Non-negative blind source separation techniques for tumor tissue typing using HR-MAS signals

A. Croitor Sava, D. M. Sima, M. C. Martinez-Bisbal, B. Celda and S. Van Huffel

Abstract—Given High Resolution Magic Angle Spinning (HR-MAS) signals from several glioblastoma tumor subjects, the goal is to differentiate between tumor tissue types by separating the different sources that contribute to the profile of each spectrum. Blind source separation techniques are applied for obtaining characteristic profiles for necrosis, high cellular tumor and border tumor tissue, and providing the contribution (abundance) of each tumor tissue to the profile of the spectra. The problem is formulated as a non-negative source separation problem. We illustrate the effectiveness of the proposed methods and we analyze to which extent the dimension of the input space could influence the performance by comparing the results on the full magnitude signals and on dimensionally reduced spaces.

I. INTRODUCTION

EX-VIVO HR-MAS (high-resolution magic angle spinning spectroscopy) is a Nuclear Magnetic Resonance (NMR) spectroscopy technique. It provides significant biochemical information on the metabolites by exhibiting peaks at frequencies specific to the molecular composition of the tissue under investigation. Since HR-MAS NMR allows the identification of an important number of metabolites, it has lately been extensively used for characterizing and diagnosing brain tumors.

Previous studies have reported that alteration in the metabolite concentrations are correlated with the brain tissue type. Hence, each tissue type can be viewed as having a characteristic metabolic profile corresponding to the chemical composition of the tissue [1].

HR-MAS signals, typically called spectra, are obtained from a small tissue biopsy sample, thus the spectra reflect the local tissue characteristics. A significant variability within the HR-MAS spectral profiles belonging to each of the main brain tissue types can be observed [2]. This is because brain tissues have the particularity of being very heterogeneous. In brain tumors, for example, the tissue under investigation might present contributions from various tumor tissue types.

A. Croitor Sava, D.M. Sima and S. Van Huffel are with Department of Electrical Engineering (ESAT-SCD) - Biomed, Katholieke Universiteit Leuven, Leuven (phone: +32 (0)16 32 17 09, fax: +32 (0)16 32 19 70, e-mail: anca.croitor@esat.kuleuven.be).

M. C. Martinez-Bisbal and B. Celda are with Departamento de Química-Física, Facultad de Química, Universitat de Valencia, Valencia, Spain and CIBER-BBN, ISC-III, Universitat de Valencia, Spain.

The observed spectra are, therefore, a combination of different constituent sub-spectra, since the measured signal is the response to the stimulation of the entire tissue sample. The overall gain with which a tissue type contributes to a spectrum is proportional to its concentration. As a result, multiple metabolites and tissue types may be present in a single HR-MAS spectrum.

We can summarize this concept by describing the spectra available from m samples, which are stacked as n-dimensional row vectors in an m by n matrix X, as in [3]:

$$X = AS + N \qquad (1)$$

where A contains the concentrations, or abundances, of the constituent pure tissue sources in each sample, and S is a k by n matrix whose rows are the unknown pure tissue spectra. N represents additive noise.

At present, histopathological analysis is the standard procedure routinely used to reveal the microheterogeneity of a tumor tissue [4]. Because histopathology is time and effort expensive, it is interesting to obtain a similar separation using HR-MAS information. When classifying HR-MAS spectra, a common procedure is to assign each spectrum to a certain tissue class without taking into consideration the tissue microheterogeneity. In this study we propose to differentiate between tumor tissue types by identifying the pure components of the different tissues, S, and estimating the concentration of each component, A. This problem is formulated as a source separation problem where an important constraint is the non–negativity of the source signals and of the mixing coefficients. This solution is motivated by the nature of the HR-MAS magnitude spectra, where one has to deal with non-negative mixtures of non-negative signals. Non-Negative Matrix Factorization (NMF) [12] and a Convex Analysis for Blind Separation of Non-Negative Sources [13] are considered.

