



COMPUTER COMPENSATION FOR NMR QUANTITATIVE
ANALYSIS OF TRACE COMPONENTS

by

Takashi Nakayama

and

Yuzuru Fujiwara *

July 22, 1981

INSTITUTE
OF
INFORMATION SCIENCES AND ELECTRONICS

UNIVERSITY OF TSUKUBA

BRIEF

A computer program has been developed to determine trace components and to separate overlapping components in NMR spectra. The accuracy of quantitative analysis has been improved by two significant figures. The program was applied to determining the abundance of ^{13}C .

ABSTRACT

A computer program has been written that determines trace components and separates overlapping components in multicomponent NMR spectra. This program uses the Lorentzian curve as a theoretical curve of NMR spectra. The coefficients of the Lorentzian are determined by the method of least squares. Systematic errors such as baseline/phase distortion are compensated and random errors are smoothed by taking moving averages, so that these processes contribute substantially to decreasing the accumulation time of spectral data. The accuracy of quantitative analysis of trace components has been improved by two significant figures. This program was applied to determining the abundance of ^{13}C and the saponification degree of PVA.

INTRODUCTION

Programs have been developed to determine trace components and to separate overlapping components in multicomponent NMR spectra. For quantitative analysis of trace components, accumulation is the only available method for instruments which are not supported with hardware compensation, and sufficient accumulation is not always feasible because of saturation of the main peak. The method of least squares is useful for determining the composition of trace component spectra (1-3). In this program, the Lorentzian curve is used as a theoretical curve of NMR spectra, and the coefficients of the Lorentzian are determined by the method of least squares. As the Lorentzian is a nonlinear model, it is reduced to a linear one by approximation so that the method of least squares can be applied (4).

It is assumed that the Lorentzian curve approximates an NMR spectrum adequately. The Lorentzian is defined:

$$f(x) = \frac{1}{2\pi} \frac{k \cdot w}{(x - x_0)^2 + w^2/4} \quad (1)$$

where k , x and w correspond to the intensity, the position of a peak center and the half width respectively. Equation (1) can be expressed:

$$f(x) = \frac{c_2 \cdot c_3}{(x - c_1)^2 + c_3^2} \quad (2)$$

The purpose is to estimate parameters c_1 , c_2 and c_3 of expression (2) by the method of least squares. As the model used is nonlinear, it is reduced to a linear one by approximation so that the method of least squares can be used. The following expression is obtained from a Taylor expansion which is terminated at a linear term:

$$f(x; c_1 + \Delta c_1, c_2 + \Delta c_2, c_3 + \Delta c_3) = f(x; c_1, c_2, c_3) + \sum_j \frac{\partial f}{\partial c_j} \Delta c_j \quad (3)$$

Supposing that the left side of expression (3) is the observed value, n equations are obtained per n observed values:

$$\sum_j \frac{\partial f}{\partial c_j} \Big|_{x_i} \cdot \Delta c_j = \Delta f \Big|_{x_i}, \quad i = 1, \dots, n$$

where

$$\Delta f|_{x_i} = f(x_i; c_1 + \Delta c_1, c_2 + \Delta c_2, c_3 + \Delta c_3) - f(x_i; c_1, c_2, c_3).$$

Applying the method of least squares, the problem is reduced to estimating Δc_j ($j=1,2,3$) which minimize the following sum of squares:

$$SS(c_1, c_2, c_3) = \sum_k \left(\Delta f|_{x_k} - \sum_j \frac{\partial f}{\partial c_j} \Big|_{x_k} \cdot \Delta c_j \right)^2$$

The following equation is obtained by $\partial SS / \partial (\Delta c_i) = 0$ ($i=1,2,3$).

$$\sum_k \frac{\partial f}{\partial c_j} \Big|_{x_k} \left(\Delta f|_{x_k} - \sum_j \frac{\partial f}{\partial c_j} \Big|_{x_k} \cdot \Delta c_j \right) = 0$$

From this:

$$\sum_j \left(\sum_k \frac{\partial f}{\partial c_i} \Big|_{x_k} \cdot \frac{\partial f}{\partial c_j} \Big|_{x_k} \right) \cdot \Delta c_j = \sum_k \frac{\partial f}{\partial c_i} \Big|_{x_k} \cdot \Delta f|_{x_k}$$

