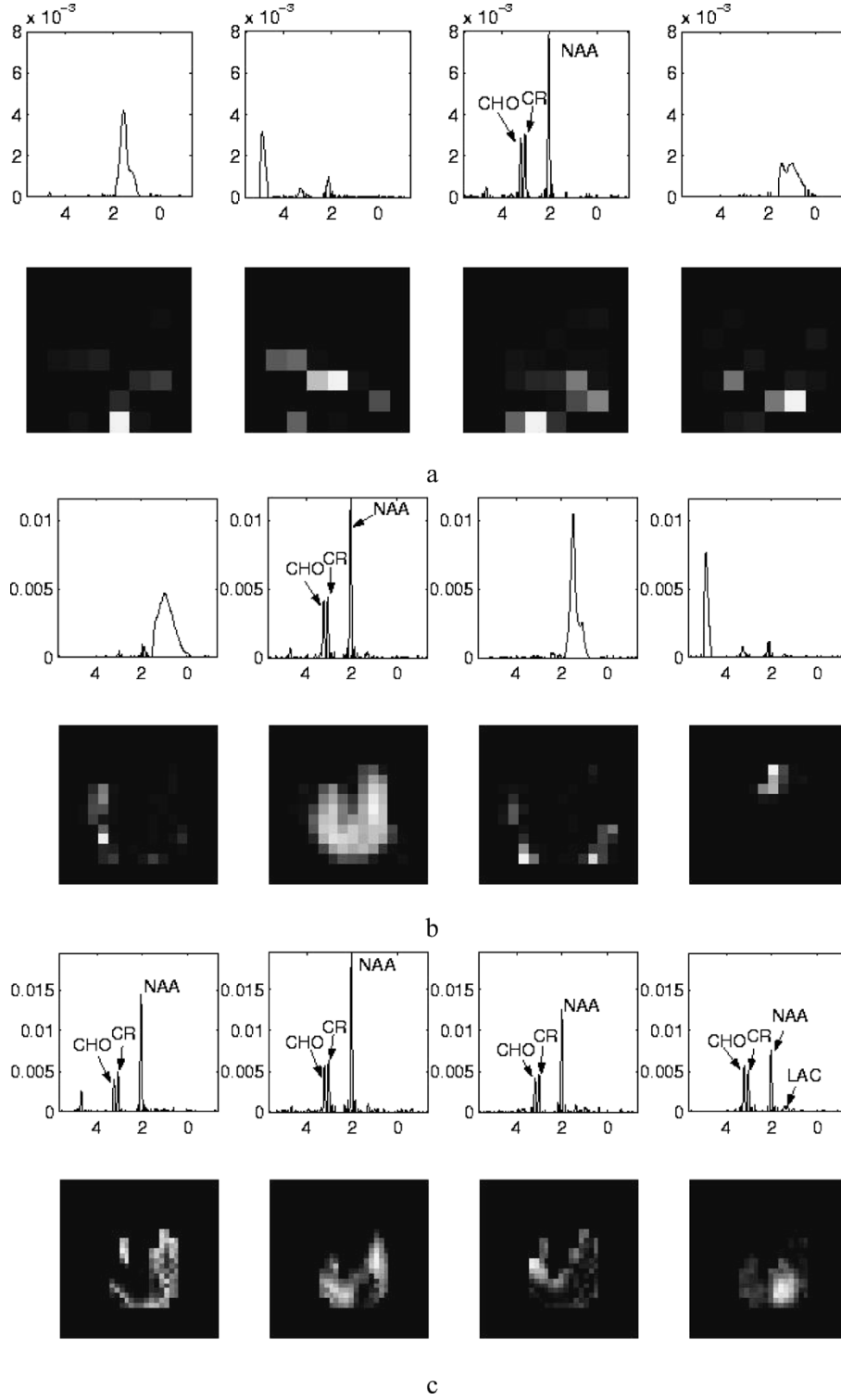
Fig. 12. Multiresolution hierarchy separation results of ^{1}H CSI human brain data case 2. (a) First-level separation results, resolution 8×8. The layout of the figure is the same as in Fig. 9. (b) Second-level separation results, resolution 16×16. (c) Third-level separation results, resolution 32×32. Low-grade brain tumor spectra (column 4).

According to multiplicative update rule in (7), the computational cost for every update of $\mathbf{A}$ and $\mathbf{S}$ is $6NML + 2NM + 2ML$ multiplications and $6NML$ additions. Since we assume the number of recovered sources is much less than the number of observations and dimensionality of the data ($M \ll N$ and $M \ll L$), we can neglect the issue of different numbers of sources at each level in the hierarchy contributing to the total computa-

tional cost. Thus, for a two-level single-resolution hierarchical decomposition, the computation cost is ($\times \equiv$ multiplications, $+ \equiv$ additions)

$$\begin{aligned} \times &= 2(6NML + 2NM + 2ML) \\ &= 12NML + 4NM + 4ML \\ + &= 2(6NML) = 12NML. \end{aligned} \tag{13}$$

