Curve Resolution of Highly-Overlapping Lorentzian Peaks
Based on the Autoregressive Model

Tetsuo IWATA and Jun KOSHOUBU

JASCO Technical Research Laboratories Corporation 2097-2, Ishikawa-cho, Hachioji-shi, Tokyo 192

(Received October 24, 1995)

Synopsis

We propose a method for separating Lorentzian peaks highly overlapped. The method is based upon the application of the autoregressive model to Fourier-transformed data of the overlapping peaks; which enables us to derive values of Lorentzian parameters such as individual positions, widths, and heights of the peaks at a time. If necessary, we can adopt those values as initial estimates in the conventional curve-fitting procedure. Results of analyses for computer-simulated and real spectral data are shown in order to demonstrate the effectiveness of the proposed method.

Keywords: Curve Resolution, Curve Fitting, Autoregressive Model, Spectral Data-Processing, Computer Applications

1. Introduction

A curve fitting (or curve resolving) method is now widely used both qualitatively and quantitatively to separate overlapping peaks in composite profiles\(^1\)-\(^4\). Usually, a simple analytic function, such as a Lorentzian and a Gaussian shape, is assumed to be the individual profiles. By use of an iterative procedure that minimizes the sum of the squares of the differences between the curve-fitted and the overlapping spectrum, appropriate initial estimates such as the individual positions, widths, and heights of the peaks are varied as parameters to arrive at a best fit. As the curve fitting method is a particular example of a nonlinear optimization problem, determination of appropriate initial estimates is very important. Inappropriate initial estimates often result in a long convergence time and a wrong solution caused by falling into a local minimum in the parameter space.

Among the initial estimates, the positions and widths can be derived from the second or the forth derivative of the overlapping peaks\(^5\). The derivative method, however, has a difficulty in applying to the overlapping peaks when the constituent peaks are situated within half width at half-maximum (HWHM) of each peak. Alternative approach to find out the initial estimates is to make use of a Fourier self-deconvolution (FSD) method\(^5,6\). The FSD method, however, also has a limitation in a practical use; the profiles and HWHM's of the individual peaks should be the same. A new approach for example linear prediction, like LOMEP (Line Shape Optimized Maximum Entropy Linear Prediction) developed by
Kauppinen et al.\textsuperscript{7,8)}, has the same difficulty as the FSD method. Furthermore, none of the methods mentioned above can give accurate initial estimates for the heights of the peaks.

Another approach is to make use of an autoregressive (AR) method\textsuperscript{9,10). In the AR model, a present value in time-series data is recursively represented by the linear combination of the finite number of past values and a noise. If we assume that the individual profiles are represented by a Lorentzian curve, Fourier-transformed data (interferogram data) of the overlapping peaks are given by the summation of exponentially-damped-sinusoid (oscillating decay) waveforms. Since a harmonic process can be generated with the AR model, the interferogram data can be fit to a part of the output of the AR model\textsuperscript{11):}

\begin{equation}
x_n = \sum_{k=1}^{M} B_k x_{n-k} + e_n,
\end{equation}

where $x_n$ is the interferogram data at discrete point $n$; $\{e_n\}$ is the Gaussian white noise process as the driving input sequence of the AR model; $M$ is the order of the model; and $B_k$ ($k=1, 2, \ldots, M$) are coefficients of the model. The coefficients are uniquely related with the parameters of the oscillating-decay waveforms and therefore with those of overlapping Lorentzian peaks. Since Eq. (1) is linear in unknown coefficients $B_k$, it can be solved by a linear least-squares fitting method whose computation time is shorter than that of the conventional nonlinear method. In addition, it might give more accurate estimates when the AR modeling is carried out correctly.

Application of the AR model to a multiexponential-decay-curve has already been reported by Sasaki et al., who have measured fluorescence lifetimes\textsuperscript{12). The method has also been used in NMR spectroscopy. An analysis of the oscillating-decay waveform had been presented by Van Blaricum et al.\textsuperscript{13) The purpose of the present paper is to apply the AR model to the interferogram data in order to estimate the values of the Lorentzian parameters from the highly overlapping peaks. In our method, unfortunately, the peak profile applicable is limited to a Lorentzian curve alone. However, all of the Lorentzian parameters can be computed simultaneously. We have found that accurate solutions can be retrieved from a completely overlapping pair of Lorentzian peaks that have slightly different HWHM's. For a spectrum of high signal-to-noise ratio (SNR), no additional procedure of the conventional curve-fitting is required. In the present paper, the principle of the proposed method is described. The effectiveness is verified by using simulated and experimentally obtained spectral data.

