Pattern recognition of the multiplet structure of NMR spectra

*S. S. Golotvin and V. A. Chertkov**

Department of Chemistry, M. V. Lomonosov Moscow State University,
Vorob'evy Gory, 119899 Moscow, Russian Federation.
Fax: 007 (095) 932 8846. E-mail: vche@nmr.chem.msu.su

A new method of analysis of high-resolution NMR spectra was suggested, and an algorithm and the PAREMUS program package for assignment of patterns of the spectral components based on pattern recognition theory were developed. High effectiveness of the method has been exemplified by determination of the values and relative signs of the inter-ring $^1H-^1H$ and $^{13}C-^1H$ spin-spin coupling constants in 1,2,3-trichloronaphthalene.

Key words: analysis of the high-resolution NMR spectra; recognition of patterns of the spectral components, inter-ring spin-spin coupling constants.

Currently, there are two principal approaches for analyzing high-resolution NMR spectra. The fundamental difference between them concerns the method of processing the spectral data. The method proposed by Castellano and Bothner-By[1,2] is based on two-level processing of spectral information. In the preliminary stage, deconvolution is used to separate spectral components (usually, individual lines) for which frequencies, amplitudes, phases, and the lineshape and linewidth parameters can be determined. The residual functional is constructed by using frequencies of the experimental spectrum, which are compared with those of theoretical transitions. The most important and time-consuming stage of solution of the problem is the assignment of the frequencies of the spectral components and the numbers of theoretical transitions. Solution is found by minimization of the root-mean-square (r.m.s.) deviation of the frequencies of the experimental lines from those of theoretical transitions, which makes it possible to estimate the best values of the parameters of the NMR spectrum, *i.e.*, chemical shifts, constants of the spin-spin and dipole-dipole interactions, *etc.*

The alternate strategy of total-lineshape analysis used in the NMRCON[3] and DAVINS[4,5] programs consists in studying the total spectral contour as a whole. In principle, this approach requires no preliminary interpretation of the data, however, it is associated with immense computational resources. Additional problems arise if the experimental spectrum contains signals of impurities or artifacts. Correctly accounting for the lineshape when it appreciably deviates from the Lorentzian shape is also a challenge. The areas of application of both procedures and their disadvantages have been previously discussed in detail.[6]

In the present work, a new method of matching the components of multiplets of experimental and theoretical spectra has been suggested in the framework of the Castellano and Bothner-By approach. The development of algorithms of matching is a classical problem of pattern recognition theory. Based on this theory, the PAREMUS[7] (Pattern Recognition of Multiplet Structure) program package for automated assignment of high-resolution NMR spectra based on the construction of precise patterns of spectral components was developed.

Experimental

The ^1H and ^{13}C NMR spectra of a 2 *M* solution of 1,2,3-trichloronaphthalene in acetone-d_6 were recorded on a Varian VXR-400 spectrometer (with an operating frequency of 400 MHz for the ^1H nuclei) at 301 K. The ^1H NMR spectrum of pyridine was simulated (see Table 1) using Lorentzian

Table 1. Spectral parameters (relative frequencies $\Delta\nu_{H(i)}$ and spin-spin coupling constants $J_{H(i)-H(j)}$/Hz) used in the analysis of the simulated ^1H NMR spectrum of pyridine

Parameter	Value	
	true	trial
$\Delta\nu_{H(1)}$	516.506	516.0
$\Delta\nu_{H(2)}$	427.435	427.0
$\Delta\nu_{H(3)}$	450.148	450.0
$\Delta\nu_{H(4)}$	427.435	427.0
$\Delta\nu_{H(5)}$	516.506	516.0
$J_{H(1)-H(2)}$	4.950	5.0
$J_{H(1)-H(3)}$	1.824	2.0
$J_{H(1)-H(4)}$	1.019	1.0
$J_{H(1)-H(5)}$	−0.043	−0.1
$J_{H(2)-H(3)}$	7.627	8.0
$J_{H(2)-H(4)}$	1.466	1.5
$J_{H(2)-H(5)}$	1.019	1.0
$J_{H(3)-H(4)}$	7.627	8.0
$J_{H(3)-H(5)}$	1.824	2.0
$J_{H(4)-H(5)}$	4.950	5.0

Translated from *Izvestiya Akademii Nauk. Seriya Khimicheskaya*, No. 3, pp. 444—451, March, 1997.

1066-5285/97/4603-0423 $18.00 © 1997 Plenum Publishing Corporation

lines of 0.05 Hz width. The level of white noise added was 15% of the unity peak amplitude, and digital resolution was 0.03 Hz per point.

1,2,3-Trichloronaphthalene was synthesized[8,9] by chlorination of α-chloronaphthalene with gaseous chlorine followed by treatment of the 1,1,2,3,4-pentachloro-1,2,3,4-tetrahydronaphthalene obtained with sodium methylate in dry acetone. The product was isolated by column chromatography (using Silpearl silica as adsorbent and hexane as eluent) with subsequent recrystallization from pentane.

The calculations using the PAREMUS program package were carried out on an IBM PC 486/DX4-100 personal computer. Typical computing time of one "long" cycle for a five-spin system was ~20 s.