This is expressed:

$$A \cdot \underline{c} = \underline{b} \tag{4}$$

where $A = (a_{ij})$, $\underline{b} = (b_1, b_2, b_3)^t$, $\underline{c} = (c_1, c_2, c_3)^t$,

$$a_{ij} = \sum_k \frac{\partial f}{\partial c_i} \Big|_{x_k} \cdot \frac{\partial f}{\partial c_j} \Big|_{x_k} \quad \text{and} \quad b_i = \sum_k \frac{\partial f}{\partial c_i} \Big|_{x_k} \cdot \Delta f|_{x_k}$$

(t means transposed).

Parameters c_j are replaced by $c_j + \Delta c_j$ by solving equation (4), and this process is repeated until the residual converges.

EXPERIMENTAL

This method of quantitative analysis was applied to several kinds of spectra. In this paper, the analysis of abundance of ^{13}C in CHCl_3 is discussed in detail as an example of the determination of a trace component. The method was applied also to the determination of the saponification degrees of PVA to show the separating process for overlapping components in multicomponent NMR spectra. NMR spectral data were collected using a VARIAN HA-100 controlled by a DATA 620/i computer, and output to paper tape.

1. Constitution of Programs. The block diagram of the procedure is shown in Figure 1. NMR spectral data stored by the DATA 620/i computer are dumped in the form of binary values to paper tape

by the program Binary Dump, and edited by the program Editing.

Editing performs the following data processing: (1) Smoothing Data. Raw data are smoothed by the weighted moving average method to reduce random error. The number of points for calculating the moving average is selected from the set of { 5,7,9, ..., 25 }. The abscissa is partitioned into several sections according to the kind of spectral curve, and the most suitable number is selected. (2) Phase compensation. To compensate for phase distortion, each peak is inverted at the center of the peak ($x = c_1$), and the inverted peak is added to the original one. That is to say, $(f(x) + f(c_1 - x)) / 2$ is used as spectral data instead of the original spectral data $f(x)$. This procedure is applied to bisymmetric spectral patterns. (3) Baseline compensation. A new baseline is set up so that the average of 50 points at the bottom of the spectral curve is to be zero. Systematic errors are reduced by this treatment. (4) Accumulation of trace components. Trace components are extracted from the original data and accumulated under the condition that the peak center

should be adjusted.

After the editing process described above, the reduced data are output to magnetic tape. The Plot program is used to plot spectra and the Lorentzian curve. The program for the Method of Least Squares performs the calculation for the algorithm described in the previous section, and the Main program controls all of these programs and generates the report.

2. Selection of Initial Values. The selection of the initial values of parameters c_1 , c_2 and c_3 (expressed here as c_{01} , c_{02} and c_{03}) is important in the application of the method of least squares. c_{01} , which corresponds to the position of the peak center, is set at the position of the maximal value of spectral data; c_{02} , which corresponds to the intensity, at the maximal value of spectral data; and c_{03} , which corresponds to the half width, at the mean of the actual values obtained from several spectra. The convergence is significantly influenced mainly by the distortion of c_{01} and is stable for c_{02} and c_{03} . It is sufficient to select initial values as described above because c_{01} is determined accurately.

3. Selection of Sampling Points for the Method of Least Squares. Selection of sampling points also significantly influences convergence. It was determined with pseudo data that good values result from arbitrarily selected points given that the points are sampled approximately symmetrically about the peak center. There is almost no change as the number of sampling points is varied within the range from 5 to 30 points (cf. Figure 4). Consequently the sampling points for the actual spectra are selected from the frequency area near the peak center where the random error seems relatively small.

RESULTS AND DISCUSSION

1. Determination of a trace component.

General. The suitability of this method for determining trace components in multicomponent NMR spectra is most clearly shown when the trace component is hidden in the shoulder of a main spectral component. The process of detecting such

a trace component is shown in Figure 2. Figure 2(a) shows the original spectral pattern which contains a trace component on its shoulder. The separated trace component is also shown in Figure 2(a), and Figure 2(b) shows the trace component after it has been accumulated 30 times; its pattern is refined and enlarged enough to approximate by the Lorentzian curve.

