Modeling the macromolecular background in Nuclear Magnetic Resonance spectroscopic signals

D.M. Sima[1], A.M. Rodríguez Díaz[1] and S. Van Huffel[1]

[1] Katholieke Universiteit Leuven, Dept. Electrical Engineering (ESAT-SCD), Leuven, Belgium

Abstract—**Metabolite quantitation of Nuclear Magnetic Resonance spectroscopic signals is hampered by the presence of a smoothly varying background signal. Quantitation methods such as QUEST [Ratiney et al., *NMR Biomed.* 18(1), 1-13. 2005] remove this background in a preprocessing step that requires signal denoising. A comparison between three denoising methods shows that one can use spline smoothing or wavelet denoising instead of subspace-based modeling of the background signal, with equally good quantitation results.**

Keywords—**Nuclear Magnetic Resonance Spectroscopy, denoising, nonparametric modeling, macromolecular background.**

I. Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy can be used as a non-invasive technique to quantify metabolite concentrations in human tissue. Metabolite quantitation is often tackled as a fitting problem between the time-domain NMR signal (or the corresponding spectrum in the frequency domain) and a theoretical mathematical model. Short echo-time NMR signals are usually severely hampered by a nuisance background, coming from some unknown macromolecules present in the tissue under investigation, for which a correct mathematical model is not available. Macromolecular signals are characterized by broad spectral lines, which often overlap in the frequency domain with metabolite components.

This study is motivated by the a time-domain quantitation method QUEST [1], which uses a metabolite basis set built up from quantum mechanically simulated spectra or from *in vitro* spectra. The use of a metabolite basis set facilitates the disentangling of overlapping resonances when the corresponding metabolite profiles also contain at least one non-overlapping resonance. However, the macromolecular background is usually strongly overlapping with some metabolite components, and a typical assumption aimed at disentangling it from the metabolites is the fact that the macromolecular signal decays much faster that the metabolites of interest; thus the macromolecular background is concentrated in the first few data points of the time-domain NMR signal.

The procedure Subtract-QUEST [1,2] goes as follows:

- Truncate an appropriate number of points from the beginning of the signal and fit the (background-free) shorter signal as a combination of metabolites from the basis set.
- Back-extrapolate the mathematical model towards the begin point.
- Subtract the modeled signal from the original one; the residual is considered a noisy version of the background signal.
- Denoise the noisy background.
- Subtract the noise-free background from the original signal.
- Fit the obtained (full-length and background-free) signal to a combination of metabolites from the basis set.

As seen above, a noisy background signal is obtained in QUEST, which needs to be denoised before the final quantitation. In [1,2] this is done by fitting several Lorentzian peaks. Here we compare 3 different methods (based on splines, wavelets and Lorentzians) for modelling the macromolecular background. We focus on the robustness with respect to hyperparameters, *i.e.*, the design parameters that must be tuned in order to obtain good performance of each method.

II. Materials and methods

A. Methods for denoising the background

In this study we consider the parameterization with Lorentzians as one of the denoising methods, but we also analyze denoising the background by means of spline and wavelets. First, a preliminary Monte Carlo study on the 3 considered denoising schemes is performed. Simulated background signals are used to estimate the performance of each denoising method with respect to various hyperparameters and a whole range of signal-to-noise ratios (SNRs). We start from a simulated noiseless complex-valued background to which we add white noise of various SNRs ranging from 5 to 30. The background spectrum was simulated as suggested in [3] as a sum of 6 overlapping Gaussians (fig.1).

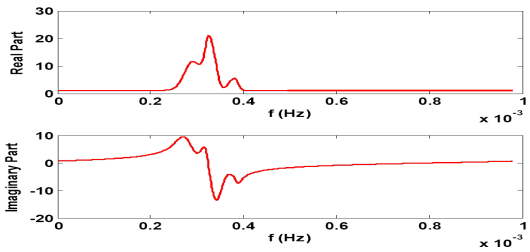

Fig. 1 Simulated noiseless background spectrum.

The criterion used for evaluating the three denoising methods (described below) is maximizing the Relative SNR (RSNR), which shows the SNR improvement. If the noiseless signal is denoted by s and the denoised signal by s_D, each being (possibly complex-valued) vectors of length N, then the RSNR is computed as

$$RSNR = 10\log_{10}\left(\frac{\sum_{n=1}^{N}|s(n)|^2}{\sum_{n=1}^{N}|s(n)-s_D(n)|^2}\right). \qquad (1)$$

Smoothing **splines** are piecewise polynomial functions that can be used to approximately fit a noisy signal as a linear combination of splines. The degree of smoothness in this fit is decided either by tuning the number of spline components or by imposing a bound on the allowed roughness of the fitted signal. We focus on the latter approach, called *penalized splines* [4], where a single hyperparameter (the smoothing parameter) must be tuned; this parameter gives the weighting between the approximation error and the roughness measure. Our study aims at finding for each SNR the range of the smoothing parameter p that gives the best RSNR.

