SPECTRUM SEPARATION OF MAGNETIC RESONANCE SPECTROSCOPY BASED ON SPARSE REPRESENTATION

Yu Guo[1], Su Ruan[1], Gilles Millon[1], Jean-Marc Constans[2]

[1]CReSTIC-Université de Reims, IUT de Troyes, France
[2] CHU de Caen, France
yu.guo@etudiant.univ-reims.fr , su.ruan@univ-reims.fr

ABSTRACT

In this paper, a novel spectrum separation technique based on sparse representation is proposed to deal with Magnetic Resonance Spectroscopy (MRS) quantification which is used to measure the levels of different metabolites in brain tissues. Since a measured MR spectrum contains the spectra of numbers of metabolites and a baseline, the separation and quantification of them becomes difficult. A nonnegative pursuit algorithm based on regularized FOCUSS algorithm is proposed here to decompose a measured spectrum with respect to an overcomplete dictionary. Benefitting from the a priori knowledge, the dictionary is built by Lorentzian and Gaussian basis functions representing different metabolites and baseline. Using this algorithm, not only the baseline is separated from the spectra of interest, but also the spectra of different metabolites are separated. The accuracy of quantification and the robustness are improved, from simulation data, compared with a commonly used estimation method [1]. The quantification on tumor metabolism with in vivo brain MR spectra is also demonstrated.

Index Terms—Magnetic resonance spectroscopy (MRS), spectrum separation, quantification, sparse representation

1. INTRODUCTION

The MRS signal produces a spectrum of resonances that correspond to different molecular arrangements of the isotope being "excited" [2]. A measured MR spectrum is composed by the spectra of several individual metabolites, and the concentration of each metabolite is proportional to its respective peak area. Accurate quantification of in vivo MR spectra (measuring peak areas) is very important to diagnose certain metabolic disorders, especially those affecting the brain. However, some serious problems make the task difficult: strongly overlapping metabolite peaks, poor knowledge about background (the baseline) originating mainly from macromolecules and lipids, and low signal-to-noise ratio (SNR) [3]. Several MRS quantification approaches have been proposed, such as AMARES [4], LCModel [5], semi-parametric QUEST [6], AQSES [7]. Generally, most of these approaches can be described as applying numbers of nonlinear optimization algorithms to estimate the parameters of a mathematical model which is used to characterize MR spectra. However, the problems mentioned above still affect the accuracy of the quantification of MR spectra.

Sparse representation of signals has received a great deal of attention in signal processing region in recent years [8]. A signal sparse representation problem can be described as sparsely representing a signal with the linear combination of several basis functions in an overcomplete dictionary.

In this paper, sparse representation is used for the separation of MR spectra. With the a priori knowledge about MRS data, an overcomplete dictionary with a series of Lorentzian and Gaussian basis functions is firstly built to represent the spectra of different metabolites and the baseline. A nonnegative pursuit algorithm, based on regularized FOCUSS (Focal Underdetermined System Solver) algorithm [9], is then proposed to decompose a measured MR spectrum to the dictionary and finally separate the different spectra of metabolites and the baseline. As the dictionary benefits maximally the priori knowledge, the baseline problem can be well dealt with and the bad influence of noisy and the severe overlap of spectral lines can be decreased. The proposed method is tested with simulated data, as well as real MR spectra. The robustness and accuracy of the method are demonstrated by comparing with a commonly used nonlinear fitting algorithm [1].

2. THEORY AND METHOD

2.1. Signal model

Generally, a mixture of Lorentzian and Gaussian functions $S(f,\boldsymbol{p})$ is used to model MR Spectra [10]. The mathematical model can be expressed in the following form:

$$S(f,\boldsymbol{p})=\sum_{k=1}^{K}\left[L_k(f)+G_k(f)\right] \tag{1}$$

with $L_k(f)=\dfrac{a_{Lk}}{1+((f-f_{Lk})/d_{Lk})^2}$ and $G_k(f)=a_{Gk}e^{-(\frac{f-f_{Gk}}{d_{Gk}})^2}$

where L_k and G_k denote Lorentzian and Gaussian functions

respectively, f is the frequency of each data point and K stands for the number of lines used to build up the spectrum. $\boldsymbol{p}_k = [a_{Lk}, d_{Lk}, f_{Lk}, a_{Gk}, d_{Gk}, f_{Gk}]$ is a model parameter vector, where a is the intensity, d the linewidth, f the central frequency. When a possible baseline contribution $\boldsymbol{B}$ and noise $\boldsymbol{e}$ which is often assumed as Gaussian distributed are considered, a measured MR spectrum can be modelled as

$$S(f, \boldsymbol{p}) = \sum_{k=1}^{N} [L_k(f) + G_k(f)] + \boldsymbol{B}(f) + \boldsymbol{e} \qquad (2)$$

