Algorithmic subtraction of high peaks in NMR spectra

Jae-Seung Lee *, A.K. Khitrin

Department of Chemistry, Kent State University, Kent, OH 44242-0001, USA

Abstract

A method of data processing, which suppresses high peaks in NMR spectra, is presented. The algorithm is universal, does not require any information about the shape of subtracted peaks, and can be applied to various types of spectra. In some cases, algorithmic subtraction of peaks in simple 1D spectra can replace more complex and time-consuming methods of experimental peaks suppression.
© 2006 Elsevier B.V. All rights reserved.

1. Introduction

The problem of observing weak signals in the presence of much stronger signals is very common in nuclear magnetic resonance (NMR). Usually, this problem is addressed by choosing an appropriate pulse sequence, which experimentally suppresses undesired large peaks in NMR spectra. Modern NMR spectrometers have large dynamic range and allow recording small signals in the presence of strong ones without distortion. Better spectrometers and better computers make it possible today to suppress undesired high peaks with data processing and avoid, in some cases, more complex and time-consuming experiments. When the shape of a peak is known, its subtraction is trivial. However, the shapes are affected by many experimental parameters and are rarely known with sufficient accuracy. Various filtering techniques, where the suppressed peak is distinguished by its frequency, have been proposed [1,2]. When the number of subtracted peaks is large, both algorithmic and experimental methods of suppression, based on selecting the peaks by their frequencies, become inconvenient. Below we describe an algorithm, which does not require information about the peaks' shapes or positions. The method automatically suppresses peaks higher than some preset value and can be used for processing various types of spectra.

Let $I(\omega)$ be the experimental spectrum. We select some value Δ, which is slightly larger than the maximum height of the peaks we want to keep intact. The main idea of the processing scheme is to find a smooth line $J(\omega)$ which lies between $I(\omega)$ and $I(\omega) - \Delta$: $I(\omega) - \Delta \leqslant J(\omega) \leqslant I(\omega)$, and subtract this smooth line from the experimental spectrum. As a result, the processed spectrum $I^*(\omega) = I(\omega) - J(\omega)$ will not contain peaks higher than Δ. At the boundaries of the spectrum, we set $J(\omega) = I(\omega)$. High peaks may have large derivatives even at considerable distances from their centers, and the overlapped small peaks produce only a small bending of the total spectrum. Therefore, the line $J(\omega)$ should avoid bending as much as possible. Mathematically, this property can be formulated in different ways. The best choice of $J(\omega)$ would incorporate the most common features of real NMR peaks. Among the explored possibilities, one was to require that $J(\omega)$ has minimal length, and another to minimize an average square of the second derivative. As we have found, a combination of these two criteria gives the best results.

2. Experimental examples

2.1. Water suppression: D-ribose

The water suppression was performed for a sample of 0.5 mmol D-ribose dissolved in 1 mL of 90% H_2O/10% D_2O with 0.05 mmol DSS as a frequency standard. ^{1}H NMR spectrum is shown in Fig. 1a. The result of algorith-

* Corresponding author.
  E-mail address: jlee2@kent.edu (J.-S. Lee).

mic water peak suppression is shown in Fig. 1b. For comparison, we performed experimental water suppression [3] by using presaturation (Fig. 1c), and also recorded the spectrum in D_2O (Fig. 1d). As one can see in Fig. 1d, the spectrum of D-ribose contains peaks between 4.8 ppm and 5.0 ppm, which are very close to the water peak at 4.77 ppm. The processed spectrum in Fig. 1b clearly shows these peaks.

