Principal Component Analysis versus Hankel Total Least Squares in Magnetic Resonance Spectroscopic Quantitation

S. Van Huffel, Y. Wang, L. Vanhamme and N. Mastronardi

ESAT Laboratory, Dept. of Electrical Engineering, Katholieke Universiteit Leuven, Kard. Mercierlaan 94, 3001 Leuven, Belgium.

INTRODUCTION

In this paper, a careful comparison between the Principal Component Analysis (PCA) and the Hankel Total Least Squares (HTLS) based methods for the automated quantitation of resonances in large sets of MR spectra is presented. S. Van Huffel et al. [1] applied the total least squares method to MRS data quantitation. Their algorithm, called HTLS, was generalized [2] to the quantitation of sets of MR spectra, resulting into the HTLSstack and HTLSsum algorithms.

PCA was introduced in MRS data quantitation by R. Stoyanova et al. [3]. It has been shown that under some conditions PCA can successfully extract quantitative metabolite information from data sets without any prior knowledge about the line shape [3][5]. By using complex data, PCA can be directly applied to unphased data sets and moreover improves peak area estimation with a factor $\sqrt{2}$ [4]. Previous studies using PCA were performed in the frequency domain and could only relatively quantify the MRS data if complex data are used [4]. By choosing an appropriate normalization of the first PC, we show here how to quantify peak areas absolutely in the frequency domain and how to perform PCA in the time domain. In addition, we show under which conditions the HTLS based methods outperform PCA and vice versa.

THEORY

Suppose we have P MRS signals $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_P$ of N data points each. Let $\bar{\mathbf{y}} = \mathbf{y}_1 + \mathbf{y}_2 + \cdots + \mathbf{y}_P$. We can apply the HTLS algorithm [1] to the sum $\bar{\mathbf{y}}$ in order to obtain the signal poles and afterwards calculate the corresponding amplitudes and phases for each signal separately. This algorithm is called HTLSsum.

The HTLSstack algorithm [2] arranges the N data points of each MRS signal $\mathbf{y}_p$ into a Hankel matrix H_p, $p = 1, 2, \cdots, P$. These P matrices are stacked into $H_{stack} = \begin{bmatrix} H_1 & H_2 & \cdots & H_P \end{bmatrix}$. Again we apply HTLS [1] to H_{stack} in order to quantify all signal poles for the entire set and then compute the corresponding amplitudes and phases for each signal separately.

PCA is usually performed by means of an eigenvalue decomposition. However, the use of the complex SVD leads to more accurate parameter estimates. This involves arranging the P signals $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_P$ as rows in a matrix D and decomposing D as $D = U\Sigma V^H = SV^H$ (^{H} denotes conjugate transpose). The columns of V contain the basic shapes of the signals, called principal components (PC). The elements in $S = U\Sigma$ are called the scores of each PC.

Now, suppose our set of P MRS signals is composed of a single resonance (peak). Each signal $\mathbf{y}_p$ is then expressed as : $\mathbf{y}_p = s_{1p}\mathbf{v}_1$, for $p = 1, \cdots, P$. $\mathbf{s}_1$ and $\mathbf{v}_1$ are respectively the first column of S and V with elements s_{i1} and v_{i1}.

Using time domain data, the first PC $\mathbf{v}_1$ needs to be normalized in such a way that it represents a unit amplitude signal. Hereto we divide $\mathbf{v}_1$ by its first element v_{11}. The modified scores $\tilde{\mathbf{s}} = v_{11}\mathbf{s}_1$ then represent the absolute amplitudes of the original signals. Using frequency domain data, obtained by applying the DFT to the rows of MRS signals, we need to normalize $\mathbf{v}_1$ so that it has unit area, i.e., $\Sigma_{i=1}^{N} v_{i1} = 1$. The modified scores $\tilde{\mathbf{s}} = (\Sigma_{i=1}^{N} v_{i1})\mathbf{s}_1$ then represent the absolute areas under the spectra, which are equal to the amplitudes of the signals in the time domain.

METHODS

First, we generate a complex signal of 128 data points composed of one exponentially damped sinusoid (frequency = 260 Hz, damping factor = 100 Hz, amplitude = 100 arbitrary units (a.u.), phase = 0°). This signal is subsequently multiplied by 1 to 91 to produce a data set which contains 91 signals having the same Lorentzian line shape. Random noise with variance 5 and mean 0 is added to the real and imaginary parts of the signal. The test is repeated 100 times.

Second, we repeat the experiment for a Gaussian line shape. The HTLS based methods approximate this line shape by a linear combination of several Lorentzians (e.g. 5, as chosen in Figure 2).

RESULTS

Figure 1 shows the average relative error of the peak area of a Lorentzian line shape as estimated by HTLS, HTLSstack, HTLSsum and PCA as a function of the signal-to-noise ratio (SNR). HTLSstack and HTLSsum perform the best. Only at low SNR, PCA is better than HTLS.

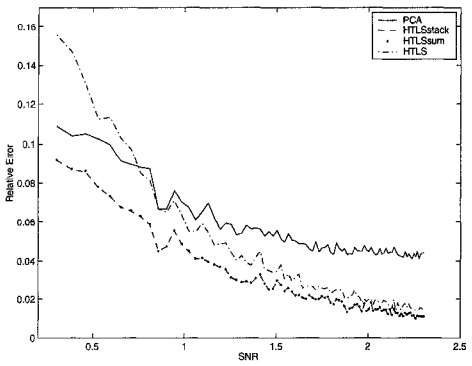

Figure 1: Lorentzian line shape

Figure 2 shows the advantage of PCA at very low SNR for a Gaussian line shape. Without any prior knowledge about the model function, HTLS based algorithms perform not as well as PCA at very low SNR, although HTLSstack and HTLSsum yield better results than PCA at high SNR.

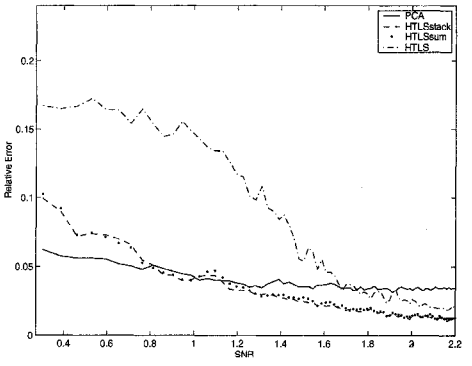

Figure 2: Gaussian line shape

CONCLUSIONS

PCA demonstrates good performance both in terms of accuracy and computational efficiency for MRS data quantitation, but is not suitable for direct quantitation of more than one line shape. HTLSstack and HTLSsum simultaneously quantify multiple line shapes and outperform PCA, in particular when the line shape is Lorentzian, since they exploit this prior knowledge. Only at low SNR, PCA outperforms the HTLS based methods if the line shape is non-Lorentzian. The same results hold for unphased sets.

REFERENCE

[1]. S. Van Huffel, et al., JMR A 110, 228-237 (1994).
[2]. L. Vanhamme, et al., Proc. SPIE 3461, 237-248 (1998).
[3]. R. Stoyanova, et al., JMR A 115, 265-269 (1995).
[4]. M. A. Elliott, et al., MRM 41, 450-455 (1999).
[5]. A. C. Kuesel, et al., NMR Biomed. 9: 93-104 (1996).