For a three-level multiresolution hierarchical model, the first level of observation voxels are downsampled t o $(1/16)N$, the second level $(1/4)N$ with the third level having the original number N, so at each update step the computational cost is

$$
\begin{aligned}
\times &= 6\left(\frac{1}{16} + \frac{1}{4} + 1\right) NML \\
&+ 2\left(\frac{1}{16} + \frac{1}{4} + 1\right) NM + 3 * 2ML \\
&= 7.875 NML + 2.625 NM + 6ML \\
&+ = 6\left(\frac{1}{16} + \frac{1}{4} + 1\right) NML = 7.875 NML. \quad (14)
\end{aligned}
$$

Comparing (13) and (14), we can see the cost of a three-level multiresolution decomposition is approximately 2/3 of a two-level single-resolution decomposition, which saves significant computational time.

For the results in Fig. 10, $N = 1024$, $M = 4$, $L = 1024$, and a two-level single-resolution decomposition model is used, so the computation cost for each step is 5.0×10^7 multiplications and additions, respectively. For the results in Fig. 12, $N = 1024$, $M = 4$, $L = 1024$, and a three-level multiresolution decomposition model is used. Here, the computational cost for each step is 3.3×10^7 multiplications and additions, which is about 66% of the cost of a two-level single-resolution hierarchical model. This computational cost savings is without any loss in separation quality, and in fact the separation results of the three-level multiresolution hierarchical decomposition are somewhat better than those of the two-level single-resolution, e.g., in Fig. 12 the tumor is more pronounced than in Fig. 10.

REFERENCES

[1] D. Kondziola, L. D. Lunsford, and A. J. Martinez, "Unreliability of contemporary imaging in evaluating suspected adult supratentorial (low-grade) astrocytoma," *J. Neurosurg.*, vol. 79, pp. 533–536, 1993.

[2] M. Bernstein and A. G. Parrent, "Complications of CT-guided stereotactic biopsy of intra-axial brain lesions," *J. Neurosurg.*, vol. 81, pp. 165–168, 1994.

[3] X. Yu, Z. Liu, Z. Tian, S. Li, H. Huang, B. Xiu, Q. Zhao, L. Liu, and W. Jing, "Stereotactic biopsy of intracranial space occupying lesions: Clinical analysis of 550 cases," *Stereotact. Funct. Neurosurg.*, vol. 75, pp. 103–108, 2000.

[4] F. Alesch, J. Pappaterra, S. Trattnig, and W. Koo, "The role of stereotactic biopsy in radiosurgery," *Acta Neurochir. Suppl. (Wien)*, vol. 63, pp. 20–24, 1995.

[5] F. A. Howe and K. S. Opstad, "^{1}H MR spectroscopy of brain tumours and masses," *NMR Biomed*, vol. 16, pp. 123–131, 2003.

[6] I. C. P. Smith and L. C. Stewart, "Magnetic resonance spectroscopy in medicine: Clinical impact," *Progr. Nucl. Magn. Reson. Spectrosc.*, vol. 40, pp. 1–34, 2002.

[7] J. P. Cousins, "Clinical MR spectroscopy: Fundamentals, current applications, and future potential," *Amer. J. Roentgenol.*, vol. 164, pp. 1337–1347, 1995.

[8] T. Brown, B. Kincaid, and K. Ugurbil, "NMR chemical shift imaging in three dimensions," *Proc. Nat. Acad. Sci.*, vol. 79, no. 11, pp. 3523–3526, 1982.

[9] W. G. Negendank, R. Sauter, T. R. Brown, J. L. Evelhoch, A. Falini, E. D. Gotsis, A. Heerschap, K. Kamada, B. C. P. Lee, M. M. Mengeot, E. Moser, K. A. Padavic-Shaller, J. A. Sanders, T. A. Spraggins, A. E. Stillman, B. Terwey, T. J. Vogl, K. Wicklow, and R. A. Zimmerman, "Proton magnetic resonance spectroscopy in patients with glial tumors: A multicenter study," *J. Neurosurg.*, vol. 84, pp. 449–458, 1996.

[10] W. G. Negendank and R. Sauter, "Intratumoral lipids in ^{1}H MRS *in vivo* in brain tumors: Experience of Siemens cooperative clinical trial," *Anticancer Res.*, vol. 16, pp. 1533–1538, 1996.

[11] J. W. Hugg, J. H. Duijn, G. B. Matson, A. A. Maudsley, J. S. Tsuruda, D. F. Gelinas, and M. W. Weiner, "Elevated lactate and alkalosis in chronic human brain infarction observed by ^{1}H and ^{31}P MR spectroscopic imaging," *J. Cereb. Blood Flow Metab.*, vol. 12, pp. 734–744, 1992.

[12] M. L. Simmons, C. G. Frondoza, and J. T. Coyle, "Immunocyto-chemical localization of N-acetyl-aspartate with monoclonal antibodies," *Neuroscience*, vol. 45, pp. 37–45, 1991.