II. MATERIALS AND METHODS

A. Materials

Brain tumor biopsies were carried out on 27 patients with glioblastoma tumor (GBM). The tissue specimens were snap-frozen in liquid nitrogen and stored at -80^0 C until the time of spectroscopic analysis. 1D PRESAT HR-MAS (pulse-and-acquire) data were acquired, following the eTUMOUR project protocols (http://www.etumour.net/) at 11.7 and 14T

(500 MHz and 600MHz for 1H) at 4 ^{0}C and 4000 Hz and 5000Hz spinning rate using BRUKER Analytik GmbH spectrometers.

After the HR-MAS study, tumor specimens were snap-frozen and submitted for quantitative histopathological examinations. 22 out of the 27 biopsy samples were followed by standard histological examination performed by an expert neuro-pathologist on tissue samples taken from the same part used in the HR-MAS study. Based on the histology, the samples were considered with variable content of tumor, border and/or necrotic tissue. The percentage of the contribution from different tissues was calculated for each sample by measuring the total area of the biopsy sections and then delineating the necrotic, high cellular or border tissue regions of interest. The histopathological analysis is further used in the study for validating our results.

The complex-valued HR-MAS time-domain signals were preprocessed as follows: signals were truncated from 8120 to the first 2048 points to reduce the computational load; the water components were removed by HLSVD-PRO [5]; baseline was corrected as follows. An apodization function, containing the broad baseline components, is computed by multiplying the signal in the time-domain with an exponentially decaying function and subsequently subtracted from the original spectrum. Magnitude spectra are then computed by taking the absolute value of the Fourier transformed time-domain signals. Contributions outside the frequency interval [0.25, 4.2] ppm were filtered out in order to keep only the contribution of the metabolites of interest; the filtered spectra were normalized (divided by the l2 norm of the spectrum between 0.5 and 4.2 ppm) and aligned.

Although all the considered 27 tissue specimens are known to come from glioblastoma tumors, the variability in the contribution of tumor, border and necrotic tissue to their content is reflected in the profile of the HR-MAS spectra. To better illustrate this problem, in Fig. 1 we visualize three glioblastoma HR-MAS spectra, which come from tissue samples that are indicated by the histopathology as consisting only of pure necrotic, high cellular or border tumor tissue. In a classical classification approach, all these spectra would have been assigned to the same tissue class, respectively, to the glioblastoma tissue class.
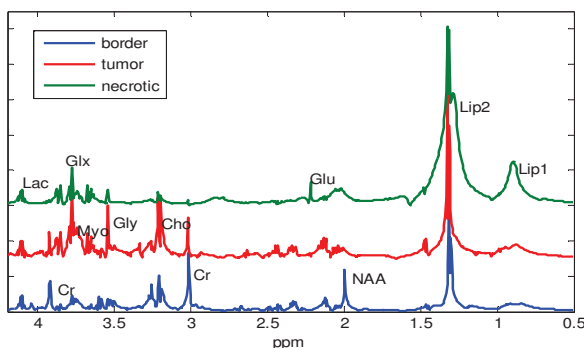

Fig. 1. HR-MAS spectra profiles of GBM tissue samples containing predominant necrotic tissue, high cellular tumor or border tumor tissue.

B. Methods

Blind separation of non-negative sources (nBSS) have been recently successfully used in many applications where the sources to be separated are of a non-negative nature, e.g biomedical imaging [6], analytic chemistry [7] and hyperspectral imaging [8]. The way to exploit the non-negative signal characteristic in nBSS is subjective to the data under analysis. Therefore numerous nBSS alternatives have been proposed. A class of nBSS methods utilizes the statistical property that the sources are mutually uncorrelated or independent; this class includes: second-order blind identification (SOBI) [9], non-negative independent component analysis (nICA)[10], Bayesian positive source separation (BPSS) [11].