2.2. Method

In most of MRS quantification approaches, quantification can be achieved by using the Levenberg–Marquardt algorithm which is a sophisticated nonlinear least-squares algorithm to estimate the nonlinear parameter vector $\boldsymbol{p}$ in the presence of baseline $\boldsymbol{B}$ and noise $\boldsymbol{e}$ [10]. This kind of nonlinear parametric estimation methods has nontrivial requirement for the accurate parametric description of the signal and with the increase of the number of model parameters the accuracy of the estimation result reduces, besides, the performance of these methods deteriorates seriously with the presence of noise and baseline. The method proposed in this paper makes use of a linear nonparametric estimation algorithm to cope with the shortcomings.

Firstly, an overcomplete dictionary $\boldsymbol{D}_{N\times M}$ which contains M basis functions is constructed. 'Overcomplete' means that $M > N$, where N is the length of a MRS signal. Normalized Lorentzian and Gaussian function with different linewidths and central frequencies are used as basis functions. In this way, refer to the signal model (1) the spectrum of single metabolite can be represented by one or several basis functions with a certain central frequency (known a priori). For taking into account the baseline contribution $\boldsymbol{B}$, some basis functions which can linearly represent the background spectra are also contained in the dictionary. In ^{1}H MRS the baseline typically includes contributions from unresolved proteins, polypeptides, residual water and subcutaneous lipids and baseline is usually much smoother than the spectra of metabolites of interest [6]. Therefore, in general Lorentzian and Gaussian functions with bigger linewidths are supposed to model baseline [6] and here they are contained in the dictionary. According to the central frequencies and the ranges of linewidth, we classify the dictionary into several groups $\{\boldsymbol{D}_1, \boldsymbol{D}_2, \cdots, \boldsymbol{D}_K, \boldsymbol{D}_{K+1}\}$. $\boldsymbol{D}_i$ ($i = 1, 2, \cdots, K$) is composed of the basis functions with the same central frequency but different linewidths to represent metabolite spectra of interest, while $\boldsymbol{D}_{K+1}$ contains the basis functions with bigger linewidths to represent baseline. If we can find $\boldsymbol{x}$ which is the solution of $\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x} = \sum_{i=1}^{P} \boldsymbol{D}_i \boldsymbol{x}_i$, where $\boldsymbol{y}_{N\times 1}$ is a measured MR spectrum, then we can decompose the spectrum into several components which relate to spectra of different metabolites and baseline respectively.

However, if only a single observed MR spectrum is available, the problem of finding $\boldsymbol{x}$ is an underdetermined problem, infinitely many solutions exist and additional criteria must be used to select a unique estimation. As most of the underdetermined problems, the sparsity constraint is available as a priori selection criterion in the problem. For getting exact determination of sparsest representations, in the past decade several efficient pursuit algorithms for getting the sparsest representation have been proposed, such as matching pursuit (MP), the orthogonal matching pursuit (OMP), basis pursuit (BP), focal underdetermined system solver (FOCUSS) and other extensive studies of these algorithms [8]. Regularized FOCUSS algorithm in [9] which is a recursive algorithm to find the localized energy solution, has good performance in noisy environments. Considering the exist of noise in the measured MR spectra and the nonnegative character of $\boldsymbol{x}$, a nonnegative pursuit algorithm based on the regularized FOCUSS algorithm is proposed in this paper to find the solution $\boldsymbol{x}$.

2.3. Optimization

Theoretically, the sparsest representation of $\boldsymbol{y}$ in an overcomplete dictionary $\boldsymbol{D}$ is the following optimization problem:

$$\min_{\boldsymbol{x}} \| \boldsymbol{x} \|_0 \quad \text{subject to } \boldsymbol{y} = \boldsymbol{D}\boldsymbol{x} \qquad (3)$$

$$\text{or } \min_{\boldsymbol{x}} \| \boldsymbol{x} \|_0 \quad \text{subject to } \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}\|_2 \le \varepsilon \qquad (4)$$

where $\|\cdot\|_0$ is the l_0 norm, and counting the nonzero entries of a vector; $\boldsymbol{x}_{M\times 1}$ is the coefficient vector of basis functions. However, it is a NP-hard problem [8]. Thus, approximate solutions are considered instead, the regularized FOCUSS minimize $l_{(p\le 1)}$ norm in place of l_0 norm to obtain sparse solution. Then, the sparse representation becomes the solution of

$$\min_{\boldsymbol{x}} \operatorname{sgn}(p) \sum_{i=1}^{M} |x(i)|^p \quad \text{subject to } \boldsymbol{y} = \boldsymbol{D}\boldsymbol{x} \qquad (5)$$