2.2. ^{13}C–^{13}C scalar couplings at natural abundance: sucrose

Natural-abundance ^{13}C NMR spectrum of 20% sucrose in D_2O, recorded with 28 000 scans, is shown in Fig. 2a. Figs. 2b–e show magnified parts of this spectrum for some selected peaks. Even though the side peaks, resulting from ^{13}C–^{13}C scalar couplings, are well resolved, large central peaks affect the positions of the side peaks and, therefore, the coupling constants obtained from those positions would not be accurate. Figs. 2b′–e′ display the corresponding subtracted spectra. Some ^{13}C–^{13}C scalar coupling constants measured from the processed spectrum are given in Table 1 and compared with the values obtained with 1D ^{13}C INADEQUATE experiment [4].
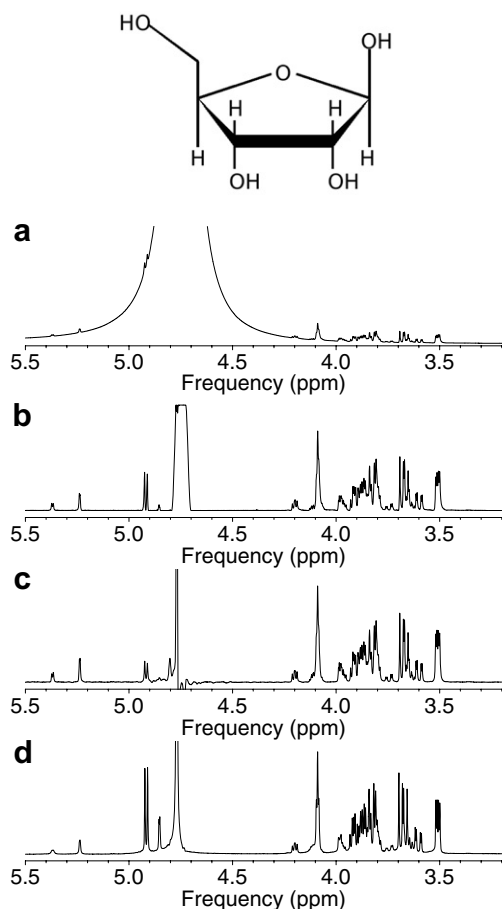

Fig. 1. (a) ^1H NMR spectrum of D-ribose in 90% H_2O/10% D_2O. (b) The same spectrum after algorithmic subtraction. (c) The spectrum with water signal experimentally suppressed by presaturation; (d) ^1H NMR spectrum of D-ribose in D_2O.

2.3. ^{13}C–^{13}C dipolar couplings in liquid crystal: 5CB

Natural-abundance ^{13}C–^{13}C dipolar couplings in liquid crystal 5CB have been studied before by using a double-quantum filtering technique [5]. Although many couplings between aromatic carbons have been measured in [5], poor resolution of the aliphatic region of the ^{13}C spectrum did not allow measuring any couplings involving aliphatic carbons. Improved resolution requires not only efficient heteronuclear decoupling [6], but long waiting time between scans to reduce the effect of radio-frequency heating [7]. In our experiment we used 1 min repetition time, which would make multi-dimensional experiments extremely time-consuming. 1D spectrum of aliphatic region with linewidths about 10 Hz is presented in Fig. 3a, where the peaks assignment [8] is also shown. Adiabatic cross polarization [9] has been used to enhance sensitivity. Magnified spectrum in Fig. 3b reveals many ^{13}C–^{13}C dipolar couplings, which are especially clearly seen in the processed spectrum in Fig. 3c. As an example, 196 Hz coupling between C_γ and C_ω is indicated in Fig. 3c.