Wavelets [5] can also be used for signal denoising. A signal is decomposed into a combination of building blocks, which are translated and/or dilated versions of the same basic function called *mother wavelet*. Some of the obtained wavelet coefficients correspond to details in the data set (such as the noise), while others correspond to the big features in the data. Wavelet *thresholding*, means setting to zero all coefficients that are less than a particular threshold, and use an inverse wavelet transformation to reconstruct the denoised data. For choosing the best wavelet from a variety of wavelet families and hyperparameters a rigorous and very extensive procedure was followed. Automatic wavelet denoising is possible if the following hyperparameters are chosen: the wavelet family (*i.e.*, the shape of the mother wavelet), the number of vanishing moments, the level of wavelet decomposition, the threshold selection rule, the choice between soft or hard thresholding, and the multiplicative threshold rescaling rule.

In order to model the background with **Lorentzians**, we use the subspace-based method HSVD, in particular the HLSVDPRO implementation [2]. HSVD requires the computation of the truncated singular value decomposition (SVD) of a Hankel matrix H, containing the noisy signal. If the signal is a sum of K exponentially decaying sinusoids (giving K Lorentzian peaks, when the signal is Fourier transformed to the frequency domain), then the data matrix H has rank exactly equal to K. Noise in the signal or a signal not satisfying the exponentially decaying model can still be approximately modeled as a sum of Lorentzians, by truncating the SVD to rank K. In this study, the aim is finding the ideal number K of Lorentzian components that are required in order to obtain the highest RSNR for each SNR.

B. Monte Carlo simulations for quantitation

The 3 different denoising approaches described above can be inserted into the background denoising step of the metabolite quantitation method QUEST. The following simulation protocol was used in order to test how the choice of denoising method influences the final quantitation results.

A basis set consisting of 11 metabolite signals (measured *in vitro*) was used to create a noiseless simulation signal as:

$$y(t) = \sum_{k=1}^{11} a_k \exp(i\varphi_k - d_k t + i f_k t)\, met_k(t) \qquad (2)$$

Here the 11 random real amplitudes a_k, phase shifts φ_k, damping perturbations d_k, and frequency shifts f_k were generated within the following values: $0 \le a_k \le 100$, $-\pi \le \varphi_k \le \pi$, $0 \le d_k \le 0.01$, $-0.01 \le f_k \le 0.01$. The amplitude values, as well as the phase values, are arbitrary real values. The damping and the frequency shifts take quite small values, to mimic the perturbations that may occur in practice; these perturbations come from the difference in the acquisition of *in vivo* measurements and of *in vitro* metabolite measurements. Finally, a simulated background signal (fig. 1) and white noise are added to the simulated signal. An illustration for SNR 10 is given in fig.2.

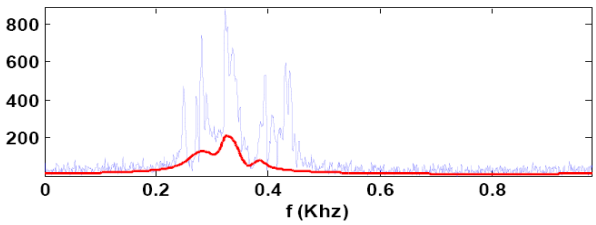

Fig. 2. Thin line: absolute value of the Fourier transform of a simulated test signal, with added background and added white noise (SNR=10). Thick line: absolute value of the Fourier transform of the simulated background signal.

To assess the performance of the quantitation, Monte-Carlo simulations are performed in which the statistics are computed using the same noise-free data set but different realizations of the noise at various SNRs. We evaluate the quality of the results by computing the relative errors between the true amplitude values and the amplitude values estimated after quantitation with Subtract-QUEST combined with each of the 3 denoising approaches.

III. Results

A. Monte Carlo simulations for denoising the background

Our **spline** study aimed at finding the range of the smoothing parameter p that gives the best RSNR (1) for each SNR. The results show that the optimal p is proportional to the SNR. Moreover, the standard deviation for the optimal p grows with the SNR. We also found that the improvement in terms of RSNR is linear. The improvement is bigger when the signal is noisier, but in general the global improvement is around 10dB.

The optimal results for each family of **wavelets** in terms of RSNR are quite similar regardless of the wavelet family. The best vanishing moment for each family of wavelets and the level of wavelet decomposition do not depend on the SNR. The best thresholding criterion is Soft Thresholding. The average improvement is also around 10dB in the denoising process.