In practice, this technique for determining a trace component was applied to the analysis of the abundance of ^{13}C in CHCl_3 . The abundance of ^{13}C is given by

$$T = s / (S + s) \quad (5)$$

where s is the true area of the satellite peak and S is the true area of the main peak. The area obtained by this program contains various errors such as systematic (instrumental) errors and random errors, and so the abundance calculated can be expressed:

$$R = (s + e) / (S + E + s + e) \quad (6)$$

where e and E are the errors associated with experiment as well as calculation. If p and q are relative errors of s and S in expression (6), then expression (5) is reduced to:

$$T = R(1+q) / (1+q+R(q-p)) \quad (7)$$

The abundance of ^{13}C is estimated by expression (7).

Error factors include (1) systematic errors resulting from instruments, (2) random errors during observation, (3) rounding errors, and (4) computational errors. The rounding error and computational error above are always included in the estimation described below, so they are not treated directly. Systematic errors are reduced by using $(f(x) + f(c_1 - x)) / 2$ instead of $f(x)$ as spectral data, which means the correction of phase distortion, and by setting up the new baseline. Random error is reduced by the smoothing process. The relative error δ is estimated by

$$\bar{\delta} - \sigma_m < \delta < \bar{\delta} + \sigma_m$$

where $\bar{\delta}$ is the mean value of relative errors of estimate and σ_m^2 is the variance of $\bar{\delta}$.

Determination of Baseline. The baseline is determined so that the intensities of both ends of the spectral curve become equal to zero after phase compensation. In practice, baseline-compensated data is obtained by subtracting a straight line which connects both ends of a spectrum from the value of phase compensated data. This processing has almost no influence on the precision of approximation if the baseline range considered is large enough for the half width of the peak. Figure 3 shows the effect of the ratio "baseline range / half width" on the accuracy of measurement of the area. Figure 3 is for calculated pseudo spectral data. For actual data, the following values were obtained:

$$\begin{aligned}\bar{\delta}_M &= -0.0012, & \sigma_M &= 0.0002 \\ \bar{\delta}_S &= -0.0013, & \sigma_S &= 0.0001\end{aligned}\tag{8}$$

where $\bar{\delta}$ is the mean of the relative error, σ^2 is its variance, and subscripts M and S represent main and satellite peaks respectively.

Random Error. As described above, raw data are smoothed by the weighted moving average method. The extent to which random error is reduced by this smoothing method has been examined for pseudo data. The distribution of relative error as a function of the number of points used for computing the moving average is shown in Figure 4. The case of the number of sampling points equal to zero is the case of random error without smoothing. Figure 4 also shows that the five-points smoothing gives the minimal relative error.

The distribution of the relative error as a function of the relative intensity of irregular noise is shown in Figure 5 for data with five-points smoothing and in Figure 6 without smoothing. For the actual spectra, the ratios "noise intensity / maximal value of peak" in Figure 5 and Figure 6 are about 0.0003 (main peak) and 0.0020 (satellite). Consequently the smoothing process is not very effective for the main peak, but fairly effective for the satellite peak. The following values are obtained:

$$\begin{aligned}\delta_M &= 0.0006, & \sigma_M &= 0.0001 \\ \delta_S &= 0.0003, & \sigma_S &= 0.0007\end{aligned}\tag{9}$$

Finally, the abundance of ^{13}C estimated by expressions (7), (8) and (9) is

$$1.1077 \pm 0.0012$$

2. Separation of overlapping components.

The saponification degree of PVA was determined by this method, as an example of separating overlapping components in multicomponent NMR spectra. The object spectral data are approximated by the sum of four Lorentzian curves, so 12 parameters are determined by the method described in the previous section. The original spectrum and the result of separation are shown in Figure 7. Table I compares the saponification degrees computed by this method with the ones determined by chemical analysis.

CONCLUSION

This program is very useful not only for determining trace components of NMR spectra, but also for separating overlapping components in a multi-component mixture and detecting a small shoulder component. Thus the program is suitable for precise analysis, such as measuring the abundance of ^{13}C , the saponification degree of PVA, the concentration of 3·4 couplings in polyisoprene, and the ratio of cis-trans form in PIP.