3. Algorithm

The main idea of the processing is to find a smooth line $J(\omega)$ for subtraction, which follows general trends imposed by high peaks but avoids bending by small peaks. $J(\omega)$ is iteratively approached, starting from the experimental spectrum $I(\omega)$. The constraints are set by the original spectrum $I(\omega)$ and the one shifted by some predefined amount Δ: $I(\omega) - \Delta \leqslant J(\omega) \leqslant I(\omega)$. In each particular case, the choice of Δ is straightforward: it is the height of the tallest peak one wants to keep intact. A 'physical' model of $J(\omega)$ with minimal length is a string placed between $I(\omega) - \Delta$ and $I(\omega)$ and pulled by its ends. $J(\omega)$ with minimized average square of the second derivative resembles behavior of elastic rod. When only the length of $J(\omega)$ is minimized, the processed spectrum $I^*(\omega) = I(\omega) - J(\omega)$ has flat baseline. However, in order to make the length shortest, $J(\omega)$ follows the lower boundary $I(\omega) - \Delta$ near the centers of the $I(\omega)$ high peaks and repeats all its bendings. As a result, the small peaks in these regions are lost in $I^*(\omega)$. When the square of the second derivative of $J(\omega)$ is minimized, $J(\omega)$ 'detaches' from the lower boundary $I(\omega) - \Delta$ at the centers of the $I(\omega)$ high peaks and the small peaks are better recovered in $I^*(\omega)$. The drawback of this choice of $J(\omega)$ is that it produces a curved baseline of $I^*(\omega)$ near high peaks of $I(\omega)$. The best results were obtained when we used, with some weighting factors, both of these criteria in the iteration scheme.

Variable part of the square of the length and square of the second derivative are, respectively, $S_1 = \sum_{k=1}^{N-1} (y_{k+1} - y_k)^2$ and $S_2 = \sum_{k=1}^{N-1} (y_{k+1} - 2y_k + y_{k-1})^2$, where y_k is the vertical coordinate of the kth point of the digitized spectrum. Suppose that the final solution minimizes the linear combination $S = \alpha S_1 + (1 - \alpha) S_2$, where $0 \leqslant \alpha \leqslant 1$. From $\partial S / \partial y_k = 0$, each point of the final solution satisfies the equation
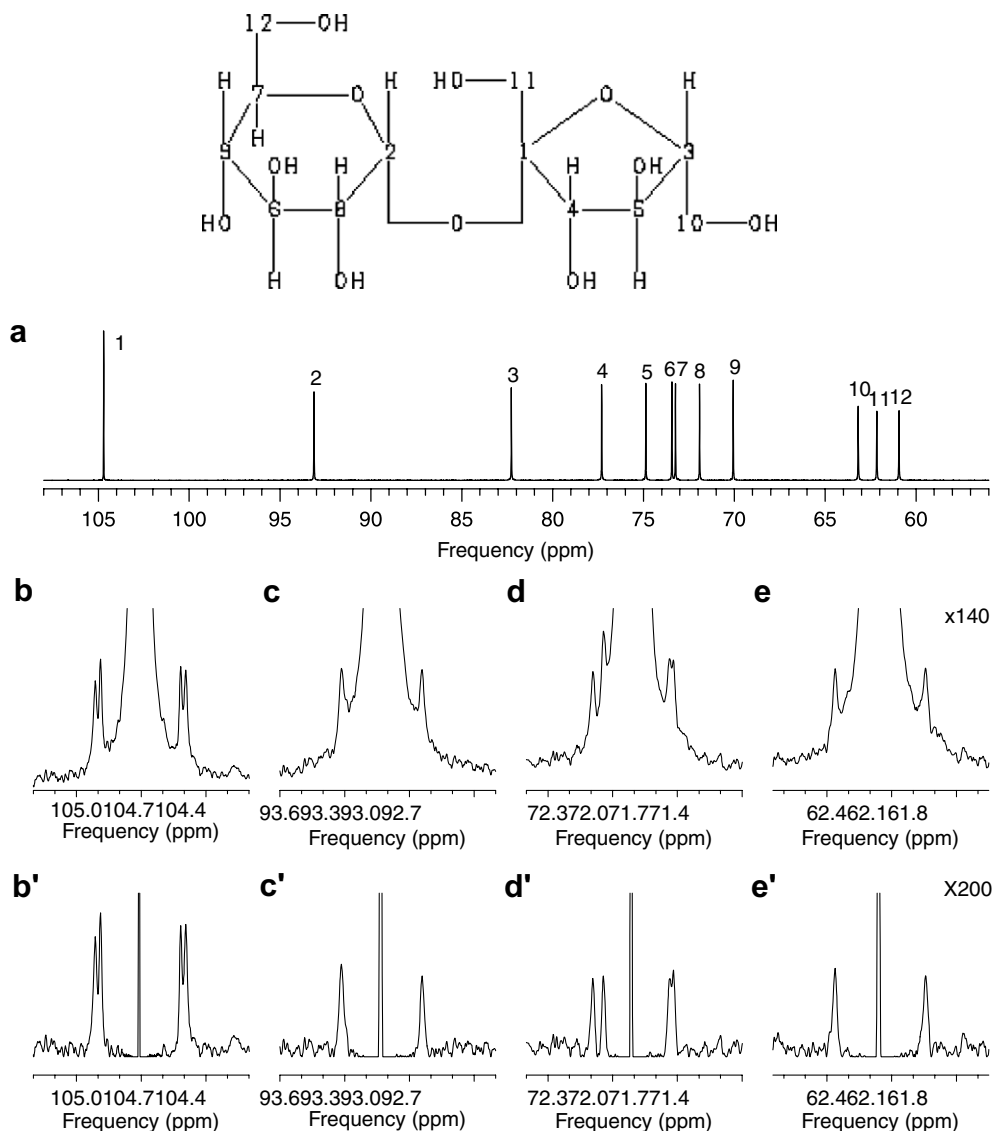
Fig. 2. (a) ^{13}C NMR spectrum of sucrose in D$_2$O. b–e Magnified spectra for selected peaks; corresponding spectra after subtraction are shown in b′–e′.