[13] M. G. Swanson, D. B. Vigneron, T.-K. C. Tran, N. Sailasuta, R. E. Hurd, and J. Kurhanewicz, "Single-voxel oversampled J-resolved spectroscopy of *in vivo* human prostate tissue," *Magn. Reson. Med.*, vol. 45, pp. 973–980, 2001.

[14] D. K. W. Yeung, W.-T. Yang, and G. M. K. Tse, "Breast cancer: *In vivo* proton MR spectroscopy in the characterization of histopathologic subtypes and preliminary observations in axillary node metastases," *Radiology*, vol. 225, pp. 190–197, 2002.

[15] H. Kugel, W. Heindel, R. I. Ernestus, J. Bunke, R. du Mesnil, and G. Friedmann, "Human brain tumors: Spectral patterns detected with localized H-1 MR spectroscopy," *Radiology*, vol. 183, pp. 701–709, 1992.

[16] M. C. Preul, Z. Caramanos, D. L. Collins, J.-G. Villemure, R. Leblanc, A. Olivier, R. Pokrupa, and D. L. Arnold, "Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy," *Nature Med.*, vol. 2, pp. 323–325, 1996.

[17] C. M. Segebarth, D. F. Baleriaux, P. R. Luyten, and J. A. Hollander, "Detection of metabolic heterogeneity of human intercranial tumors *in vivo* by ^{1}H NMR spectroscopic imaging," *Magn. Reson. Med.*, vol. 13, pp. 62–76, 1990.

[18] M. Stubbs, R. L. Veech, and J. R. Griffiths, "Tumor metabolism: the lessons of magnetic resonance spectroscopy," *Adv. Enzyme Regul.*, vol. 35, pp. 101–115, 1995.

[19] S. Mierisova and M. Ala-Korpela, "MR spectroscopy quantitation: A review of frequency domain methods," *NMR Biomed.*, vol. 14, no. 4, pp. 247–259, 2001.

[20] L. Vanhamme, T. Sundin, P. V. Hecke, and S. V. Huffel, "MR spectroscopy quantitation: A review of time-domain methods," *NMR Biomed.*, vol. 14, no. 4, pp. 233–246, 2001.

[21] 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 Biomed.*, vol. 10, pp. 25–30, 1997.

[22] M. J. Fulham, A. Bizzi, M. J. Dietz, H. H. Shih, R. Raman, G. S. Sobering, J. A. Frank, A. J. Dwyer, J. R. Alger, and G. Chiro, "Mapping of brain tumor metabolites with proton MR spectroscopic imaging: Clinical relevance," *Radiology*, vol. 185, pp. 675–686, 1992.

[23] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice-Hall, 1974.

[24] L. Vanhamme, S. V. Huffel, P. V. Hecke, and D. V. Ormondt, "Time domain quantification of series of biomedical magnetic resonance spectroscopy signals," *J. Magn. Reson.*, vol. 140, pp. 120–130, 1999.

[25] D. Nuzillard, S. Bourg, and J.-M. Nuzillard, "Model-free analysis of mixtures by NMR using blind source separation," *J. Magn. Reson.*, vol. 133, pp. 358–363, 1998.

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

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

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

[29] ——, "Algorithms for nonnegative matrix factorization," in *Advances in Neural Information Processing Systems 13*. Cambridge, MA: MIT Press, 2001, pp. 556–562.

[30] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *Proc. 37th Annu. Allerton Conf. Communication, Control and Computing*, 1999, pp. 368–377.

[31] N. Friedman, O. Mosenzon, N. Slonim, and N. Tishby, "Multivariate information bottleneck," in *Proc. 17th Conf. Uncertainty in Artificial Intelligence (UAI)*, 2001, pp. 152–161.

[32] P. Sajda, S. Du, T. R. Brown, L. C. Parra, and R. Stoyanova, "Recovery of constituent spectra in 3D chemical shift imaging using nonnegative matrix factorization," in *Proc. 4th Int. Symp. Independent Component Analysis and Blind Source Separation*, 2003, pp. 71–76.

[33] 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.*, vol. 154, pp. 163–175, 2002.

[34] D. J. C. Mackay, "Information-based objective functions for active data selection," *Neural Comput.*, vol. 4, no. 4, pp. 590–604, 1992.

[35] P. O. Hoyer, "Non-negative sparse coding," in *Neural Networks for Signal Processing XII (Proc. IEEE Workshop Neural Networks for Signal Processing)*, Martigny, Switzerland, 2002, pp. 557–565.

[36] J. H. Duyn, J. Gillen, G. Sobering, P. C. van Zijl, and C. T. Moonen, "Multisection MR spectroscopic imaging of the brain," *Radiology*, vol. 188, pp. 277–282, 1993.

[37] L. Hofmann, J. Slotboom, B. Jung, P. Maloca, C. Boesch, and R. Kreis, "Quantitative ^{1}H-magnetic resonance spectroscopy of human brain: Influence of composition and parameterization of the basis set in linear combination model-fitting," *Magn. Reson. Med.*, vol. 48, pp. 440–453, 2002.