

一种基于复数主成分分析的磁共振波谱量化预处理方法

曾卫明 梁晓云 罗立民

(东南大学生物医学工程系 南京 210096)

摘要: 该文提出了一种应用复数主成分分析方法(CPCA)来获取波谱基本波型, 自动校正频率和衰减系数偏移的方法。该方法首先对频率进行校正, 然后拓宽 CPCA 的应用范围, 把它使用到衰减系数偏移校正领域。文中的实现结果证明: 该方法较好地实现了频率和衰减系数的校正, 在同实域 PCA 方法性能比较上, 前者的计算速度和估计精度明显占优。

关键词: 复数主成分分析, 磁共振波谱, 频率校正, 衰减系数校正

中图分类号: TN391.4, R8 **文献标识码:** A **文章编号:** 1009-5896(2005)01-0051-05

A Preprocessing Method for Quantitation of Magnetic Resonance Spectroscopy Based on Complex Principal Component Analysis

Zeng Wei-ming Liang Xiao-yun Luo Li-min

(Department of Biomedical Engineering, Southeast University, Nanjing 210096, China)

Abstract This paper proposes a method that can automatically obtain common lineshape in the spectrum data set based on Complex Principal Component Analysis(CPCA), which can correct frequency shifts and damping factor shifts effectively. This method aligns frequency at first, then extends CPCA to the correction of damping factor shifts in a single resonant peak across spectra. Through this method, frequency and damping factor alignment are successfully achieved. In terms of computation time and estimation precision this method is significantly better than that of PCA.

Key words Complex Principal Component Analysis(CPCA), Magnetic Resonance Spectroscopy(MRS), Frequency alignment, Damping factor alignment

1 引言

磁共振波谱(MRS)可用于无创伤地检测身体的主要代谢产物, 对预防和控制疾病, 探测大脑等许多方面有很重要的应用^[1-3]。但是, 由于涉及到的研究领域非常广泛, 实际中对 MRS 的分析和处理相当复杂。本文就是对其中的一个重要领域, 即谱模型参数的校正展开研究。MRS 的时间系列信号通常可用如下 Lorentz 函数表达^[4,5], 即

$$y_n = \sum_{k=1}^K a_k e^{[-d_k t'_n + j(2\pi f_k t'_n + \phi_k)]} + e_n, \quad n = 0, 1, \dots, N-1 \quad (1)$$

其中 $j = \sqrt{-1}$, e_n 是高斯噪声, f_k, d_k, a_k 及 ϕ_k 是 MRS 中重要的 4 个参数, 分别对应谱中 k 峰的频率、衰减系数、幅度以及相位。这些参数代表了组织器官重要的分子信息, 对它们的精确量化, 是充分发挥 MRS 作用的关键。然而, 实际中由于磁场不均匀, 设备不完善, 参数值往往出现较大的偏移失真, 因此, 最大限度地除去这些偏移失真就成为 MRS 信号量化中的一项重要预处理任务。本文就是基于复数主成分分析(CPCA)方法对此进行论述。

PCA 是一种对多变量巨型数据非常有名的统计分析技术^[6]。1995 年, Stoyanova 等^[6]开始把 PCA 应用到磁共振波谱量化领域; Brown、Witjes 等^[7-10]进一步拓宽了 PCA 的应用, 实现了利用 PCA 对磁共振谱参数的校正; Elliott、Huffeld 等^[11,12]又把 PCA 的应用推广到了复数域, 并说明了复数 PCA(CPCA)分析比 PCA 更为有效。以上文献都在一定范围内对 MRS 的后处理及分析发挥着重要作用, 但基于 PCA 的分析应该延伸到复数域。因为复数域多了一个虚部, 就如同 MRS 数据集中谱的数量增加了一倍, 显然, 谱越多就越容易发挥 PCA 分析的作用; 另外, 考虑到 MRS 模型参数中, 幅度是线性参数, 可以在其它参数校正后直接量化得到, 又复数域分析对相移不敏感^[11], 所以在复数域量化过程中, 相位一般可以不必校正^[12], 但频率和衰减系数通常都必须校正处理, 而现有文献中衰减系数的偏移校正很少有具体的研究, 我们认为这需要改进, 主要原因是衰减系数的校正最为复杂, 而决不是文献[10, 11]等涉及到的频率和相位校正的简单推广。 $e^{-d_k t}$ 的模不为 1, 它有放大噪声和偏