When the noise exists, an exact minimum norm solution of (5) can not be sought. Instead, a solution that minimizes $l_{(p\le 1)}$ norm and approximately satisfies the set of constraints is found. The solution is:

$$\boldsymbol{x} = \arg\min_{x} J(\boldsymbol{x}) \quad \text{where } J(\boldsymbol{x}) = \left[\left\| \boldsymbol{D}\boldsymbol{x} - \boldsymbol{y} \right\|^2 + \gamma \mathrm{E}^{(p)}(\boldsymbol{x}) \right] \qquad (6)$$

For the problem in this paper, the basis functions in $\boldsymbol{D}$ and the MR spectrum $\boldsymbol{y}$ are nonnegative, so the coefficient vector must be nonnegative. Then, a nonnegative constraint is added to (6), and the nonnegative sparse representation is

given by

$$\min_{\boldsymbol{x}} \operatorname{sgn}(p)\sum_{i=1}^{m}\left|x(i)\right|^{p} \quad \text{subject to } \boldsymbol{y}=\boldsymbol{D}\boldsymbol{x} \text{ and } \forall i: x_i \geq 0 \quad (7)$$

At each iteration step, we set the negative values of the solution to zero for guaranteeing the nonnegative character of coefficient vector $\boldsymbol{x}$. Based on the basic iterative form of the regularized FOCUSS algorithm, the iterative form of the nonnegative pursuit algorithm used in the paper is as follow:

(a) $\boldsymbol{W}_{k+1} = \operatorname{diag}(x_k(i)^{1-(p/2)})$, $\boldsymbol{D}_{k+1} = \boldsymbol{D}\boldsymbol{W}_{k+1}$

(b) $\boldsymbol{x}_{k+1} = \boldsymbol{W}_{k+1}\boldsymbol{D}_{k+1}^{T}(\boldsymbol{D}_{k+1}\boldsymbol{D}_{k+1}^{T}+\lambda\boldsymbol{I})^{-1}\boldsymbol{y}$

(c) $x_{k+1}(i) = \begin{cases} 0 & \text{if } x_{k+1}(i) < 0 \\ x_{k+1}(i) & \text{if } x_{k+1}(i) \geq 0 \end{cases}$

The parameter λ controls the trade-off between quality of fit $\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}\|$ and the degree of sparsity, and the value of λ should increase with the level of noise.

When the solution of $\boldsymbol{x}$ is obtained, we can compute spectra of each metabolite and baseline by computing $\boldsymbol{D}_i\boldsymbol{x}_i$. In this way, spectra of different metabolites and baseline can be separated. By computing the peak area of separated spectra of interest, the quantification can be achieved.

3. EVALUATION AND RESULT

3.1. Evaluation with simulated spectra

The approach was first tested with simulated ^{1}H human brain MRS signals. Each simulated spectrum with 512 data points is composed of eleven main metabolite spectra modelled as Gaussian functions, a baseline mixing of nine models and various levels of Gaussian noise. The parameters of the main metabolite spectra and the baseline are summarized in Tab.1. For each noise level, a set of 100 MR spectra was created in order to give reliable values for the uncertainties in the estimation result and check the robustness of the method. The value of the SNR is set as the ratio of the amplitude of the NAA resonance to the standard deviation of noise $\boldsymbol{e}$.

As an a priori knowledge, the basis functions supposed to represent spectra of main metabolites have the same central frequency with main metabolite spectra, see Tab.1, and the linewidths taken from the range $[d_1, d_2]$ with the sample step Δ_{d1}. Because there is no a priori knowledge about the baseline, the basis functions, with the central frequencies and the linewidths taken from the range $[f_1, f_2]$ with a desired sample step Δ_f and the range $[d_3, d_4]$ with a desired sample step Δ_{d2} respectively, are designed to represent baseline. For the simulated ^{1}H MR spectrum, the parameters $f_1, f_2, \Delta_f, d_1, d_2, \Delta_{d1}, d_3, d_4, \Delta_{d2}$ are respectively set as 0, 400, 5, 1, 10, 0.5, 50, 350, 10.

