

# A Bayesian approach for the alignment of high-resolution NMR spectra

Seoung Bum Kim · Zhou Wang · Basavaraj Hiremath

Published online: 18 March 2008  
© Springer Science+Business Media, LLC 2008

**Abstract** Metabolic analysis with high-resolution nuclear magnetic resonance (NMR) enables simultaneous investigation of numerous chemical species in response to biochemical changes in subjects. When the analysis involves comparing two or more NMR spectra, it is essential to properly align them because small variations across different spectra influence the alignment and thus, interfere with direct comparisons between samples. We propose a new alignment method within the Bayesian modeling framework. The proposed method allows us to estimate the amplitude and phase shifts simultaneously and to obtain robust results in the existence of noise. Effectiveness of our proposed method is demonstrated through real NMR spectra in human plasma and a comparison study with dynamic time warping and correlated optimized warping, two widely used alignment methods in spectral data.

**Keywords** Alignment · Amplitude (baseline intensity) variation · Bayesian method · Nuclear magnetic resonance (NMR) · Phase (spectral) shift

## 1 Introduction

Metabolomics approaches that use high-resolution nuclear magnetic resonance (NMR) spectroscopy have been used to characterize metabolic variations in response to physiological alternation, disease states, genetic modification, and nutrition intake (Nicholson et al.

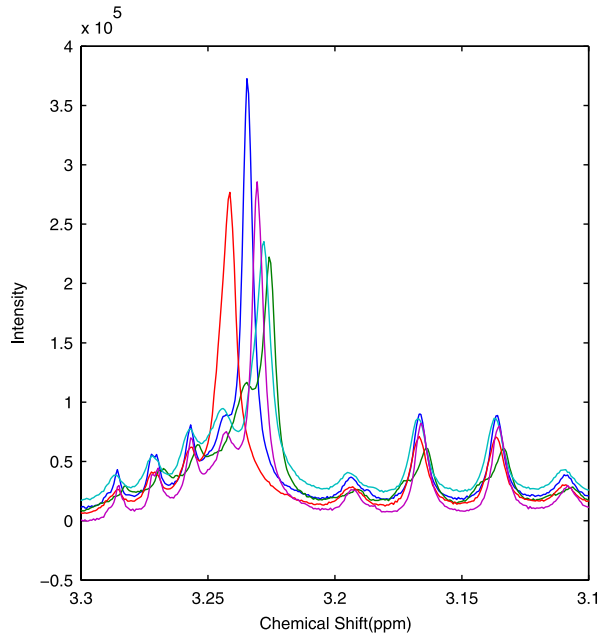
---

S.B. Kim (✉)  
Department of Industrial and Manufacturing Systems Engineering, University of Texas at Arlington,  
Arlington, TX 76019, USA  
e-mail: [sbkim@uta.edu](mailto:sbkim@uta.edu)

Z. Wang  
Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1,  
Canada  
e-mail: [zhouwang@ieec.org](mailto:zhouwang@ieec.org)

B. Hiremath  
CrossRate Technology, Standish, ME 04084, USA  
e-mail: [bhiremath@crossrate.com](mailto:bhiremath@crossrate.com)

**Fig. 1** A portion of five NMR spectra varying in both phase and amplitude



1999, 2002). A variety of techniques are available for studying metabolomics, of these, high-resolution NMR spectroscopy has advantages in terms of little sample preparation, minimal cost, and minimal invasiveness (Lindon 2004).

Analysis of NMR spectra is often devoted to a comparison of a set of spectra from different subjects, condition, or time points. Such combinations of multiple samples, each with a large number of features lead to a huge number of data points and a situation that poses a great challenge to analytical and computational capabilities. Numerous statistical data mining approaches, such as unsupervised and supervised learning methods, can reduce such complexity, and thus, facilitate the extraction of implicit patterns and help discriminate the spectra according to their biological/experimental conditions (Beckonert et al. 2003; Holmes et al. 2001; Lindon et al. 2001).

In the analysis and modeling of NMR spectra, preprocessing is an important step because inadequate preprocessing of NMR spectra can make it difficult to extract meaningful biological interpretations. Specifically, when multiple spectra are being considered, small variations in spectra caused by concentration, pH, and temperature, may influence the spectral alignment and thus can interfere with direct comparisons between spectra. Hence, it is crucial to align spectra before applying any subsequent statistical analyses. Figure 1 shows the portion of five NMR spectra varying in both phase (spectral shift) and amplitude (baseline intensity variation). This clearly demonstrates the necessity of spectral alignment.

A number of methods have been proposed for the alignment of multiple spectra. Dynamic time warping (DTW) and correlation optimized warping (COW) are widely used methods for alignment of spectral data sets (Pravdova et al. 2002; Tomasi et al. 2004). A brief review of the DTW and COW methods is provided in Sect. 2. Other methods include genetic algorithm-based methods (Forshed et al. 2002; Lee and Woodruff 2004), partial linear fit (Vogels et al. 1996), reduced set mapping (Torgrip et al. 2003), and principal component analysis-based methods (Stoyanova et al. 2004). One common problem with the existing methods is that they often disregard the existence of noise, which is typically observed in

NMR spectra. Furthermore, almost all existing methods are based on numerical optimizations that usually require significant computation effort. In practice, however, due to the high-throughput nature of NMR spectra, faster alignment algorithms are highly desirable.

In the present study we propose a new approach for spectral alignment based on Bayesian statistical modeling. By formulating the problem within a Bayesian framework, prior knowledge about the shift as well as the noise effect are taken into account in a natural way. Moreover, the proposed method can simultaneously estimate the spectral shift and baseline intensity variation in a closed form.