Results and Discussion

Construction of patterns of the spectral components. In the framework of formal pattern recognition theory,[10] information on an object can be represented using its pattern, *i.e.*, a vector $\mathbf{X}(x_1, x_2,..., x_n)$ where x_i are certain characteristics of the object. Patterns can be treated as points in the n-dimensional space of the parameters. In high-resolution NMR spectra, the objects under consideration are spectral components (SC), which are either individual resolved lines or groups of overlapped lines. No characteristics of the individual lines can be distinguished in the latter case, therefore it is reasonable to consider patterns of the spectral components (it should be remembered that the concepts of spectral components and individual lines are identical only in the case of completely resolved spectra). The traditional procedure used by Castellano and Bothner-By[2] requires relatively well-resolved experimental spectra, in which the spectral components contain no more than two or three individual lines. The pattern of SC is constructed taking into account such measurable parameters as frequency, intensity, phase, *etc.* In the case of poorly resolved multiplets, the parameters of individual lines can contain considerable systematic errors, and such spectra must be analyzed using total-lineshape analysis.

The above mentioned characteristics are transduced into a vector $\mathbf{X}(x_1, x_2,..., x_n)$.[10] To construct a more precise pattern of an object and ensure ease and reliability of its recognition, a perfect transducer should register all the information that can be measured. In this connection, the limitations due to the use of only frequencies and amplitudes in constructing patterns of the SCs of NMR spectra are quite obvious. In our opinion, the data on the relative position of a SC in the spectrum, *i.e.*, its presence in a certain multiplet and the character of its position there, the type of the nearest environment, *etc.*, are important as well and should be adequately taken into account. The methodology of using such information is based on integral transforms (IT).[11] An integral transform of a finite number of SCs with frequency ν_i and amplitude A_i is calculated by the formula

$$\mathrm{IT}_n(\nu_i) = \sum_{j=1}^{N} A(\nu_j) S_n(\nu_i, \nu_j)\,, \tag{1}$$

where N is the total number of SCs in the spectrum, and $S_n(\nu_i, \nu_j)$ is a continuous quadratically normalized function. It is required that the function $S_n(\nu_i, \nu_j)$ gradually decrease, since the contribution to the integral transforms from the nearest SCs should be larger than those from the outermost ones. The simple analytical form of the function is also of practical importance. We tested the functions describing the absorption and dispersion signals for correspondence with these criteria. It was found that the dispersion function (2) with the parameter α has the most discriminative properties:

$$S_n(\nu_i, \nu_j) = k\left\{\frac{\alpha^2(\nu_i - \nu_j)}{1 + (\nu_i - \nu_j)^2\alpha^2}\right\}, \tag{2}$$

where k is the normalizing factor. The asymmetry of this function allows one to take into account contributions from the left and right neighbors of a given SC with opposite signs.

The highest sensitivity of the dispersion function to neighboring SCs is achieved at its extrema, $\nu = -1/\alpha$ and $\nu = +1/\alpha$. The maximum value of parameter α can be estimated from the value of the minimum repeating spectral interval $\Delta_{\min}$, which can be approximated in the case of a first-order spectrum either by the minimum resolved coupling constant for a given spectral interval or by the pertinent sum (or difference) of the spin-spin coupling constants with approximately equal moduli. Thus, it is believed that for the first-order spectra

$$\alpha_{\max} \approx 1/\Delta_{\min}. \tag{3}$$

The minimum value of parameter α ($\alpha_{\min}$) can be estimated by the total width of the multiplet with allowance for the condition of maximum sensitivity of IT to the outermost SCs in the multiplet. Hence,

$$\alpha_{\min} \approx 1/\Delta\mathrm{MW}, \tag{4}$$

where ΔMW is the width of the multiplet. Other repeating spectral intervals can also be used in determining the parameters of integral transforms in the construction of precise patterns of SCs.

The aforesaid can be illustrated by the following example. In Fig. 1, two model pairs of doublets (see their parameters in the legend to Fig. 1) are shown. The integral transforms obtained for each of the components of the multiplet by using function (2) for three values of parameter α determined from expressions (3) and (4) are given in Table 2. It can be seen that the integral transforms of the components of the doublets with constants J = 10.0 Hz (*e.g.*, A and C, and B and D for the upper multiplet) are of opposite sign and have larger moduli. This means that the dispersion function with α = 0.1 s (a wide-band filter) provides the best way of distinguishing the spectral components at the opposite ends of the multiplet. To distinguish the SCs within the doublets with smaller constants, one should use a filter with a narrower band using formula (4). In this case, the best

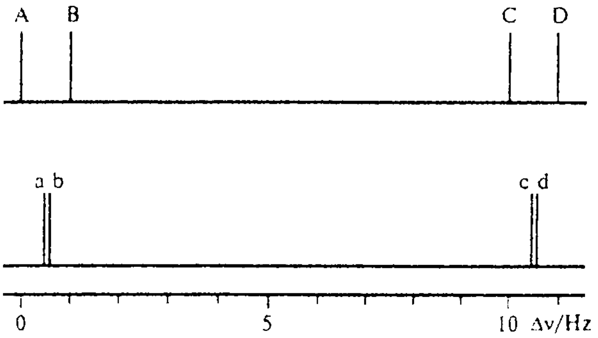

Fig. 1. Model pairs of doublets with splitting of 1.0 and 10.0 Hz (lines A, B, C, and D), and 0.1 and 10.0 Hz (lines a, b, c, and d). All components have unity intensities.