Fig.1 plots the RRMSE (Relative Root Mean Square Error) of estimated metabolite peak areas in different noise levels to show the robustness. RRMSE is defined as the ratio of RMSE (Root Mean Square Error) of estimated result to the real data. As demonstrated in Fig.1, the quantification accuracy does not deteriorate with the increase of noise and the quantification results stay very reliable for the metabolites with big peak amplitudes. Fig.2 displays the simulated MR spectra with the SNR=17dB as in vivo conditions, and the corresponding separation results. The comparison of separation results and true spectra in Fig.2 shows that baseline was well separated from spectra of interest.

Tab.1 The parameter values of the simulated spectra of twelve metabolites and the simulated baseline.

Spectra parameters				Baseline parameters			
Metabolite	f_k (ppm)	d_k	a_k	k	f_k (ppm)	d_k	a_k
Cr	3.91	8.0	8.0	1^G	2.02	85	2.33
Glu/ Gln	3.74	4.0	4.0	2^G	2.35	100	0.33
mI	3.56	4.0	5.5	3^G	2.50	70	0.67
Tau	3.42	3.0	3.2	4^G	3.00	70	0.67
Cho	3.22	6.0	2.0	5^G	3.29	100	1.00
Cr/PCr	3.03	6.0	4.0	6^G	3.50	80	2.00
GABA	2.37	4.0	1.5	7^G	4.00	70	0.67
Glu/ Gln	2.12	6.0	9.5	8^L	0.90	300	1.33
NAA	2.02	5.0	13.0	9^L	1.30	300	1.67
Lac	1.33	5.0	1.5				
Lac	1.26	3.0	1.5				

^{G} Gaussian model. ^{L} Lorentzian model

3.2. Comparison

Using simulated MR spectra with SNR=17dB, the method in this paper is compared with the commonly used frequency-domain MRS quantification method in which Levenberg–Marquardt algorithm is used to estimate the nonlinear model parameters of metabolite spectra and a wavelet filter is used to remove the baseline component in an iterative subtraction manner [1]. As for the nonlinear method, the choice of the initial parameters can deeply influence the estimation result and in simulation experiment, we use the same prior knowledge in the two methods. The RRMSE of the estimated metabolite peak areas and the error bars in both of the methods are computed, see Fig.3. The comparison shows the superiority of our method.

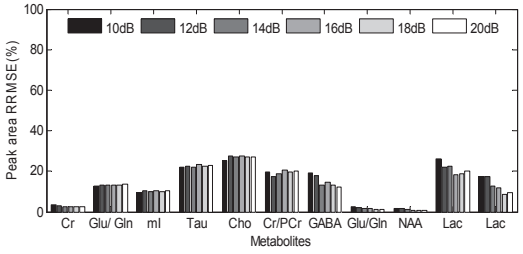

Fig.1 The RRMSE of estimated metabolite peak areas in different noise levels.

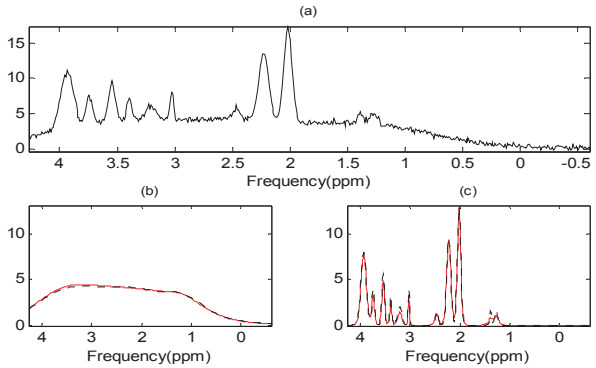

Fig.2 Separation results of a simulated MR spectrum (SNR=17dB). (a) raw spectrum; (b) the estimated and the true baseline (dashed line); (c) the estimated and the true spectra (dashed line).

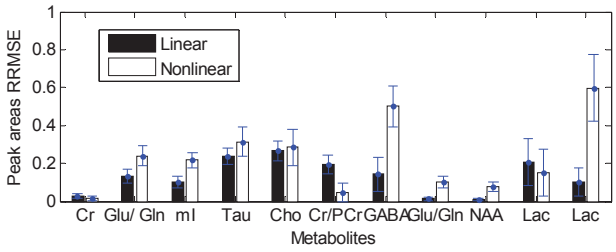

Fig.3 Comparison results. The RRMSE of estimated metabolite peak areas and the error bars in nonlinear method in [1] and the proposed method (SNR =17dB).