The rest of this paper is organized as follows. Section 2 briefly describes the DTW and COW algorithms. In Sect. 3, the proposed method is presented in detail. Section 4 presents the experimental results with real NMR spectra and compares the proposed algorithm with the DTW and COW algorithms. Finally, concluding remarks are given in Sect. 5.

## 2 Dynamic time warping and correlation optimized warping

### 2.1 Dynamic time warping

DTW is originally developed as an alignment method for speech recognition (Sakoe and Chiba 1971). The DTW algorithm calculates a cumulative distance function that measures the similarity of two signals which may vary in time. An optimal path can then be found by dynamic programming. Consider two spectra,  $T$  (length  $|T|$ ) for a reference spectrum and  $R$  (length  $|R|$ ) for a sample spectrum that needs to be aligned.

A grid plot is constructed with size  $|T| \times |R|$ . Then a set of warping paths  $\mathbf{P}$  of  $M$  points is defined as follows:

$$\mathbf{P} = \{[T_t^{(m)}, R_r^{(m)}], m = 1, 2, \dots, M\},$$

where  $t$  and  $r$  denote the indices of  $T$  and  $R$ , respectively. Then an optimal warping path is determined such that a cumulative distance between the two spectra is minimized. The cumulative distance between  $T_t$  and  $R_r$ ,  $D(t, r)$  is calculated as follows:

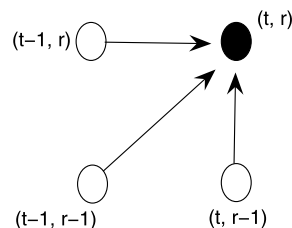
$$D(t, r) = d(t, r) + \min[D(t-1, r), D(t-1, r-1), D(t, r-1)],$$

where  $d(t, r)$  is the local distance between  $T_t$  and  $R_r$ . More precisely, the cumulative distance between  $T_t$  and  $R_r$  is defined as the distance between  $T_t$  and  $R_r$  plus the minimum cumulative distance among three allowable predecessors illustrated in Fig. 2.

### 2.2 Correlation optimized warping

The COW algorithm is introduced by Nielsen et al. (1998) for correcting misalignment of chromatographic spectra. Compared with DTW that identifies the optimal alignment by

**Fig. 2** Three allowable paths to compute the  $D(t, r)$  in the DTW algorithm



minimizing the cumulative distance between the reference and sample spectra, COW finds the optimal warping path by maximizing the overall correlation between the reference and sample spectra. Figure 3 shows the graphical illustration of the COW algorithm.  $T$  is a reference spectrum with length  $L_T$  and  $R$  is a sample spectrum with length  $L_R$ . Initially, the sample spectrum ( $R$ ) is divided into  $N$  sections with the length of  $m$ . Similarly, the reference spectrum ( $T$ ) is also divided into  $N$  sections. When the length of  $T$  and  $R$  differs, each section in  $R$  is stretched or shrunk by linear interpolation within the range  $(\Delta - t, \Delta + t)$ , so that the spectrum  $R'$  with the same length as  $T$  is obtained (see Fig. 3a). The parameter  $t$ , so called “slack parameter” defines the maximum length increase  $(+t)$  or length decrease  $(-t)$ .  $\Delta$  characterizes the length difference between the reference and sample spectra and is obtained by  $\Delta = (L_T/N) - m$ .

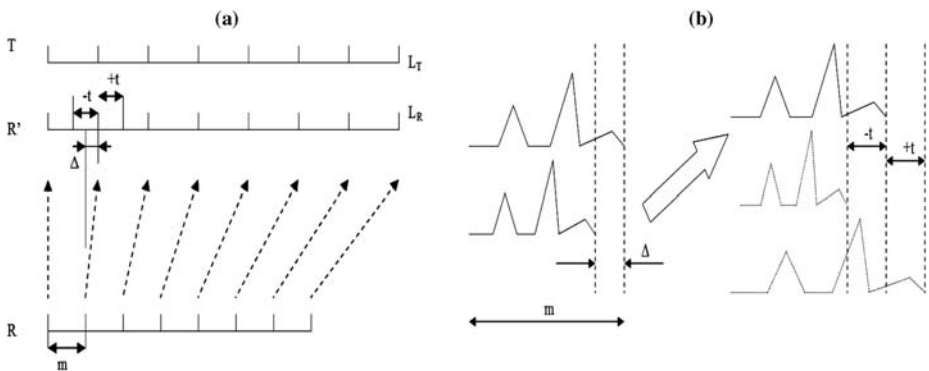
For implementation, two matrices (i.e.,  $\mathbf{M}$  and  $\mathbf{N}$ ) of size  $(L_T + 1) \times (N + 1)$  are constructed. The first matrix  $\mathbf{M}$  is initialized by assigning zero to the element  $(L_T + 1, N + 1)$  and minus infinity to the rest of the elements. Correlation between the section belonging to the reference spectrum, denoted as  $\zeta_T$ , and interpolated section of the sample spectrum, denoted as,  $\zeta_{R'}$  is computed from the following equation:

$$\rho = \frac{(\zeta_T - \bar{\zeta}_T)^T (\zeta_{R'} - \bar{\zeta}_{R'})}{\sigma_{\zeta_T} \sigma_{\zeta_{R'}}}, \tag{1}$$

where  $\bar{\zeta}_T$  and  $\sigma_{\zeta_T}$  are mean and standard deviation of  $\zeta_T$ . Similarly,  $\bar{\zeta}_{R'}$  and  $\sigma_{\zeta_{R'}}$  are mean and standard deviation of  $\zeta_{R'}$ .

In order to find the optimal path, the correlation values for all the possible warpings in the range  $[\Delta - t, \Delta + t]$  are calculated in each section.