results are obtained at $\alpha = 1.0$ s for the constant $J = 1.0$ Hz, and at $\alpha = 10.0$ s for the constant $J = 0.1$ Hz (see Table 2). An analysis of relations (1) and (2) shows that the integral transforms characterize the positions of SCs within a multiplet but are insensitive to the relative frequency shifts of the multiplets (which are as a rule due to the errors in determining the chemical shifts). The properties of the integral transforms taken altogether make them almost invariant characteristics of the spectral components. Thus, a set of filters "fitted" to a multiplet allows one to construct a precise pattern of the *i*th SC:

$$\mathbf{X}_i = \mathbf{X}\{(\nu_i, A_i, \mathrm{IT}_1(\nu_i), \ldots, \mathrm{IT}_n(\nu_i)\}. \qquad (5)$$

Each individual *i*th SC determines an X_i, *i.e.*, a pattern to which, as mentioned above, a number of unresolved lines can pertain. A different situation takes place for transitions in the theoretical spectrum. Strictly speaking, they are infinitely narrow. Therefore, a spectral component in the theoretical spectrum is composed of all transitions with the same frequencies, whose intensities are added. A vector of parameters Y_j can be written for each *j*th theoretical line:

$$\mathbf{Y}_j = \mathbf{Y}\{(\nu_j, A_j, \mathrm{IT}_1(\nu_j), \ldots, \mathrm{IT}_n(\nu_j)\}, \qquad (6)$$

where the notations are analogous to those in Eq. (5).

In terms of pattern recognition theory, the "line assignment" problem consists in assigning each pattern of a theoretical line Y_j to a certain individual pattern X_i, $i = 1, 2, \ldots, N$.[10] If the experimental spectrum contains no signals of impurities and artifacts, and the theoretical model chosen is correct, then at least one pattern of a theoretical line should be assigned to each pattern X_i. If a spectral component consists of a number of unresolved experimental lines, then the same number of theoretical lines should correspond to it.

The actual assignment (or classification)[10] procedure deals with finding the correspondence between each set of values of the parameters at the transducer input and one of the patterns X_i based on the predetermined criteria. In this work, the minimum distance criterion was used, since it is relatively simple and illustrative. According to this criterion, the degree of similarity of the patterns treated as points in multidimensional data space is compared with the proximity of these points.

It should be particularly emphasized that the pattern space is anisotropic, *i.e.*, not all directions in this space are equivalent. Let us consider, for instance, the frequency and amplitude axes. According to information theory,[12] the informativities of these parameters H_ν and H_A are defined as follows:

$$H_\nu = \log_2(\Delta\nu/\delta\nu) \qquad (7)$$

and

$$H_A = \log_2(\Delta A/\delta A), \qquad (8)$$

where ΔA and $\Delta\nu$ are the ranges of the changes in the amplitude and frequency, respectively, and δA and $\delta\nu$ are the degrees of experimental errors of the parameters, respectively. It is obvious that because of the relatively low accuracy in measuring the intensities, the frequencies of the spectral components are much more informative than their amplitudes. The informational significance of integral transforms, which are linear combinations of frequencies and amplitudes, is of intermediate value.

The distances between vectors $\mathbf{X}_i$ and $\mathbf{Y}_j$ in the anisotropic space are defined by the square of the generalized functional of residuals Δ_{ij}^2:

$$\begin{aligned}\Delta_{ij}^2 = a_1\left\{\nu_i^{\mathrm{exp}} - \nu_j^{\mathrm{th}}\right\}^2 + a_2\left\{A_i^{\mathrm{exp}} - A_j^{\mathrm{th}}\right\}^2 + \\ + a_3\left\{\mathrm{IT}_1\left(\nu_i^{\mathrm{exp}}\right) - \mathrm{IT}_1\left(\nu_j^{\mathrm{th}}\right)\right\}^2 + \ldots + \\ + a_n\left\{\mathrm{IT}_{n-2}\left(\nu_i^{\mathrm{exp}}\right) - \mathrm{IT}_{n-2}\left(\nu_j^{\mathrm{th}}\right)\right\}^2 .\end{aligned} \qquad (9)$$

Table 2. Integral transforms of the components of the quartets shown in Fig. 1, obtained by using a function of the type (2)

α/s	Component							
	A	B	C	D	a	b	c	d
0.1	−0.698	−0.572	0.572	0.698	−0.643	−0.630	0.630	0.643
1.0	−0.439	0.185	−0.185	0.439	−0.189	−0.064	0.064	0.189
10.0	−0.075	0.050	−0.050	0.075	−0.331	0.306	—0.306	0.331

with parameters a_i normalized:

$$\sum_{i=1}^{n} a_i^2 = 1 \,. \tag{10}$$

Thus, the coefficients a_i determine the relative significance of the different characteristics of an SC in its pattern. It can be seen from relations (7) and (8) that the relative informativities of the parameters of the spectral components from which the pattern is constructed are different. It is also very important that their relative significance depends on the *a priori* information, *i.e.*, on the accuracy of the initial approximation of the spectral parameters. If only rough estimates of their values are known, distinctions between the frequencies of the experimental spectral components and their theoretical analogs are of little importance if the model of the spin system is correctly chosen. On the contrary, when the initial spectral parameters are known with a high degree of accuracy, the rms deviation criterion is of paramount importance.