差的副作用,因此不宜采用迭代的方法,校正时应该居于最后步骤。根据上面的分析,我们在前人的基础上,提出了一种基于CPCA的频率和衰减系数自动校正的方法,该方法主要分为两部分:利用CPCA迭代校正频率的偏移;再利用CPCA一次实现对衰减系数偏移的校正。

2 频率校正

设MRS数据集为 D ,经奇异值分解,可以表示为^[6,7,10]

$$D = SP \quad (2)$$

其中, $P = \{P_1, P_2, \dots\}$ 为CPCA方法中各主成分组成的集合, P_1 为第一主成分,对应谱系列中的基本线型函数,即参数未发生偏移时的波谱线型, P_2 等主成分对应谱中的偏差和噪声成分,而 S 为对应的权值。

令实际中频率和衰减系数偏移分别为 Δw_n 和 Δd_n ,则式(2)可改写为

$$D = Cf(w_k + \Delta w_n, d_n + \Delta d_n) \quad (3)$$

其中 C 为系数, w_k 为频率, d_n 为谱 n 时的衰减系数, $f(w_k + \Delta w_n, d_n + \Delta d_n)$ 为有参数偏移时的线型函数。式(3)泰勒展开为

$$D = C \left(f(w_k, d_n) + \frac{df(w_k, d_n)}{d(w)} \Delta w + \frac{df(w_k, d_n)}{d(d)} \Delta d + \dots \right) \quad (4)$$

考虑频率和衰减系数的偏移通常较小,于是只取式(4)中的低价展开式,即得

$$D' = C \left(f(w_k, d_n) + \frac{df(w_k, d_n)}{d(w)} \Delta w + \frac{df(w_k, d_n)}{d(d)} \Delta d \right) \quad (5)$$

结合式(2)得

$$S'P' = C \left(f(w_k, d_n) + \frac{df(w_k, d_n)}{d(w)} \Delta w + \frac{df(w_k, d_n)}{d(d)} \Delta d \right) \quad (6)$$

其中 $P' = [P_1, P_2, P_3]^T$ 为主成分的集合, $S' = [S_1, S_2, S_3]$ 为对应的权值,则式(6)为

$$[S_1, S_2, S_3][P_1, P_2, P_3]^T = C[1, \Delta w, \Delta d] \cdot \left[f(w_k, d_n), \frac{df(w_k, d_n)}{d(w)}, \frac{df(w_k, d_n)}{d(d)} \right]^T \quad (7)$$

定义

$$F = \left[f(w_k, d_n), \frac{df(w_k, d_n)}{d(w)}, \frac{df(w_k, d_n)}{d(d)} \right]^T \quad (8)$$

$$B = [S_1, S_2, S_3][P_1, P_2, P_3]^T$$

则式(7)解为

$$[1, \Delta w, \Delta d] = \frac{1}{C} BF^T (FF^T)^{-1} \quad (9)$$

令 $Z = (1/C) BF^T (FF^T)^{-1}$, 则

$$\Delta w = Z_2 / Z_1 \quad (10)$$

其中, Z_1, Z_2 分别为 Z 的第1,第2列列向量。

3 衰减系数的校正

设经过频率校正后,MRS的频谱数据集变为 \bar{D} 。实际中,当各谱的衰减系数偏移为 Δd_n 时,则式(3)可变为

$$\bar{D} = \bar{C} f(d_n + \Delta d_n) \quad (11)$$

式(11)泰勒展开为

$$\bar{D} = \bar{C} \left(f(d_n) + \frac{df(d_n)}{d(d)} \Delta d + \frac{d^2 f(d_n) (\Delta d)^2}{d(d^2) 2} + \dots \right) \quad (12)$$

这里也只取一价和二价的低价泰勒展开式,即得

$$\bar{D}' = \bar{C} \left(f(d_n) + \frac{df(d_n)}{d(d)} \Delta d + \frac{d^2 f(d_n) (\Delta d)^2}{d(d^2) 2} \right) \quad (13)$$

对应式(6),上式变为

$$\bar{S}' \bar{P}' = \bar{C} \left(f(d_n) + \frac{df(d_n)}{d(d)} \Delta d + \frac{d^2 f(d_n) (\Delta d)^2}{d(d^2) 2} \right) \quad (14)$$