For the third approach, we focused on choosing the number of **Lorentzian** components. Except for the simulations with huge noise, the RSNR increased with increasing number of Lorentzian components. Thus we devised the rule for choosing an adequate number of Lorentzians as taking the lowest number that delivers higher RSNR than the SNR considered in each case. Lorentzians fit perfectly the flat regions in the spectrum when the noise is not huge, but they do not fit the non-flat regions as good as splines and wavelets. Their average improvement in terms of RSNR is 5 dB.

B. Monte Carlo results for quantitation

We have performed two complete sets of Monte Carlo simulations for quantitating the signals described in II.B. Each set consisted of 100 noisy signals for each SNR level ranging from 5 to 30. For the first set of simulations we fixed the hyperparameters at the optimal values predicted for each SNR by the results in III.A, while for the second set of simulations we fixed the hyperparameters at non-optimal values.

When hyperparameters are optimally chosen, we observe for all denoising methods that the relative errors on the amplitudes are reasonably low, and, as expected, the trend is having smaller average relative errors for higher values of SNR. For illustration, at SNR = 30, the median of the relative errors of 100 simulations varies between 1% and 10%, while at SNR = 5, the median of the relative errors varies between 10% and 20%.

Slight variations in the values of the hyperpararameters do not influence very severely the quantitation results. We illustrate the following non-optimal choice of hyperparameters: for splines, the regularization parameter was chosen 10 times larger than the considered optimal value; for wavelets, the number of vanishing moments was half compared to the optimal value, a low decomposition level was fixed, and hard thresholding was used; for Lorentzians, we considered a fixed number of 10 components.

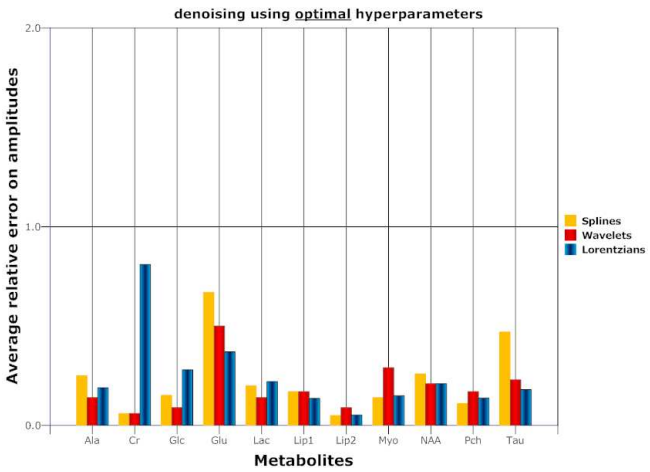

Fig. 3: Quantitation results using optimal denoising hyperparameters.

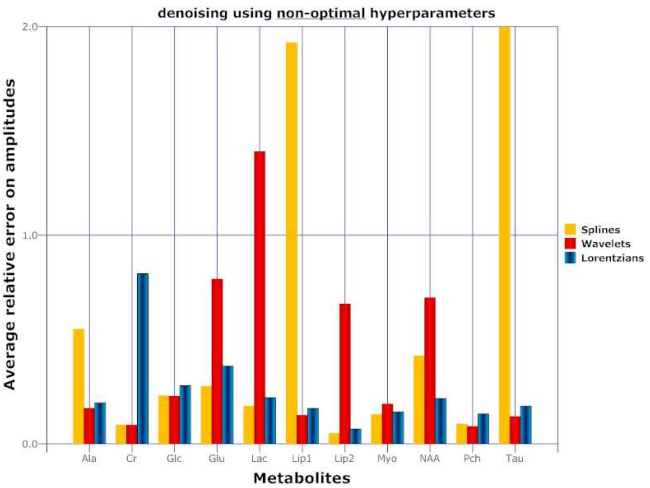

Fig. 4: Quantitation results using non-optimal denoising hyperparameters.

Figures 3 and 4 show quantitation results in terms of relative errors on the amplitudes for each metabolite, averaged over 100 noisy simulations at SNR=15. The optimal smoothing spline parameter is 0.005, the chosen wavelet family is *symlets* with 10 vanishing moments, and the optimal number of Lorentzians is 16.

IV. Discussion

Various preprocessing approaches have been developed for the problem of removing the macromolecular background from short echo-time NMR signals. The simplest one, based on the fact that the macromolecular components decay more rapidly than the metabolites, is to truncate some of the initial points in the time domain (also called the 'Trunc' method by Ratiney et al. [2]). This technique presents some drawbacks: the useful information is partially lost, and selecting the number of points to be truncated is difficult. More advanced techniques consist of subtracting a non-parametrically modeled macromolecular background in the frequency domain from the original spectrum. Although these methods have been shown to be rather successful, they require an additional step prior to quantitation. The background can also be modeled in the quantitation step. Profiles of macromolecular components obtained from measurements or from theoretical considerations are added to the database of metabolites; then one can optimize background and metabolites in a single pass.