The informational significance of the integral transforms changes essentially as the values of the spectral parameters approach their true values. Therefore, we made it possible to vary the ratio between the coefficients a_i in the equation of the generalized residual (9) in the algorithm. It should be noted that according to the theory,[10] the classifier constructed according to the principle of minimum distance calculates the distance in multidimensional space between vectors $\mathbf{X}_i$ and $\mathbf{Y}_j$ and assigns pattern Y_j to the nearest pattern X_k if the following condition is fulfilled

$$\Delta_{jk}^2 < \Delta_{ij}^2 \tag{11}$$

for all $j \neq k$.

The PAREMUS program package. The principles set forth in the previous section form the basis for the method we developed for automated assignment of the lines in the NMR spectra. The corresponding PAREMUS program package is written in the algorithmic language FORTRAN on the basis of a modified LAOCOON5 program.[13] It works according to the "long" cycle scheme as follows.

Input data are the type and values of the parameters of the spin system as well as the frequencies and intensities of the spectral components of the experimental spectrum. Based on these data, the first trial spectrum is calculated and automatically matched with the experimental spectrum. The operator must only make an approximate estimate of the expected value of the r.m.s. deviation. The correspondence between the frequencies of the experimental spectral components and the numbers of theoretical transitions obtained allows one to calculate the second approximation of the experimental spectrum. If the value of the r.m.s. deviation obtained is reasonable, the procedure is terminated, otherwise a new "long" cycle begins, in which the lines are reassigned, *etc*.

The correlation between the values of coefficients a_i in the equation of the generalized residual (9) and the value of the r.m.s. deviation obtained is found in the iterative calculations. In the course of the procedure of automated assignment the user can manually edit the assignment using a graphic interface. All intermediate results are stored in individual files.

Practical aspects of the automated assignment of the NMR spectra. In the general case, spectra containing signals from impurities, artifacts, and unresolved individual lines are analyzed. The manner of interpretation of unresolved components is not always clear at the deconvolution stage. Broadening can be caused by both inadequate resolution and the short relaxation time of the nucleus due to the exchange processes or interactions with the quadrupole nuclei. Therefore, parasite and broadened lines are ignored when rough fitting of the initial parameters is performed. The details of the multiplet structure become clear as experimental and theoretical spectra approach each other in the course of the iterative procedure. This allows one to distinguish signals from impurities with higher reliability and make reassignments in accordance with the refined model.

The analysis of the ^{1}H NMR spectrum of the five-spin system of pyridine performed in the classical work by Castellano and Bothner-By[2] was the first test for programs of the LAOCOON type. The ^{1}H NMR spectrum of pyridine is a typical strongly-coupled spectrum of aromatic protons. Non-automated assignment of all lines (about 100) is extremely time-consuming, and the use of a graphic interface does not solve the problem completely. To estimate the effectiveness of the PAREMUS program package, it was of interest to perform an automated analysis of the multiplet structure of a spectrum of such type. For this purpose, the spectrum of the AA′BB′C type spin system was simulated (see Fig. 2 and the parameters in Table 1) containing Lorentzian lines (with a half-width of 0.05 Hz) and additional white noise. The frequencies and amplitudes of the spectral components were determined from the top-peak positions.

The initial values of the parameters from the paper by Castellano and Bothner-By[2] were used as trial values of the spectral parameters (see Table 1). The values of these coupling constants are typical of the *ortho*-, *meta*-, and *para*-protons of aromatic compounds. It can be seen from Fig. 2 that the multiplet structures of the "experimental" and theoretical spectra are similar, while no detailed coincidence is observed. Equation (9) containing five terms was used in the calculations, and coefficients a_i had close values. With the aid of the PAREMUS procedure, approximately two thirds of all "experimental" lines were assigned in the course of the first "long" cycle, and the second cycle resulted in the complete assignment of the whole "experimental" spectrum. The final value of the r.m.s. deviation (0.010 Hz) corresponded to the preset level of digital resolution.