其中 $\bar{P}' = [\bar{P}_1, \bar{P}_2, \bar{P}_3]^T$, \bar{S}' 为对应的幅度,则式(14)变为

$$[\bar{S}_1, \bar{S}_2, \bar{S}_3] [\bar{P}_1, \bar{P}_2, \bar{P}_3]^T = \bar{C} \left[1, \Delta d, \frac{(\Delta d)^2}{2} \right] \cdot \left[f(d_n), \frac{df(d_n)}{d(d)}, \frac{d^2 f(d_n)}{d(d^2)} \right]^T \quad (15)$$

定义

$$\bar{F} = \left[f(d_n), \frac{df(d_n)}{d(d)}, \frac{d^2 f(d_n)}{d(d^2)} \right]^T \quad (16)$$

$$\bar{B} = [\bar{S}_1, \bar{S}_2, \bar{S}_3][\bar{P}_1, \bar{P}_2, \bar{P}_3]^T$$

则式(15)解为

$$\left[1, \Delta d, \frac{(\Delta d)^2}{2} \right] = \frac{1}{\bar{C}} \bar{B} \bar{F}^T (\bar{F} \bar{F}^T)^{-1} \quad (17)$$

令 $\bar{Z} = \frac{1}{\bar{C}} \bar{B} \bar{F}^T (\bar{F} \bar{F}^T)^{-1}$, 则

$$\Delta d = \frac{\bar{Z}_2}{\bar{Z}_1} \quad (18)$$

其中 \bar{Z}_1, \bar{Z}_2 分别为 \bar{Z} 的第 1, 第 2 列列向量。

4 方法的实现

上面两节分别给出了频率和衰减系数校正的数学原理，根据式(9), (10)及式(17), (18)，求取频率和衰减系数偏移的关键就是获得 F 和 \bar{F} 。

根据式(8)， F 由 3 个待求成分组成， $f(w_k, d_n)$ 可由 D 的归一化 P_i 来近似，即^[6]

$$f = \frac{P_i}{\sum_{j=1}^k P_{ij}} \tag{19}$$

$df(w_k, d_n)/d(w)$ 可用平均差分法求得；而剩下的 $df(w_k, d_n)/d(d)$ ，其计算方法类似文献[10]，但由于推广为复数域，故需要作较大的改进，其步骤为：对 f 进行快速傅里叶逆变换，得时域信号 t_1 ；再结合式(1)，应用 1Hz 洛仑茨 (Lorentz) 滤波得 t_2 ^[10]；然后计算时域一价变化成分 t ；最后对 t 进行快速傅里叶变换而得。 \bar{F} 的求法类似 F 。

5 数据论证

5.1 模拟数据

式(1)中，取 $K=1$ ，抽样点数为 100，幅度取 30 个单位，相位定为 0，频率和衰减系数都取 50Hz 并随机有 -10 ~ 10Hz 的偏移，然后添加复数标准正态噪声，并依次重复 256 次得到有 256 个时间系列的数据集。下面给出本文方法对该数据集实现的结果及有关分析。

图 1 中(a), (b)和(c)分别为模拟数据中的前 3 个谱线，频率校正后对应的谱线以及频率和衰减系数校正后对应的谱线。

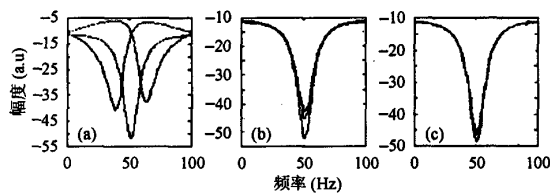


图 1

从图 1(a), 图 1(b)可以看出，模拟数据谱线之间存在较大的频率偏移，而经过基于 CPCA 方法的频率校正后，该类偏移几乎完全得到消除。图 2 则给出了更直观的结果。图 2 中(a)-(h)分别表示第 1~8 次迭代中实际的频率偏移值(x 坐标)与其估计值(y 坐标)对应的情况，所以经过迭代收敛后，偏移估计已相当正确。下面，为进一步与 PCA 方法比较，我们给出如图 3 所视的实现结果。

由图 3 知，PCA 方法要经过 17 次迭代计算才能收敛，所以其计算速度比 CPCA 方法显然要慢；另外，由表 1，表 2 知，PCA 方法的估计精度明显要低于 CPCA 方法，由此看出本文立足 CPCA 方法研究的合理性。