LITERATURE CITED

- (1) Barnett, H.A.; Bartoli, A. Anal. Chem. 1960, 32, 1153-1156.
- (2) Blackburn, J.A. Anal. Chem. 1965, 37, 1000-1003.
- (3) Roberts, S.M.; Wilkinson, D.H.; Walker, L.R. Anal. Chem. 1970, 42, 886-893.
- (4) Physical Society of Japan "Data Processing of Physical Experiment by Computer", Saiensu-sha, 1973.

FIGURE LEGENDS

- Figure 1. Block diagram of quantitative analysis program. — represents the program control and == represents the data flow.
- Figure 2. Accumulation of a detected trace component on a shoulder of a main peak. The lower peak extracted from the main peak (shown in (a)) enlarges to (b) after accumulation 30 times.
- Figure 3. The effect of the ratio "baseline range / half width" on the accuracy of measurement of the area.
- Figure 4. Distribution of random errors by the number of sampling points for moving averages.
- Figure 5. Distribution of relative errors of estimated areas for irregular noise (with five-point smoothing).
- Figure 6. Distribution of relative errors of estimated areas for irregular noise (without smoothing).
- Figure 7. Separation of overlapping components.

Saponification degrees (%)

sample *)	saponification degree
87.6	85.8
96.5	96.1
98.6	98.2
99.3	99.5
99.7	99.6
99.9	100.0

*)Values determined by chemical analysis.