2. Theory

2.1 Application of the AR model

The basic idea of the proposed method is to apply the AR model to the interferogram data. In a Fourier domain, a Lorentzian peak is transformed to an exponentially-damped-sinusoid (oscillating decay) waveform. Hence, a pair of overlapping Lorentzian peaks whose individual positions, widths, and heights differ slightly is transformed to a synthetic waveform consisting of two oscillating decay waveforms whose initial amplitudes, decay constants, and oscillating frequencies differ slightly. The interferogram data in the Fourier domain, therefore, can be expressed with the AR model.

A Lorentzian peak (height, $h$; HWHM, $\sigma$; position, $v_0$) digitized at $N$ discrete points in Fig. 1(a) is transformed to an oscillating decay waveform (initial amplitude, $A$; decay constant, $\gamma$; oscillating frequency, $\omega$) presented in Fig. 1(b). Then, the relation between those parameters are given by the equations

\begin{equation}
\begin{aligned}
\sigma &= \frac{\gamma N}{2\pi} \\
v_0 &= \frac{\omega N}{2\pi} \\
h &= \frac{AN}{2\pi\omega} = \frac{AN^2}{4\pi^2v_0}.
\end{aligned}
\end{equation}

Therefore, if we obtain the values of $\gamma$, $\omega$, and $A$

\begin{subequations}
\begin{align}
Y(n\Delta) &= \text{Aexp}(-\gamma n\Delta)\cos(n\omega \Delta) \\
\sigma &= \frac{h}{(h^2-v_0^2)^{1/2}+\sigma^2}.
\end{align}
\end{subequations}

Fig. 1 A Fourier transform pair: (a) a Lorentzian profile whose height, position, and HWHM are $h$, $v_0$, and $\sigma$, respectively; (b) an oscillating-decay waveform whose decay constant, oscillating frequency, and initial amplitude are $\gamma$, $\omega$, and $A$, respectively.
by the AR-model fitting to the oscillating-decay waveform, we can derive the Lorentzian parameters of $\sigma$, $v_0$, and $h$ from Eq. (2). The mathematical details to obtain the values of $\gamma$, $\omega$, and $A$ from the AR method are described in Appendix. In the course of computation, no iteration is required. The order $M$ of the AR model is the first adjustable parameter in the proposed method. When the overlapping peaks consist of $m$ Lorentzian components without noise, $M=2m$ in principle. We, however, have empirically found that accurate estimates are obtained when $M$ is slightly larger than $2m$.

2.2 Noise reduction

In above discussion, a noise component was not considered at all. The random noise contained in the overlapping peaks degrades the quality of estimates of Lorentzian parameters. In order to reduce the noise component, we introduced an apodization function to the interferogram data. By means of the apodization, the accuracy in parameter estimation was enhanced by a factor of 10. This is because the AR model is applied only to a low frequency part of the interferogram data under high-SNR condition. Among various apodization functions, a rectangular function gave the best result because it brought about no distortion to the interferogram data. The cutoff frequency $P$ of the rectangular function is the second adjustable parameter in this method. For a low-SNR spectrum, the value of $P$ should be small. It should be determined to become within the range of $M<P\leq N$.

2.3 Procedure of curve resolution

Figure 2 shows a procedure of the new curve-resolution method.

(1) Overlapping peaks digitized at $N$ sampled points in Fig. 2(a) are symmetrically folded to make $2N$-point data as shown in Fig. 2(b).

(2) The $2N$-point data are Fourier transformed; the real part is a symmetric interferogram as shown in Fig. 2(c) whereas the imaginary part zero.

(3) As shown in Fig. 2(d), an $N$-point one-sided interferogram is extracted from the symmetric interferogram. For a noisy spectrum, the rectangular apodization function, width of which is $P(<N)$, is introduced to the interferogram.

(4) The AR model is applied to the one-sided interferogram. By using the AR coefficients, a polynomial equation is created. Solving the equation, we can derive all of the Lorentzian parameters at a time (see Appendix).

(5) If necessary, we can perform the conventional curve-fitting procedure with its initial estimates being the Lorentzian parameters obtained above.