One of the traditionally nontrivial problems of NMR spectroscopy is the determination of the unknown signs

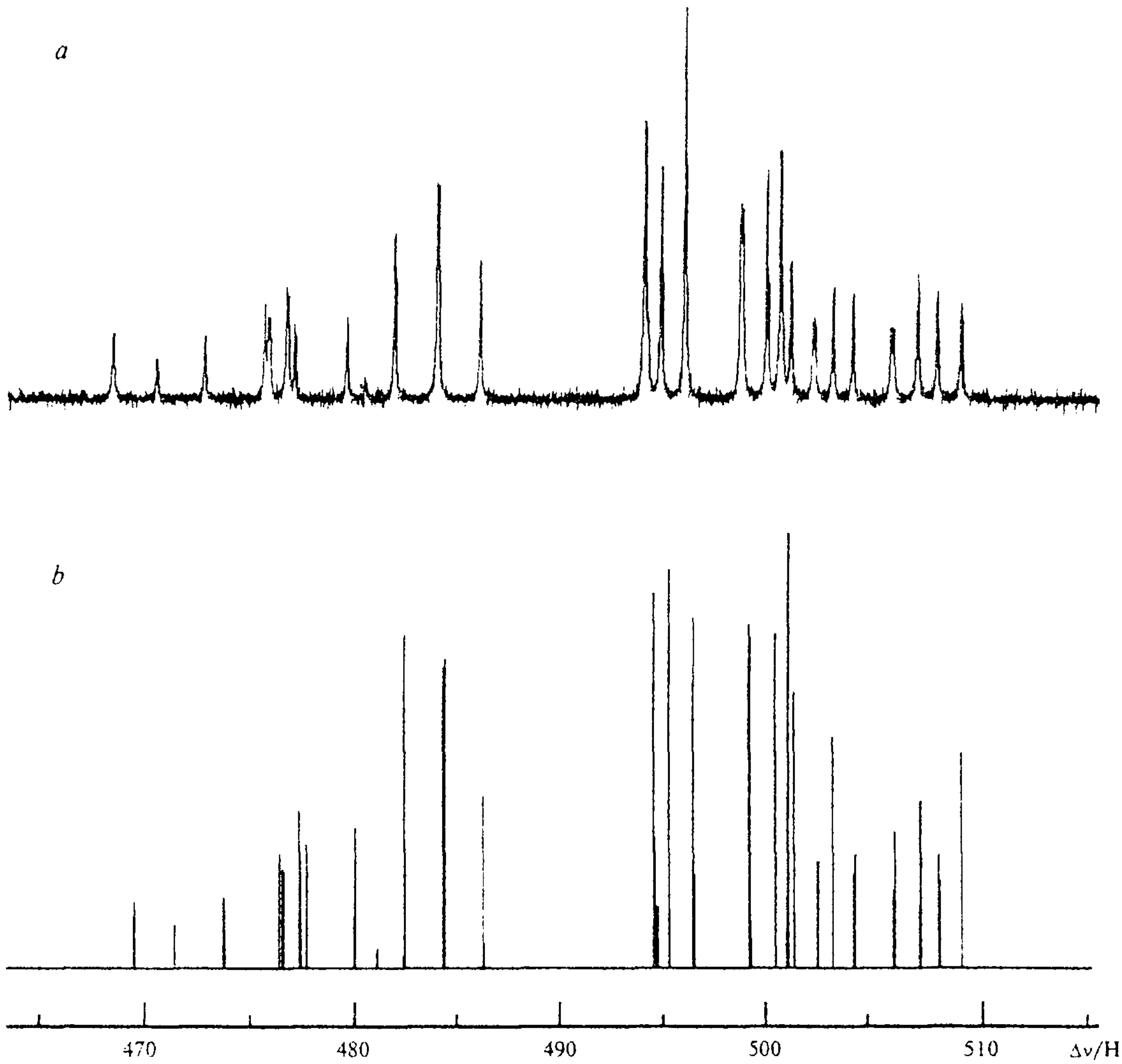

Fig. 2. ^{1}H NMR spectrum of pyridine (the region of the α-H and β-H protons): *a*, simulated "experimental", *b*, calculated using the parameters from Table 2, see text.

of the spin-spin coupling constants. The sign of the constant can be directly deduced from the shape of a strongly-coupled NMR spectrum. First-order spectra require additional information obtained from either special double-resonance (tickling or INDOR, see Refs. 14 and 15) or two-dimensional (SOFTCOSY or E-COSY, see, for instance, Ref. 15) experiments. Perturbations caused by strong coupling can be insignificant in spectra of this type, and a complete and precise analysis of their multiplet structure[16,17] is required to reach demonstrative conclusions on the signs of the coupling constants. If several constants simultaneously have unknown signs, calculations become time-consuming due to the large number of variants of the assignment.

This is just the situation that arises in the interpretation of the strongly-coupled ^{1}H NMR spectra of naphthalene derivatives with allowance for the long-range inter-ring coupling constants. In this regard, the analysis of the ^{1}H NMR spectrum of 1,2,3-trichloronaphthalene, performed in the formalism for the five-spin ABCDE type system with no allowance for the effects of the orientation of the molecules in a strong magnetic field, is typical. The experimental spectrum is close to the first-order spectrum, and its multiplet structure is completely resolved. It contains all 80 spectral components (see Fig. 3). The minimum splitting was about ~0.1 Hz at a digital resolution of ~0.005 Hz per point.