接下来，我们又给出本文方法实现的衰减系数校正结

果，如图 1(c)所示。所以，从图(1)的 3 幅小图比较看，CPCA 方法也较好地实现了衰减系数偏移的校正。

5.2 活体 MRS 数据

图 4 展示了一幅 MR (Magnetic Resonance) 图像。中间的 16×16 方格指示出获取 MRSI (Magnetic Resonance Spectroscopy Imaging) 信号的位置。由于水成分、噪声等影响，其中仅在方框内可能有需要的信息。实验中，我们是以其中未被较大破坏的 23 个谱信号作为研究对象。数据抽样间隔为 1ms，由 1.5T Siemens Vision System 生成(MR 图像及其原始 MRSI 时间系列数据由比利时的 SISTA (Signals, Identification, System Theory and Automation) 的 Leentje Vanhamme 博士提供)。

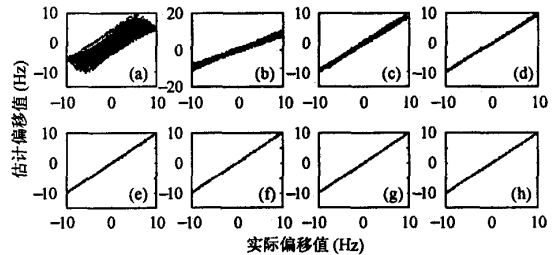


图 2 CPCA 实现的频率估计偏移值和实际偏移值((a)-(h)分别对应第 1~8 次迭代结果)

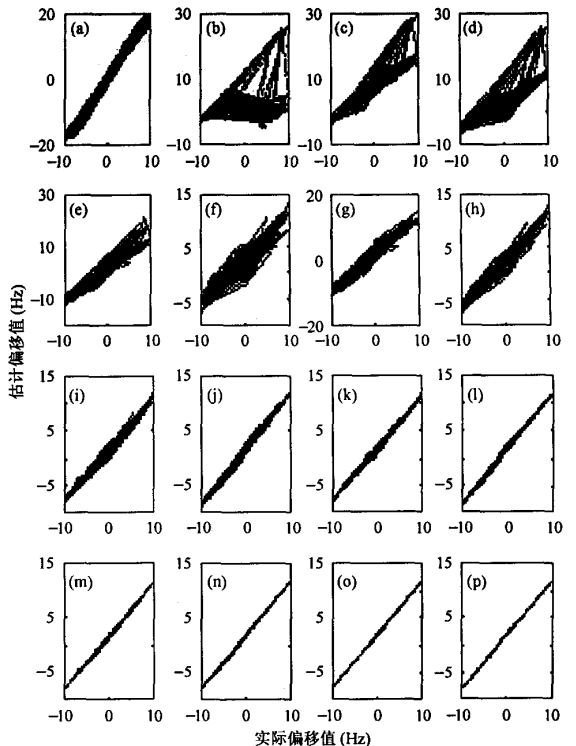


图 3 PCA 实现的频率估计偏移值和实际偏移值 ((a)-(h)分别对应第 1~8 次迭代结果；(i)-(p) 分别对应第 10~17 次迭代结果)

表1 CPCA 校正的频率估计值与实际偏移值的误差表

迭代	1	2	3	4	5	6	7	8
误差	0.1641	0.0472	0.0203	0.0120	0.0096	0.0089	0.0087	0.0085

表2 PCA 校正的频率估计值与实际偏移值的误差表(只取前后各4次迭代的误差值)

迭代	1	2	3	4	14	15	16	17
误差	0.4003	0.4452	0.3320	0.2832	0.1050	0.1047	0.1046	0.1044

$$\text{附: 误差} = \sqrt{\sum (\text{真值} - \text{估计值})^2 / \text{谱数}}$$

图5是活体MRSI数据, 经过HSVD方法除去水干扰^[13], 再应用式(1)的Lorentz模型化后的23个谱的氮-乙酰天门冬氨酸(N-acetylaspartate, NAA)的波峰信号, 图6及图7分别为本文方法实现的频率校正及其随后的衰减系数校正结果。从图5, 图6和图7比较看, 基于CPCA的方法较好地实现了活体MRS信号的频率和衰减系数的偏移校正。