$$y_k = [(4 - 3\alpha)(y_{k+1} - y_{k-1}) - (1 - \alpha)(y_{k+2} - y_{k-2})]/[2(3 - 2\alpha)]. \quad (1)$$

Using these equations, the solution can be approached iteratively, starting from the original spectrum, by updating each vertical coordinate as

$$y_k^{n+1} = y_k^n + C\left[\frac{(4 - 3\alpha)(y_{k+1}^n - y_{k-1}^n) - (1 - \alpha)(y_{k+2}^n - y_{k-2}^n)}{2(3 - 2\alpha)} - y_k^n\right], \quad (2)$$

where n denotes the nth step of the iteration, $\{y_k^0\}$ is the original spectrum, and C is a constant. The constant C should be optimized because the calculation will be slow if C is small or will not converge if C is larger than about 0.7, depending on the spectrum. We used the value $C = 0.3$ which guaranteed convergence for all types of spectra.

For the examples shown above, the calculation was done with a Pentium-4 PC (2.53 GHz) using MATLAB software, with α set to 0.5. The results are not very sensitive to α, unless it is close to one of the extremes: 0 or 1. The value $\alpha = 0.5$ worked well for all of the spectra. It took about 8 h and 2.4×10^7 iterations for the ^{1}H NMR spectrum of D-ribose in water with 6.5 k spectral points and about 30 h and 6.8×10^7 iterations for the ^{13}C NMR spectrum of 5CB with 10 k points to finish the calculation. The