Calculation starts from the last section and moves toward to the first section. From the second last section, all possible correlation values are cumulated to the correlation values obtained from the previous step. The second matrix  $\mathbf{N}$  is constructed, keeping only the highest cumulative correlation value in each section. Once  $\mathbf{N}$  is completely determined, the optimal path is selected by backtracking.



**Fig. 3** Graphical illustration of COW

### 3 Proposed method

#### 3.1 Bayesian alignment method

Let  $x(\omega)$  and  $y(\omega)$  be two spectra to be aligned, where  $\omega$  is the frequency index of the spectra. In the ideal case, the two spectra represent the same spectral structure but are shifted versions of each other in both the frequency and the intensity directions. Thus, we can write

$$y(\omega) = x(\omega + \Delta\omega) + \Delta a, \quad (2)$$

where  $\Delta\omega$  and  $\Delta a$  the spectral shift and the baseline intensity variation, respectively. A Taylor series expansion of the right hand side at  $\omega_0$  yields

$$y(\omega_0) = x(\omega_0) + \Delta\omega \left. \frac{dx}{d\omega} \right|_{\omega_0} + \frac{(\Delta\omega)^2}{2!} \left. \frac{d^2x}{d\omega^2} \right|_{\omega_0} + \dots + \Delta a. \quad (3)$$

In practice, the amount of the spectral shift and the baseline intensity variation are typically not fixed, but varies smoothly along the frequency axis. Therefore, (3) is only approximately true for a local spectral region-of-interest (SROI). In addition, the NMR spectral data acquired is discrete along the frequency axis. Assume that there are  $N$  discrete points within an SROI from the two spectra. We denote them as  $\{x(\omega_1), x(\omega_2), \dots, x(\omega_N)\}$  and  $\{y(\omega_1), y(\omega_2), \dots, y(\omega_N)\}$ , respectively. Also assume that the frequency shift  $\Delta\omega$  is small, so that the second and higher order terms can be ignored. We can then write

$$\mathbf{y} = \mathbf{x} + \Delta\omega \mathbf{x}' + \Delta a \mathbf{1}, \quad (4)$$

where  $\mathbf{x} = [x(\omega_1), x(\omega_2), \dots, x(\omega_N)]^T$ ,  $\mathbf{y} = [y(\omega_1), y(\omega_2), \dots, y(\omega_N)]^T$ ,  $\mathbf{x}' = [\left. \frac{dx}{d\omega} \right|_{\omega_1}, \left. \frac{dx}{d\omega} \right|_{\omega_2}, \dots, \left. \frac{dx}{d\omega} \right|_{\omega_N}]^T$ , and  $\mathbf{1}$  is an  $N$  dimensional column vector with all entries equaling 1. Reorganizing (4) into a matrix operation format, we obtain

$$\Delta \mathbf{x} = \mathbf{A} \mathbf{c}, \quad (5)$$

where  $\Delta \mathbf{x} = \mathbf{y} - \mathbf{x}$ ,  $\mathbf{A} = [\mathbf{x}' \ \mathbf{1}]$ , and  $\mathbf{c} = [\Delta\omega \ \Delta a]^T$  is a column vector containing the amount of shifts we would like to estimate.

Motivated by the Bayesian approach in optical flow estimation (Simoncelli et al. 1991), to account for the noise effect in a stochastic framework, we model

$$\epsilon = \Delta \mathbf{x} - \mathbf{A} \mathbf{c} \quad (6)$$

as a zero-mean Gaussian random vector, in which all entries are independently and identically distributed Gaussian random variables. The covariance matrix of  $\epsilon$  is thus diagonal and is denoted as  $\Lambda_n \mathbf{I}$ , where  $\Lambda_n$  is the noise variance and  $\mathbf{I}$  is the identity matrix. We can then write the probability density function (pdf) of  $\epsilon$  for a given  $\mathbf{c}$  as

$$p(\epsilon | \mathbf{c}) \propto \exp \left\{ -\frac{(\Delta \mathbf{x} - \mathbf{A} \mathbf{c})^T (\Delta \mathbf{x} - \mathbf{A} \mathbf{c})}{2\Lambda_n} \right\}. \quad (7)$$

Here we have used the fact that the covariance matrix is diagonal and ignored the constant in front of the Gaussian pdf (because the constant has no effect on the final solution). Based on Bayes' rule, we have

$$p(\mathbf{c} | \epsilon) \propto p(\epsilon | \mathbf{c}) p(\mathbf{c}). \quad (8)$$

For the prior distribution  $p(\mathbf{c})$ , we model it using a zero-mean Gaussian with a diagonal covariance matrix  $\Lambda_p$ :

$$p(\mathbf{c}) \propto \exp\left\{-\frac{1}{2}\mathbf{c}^T \Lambda_p^{-1} \mathbf{c}\right\}. \tag{9}$$

Finally, the pdf of the posterior distribution is computed as

$$\begin{aligned} p(\mathbf{c}|\epsilon) &\propto \exp\left\{-\frac{(\Delta\mathbf{x} - \mathbf{A}\mathbf{c})^T(\Delta\mathbf{x} - \mathbf{A}\mathbf{c})}{2\Lambda_n}\right\} \exp\left\{-\frac{1}{2}\mathbf{c}^T \Lambda_p^{-1} \mathbf{c}\right\} \\ &= \exp\left\{-\frac{1}{2}\left[\mathbf{c}^T \left(\frac{\mathbf{A}^T \mathbf{A}}{\Lambda_n} + \Lambda_p^{-1}\right) \mathbf{c} + \frac{2\mathbf{A}^T \Delta\mathbf{x}}{\Lambda_n} \mathbf{c} + \frac{\Delta\mathbf{x}^T \Delta\mathbf{x}}{\Lambda_n}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\mathbf{c} - \mathbf{m}_c)^T \Lambda_c^{-1} (\mathbf{c} - \mathbf{m}_c)\right\}, \end{aligned} \tag{10}$$

where

$$\mathbf{m}_c = \Lambda_c \frac{\mathbf{A}^T \Delta\mathbf{x}}{\Lambda_n}, \Lambda_c = \left(\frac{\mathbf{A}^T \mathbf{A}}{\Lambda_n} + \Lambda_p^{-1}\right)^{-1}. \tag{11}$$