Table I

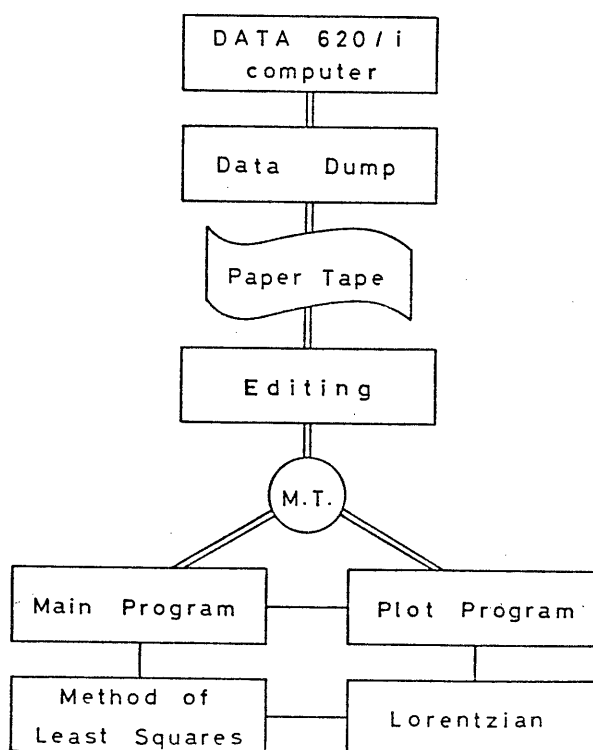
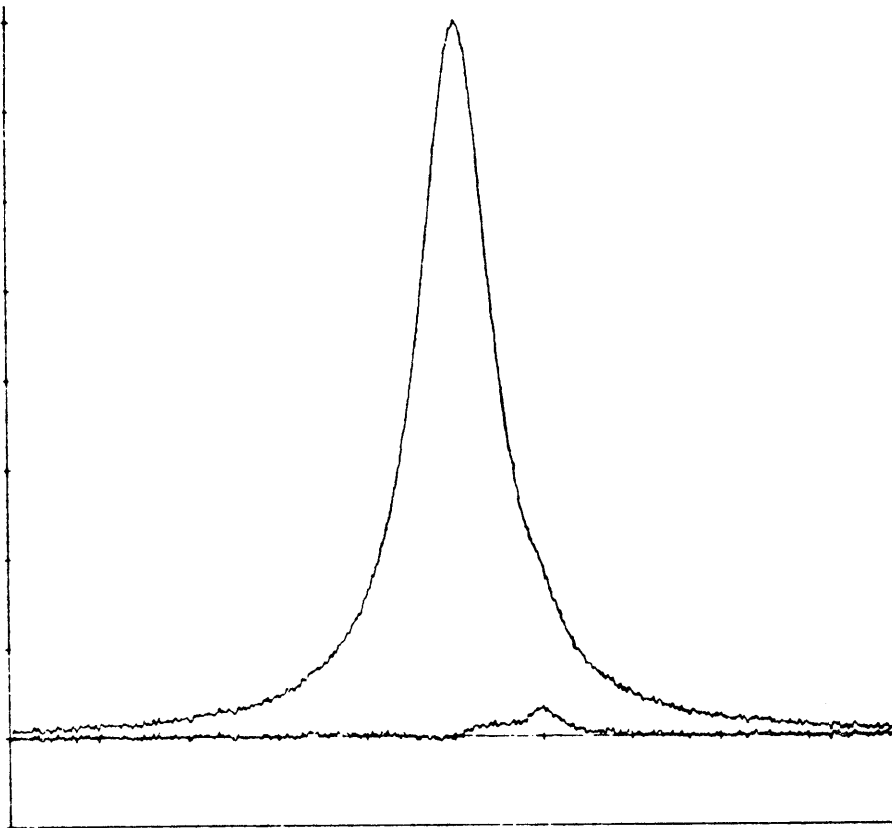
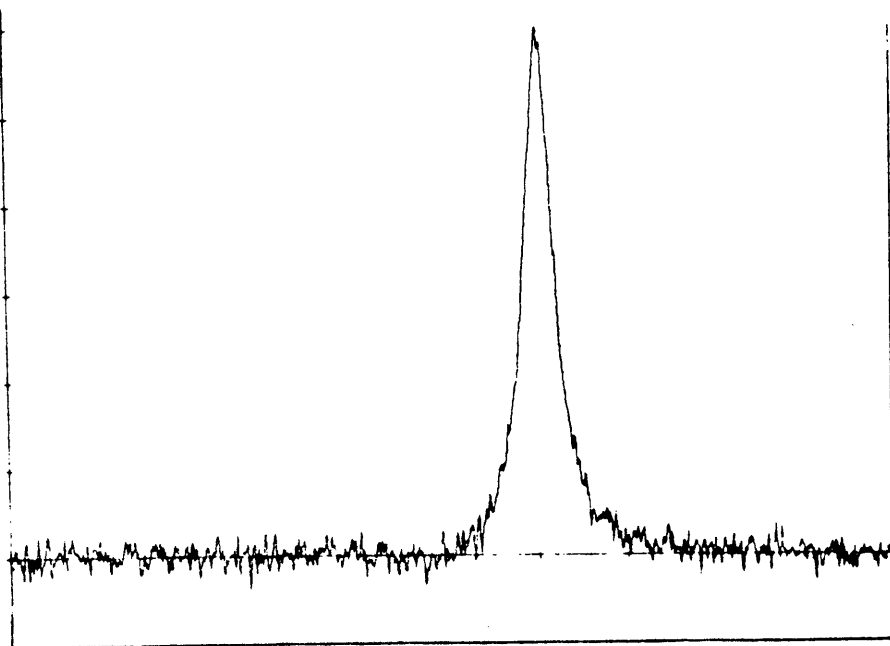


Figure 1



(a)



(b)