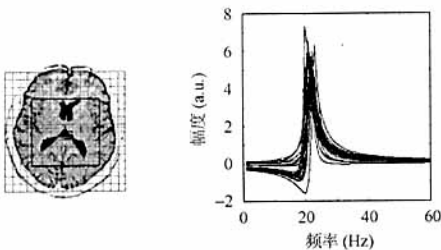


图4 MR图像 图5 Lorentz模型化的原始谱信号

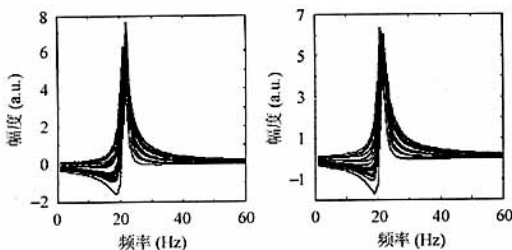


图6 频率校正后的谱信号 图7 衰减系数校正后的谱信号

6 结束语

本文在前人理论上, 提出了一种基于复数PCA校正MRS频率和衰减系数方法。该方法突出特点表现在: 立足于复数PCA, 对频率校正进行了深入的研究; 并拓宽了复数PCA的应用范围, 把它应用到了很少有具体研究的MRS衰减系数校正领域。从数据的实现结果分析, 本文方法较好地实现了频率和衰减系数的校正, 并在计算速度和估计精度上, 比基于实域PCA方法明显占优, 从而为更好地发挥MRS的价值创造有利条件。

参考文献

- [1] Lenkinski R E. MR spectroscopy: Clinical tool or research probe? (Revisited). *Academic Radiology*, 2001, 8(7): 567 - 570.
- [2] Lens P N L, Hemminga M A. Nuclear magnetic resonance in environmental engineering: Principles and applications. *Biodegradation*, 1998, 9(6): 393 - 409.
- [3] Mierisova S M R. Spectroscopy quantitation: a review of frequency domain methods. *NMR in Biomedicine*. 2001, 14(4): 247 - 259.
- [4] Chen H, Van Huffel S, Vandewalle J. Improved methods for exponential parameter estimation in the presence of known poles and noise. *IEEE Trans. on Signal Processing*, 1997, 45(5): 1390 - 1393.
- [5] Vanhamme L, Van Huffel S, Van Hecke P, Van Ormondt D. Time-domain quantification of series of biomedical magnetic resonance spectroscopy signals. *Journal of Magnetic Resonance* 1999, 140(1): 120 - 130.
- [6] Stoyanova R, Kuesel A C, Brown T R. Application of principal-component analysis for NMR spectral quantitation. *Journal of Magnetic Resonance, Series A*, 1995, 115(2): 265 - 69.
- [7] Brown T R, Stoyanova R. NMR spectral quantitation by principal component analysis. II determination of frequency and phase shifts. *Journal of Magnetic Resonance, Series B*, 1996, 112(1): 32 - 43.
- [8] Witjes H, Melsens W J, Zandt H J A, van der Graaf M, Heerschap A, Buydens L M C. Automatic correction for phase shifts, frequency shifts, and lineshape distortions across a series of single resonance lines in large spectral data sets. *Journal of Magnetic Resonance*, 2000, 144(1): 35 - 44.
- [9] Witjes H, Peppers M, Melsens W J, Buydens L M C. Modelling phase shifts, peak shifts and peak width variations in spectral data sets: its value in multivariate data analysis. *Analytica Chimica*

- Acta*, 2001, 432(1): 113 – 124.
- [10] Stoyanova R, Brown T R. NMR spectral quantitation by principal component analysis. A generalized procedure for determination of lineshape variations. *Journal of Magnetic Resonance*, 2002, 154 (2): 163 – 175.
- [11] Elliott M A, Walter Glenn A, Swift A, Vandenborne K, Schotland J C, Leigh J S. Spectral quantitation by principal component analysis using complex singular value decomposition. *Magnetic Resonance in Medicine*, 1999, 41(3): 450 – 455.
- [12] Van Huffel S, Wang Y, Vanhamme L, Van Hecke P. Automatic frequency alignment and quantitation of single resonances in multiple magnetic resonance spectra via complex principal component analysis. *Journal of Magnetic Resonance*, 2002, 158 (1): 1 – 14.
- [13] Vanhamme L, Fierro R D, Van Huffel S, de Beer R. Fast removal of residual water in proton spectra, *Journal of Magnetic Resonance*, 1998, 132(2): 197 – 203.
- 曾卫明: 男; 1971 年生, 博士生, 主要研究方向为功能成象和 MRS、图像处理及模式识别.
- 梁晓云: 男, 1973 年生, 博士生, 主要研究方向为运动分析、图像处理及模式识别.
- 罗立民: 男, 1956 年生, 博士生导师, 特聘教授, IEEE 高级会员, IEEE EMB 编委, 长期从事医学图像处理和生物医学工程领域的研究工作.