This posterior pdf is also Gaussian with a mean vector  $\mathbf{m}_c$  and a covariance matrix  $\Lambda_c$ . Therefore, the Bayes least square (BLS) as well as the Bayes maximum a posterior (MAP) estimators of  $\mathbf{c}$  are both the centroid of the distribution given by

$$\begin{aligned} \hat{\mathbf{c}}_{BLS} = \hat{\mathbf{c}}_{MAP} = \mathbf{m}_c &= \Lambda_c \frac{\mathbf{A}^T \Delta\mathbf{x}}{\Lambda_n} \\ &= \left(\frac{\mathbf{A}^T \mathbf{A}}{\Lambda_n} + \Lambda_p^{-1}\right)^{-1} \frac{\mathbf{A}^T \Delta\mathbf{x}}{\Lambda_n} \\ &= (\mathbf{A}^T \mathbf{A} + \Lambda_n \Lambda_p^{-1})^{-1} \mathbf{A}^T \Delta\mathbf{x}. \end{aligned} \tag{12}$$

It is interesting to note that the Bayes’ estimators are consistent with the least-square estimator, which can be simply obtained by the pseudo-inverse of (5):

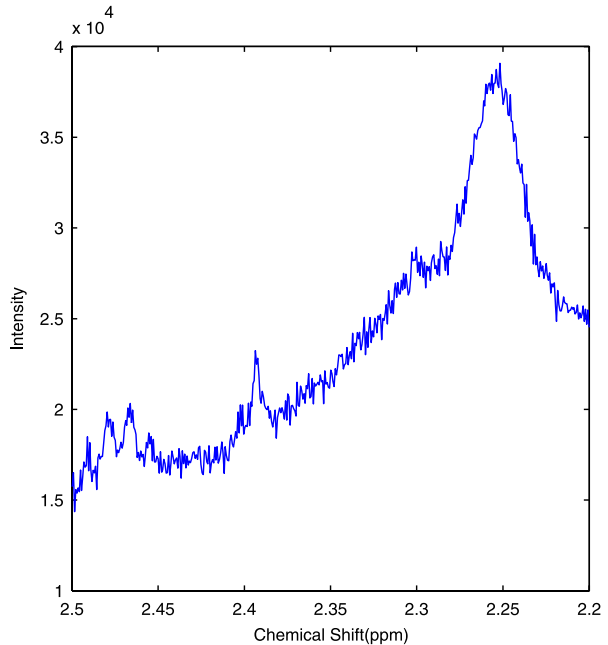
$$\hat{\mathbf{c}}_{LS} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \Delta\mathbf{x}. \tag{13}$$

The Bayesian estimators (12) and the least-squares estimator (13) are exactly the same when the spectral measurements are noise-free, i.e.,  $\Lambda_n = 0$ . In practice, however, NMR spectra are frequently contaminated by noise (see Fig. 4).

The main advantages of the Bayesian approach can be summarized as follows: (1) The effect of noise can be automatically accounted for. (2) In the case that the matrix  $(\mathbf{A}^T \mathbf{A})$  is not of full rank, the added diagonal matrix  $\Lambda_n \Lambda_p^{-1}$  in (12) makes it nonsingular, thus the solution is more robust than the least-squares solution. (3) Prior knowledge about the spectral shift and the baseline intensity variation can be included in a natural way.

Two remarks on the Bayesian alignment method are given below.

*Remark 1* In the formulation of the prior distribution (9), we have assumed that the spectra shift and the baseline intensity variation are uncorrelated. This is intuitively sensible because they are likely to be caused by independent physical reasons. We have also imposed a preference for small spectral shift and small baseline intensity variation because large spectral shift of baseline variations are usually unexpected in NMR spectra.