A comparison between wavelets and splines for modeling the background has already been done in [7], with no significant differences found between the two. Choosing the best wavelet family for NMR signal denoising was studied in [8]. Our Monte Carlo procedure in III.A is similar the one in [8], but we have extended it to a wider range of possible hyperparameters.

Each considered denoising method can either smoothen well or smoothen badly the noisy background, depending on the choice of hyperparameters. With optimally chosen hyperparameters, the considered denoising methods give almost equally accurate quantitation results in combination with Subtract-QUEST. Misspecification of hyperparameters has a negative effect on the final quantitation, but we notice the following issues: The smoothing parameter of splines has an easily predictable range of optimal values that is a function of the SNR. Wavelets seem to have a huge number of hyperparameters, but many of them could be eliminated since they are not influential (*e.g.*, several wavelet families, or thresholding rules are equally good). For all methods, smoothening too much is worse than smoothening too little, thus overestimating the number of Lorentzians is acceptable. Finally, optimizing the denoising step should be coupled with optimizing the number of truncated points in the first step of QUEST, which is an essential choice for good quantitation.

V. Conclusions

Our study proves that spline- or wavelet-based modeling of the macromolecular background can replace the Lorentzian-based modeling in a quantitation method such as QUEST.

Acknowledgment

Dr. Diana M. Sima is a postdoctoral fellow of the Fund for Scientific Research-Flanders. Dr. Sabine Van Huffel is a full professor at the Katholieke Universiteit Leuven, Belgium. Research supported by:
Research Council KUL: GOA-AMBioRICS, CoE EF/05/006 Optimization in Engineering (OPTEC), IDO 05/010 EEG-fMRI, IOF-KP06/11 FunCopt, several PhD/postdoc & fellow grants; Flemish Government: FWO: PhD/postdoc grants, projects, G.0407.02 (support vector machines), G.0360.05 (EEG, Epileptic), G.0519.06 (Noninvasive brain oxygenation), FWO-G.0321.06 (Tensors/Spectral Analysis), G.0302.07 (SVM), G.0341.07 (Data fusion), research communities (ICCoS, ANMMM); IWT: TBM070713-Accelero, TBM-IOTA3, PhD Grants; Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, `Dynamical systems, control and optimization', 2007-2011); EU: BIOPATTERN (FP6-2002-IST 508803), ETUMOUR (FP6-2002-LIFESCIHEALTH 503094), Healthagents (IST–2004–27214), FAST (FP6-MC-RTN-035801), Neuromath (COST-BM0601); ESA: Cardiovascular Control (Prodex-8 C90242)

References

[1] Ratiney H, Sdika M, Coenradie Y, Cavassila S, van Ormondt D, Graveron-Demilly D (2005), Time-domain semi-parametric estimation based on a metabolite basis set. NMR Biomed 18:1-13

[2] Ratiney H, Coenradie Y, Cavassila S, van Ormondt D, Graveron-Demilly D (2004) Time-domain quantitation of ^{1}H short echo-time signals: Background accommodation. MAGMA 16:284-296

[3] Seeger U, Klose U, Mader I, Grodd W, Nagele T (2003) Parameterized evaluation of macromolecules and lipids in proton MR spectroscopy of brain diseases. Magn Res Med 49:19-28

[4] Eilers P, Marx B (1997) Flexible smoothing with B-splines and penalties. Stat Sci 11:89-121

[5] Daubechies I (1992) 10 Lectures on Wavelets. SIAM

[6] Laudadio T, Mastronardi N, Vanhamme L, Van Hecke P, Van Huffel S (2002) Improved Lanczos algorithms for blackbox MRS data quantitation. J Magn Reson 157:292-297

[7] Soher B, Young K, Maudsley A (2001) Representation of strong baseline contributions in ^{1}H MR spectra, Magn Reson Med 45:966-972

[8] Cancino-De-Greiff H, Ramos-Garcia R, Lorenzo-Ginori J (2002) Signal de-noising in magnetic resonance spectroscopy using wavelet transforms. Concepts Magn Reson 14: 388-401

Author: D.M. Sima
Institute: K.U. Leuven, Dept. E.E. ESAT-SCD
Street: Kasteelpark Arenberg 10, bus 2446, B-3001 Heverlee
City: Leuven
Country: Belgium
Email: Diana.Sima@esat.kuleuven.be