Table 1
^{13}C–^{13}C scalar coupling constants in sucrose

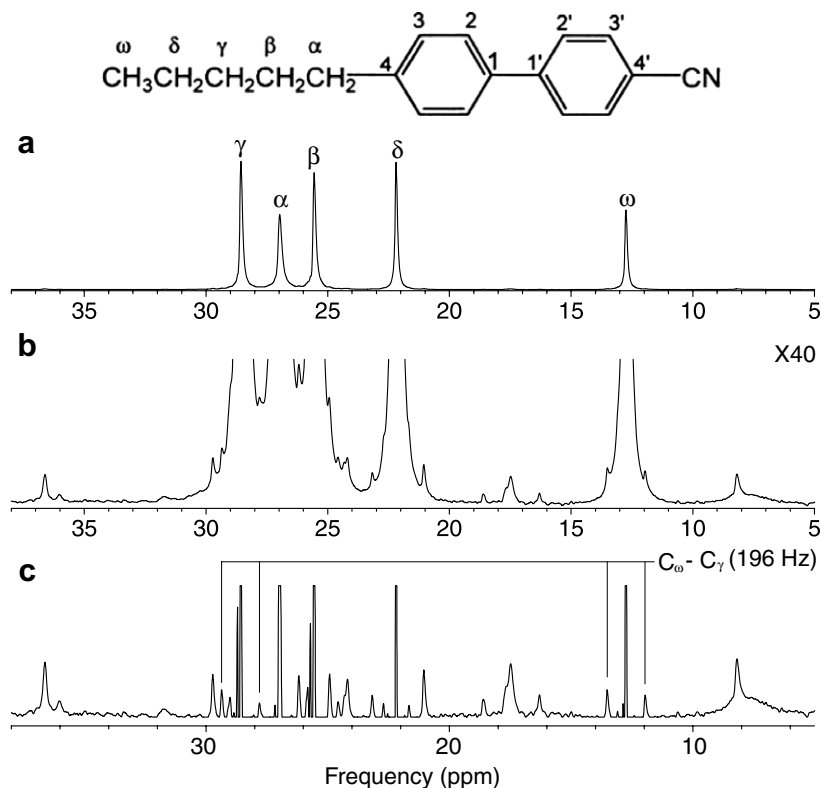| Coupled nuclei	Scalar coupling constant (Hz)	Values from Ref. [4] (Hz)
C1, C4	46.5	46.1
C1, C11	52.5	52.6
C2, C8	46.8	
C4, C5	39.4	
C3, C10	42.2	42.5
C6, C8	38.5	
C7, C12	42.8	

Fig. 3. (a) ^{13}C NMR spectrum of the aliphatic carbons in liquid crystal 5CB. Adiabatic cross-polarization [9] was used for signal enhancement. Protons were decoupled with SPINAL-64 sequence [6] at the decoupling power $\gamma B_2/2\pi = 35$ kHz. The number of scans was 1680. (b) The same spectrum with 40-fold magnification. (c) The result of algorithmic subtraction.

calculation for the ^{13}C NMR spectrum of sucrose with 50 k points was stopped after 24 h and 1.5×10^7 iterations, when the solution has been reached for all spectral regions of interest. The computation time grows as cube of the number of points. For the spectra or their fragments, digitized by a few thousand of points, the computation is very fast.

The program is available upon e-mail request.

References

[1] A. Coron, L. Vanhamme, J.-P. Antonie, P. Van Hecke, S. Van Huffel, J. Magn. Reson. 152 (2001) 26.

[2] U.L. Gunther, C. Ludwig, H. Ruterjans, J. Magn. Reson. 156 (2002) 19.

[3] M. Gueron, P. Plateau, in: D.M. Grant, R.K. Harris (Eds.), Encyclopedia of Nuclear Magnetic Resonance, Wiley, Chichester, 1996, p. 4931.

[4] J.M. Duker, A.S. Serianni, Carbohydr. Res. 249 (1993) 281.

[5] D. Sandström, M.H. Levitt, J. Chem. Phys. 118 (1996) 6966.

[6] B.M. Fung, A.K. Khitrin, K. Ermolaev, J. Magn. Reson. 142 (2000) 97.

[7] B.M. Fung, D.S.L. Mui, I.R. Bonnell, E.L. Enwall, J. Magn. Reson. 58 (1984) 254, 1969.

[8] B.M. Fung, K. Ermolaev, Y. Yu, J. Magn. Reson. 138 (1999) 28.

[9] J.-S. Lee, A.K. Khitrin, J. Magn. Reson. 177 (2005) 152.