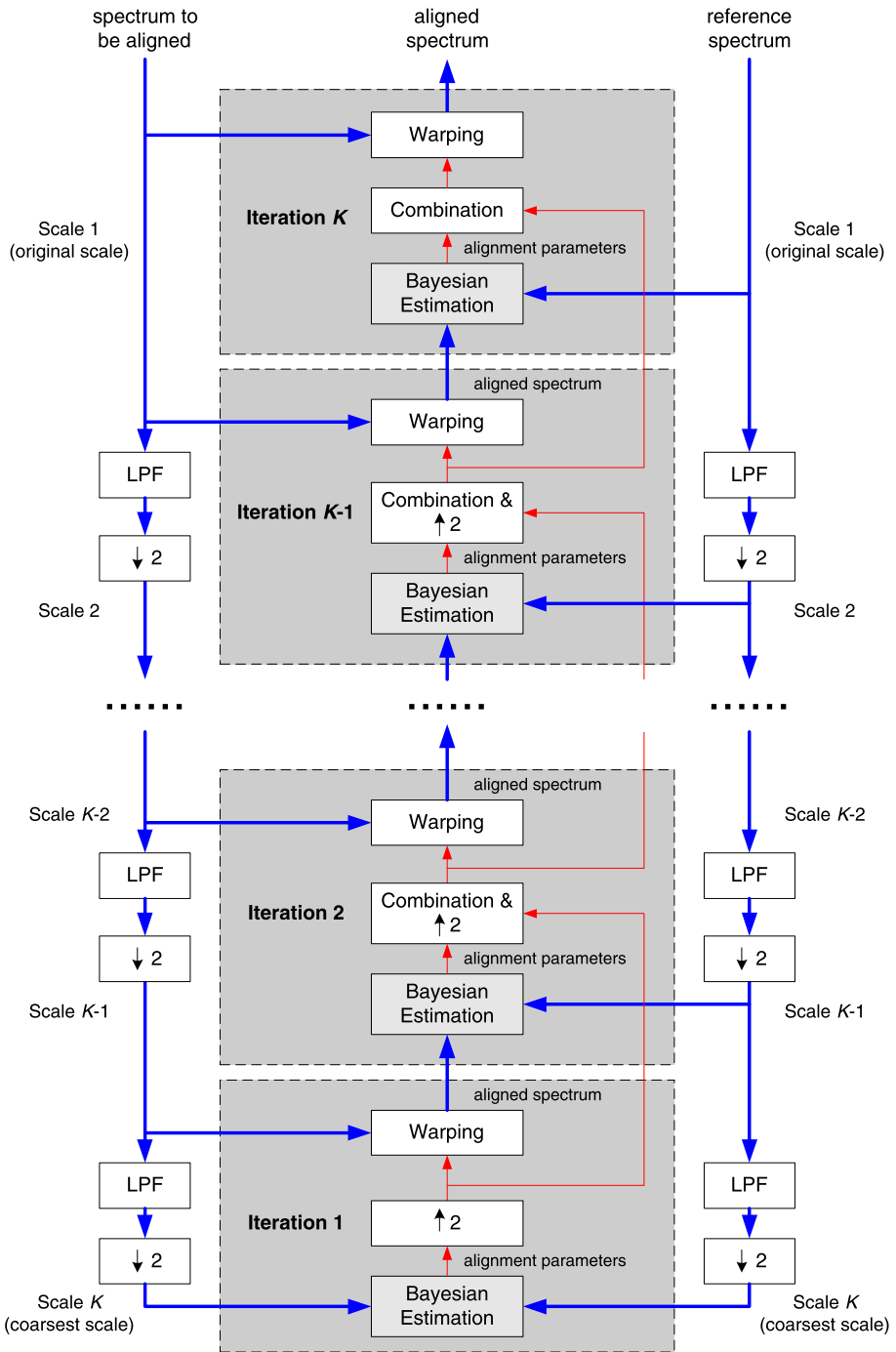
**Fig. 4** A portion of noisy NMR spectrum

*Remark 2* The derived Bayesian estimator is analogous to the estimator obtained from certain regularization-based methods. For example, the ridge regression method imposes a penalty that minimizes a residual sum of squares to obtain reliable solutions by reducing the risk of inverting a potentially singular matrix. The Bayesian formulation presented in this paper provides an intuitive understanding about the noise effect in the estimation problem, and the parameters in the Bayesian models can be easily linked with physical explanations (prior signal models and noise variance), although these parameters should be properly estimated.

### 3.2 Multiscale Bayesian alignment method

The linearization in the derivation of the Bayesian approach described in Sect. 3.1 relies on the assumption that the spectral misalignment is small, such that the higher-order terms in the Taylor series expansion in (3) can be ignored. In reality, this assumption may not hold. To overcome this problem and make the proposed algorithm generally applicable, here we use a multiscale approach that applies the Bayesian estimation method *iteratively* from coarse to fine. The general idea is to align the spectra at the coarser scales first (where larger misalignment can be handled, but with reduced accuracy) and refine the alignment parameters when moving towards the finer scales.

The diagram of the Bayesian multiscale alignment method is illustrated in Fig. 5. Before the iteration begins, we first apply a series of lowpass filtering and downsampling procedures repeatedly to create the spectra at different scales. This is illustrated by the left and right columns in Fig. 5, where the left is the spectrum to be aligned and the right is the reference spectrum. The purpose of lowpass filtering (where a Gaussian filter is employed in our implementation) is to reduce the aliasing effect in the subsequent downsampling process. As



**Fig. 5** Diagram of the Bayesian multiscale alignment method. LPF: low-pass filtering; ↓ 2: downsampling by a factor of 2; ↑ 2: upsampling (with interpolation) by a factor of 2



in most practical multi-resolution signal analysis, for the convenience of software/hardware implementation and easy indexing, we choose a factor of 2 in all the downsampling operations.

The iteration process starts from the coarsest scale (Scale  $K$ ) at the bottom of the central column in Fig. 5. The Bayesian estimation approach described in Sect. 3.1 is applied to the coarsest scale spectra, resulting in the BLS estimates of the alignment parameters. These parameters are not used to align the spectra at the coarsest scale. Instead, they are upsampled (with interpolation) by a factor of 2 to have the same length as in the second coarsest scale (Scale  $K - 1$ ). We can then use these interpolated alignment parameters to warp one of the spectrum towards the other (reference) spectrum, resulting in an aligned spectrum at Scale  $K - 1$ .

In the  $n$ -th iteration (where  $n \in [2, K - 1]$ ), the aligned spectrum from the previous iteration, together with the reference spectrum at the current scale (Scale  $K - n + 1$ ), is treated as a “regular” pair of spectra, and the same Bayesian estimation approach described in Sect. 3.1 is employed to generate the alignment parameters at the current scale. These alignment parameters are then combined with (being added to) the alignment parameters relayed from the previous iteration. This way, the amount of true alignment is accumulated and refined over iterations. Next, these alignment parameters are upsampled by a factor of 2. Finally, the spectrum to be aligned at Scale  $K - n$  is warped based on the upsampled parameters. Such an iteration process continues until the last iteration (Iteration  $K$ ) at the finest scale (Scale 1).