To demonstrate the potential of the method, we performed a two-stage analysis of this spectrum. To study the domain of convergence of the method, two series of test calculations of the multiplet structure of this spectrum with various initial values of spectral

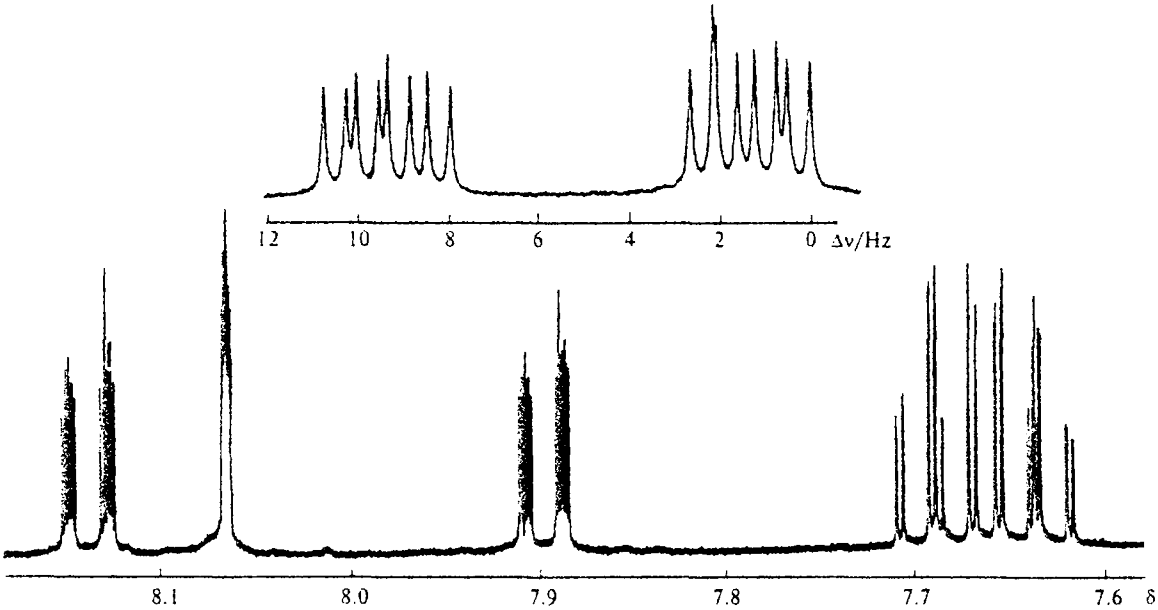

Fig. 3. ^{1}H NMR spectrum of 1,2,3-trichloronaphthalene, the region of the H(5) proton is shown in the inset.

parameters were carried out at the first stage. The signals of all protons were preliminarily assigned using the double resonance technique. The first set of parameters (see version A in Table 3) contained the values of the coupling constants obtained by Laatikainen for naphthalene[18] and chemical shifts determined from the first-order analysis of the spectrum. Version B contained a "less precise" set of parameters. In this case all vicinal coupling constants corresponded to their upper limit for the naphthalene derivatives (10.0 Hz). All coupling constants for the *meta*- and *para*-protons were assumed to be equal to 1.0 Hz. In addition, the chemical shifts of the H(6) and H(7) protons were brought close together. The assignment obtained in the course of the calculations was not edited.

Only one "long" cycle was performed to obtain the final data (see Table 3) using the parameters of version A, whereas three "long" cycles were required to obtain the same final solution using the parameters of version B. In the latter case, a set of parameters had arbitrary chosen values, which, however, were within the limits of possible values for naphthalene derivatives. Therefore version B can serve as a reasonable estimate of the upper limit of the domain of convergence of the method.

The relative signs of the ^{1}H—^{1}H coupling constants were determined at the second stage. For this purpose, we performed iterative calculations for all alternate sets of the spectral parameters corresponding to all sixteen possible combinations of the relative signs of the proton inter-ring constants. Since the structure of the ^{1}H NMR spectrum appreciably differed from that of the first-order spectrum, one could expect that the spectral parameters would be sensitive to changes in the relative signs of the inter-ring constants. The values of the coupling constants, the r.m.s. deviations, and the Fisher criterion for the sets of the signs of the inter-ring constants studied are listed in Table 4. A comparison of the experimental values of the Fisher criterion with the tabulated ones (F = 1.69 at the confidence level α = 99% for 80 degrees of freedom[19]) shows that a statistically significant minimum for the r.m.s. deviation is observed for only one set of inter-ring coupling constants containing negative values of $J_{H(4)-H(5)}$ and $J_{H(4)-H(7)}$ (variant G, see Table 4). According to the

Table 3. Parameters of the ^{1}H NMR spectrum (relative frequencies $\Delta\nu_{H(i)}$ and spin-spin coupling constants $J_{H(i)-H(j)}$/Hz) of 1,2,3-trichloronaphthalene (Varian VXR-400 spectrometer, 2 *M* solution in acetone-d$_6$ at 301 K)