Another class of nBSS methods is represented by implementations which require no assumption on source independence or zero correlations. One such nBSS approach is the NMF. This method explicitly imposes source non-negativity and even mixing matrix non-negativity. Taking into consideration the nature of HR-MAS spectra separation, where strong correlations in the metabolic profiles may still be present between spectra form different tissue types, we believe such a method is more suitable for our problem.

NMF is a statistical technique that reveals hidden factors within a dataset of signals. Given a non-negative matrix X of size m x n (in our case $m = 27$ observations and n is the dimension of each observation), NMF finds two matrices A and S with non-negative elements that minimize the function:

$$f(A, S) = \frac{1}{2}\|X - AS\|^2, \text{ with } A, S \geq 0 \qquad (2)$$

If we require S to have 3 rows, then these rows should ideally represent the constituent sources for necrotic, high cellular and border tumor tissue. A contains the coefficients of the linear combinations of the found sources and reflects the abundance of the obtained sources within each sample.

Since NMF is not a unique decomposition, which may result in indeterminacy of the sources and of the mixing matrix, we apply for this study an alternating non-negativity constrained least squares (ANLS) implementation, with a sparsity constraint on the sources as described in [12]. We will further refer to the method proposed in [12] as NNMFSC.

Another recently developed nBSS framework that accounts for sparsity is the convex analysis of mixtures of non-negative sources (CAMNS) [13]. CAMNS is deterministic, requiring no source independence assumption, the premise which can be found in many existing (usually statistical) BSS frameworks. The development is based on a special assumption called local dominance. Under local dominance, convex analysis is applied to establish a new BSS criterion. Thus, the source signals can be perfectly identified (in a blind fashion) by finding the extreme points

of an observation-constructed polyhedral set [13].

Considering as input matrix X, the 27 measured HR-MAS spectra, or features extracted from these spectra, we further analyze and compare the performance obtained with the two proposed algorithms: CAMNS and NNMFSC in identifying the pure tissue profile for necrotic, high cellular and border tumor tissue; and in correctly describing the abundance of each source within a spectrum.

C. Dimension of the input space

The proposed nBSS techniques are applied on the magnitude HR-MAS spectra and on sets of features obtained from the spectra.

For the spectra, we used $n=716$ points representing the HR-MAS spectra in the region of interest between 0.25 ppm and 4.2 ppm. For the feature case, we considered either $n=19$ features, representing the concentration of 19 most visible metabolites, or $n=8$ features representing the concentration of the metabolites considered as the most representative ones for separating between the three tissue classes [2]. Peak integration, a feature reduction method typical in NMR analysis, was used for extracting the concentration of the considered metabolites. Namely, the highest point in the area to be integrated was identified for each metabolite, then the area bounds were fixed for each metabolite individually to those ppm values at which peak slopes return to baseline, but keeping symmetric intervals with respect to the highest point.

D. Performance measurement

To evaluate the accuracy of the proposed nBSS methods in identifying the pure tumor tissue sources, a measure of the separation quality is performed. The nBSS results on the GBM group are validated by comparing them to the histology findings (the standard reference to which diagnosis is based nowadays). To this aim we compute the correlation coefficient between the sources obtained with NNMFSC and CAMNS, respectively, and the reference HR-MAS tissue models from Fig. 1, for all three input spaces mentioned above.

III. RESULTS AND DISCUSIONS

The correlation coefficients between nBSS sources and the reference spectra are presented in Table 1. The correlation coefficient takes a value between -1 and 1, where a value close to -1 indicates a negative correlation, close to 0 indicates that sources are uncorrelated and close to 1 that the sources are highly correlated with the reference tissue.

A very high correlation was obtained with both methods for the necrotic tissue, revealing the power of NNMFSC and CAMNS in accurately identifying necrotic tissue. For extracting the border and tumor tissue source, NNMFSC shows to perform overall better than CAMNS. In particular, for border tissue, with NNMFSC the correlation with the histopathology reaches more than 0.9 out of 1.