In the last iteration, the same Bayesian estimation approach is applied to the aligned spectrum from the previous scale and the reference spectrum in the finest scale, and the estimated parameters are combined with the parameters relayed from the previous iteration. The difference from all previous iterations is that the alignment parameters do not need to be upsampled, because they have already reached the same level and have the same length as the original spectra. In the last step, the warping process is applied to the original spectrum to be aligned with the final set of alignment parameters, which have been cumulated and refined step by step from all iterations.

### 3.3 Implementation issues

Several implementation issues need to be resolved before the proposed approach is applied. First, to obtain the matrix  $\mathbf{A}$ , we need to compute the derivative  $\mathbf{x}'$  of an input spectral signal  $\mathbf{x}$ . This is not trivial because of the existence of noise. A method that is often used is to apply a linear smooth filter to the signal before the differentiation operation. This is equivalent to convolving the signal with the derivative of the smooth filter. Specifically, we use a differentiation of Gaussian filter to compute  $\mathbf{x}'$ .

Second, several parameters need to be determined, which are the diagonal entries of the covariance matrices of the noise distribution and the prior distribution. In our experiment, we estimate them empirically from the acquired NMR spectral data. In particular, we estimate the noise variance  $\Lambda_n$  from the variance of the relatively inactive regions in the spectra (i.e., the regions that do not contain significant signal peaks). The two parameter values in the diagonal entries of  $\Lambda_p$  reflect the variances of the spectral shift and the baseline intensity variation, respectively (and the other entries are zeros). It is difficult to provide appropriate values without sufficient prior knowledge of the data being processed. In our implementations, such knowledge was gained by experimenting with these parameter values, and we finally selected the ones that lead to reasonable alignment results in terms of root mean squared error measurement. Although this may potentially result in overfitting, we found in

our experiments that the overall performance of our algorithm is insensitive to small variations of these parameters.

Third, as mentioned earlier, the spectral shift and the baseline variation of intensity are approximately constant only in a relatively small SROI and may vary smoothly along the frequency axis. Therefore, we apply the Bayesian estimation approach locally within a sliding window that moves sample by sample across the frequency axis. This results in two sequences of estimated parameters as functions of frequency, one for local spectral shift and the other for local baseline intensity variation. To align the two spectra, we keep one of them fixed and warp the other locally based on the estimated parameters. The size of the sliding window is selected such that the window can sufficiently cover twice of the maximal absolute spectral shift (since the shift can be either to the left or to the right). To avoid sharp changes of the estimated parameters along the spectral axis (typically around sharp peaks in the spectrum), it is useful to employ a sliding window that has smooth edges. In our implementation, we used a Hanning window, which is smooth and gives more weights to the center part of the sliding window during estimation. We also found that other smooth windows such as Gaussian and Hamming windows lead to similar results.

Finally, to avoid overfitting between the spectra, we restrict the joint spectral shift and baseline variation estimation to be applied at the coarser scales only. This can be understood as imposing additional smoothness constraints on the baseline variation in the finer scales because allowing fine tuning of the baseline levels at the fine levels could result in local structural change of the spectrum, which is not desirable. At the fine scales, to restrict the spectral alignment to spectral shift only, we simplify the Bayesian estimation algorithm in Sect. 3.1. Specifically, the matrix  $\mathbf{A}$  becomes a column vector  $\mathbf{x}'$ , the shifting parameter vector  $\mathbf{c}$  becomes a scalar  $\Delta\omega$ , and the prior covariance matrix  $\mathbf{\Lambda}_p$  becomes a scalar variance parameter about the spectral shift  $\Lambda_s$ . We can then rewrite the BLS or MAP estimator in (12) as

$$\Delta\hat{\omega}_{\text{BLS}} = \Delta\hat{\omega}_{\text{MAP}} = \frac{(\mathbf{x}')^T \Delta \mathbf{x}}{(\mathbf{x}')^T \mathbf{x}' + \Lambda_n / \Lambda_s}. \quad (14)$$

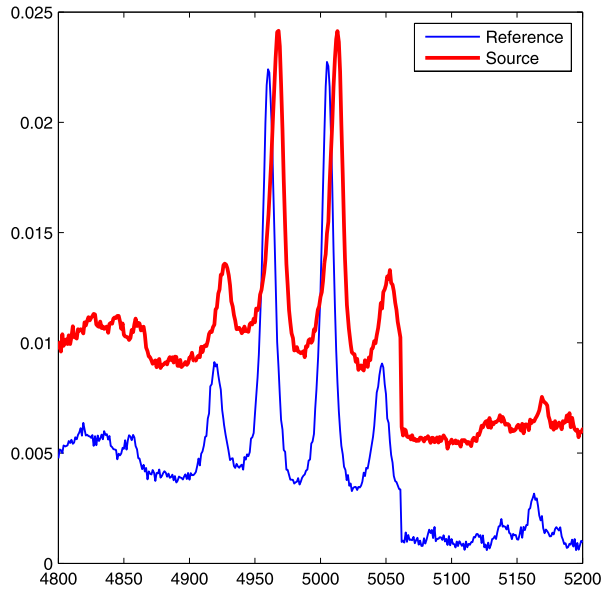
The alignment task is then simply implemented by warping one spectrum to the other using the estimated spectral shift parameter  $\Delta\hat{\omega}$ .