3.3. Real MRS data

The method was also tested on in vivo ^{1}H human brain MR spectra. Human brain MR spectra from a brain tumor patient obtained with PRESS and an echo-time of 35 ms were quantified by our method. Fig.4 (a) and Fig.4 (b) represent the MR spectra of normal tissue and tumor tissue respectively. Fig.4 (c) and Fig.4 (d) show the separation results of the baseline and the spectra of interest. According to the spectrum separation result, the peak areas of the spectra of main metabolites were computed and the difference of the peak areas of Lac in the two MR spectra shows the type of the cancer which is identical with the diagnoses of doctor.

4 CONCLUSIONS

The new MRS quantification method proposed in the paper uses a linear nonparametric algorithm based on sparse representation to separate different spectra of interest and baseline, and realize the quantification of MRS signals. The method well deals with the challenging baseline problem, effectively uses the prior knowledge and can achieve a satisfied accurate estimation. Additionally, compared with commonly used nonlinear parametric algorithm, the quantification results are more stable and the choice of initial parameters in the method has less influence on the final result. In the future work we will work at finding a more general dictionary design method and ameliorate pursuit algorithm for achieving more accurate and robust quantification.

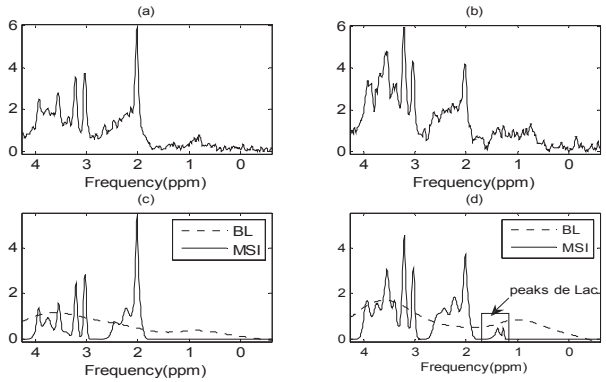

Fig.4 The separation results of in vivo ^{1}H human brain MR spectra. (a) spectrum from normal tissue; (b) spectrum from tumor tissue; (c) separation results of the spectrum in (a); (d) separation results of the spectrum in (b) (BL=baseline, MSI =spectra of interest).

5. REFERENCES

[1] P. Gillies, I. Marshall, et al., “Quantification of MRS data in the frequency domain using a wavelet filter, an approximated Voigt lineshape model and prior knowledge”. *NMR Biomed*, pp. 617–626, 2006.

[2] M.E. Moseley, Y. Cohen, et al., "Early detection of regional cerebral ischemia in cats: comparison of diffusion- and T2-weighted MRI and spectroscopy". *Magn Reson Med,* pp. 330–346, 1990.

[3] R. Bartha, D.J. Drost and P.C. Williamson, "Factors affecting the quantification of short echo in-vivo 1H MR spectra: prior knowledge, peak elimination, and filtering". *NMR Biomed*,pp. 205–216, 1999.

[4] L. Vanhamme, A. van den Boogaart and S. Van Huffel, "Improved method for accurate and efficient quantification of MRS data with use of prior knowledge". *J. Magn. Reson.*, pp. 35–43, 1997.

[5] S.W. Provencher, "Automatic quantitation of localized in vivo ^{1}H spectra with LCModel". *NMR Biomed*, pp.260–264, 2001.

[6] H. Ratiney, M. Sdika, and Y. Coenradie, "Time-domain semi-parametric estimation based on a metabolite basis set". *NMR Biomed*, pp.1–13, 2005.

[7] J. Poullet, D.M. Sima, A.W. Simonetti, et al., "An automated quantitation of short echo time MRS spectra in an open source software environment: AQSES".*NMR Biomed*, pp.493–504,2007

[8] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation". *IEEE Trans. Sig. Process.*, pp.4311-4322, 2006.

[9] B.D. Rao, K. Engan, S.F. Cotter, et al., “Subset selection in noise based on diversity measure minimization”. *IEEE Trans. Sig. Process.*, pp. 760–770, 2003.

[10] S. Mierisova, M. Ala-korpela, "MR spectroscopy quantitation: a review of frequency domain methods". *NMR Biomed*, pp.247–259, 2001.