Figure 2

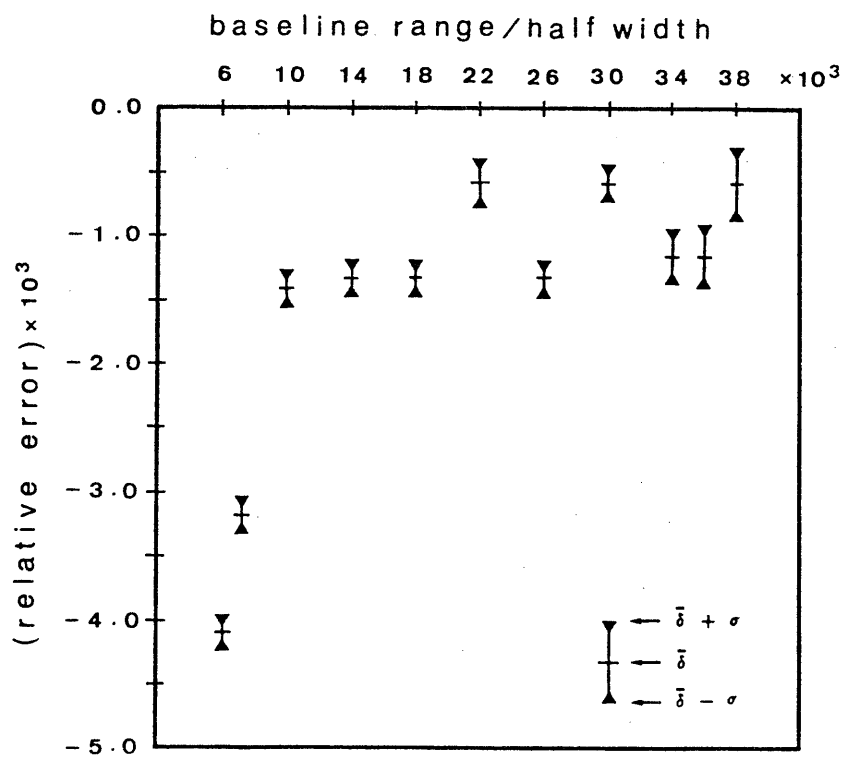


Figure 3

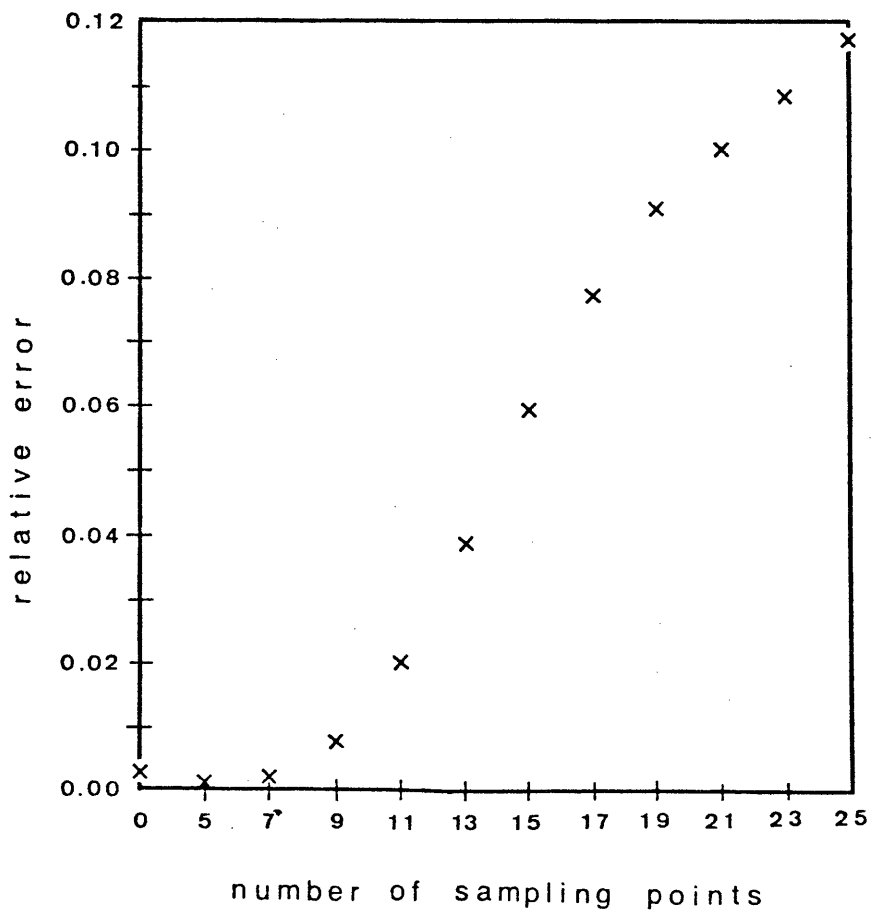


Figure 4

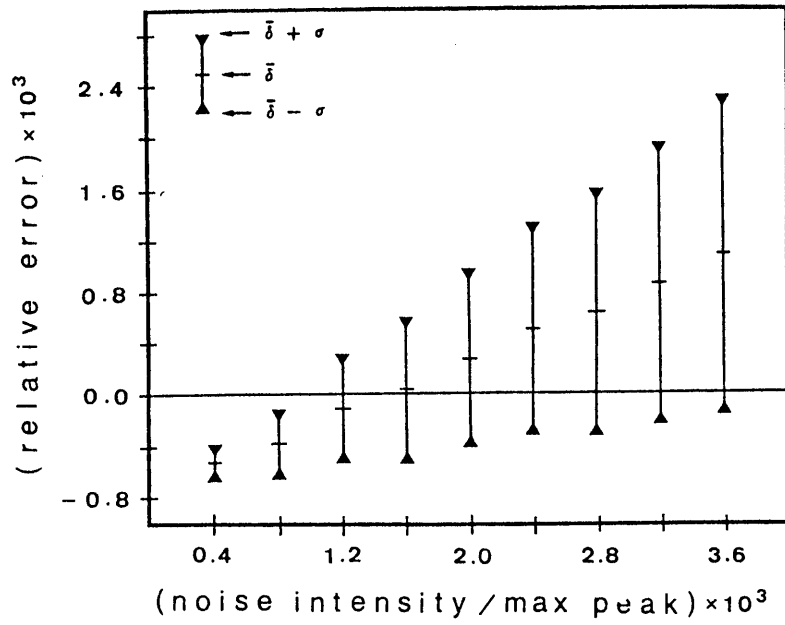


Figure 5

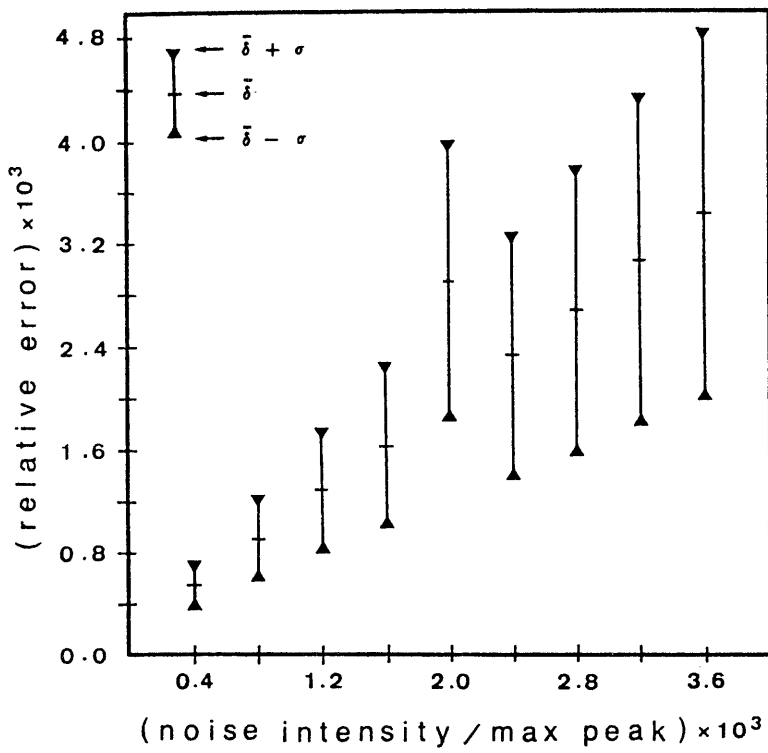


Figure 6

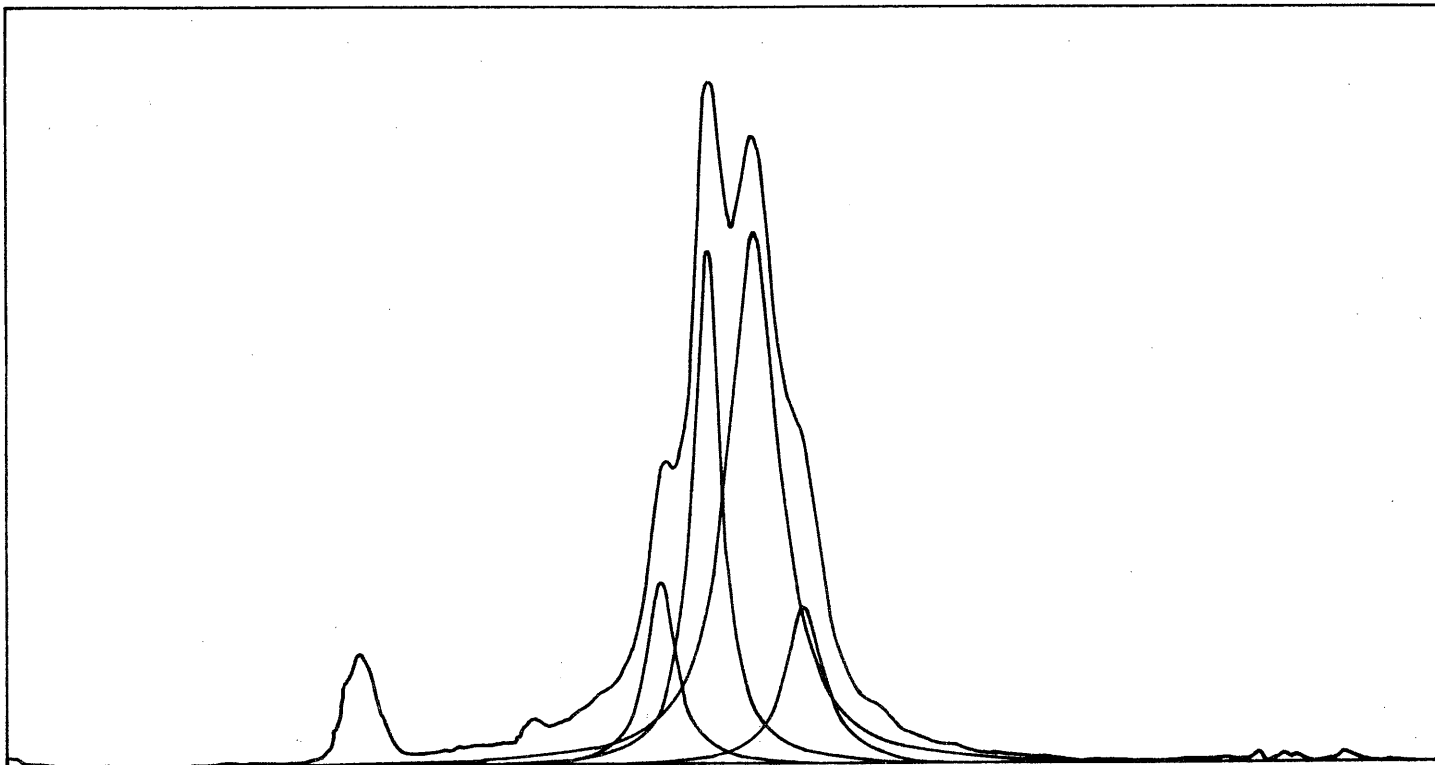


Figure 7

INSTITUTE OF ELECTRONICS AND INFORMATION SCIENCE
 UNIVERSITY OF TSUKUBA
 SAKURA-MURA, NIIHARI-GUN, IBARAKI JAPAN

REPORT DOCUMENTATION PAGE	REPORT NUMBER ISE-TR-81-21
TITLE Computer Compensation for NMR Quantitative Analysis of Trace Components	
AUTHOR(s) Takashi Nakayama [Central Research Laboratory, Kuraray Co. Ltd.] Yuzuru Fujiwara [Institute of Information Sciences, University of Tsukuba]	
REPORT DATE July 22, 1981	NUMBER OF PAGES 23
MAIN CATEGORY 3 Applications	CR CATEGORIES 3.19, 3.29
KEY WORDS Signal processing/Signal smmothing/Spectral simulation	
ABSTRACT A computer program has been written that determines trace components and separates overlapping components in multicomponent NMR spectra. This program uses the Lorentzian curve as a theoretical curve of NMR spectra. The coefficients of the Lorentzian are determined by the method of least squares. Systematic errors such as baseline/phase distortion are compensated and random errors are smoothed by taking moving averages, so that there processes contribute substantially to decreasing the accumulation time of spectral data. The accuracy of quantitative analysis of trace components has been improved by two significant figures. This program was applied to determining the abundance of ^{13}C and the saponification degree of PVA.	
SUPPLEMENTARY NOTES	