## 4 Experimental results

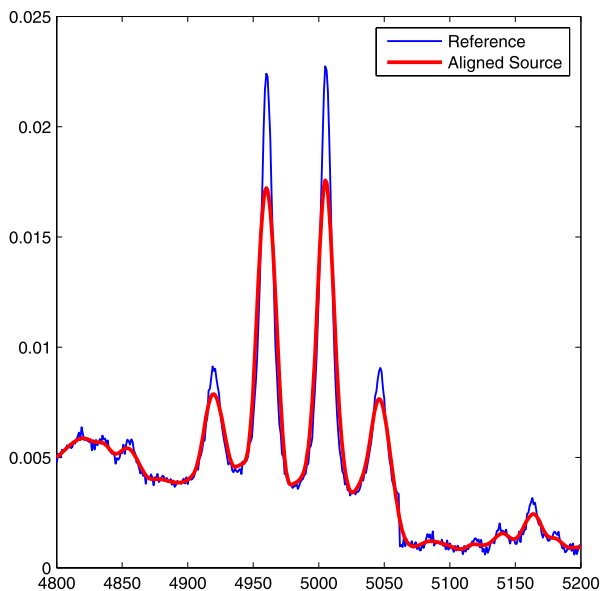
We use experimental data from 158 plasma samples obtained from five healthy subjects under controlled metabolic conditions in the Emory General Clinical Research Center. The subjects signed an informed consent approved by the Emory Institutional Review Board. 158 proton NMR spectra (each with 11,000 data points) are obtained from the collected plasma samples using a Varian INOVA 600 MHz instrument, a high-resolution NMR spectrometer. The ultimate goal of this experiment is to characterize the metabolic variations in response to sulfur amino acids intake in human plasma.

To visualize the performance of the multiscale Bayesian method, we randomly pick two spectra and align them. Figure 6 displays a small portion of the two NMR spectra (for better visualization) before alignment. It can be clearly observed that the two spectra are misaligned in both phase (spectral) and amplitude (baseline intensity). The multiscale Bayesian method is applied to align these two spectra. Figure 7 shows the result of alignment after correcting spectral shifts and baseline intensity variations. It can be seen that the two spectra are well aligned, demonstrating the effectiveness of the proposed method.

**Fig. 6** A portion of NMR spectra from two samples before alignment

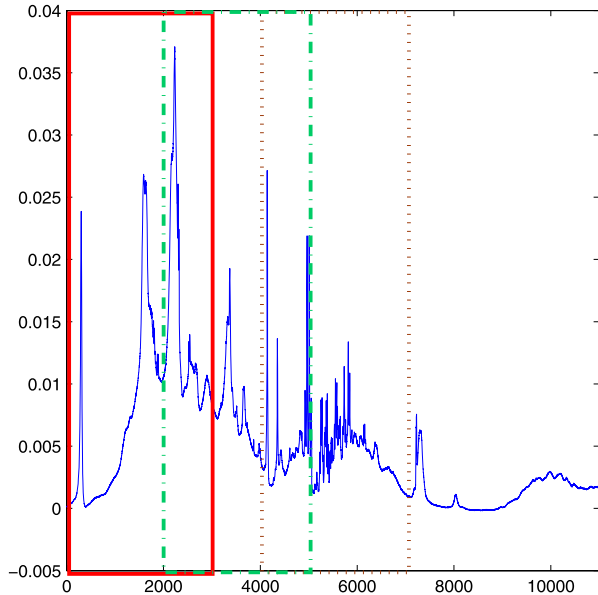


**Fig. 7** A portion of NMR spectra from two samples after alignment using the multiscale Bayesian method



For the purpose of performance comparison, four algorithms, including DTW, COW, and the proposed methods (Bayesian and multiscale Bayesian), are applied to align each of the 158 spectra with a reference spectrum computed by their median. In general, the median or mean of the spectra is used as the reference spectrum unless an experimentally verified reference spectrum is available. The reference spectrum is shown in Fig. 8. It should be noted that the large length of the spectra being tested makes the DTW and COW algorithms

**Fig. 8** The reference NMR spectrum with three segments



computationally extremely expensive. Therefore, we had to segment the first 7,000 points in a whole spectral region into three overlapped subregions, each with 3,000 data points. The spectral region after 7,000th point is not considered because this region contains only few significant metabolite features. The segmentation is shown in Fig. 8.

The first subregion (1–3,000) contains mostly large peaks, the third subregion (4,001–7,000) contains mostly small peaks, while the second subregion (2,001–5,000) contains both large and small peaks. Tables 1–3 show the resulting average root mean squares (RMS), average correlation coefficients, and computational time for four alignment methods in three subregions, respectively. The RMS value is computed as:

$$\text{RMS} = \sqrt{\frac{1}{M} \sum_{i=1}^M (s_i - r_i)^2},$$

where  $M$  is the number of points (3,000) in each subregion, and  $s_i$  and  $r_i$  are the  $i$ th point in the two spectra (one of which is the reference spectrum) being compared. Each alignment method is executed by MATLAB version 7 ([www.mathworks.com](http://www.mathworks.com)) on a 1.86 GHz Pentium M Processor. The total computational time to align all 158 spectra is recorded for each algorithm and is given in Tables 1–3.

The average RMS and the average correlation coefficient of the 158 spectra in all three subregions show that the proposed algorithms perform better than DTW and COW. The proposed algorithms also demonstrate a significant reduction of execution time, in comparison with the DTW and COW algorithms. The multiscale Bayesian method yields smaller RMS and larger correlation coefficients than the Bayesian method, although not to a significant degree. This indicates that incorporating the multiscale process into the Bayesian method improves the overall alignment performance with moderate additional computational cost.