Parameter	Obtained values[a]	Trial values	
		Version A	Version B
$\Delta\nu_{H(4)}$	179.538	180.0	177.0
$\Delta\nu_{H(5)}$	112.965	112.0	112.0
$\Delta\nu_{H(6)}$	9.728	10.0	11.0
$\Delta\nu_{H(7)}$	29.231	29.0	28.0
$\Delta\nu_{H(8)}$	208.107	208.0	208.0
$J_{H(4)-H(5)}$	−0.492	−0.44	−1.0
$J_{H(4)-H(6)}$	0.272	0.22	0.1
$J_{H(4)-H(7)}$	−0.163	−0.16	−0.1
$J_{H(4)-H(8)}$	0.846	0.85	1.0
$J_{H(5)-H(6)}$	8.296	8.31	10.0
$J_{H(5)-H(7)}$	1.236	1.24	1.0
$J_{H(5)-H(8)}$	0.722	0.76	1.0
$J_{H(6)-H(7)}$	6.947	6.86	10.0
$J_{H(6)-H(8)}$	1.135	1.24	1.0
$J_{H(7)-H(8)}$	8.637	8.31	10.0

[a] The r.m.s. deviation for 80 lines assigned is 0.0035 Hz.

Table 4. Results of the analysis of the ^1H NMR spectrum of 1,2,3-trichloronaphthalene for different variants (A—P) of signs of the inter-ring spin-spin coupling constants $J_{H(i)-H(j)}$/Hz[a]

Variant	$J_{H(4)-H(5)}$	$J_{H(4)-H(6)}$	$J_{H(4)-H(7)}$	$J_{H(4)-H(8)}$	r.m.s.[b]	F[c]
A	0.491	0.259	0.137	0.847	0.0061	3.87
B	−0.492	0.261	0.137	0.847	0.0058	3.50
C	0.492	−0.272	0.156	0.847	0.0072	5.40
D	0.491	0.270	−0.162	0.846	0.0066	4.53
E	0.491	0.259	0.139	−0.847	0.0065	4.40
F	−0.492	0.260	0.139	−0.847	0.0063	4.13
G	−0.492	0.272	−0.163	0.846	0.0035	1.00
H	−0.492	−0.271	0.156	0.847	0.0066	4.53
I	0.492	−0.272	0.158	−0.846	0.0066	3.56
J	0.492	−0.261	−0.143	0.846	0.0064	3.34
K	0.491	0.269	−0.161	−0.845	0.0089	6.47
L	−0.492	−0.259	−0.144	0.847	0.0066	4.53
M	−0.492	−0.271	0.158	−0.846	0.0065	4.40
N	−0.492	0.271	−0.161	−0.847	0.0070	5.10
O	0.492	−0.261	−0.142	−0.847	0.0058	3.10
P	−0.492	−0.259	−0.143	−0.847	0.0062	4.00

[a] The values of the chemical shifts and intra-ring coupling constants correspond to those listed in Table 3.

[b] The r.m.s. deviation between the frequencies of the experimental and calculated spectra (Hz); 80 transitions assigned.

[c] The Fisher criterion.

Table 5. The $J_{C(i)-H(j)}$ spin-spin coupling constants (Hz) for the **a** (^{13}C-6) isotopomer of 1,2,3-trichloronaphthalene (Varian VXR-400 spectrometer, 2 *M* solution in acetone-d_6 at 301 K)

Variant	$J_{C(6)-H(4)}$	$J_{C(6)-H(5)}$	$J_{C(6)-H(6)}$	$J_{C(6)-H(7)}$	$J_{C(6)-H(8)}$	r.m.s.[a]	F[b]
A	−0.150	0.559	162.406	1.462	8.[illegible]9	0.0281	1.00
B	0.155	0.572	162.417	1.471	8.440	0.0436	2.41

[a] The r.m.s. deviation between the frequencies of the experimental and calculated spectra (Hz); 30 transitions assigned. [b] The Fisher criterion.

Table 6. The $J_{C(i)-H(j)}$ coupling constants (Hz) for the **b** (^{13}C-7) isotopomer of 1,2,3-trichloronaphthalene (Varian VXR-400 spectrometer, 2 *M* solution in acetone-d_6 at 301 K)

Variant	$J_{C(7)-H(4)}$	$J_{C(7)-H(5)}$	$J_{C(7)-H(6)}$	$J_{C(7)-H(7)}$	$J_{C(7)-H(8)}$	r.m.s.[a]	F[b]
A	−0.225	8.461	1.532	162.525	0.208	0.0127	1.00
B	0.254	8.469	1.546	162.544	0.207	0.0223	3.08

[a] The r.m.s deviation between the frequencies of the experimental and calculated spectra (Hz); 20 transitions assigned.

[b] The Fisher criterion.

published data for naphthalene derivatives,[18,20,21] only these coupling constants have a negative sign. The example discussed above demonstrates that the algorithm is highly effective in determining the relative signs of the constants in spectra with pronounced manifestation of strong coupling.