3. Results and Discussion

3.1 Computer-simulated data

In order to demonstrate the effectiveness of the proposed method, it was applied to two kinds of computer-simulated data. First, a synthetic spectrum consisting of the same, ten Lorentzian peaks without a noise was analyzed. True and estimated values are summarized in Table I and are shown in Fig. 3. The number $N$ of total data points was 128. The order $M$ of the AR model and the cutoff frequency $P$ of the apodization function were 38.
Table I True and estimated values of Lorentzian parameters (see Fig. 3)

<table>
<thead>
<tr>
<th>peak number</th>
<th>True Values</th>
<th>Estimated Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>position</td>
<td>height</td>
</tr>
<tr>
<td>1</td>
<td>40.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>45.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>50.00</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>55.00</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>60.00</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>65.00</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>70.00</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>75.00</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>80.00</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>85.000</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Figure 4 shows the second results derived from synthetic peaks consisting of (a) two and (b) three Lorentzian components. For both peaks, Gaussian distributed random noise of 1% was added. True and estimated values are summarized in Table II. The number N of total data points was 128. In the case of (a), the order M of the AR model and the cutoff frequency P of the apodization function were 14 and 50, respectively. In the case of (b), M=16 and P=50. By introducing the apodization function with appropriate cutoff frequency, we were able to obtain reasonable estimates even when a noise component was presented.

3.2 Experimentally obtained data

A Raman spectrum of a mixture of cyclohexanol, cyclohexane, and toluene with 10:2:1 by volume was measured by a Raman spectrophotometer (NR-1800, JASCO Co.) with a spectral bandwidth of 1.96 cm\(^{-1}\). Figure 5(a) shows the Raman spectrum in the vicinity of 800 cm\(^{-1}\). Also, shown are three peaks separated by the proposed method and their composite profile. Here, M=14 and P=300 for N=351. Three peaks of 787.6 cm\(^{-1}\) (toluene), 803.0 cm\(^{-1}\) (cyclohexane), and 792.7 cm\(^{-1}\) (cyclohexanol) were restored to some extent. In order to obtain a best fit, we carried out the conventional curve-fitting procedure with its initial estimates being the Lorentzian parameters obtained in (a). A result presented in Fig. 5(b) would indicate that the proposed method provides us with appropriate initial estimates for the curve fitting. For this spectrum, the derivative method gave us no information on the smallest peak because of the severe overlapping of peaks. Without the proposed method, we had to find out all of the

and 128 (\(N\) ), respectively. This result indicates high potential in estimating the individual Lorentzian parameters. No other method, such as the derivative or FSD method, worked well because of the extensive overlap between the spectral profiles of the components. From our another simulation, we have found that the proposed method is applicable to a severely overlapping pairs: the positions of the two are the same, their HWHM’s differ by only one sampled point, and the intensity ratio is more than 10.
Iwata and Koshoubu: Curve Resolution Based on the Autoregressive Model

(a) An original profile from two Lorentzian peaks and 1% Gaussian noise, and two peaks separated by the proposed method together with their composite profile. The plot of the composite profile is superimposed on that of the original profile. Upper plot is a difference between the two profiles. (b) same as (a) but the number of Lorentzian peaks is three.

Table II  True and estimated values of Lorentzian parameters for (a) two and (b) three components (see Fig. 4)

<table>
<thead>
<tr>
<th>peak number</th>
<th>True Values</th>
<th>Estimated Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>position</td>
<td>height</td>
</tr>
<tr>
<td>(a) 1</td>
<td>65.000</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>70.000</td>
<td>0.700</td>
</tr>
<tr>
<td>(b) 1</td>
<td>65.000</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>70.000</td>
<td>0.700</td>
</tr>
<tr>
<td>3</td>
<td>82.000</td>
<td>0.800</td>
</tr>
</tbody>
</table>

initial estimates through trial and error, which often introduces significant errors and makes it difficult to carry out the FSD and the curve fitting.

A major restriction of the proposed method is that the peak profile applicable is limited to a Lorentzian curve alone. Optimization of the adjustable parameters of $M$ and $P$ should be carried out through trial and error to some extent. Nevertheless, we believe that it would be one of invaluable tools to analyze spectroscopic curves thanks to its high separation capability.

4. Conclusion

We have proposed a method for separating highly-overlapping Lorentzian peaks. By using the AR model, Lorentzian parameters such as the individual positions, widths, and heights of the peaks could be derived at a time. In order to cope with a noisy spectrum, we introduced a rectangular apodization function. The order $M$ of the AR model and the cutoff frequency $P$ of the apodization function were adjustable parameters of the
method. We have found that $M$ should be larger than $2m$ when the overlapping peaks consist of $m$ components. Taking into account of the SNR of the spectrum, we should choose the value of $P$ to become within the range of $M < P \leq N$, where $N$ is the number of total data points. The method was applied to computer-simulated and experimentally obtained data. The Lorentzian parameters obtained were also useful as initial estimates for the conventional curve-fitting procedure.