**Table 1** A comparison of alignment methods (region 1–3,000)

Methods	RMS (before)	RMS (after)	Corr. (before)	Corr. (after)	Time (s)
DTW	0.002638	0.002239	0.9789	0.9821	4,520
COW	0.002638	0.002946	0.9789	0.9746	7,681
Bayesian	0.002638	0.000583	0.9789	0.9969	347
Multiscale Bayesian	0.002638	0.000196	0.9789	0.9997	771

**Table 2** A comparison of alignment methods (region 2,001–5,000)

Methods	RMS (before)	RMS (after)	Corr. (before)	Corr. (after)	Time (s)
DTW	0.002951	0.002636	0.9747	0.9764	4,492
COW	0.002951	0.002993	0.9747	0.9752	7,676
Bayesian	0.002951	0.000764	0.9747	0.9919	346
Multiscale Bayesian	0.002951	0.000366	0.9747	0.9983	733

**Table 3** A comparison of alignment methods (region 4,001–7,000)

Methods	RMS (before)	RMS (after)	Corr. (before)	Corr. (after)	Time (s)
DTW	0.001850	0.001713	0.9346	0.9478	5,220
COW	0.001850	0.001812	0.9346	0.9522	7,739
Bayesian	0.001850	0.000790	0.9346	0.9547	293
Multiscale Bayesian	0.001850	0.000448	0.9346	0.9865	751

## 5 Conclusions

We have proposed a new method for spectral alignment within the Bayesian statistical modeling framework. The basic idea of our proposed method is to represent a reference spectrum as a shifted function of the spectrum to be aligned. This function is approximated based on a Taylor series expansion. The parameters of spectral shift and the baseline intensity variation are simultaneously estimated by a Bayesian approach. The estimated parameters are then used to align spectra. In addition, we have used a multiscale Bayesian alignment method to overcome the limitation posed by the first-order Taylor expansion approximation. This multiscale approach enables us to correct relatively large spectral shift by applying the Bayesian estimation procedure iteratively from coarse to fine.

The major advantage of the proposed method is three-fold: (1) The prior knowledge about the shift parameters and the noise effect are naturally incorporated into the same estimation framework; (2) The spectral shift and the baseline intensity variation are estimated simultaneously; and (3) A simple and robust closed-form estimator is obtained for the problem.

Our test with real high-resolution NMR spectra demonstrates the effectiveness of the proposed method and shows clear advantages over DTW and COW, which are two widely used methods for spectral alignment. It is also noteworthy that the performance of our algorithms is achieved with significantly reduced computational effort.

**Acknowledgements** We would like to thank the associate editor and the anonymous referees for the constructive comments and suggestions, which greatly improved the quality of the paper. We are grateful to Dean P. Jones and Thomas R. Zeigler in the Emory University Medical School for their useful comments.

## References

- Beckonert, O., Bollard, M. E., Ebbels, T. M. D., Keun, H. C., Antti, H., Holmes, E., Lindon, J. C., & Nicholson, J. K. (2003). NMR-based metabonomics toxicity classification: hierarchical cluster analysis and k-nearest-neighbour approaches. *Analytica Chimica Acta*, *490*, 3–15.
- Forshed, J., Schuppe-Koistinen, I., & Jacobsson, S. P. (2002). Peak alignment of NMR signals by means of a genetic algorithm. *Analytica Chimica Acta*, *487*, 189–199.
- Holmes, E., Nicholson, J. K., & Tranter, G. (2001). Metabonomic characterization of genetic variations in toxicological metabolic responses using probabilistic neural networks. *Chemical Research in Toxicology*, *14*, 181–191.
- Lee, G.-C., & Woodruff, D. L. (2004). Beam search for peak alignment of NMR signals. *Analytica Chimica Acta*, *513*, 413–416.
- Lindon, J. C. (2004). Metabonomics—techniques and applications. *Business Briefing: Future Drug Discovery* 1–6.
- Lindon, J. C., Holmes, E., & Nicholson, J. K. (2001). Pattern recognition methods and applications in biomedical magnetic resonance. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *39*, 1–40.
- Nicholson, J. K., Connelly, J., Lindon, J. C., & Holmes, E. (2002). Metabonomics: a platform for studying drug toxicity and gene function. *Nature Review Drug Discovery*, *1*, 153–161.
- Nicholson, J. K., Lindon, J. C., & Holmes, E. (1999). Metabonomics: Understanding the metabolic response of living systems to pathophysiological stimuli via multi-variate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, *29*, 1181–1189.
- Nielsen, N. P. V., Carstensen, J. M., & Smedsgaard, J. (1998). Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimized warping. *Journal of Chromatography A*, *805*, 17–35.
- Pravdova, V., Walczak, B., & Massart, D. L. (2002). A comparison of two algorithms for warping of analytical signals. *Analytica Chimica Acta*, *456*, 77–92.
- Sakoe, H., & Chiba, S. (1971). In *Proceedings of the international congress of acoustics*, Budapest, Paper 20 C13.
- Simoncelli, E. P., Adelson, E. H., & Heeger, D. J. (1991). Probability distributions of optical flow. In *IEEE inter. conf. computer vision & pattern recognition* (pp. 310–315). Maui, Hawaii.
- Stoyanova, R., Nicholls, A. W., Nicholson, J. K., Lindon, J. C., & Brown, T. R. (2004). Automatic alignment of individual peaks in large-high resolution spectral data sets. *Journal of Magnetic Resonance*, *170*, 329–335.
- Tomasí, G., Berg, F. V. D., & Anderson, C. (2004). Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics*, *18*, 231–241.
- Torgrip, R. J. O., Aberg, M., & Jacobsson, S. P. (2003). Peak alignment using reducing set mapping. *Journal of Chemometrics*, *17*, 573–582.
- Vogel, J. T. W. E., Tas, A. C., Venekamp, J., & van der Greef, J. (1996). Partial linear fit: A new NMR spectroscopy preprocessing tool for pattern recognition applications. *Journal of Chemometrics*, *10*, 425–438.