Another aspect of the study is to analyze to which extent the dimension of the input space could influence the results. This part of the study provides a closer inside into the question which metabolites are most representative in separating between the considered classes. Since each source is important in our problem we looked at the overall performance of the considered algorithms. As can be seen in Table 1, the dimension of the input space influences the performance of the methods. The sources obtained with NNMFSC on the full magnitude spectra (n=716) are clearly separated and are conform to the conclusions drawn in the literature [4]. As can be seen from Fig. 2, the obtained necrotic tumor tissue source is characterized by elevated peaks of lipids (Lip1, Lip2), while the rest of the metabolites are present in very low concentrations. Border tissue source presents characteristic high peaks of N-acetyl aspartate (NAA) and creatine (Cr), while, for the high cellular tumor, the alanine (Ala) and total choline (tCho) group are more elevated compared to the other sources. We can clearly see that the obtained NNMFSC sources are very similar to the profile of a pure tumor tissue as plotted in Fig.1.

TABLE I
THE CORRELATION BETWEEN THE OBTAINED TISSUE SOURCES AND THE
REFERENCE TISSUE MODELS FOR THREE INPUT SPACES

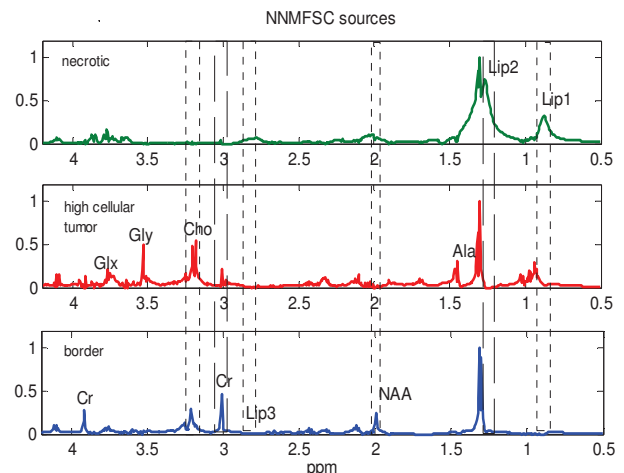|	n=716	n=19	n=8
NNMFSC			
necrotic	0.97	0.97	0.99
border	0.91	0.92	0.91
tumor	0.69	0.65	0.72
CAMNS			
necrotic	0.98	0.97	0.99
border	0.88	0.51	0.82
tumor	0.62	0.61	0.68

Fig. 2. Tumor tissue sources obtained with NNMFSC when applied on the magnitude spectra.

When considering the feature vectors coming from all the visible metabolites ($n=19$), the metabolite profiles extracted by NNMFSC for necrosis and border tissue highly correlate with the reference tissue, see Fig. 3. This input space brings

relative low performance with CAMNS for the border tissue, since the obtained source presents high contributions from metabolites expected in low concentrations. The overall performance is higher when considering only 8 features coming from the most representative metabolites. These results confirm that these metabolites are representative for solving this typical classification problem and adding extra information will affect the performance of the considered nBSS method. Additionally, by increasing the dimensionality of the matrices (case n=716), the problem becomes more ill-conditioned and might provide a solution that is less meaningful for the given problem.

When applying NNMFSC, we obtain for each spectrum the mixing coefficients representing the contribution (abundance) of each source to the spectra. Thus, we can assign each case, based on the highest abundance, to a predominant tissue class. The classification results were compared with the histopahological study and they show that NNMFSC can accurately identify the predominant tissue class. For 19 out of the 22 histopathologically confirmed cases, the same class was indicated predominant by NNMFSC and histopathology.
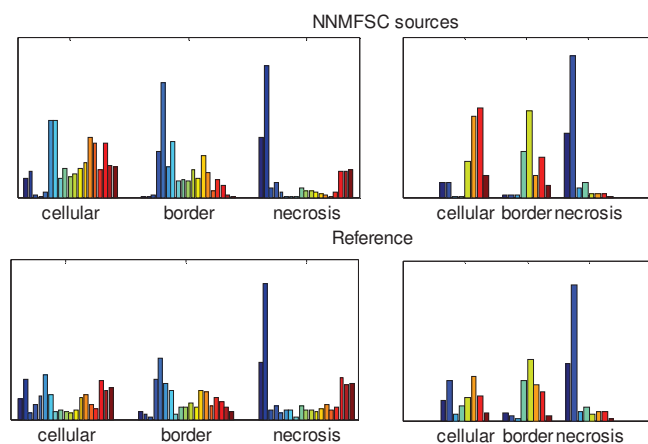