Little data on the values and signs of the inter-ring $^{13}C—^1H$ coupling constants of the naphthalene derivatives has been reported, and only relatively large vicinal constants[22] are reliable. A systematic study of the dependence of the long-range inter-ring carbon-proton coupling constants on the structural factors, neglected so far, would be of considerable interest. In this connection it is important to determine the signs and values of the inter-ring $^{13}C—^1H$ coupling constants in 1,2,3-trichloronaphthalene. To solve this problem, the approach we approved was used in the analysis of the ^1H NMR spectrum of this compound. Two variants of the spectral parameters differing in the relative signs of the inter-ring $J_{C(6)-H(4)}$ and $J_{C(7)-H(4)}$ spin-spin coupling constants (see Table 5, and Fig. 4 and data in Table 6, respectively) were calculated for two isotopomers, **a** ($^{13}C(6)$) and **b** ($^{13}C(7)$). In both cases a statistically significant minimum for the r.m.s. deviation (the tabulated

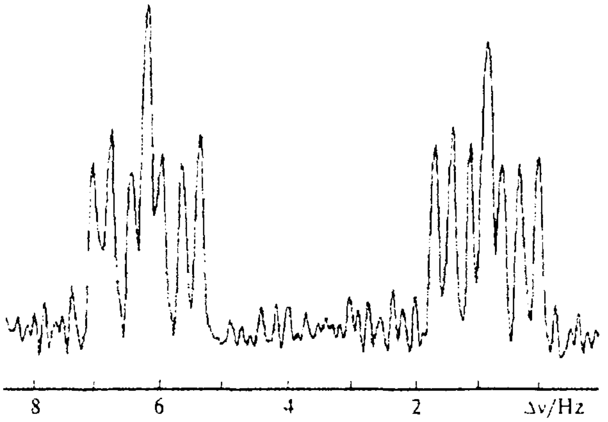

Fig. 4. High-field half of the ^{13}C NMR spectrum of the **b** (^{13}C-7) isotopomer of 1,2,3-trichloronaphthalene.

value of the Fisher criterion F = 2.34 at the confidence level α = 99% for 30 degrees of freedom and F = 2.94 for 20 degrees of freedom)[19] was observed for the set of spectral parameters with negative inter-ring spin-spin coupling constants (Table 5 and 6, variant A).

Thus, a new method for correctly assigning the components of experimental and theoretical NMR spectra based on pattern recognition theory has been developed. Along with the information on frequencies and intensities traditionally used, valuable information is also provided by the parameters characterizing the relative positions of the spectral components. The method proposed has a wide domain of convergence and allows one to perform automated matching of components in the analysis of the multiplet structure of complicated NMR spectra.

This work was financially supported by the Russian Foundation for Basic Research (Project No. 95-03-09639).

References

1. S. M. Castellano and A. A. Bothner-By, *J. Chem. Phys.*, 1964, **41**, 3863.
2. A. A. Bothner-By and S. M. Castellano, in *Computer Progress for Chemistry*, Ed. D. F. De Taar, New York, Benjamin, 1968, **1**, 10.
3. J. Heinzer, *J. Magn. Res.*, 1977, **26**, 301.
4. D. S. Stephenson and G. Binsch, *J. Magn. Res.*, 1980, **37**, 395.
5. D. S. Stephenson and G. Binsch, *J. Magn. Res.*, 1980, **37**, 409.
6. R. L. Laatikainen, *J. Magn. Res.*, 1991, **92**, 1.
7. S. S. Golotvin and V. A. Chertkov, *12 th European NMR Conference*, Oulu, Finland, 1994, Abstracts, 295.
8. E. G. Jurner and W. P. Wynne, *J. Chem. Soc.*, 1941, 243.
9. K. R. Bedford, G. W. Burton, and P. B. De la Mare, *Tetrahedron Lett.*, 1973, **334**, 3205.
10. J. T. Tou and R. C. Gonzales, *Pattern Recognition Principles*, Addison-Wesley Publishing Company Inc., London, 1974.
11. P. Diehl, S. Sycora, and J. Vogt, *J. Magn. Res.*, 1975, **19**, 67.
12. C. Shannon and W. Weaver, *The Mathematical Theory of Communication*, Urbana, University of Illinois Press, 1949, 149.
13. M. Attimonelli and O. Sciacovelli, *Org. Magn. Reson.*, 1980, **13**, 277.
14. H. Günter, NMR-*Spectroskopie*, Georg Thieme Verlag, Stuttgart, 1973.
15. R. R. Ernst, G. Bodenhausen, and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, 1987.
16. V. A. Chertkov, Yu. K. Grishin, and N. M. Sergeyev, *J. Magn. Res.*, 1976, **24**, 275.
17. V. A. Chertkov and Yu. K. Grishin, *Zh. Strukt. Khim.*, 1977, **18**, 776 [*J. Struct. Chem.*, 1977, **18** (Engl. Transl.)].
18. R. L. Laatikainen, *J. Magn. Res.*, 1988, **78**, 127.
19. B. L. Van der Waerden, *Mathemathische Statistik*, Springer Verlag, Berlin, 1974.
20. M. Attimonelli and O. Sciacovelli, *Org. Magn. Reson.*, 1977, **9**, 601.
21. M. Attimonelli and O. Sciacovelli, *Org. Magn. Reson.*, 1980, **13**, 277.
22. P. E. Hansen, *Progress in NMR Spectroscopy*, 1981, **14**, 175.

Received April 8, 1996;
in revised form November 13, 1996