Fig. 3. Tumor tissue profiles obtained with NNMFSC and on feature vectors (*n*=19 left column, *n*=8 right column). The lower plots illustrate features extracted from pure tissue samples, as indicated by histopathology.

IV. CONCLUSION

The nBSS methods proposed for obtaining characteristic profiles for each tissue subtype and the abundance of each source within a spectrum can reliably answer the problem of source separation when analyzing HR-MAS data. This finding can provide relevant additional information for a better interpretation and classification of brain tumor tissue in ex vivo high resolution magnetic resonance spectroscopy. Furthermore, a better understanding of brain tumor tissue classification problems arising from in vivo magnetic resonance spectroscopy can benefit from the same approach. A reduction of the dimension of the input space could act as an added value to this classification problem. This will bring a two-fold advantage. On one side, it reduces the computational time, and, secondly, we avoid bringing irrelevant information into the problem.

REFERENCES

[1] A. Devos, *Quantification and classification of Magnetic Resonance Spectroscopy data and applications to brain tumour recognition*, PhD thesis, Faculty of Engineering, K.U.Leuven (Leuven, Belgium), 2005, p. 217, Lirias number: 49.

[2] A. R. Croitor Sava, M. C. Martinez-Bisbal, B. Celda, J. M. Cerda, S. Van Huffel, "Tissue typing within glial tumours", Online proceeding *ESMRMB*, 2009, p.655.

[3] P. Sajda, S. Du, T.R. Brown, R.S.D.C. Shungu, X. Mao, L.C Parra, "Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain", *IEEE Trans. Medical Imaging*. v23 i12, 2004, pp.1453-1465.

[4] L. L. Cheng, L.W Chang, D. N. Louis, G. Gonzalez, "Correlation of high-resolution magic angle spinning proton magnetic resonance spectroscopy with histopathology of intact human brain tumor specimens". *Cancer Res*. 1998; 58: pp.1825–1832.

[5] T. Laudadio, N. Mastronardi, L. Vanhamme, P. Van Hecke, S. Van Hufel, "Improved Lanczos algorithms for blackbox MRS data quantitation", *J Magn Res,* 2002; 157: pp. 292-297.

[6] D. Nuzillard and J.M. Nuzillard, "Application of blind source separation to 1-D and 2-D nuclear magnetic resonance spectroscopy," *IEEE Trans. Signal Process. Lett,* 1998, vol. 5, no. 8, pp. 209–211.

[7] E. R. Malinowski, *Factor Analysis in Chemistry*. New York: John Wiley, 2002.

[8] J. M. P. Nascimento and J. M. B. Dias, "Does independent component analysis play a role in unmixing hyperspectral data?" *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 1, pp. 175–187, Jan. 2005.

[9] A. Belouchrani, K. Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[10] M. D. Plumbley, "Algorithms for non-negative independent component analysis," *IEEE Trans. Neural Netw.*, vol. 14, no. 3, pp. 534–543, 2003.

[11] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, "Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4133–4145, Nov. 2006.

[12] H. Kim and H. Park, "Sparse Non-negative Matrix Factorizations via Alternating Non-negativity-constrained Least Squares for Microarray Data Analysis", *Bioinformatics*, 23-12:1495-1502, 2007.

[13] H. Cha, W. K. Ma, C. Y. Chi and Y. Wang, "A convex analysis framework for blind separation of non-negative sources", *IEEE Trans. Signal Process.* 56 (10) (2008), pp. 5120–5134.