References


Appendix

An oscillating-decay waveform consisting of $m$ components can be represented as a function of $t$ in the form

$$x(t) = \sum_{i=1}^{m} A_i \exp\left(\frac{-\gamma_i t}{\cos(\omega_i t)}\right), \quad \text{(A1)}$$

where $\gamma_i$, $\omega_i$, and $A_i$ are the decay constant, the oscillating frequency, and the initial amplitude of the $i$-th component, respectively. Equation (A1) can be rewritten in the discrete form

$$x(nA_t) = \sum_{i=1}^{m} A_i \exp\left(\frac{-\gamma_i nA_t}{\cos(\omega_i nA_t)}\right), \quad \text{(A2)}$$

where $A_t$ is the sampling interval.

Our purpose is to express the waveform given by Eq. (A2) with an autoregressive (AR) model. First, we define values of $Q_i$ and $Q_i^*$ by

$$Q_i = \exp\left(\frac{-\gamma_i + j\omega_i A_t}{\cos(\omega_i A_t)}\right),$$

$$Q_i^* = \exp\left(\frac{-\gamma_i - j\omega_i A_t}{\cos(\omega_i A_t)}\right) \quad i = 1, 2, \ldots, m,$$

$$\text{(A3)}$$
where \( j = \sqrt{-1} \) and the symbol * stands for the complex conjugate. Then, Eq. (A2) can be rewritten as
\[
x(n \Delta t) = \sum_{i=1}^{m} \frac{A_i}{2} (Q_i^* + Q_i^{**}). \tag{A4}
\]

Although \( Q_i^* \) is a complex conjugate of \( Q_i \), we will discriminate between them. Replacing \( Q_i \) and \( Q_i^* \) by \( Q_{2i-1} \) and \( Q_{2i} \), respectively, we obtain
\[
x(n \Delta t) = \sum_{i=1}^{2m} \frac{A_i}{2} Q_i^*, \tag{A5}
\]
where \( A_{2i-1} = A_{2i} \).

Using z-transform, Eq. (A5) becomes
\[
X(z) = \sum_{k=0}^{2m} \left( \frac{2m}{2} \sum_{i=1}^{2m} A_i \prod_{j=1}^{2m} (1 - Q_j z^{-1}) \right) z^{-k}
\]
\[
= \frac{2m}{2} \sum_{i=1}^{2m} A_i \prod_{j=1}^{2m} (1 - Q_j z^{-1})
\]
\[
= \prod_{i=1}^{2m} (1 - Q_i z^{-1}), \tag{A6}
\]
where \( X(z) \) is a z-transform of \( x(n \Delta t) \).

Now, if we define coefficients \( B_k \) and \( C_k \) as
\[
\prod_{i=1}^{2m} (1 - Q_i z^{-1}) = 1 - \sum_{k=1}^{2m} B_k z^{-k}, \tag{A7}
\]
\[
\sum_{i=1}^{2m} \frac{A_i}{2} \prod_{j=1}^{2m} (1 - Q_j z^{-1}) = \sum_{k=1}^{2m} C_k z^{-k-1}, \tag{A8}
\]
Eq. (A6) can be expanded as follows:
\[
X(z) = \sum_{k=1}^{2m} B_k z^{-k}X(z) = \sum_{k=1}^{2m} C_k z^{-k-1}. \tag{A9}
\]

The inverse z-transform of Eq. (A9) is
\[
x(n \Delta t) = \sum_{k=1}^{2m} B_k x[(n-k) \Delta t]
+ \sum_{k=1}^{2m} C_k \delta[(n-(k-1)) \Delta t], \tag{A10}
\]
where \( \delta \) is the Kronecker's delta function. By putting \( \Delta t = 1 \), Eq. (A10) is rewritten as
\[
x_n = \sum_{k=1}^{2m} B_k x_{n-k} + \sum_{k=1}^{2m} C_k \delta_{n-(k-1)}, \tag{A11}
\]
where \( x_n \) and \( \delta_n \) are aberrations of \( x(n \Delta t) \) and \( \delta(n \Delta t) \), respectively. Now, we have found that the oscillating-decay waveform of Eq. (A1) can be expressed by the \( 2m (=M) \)-th order AR model. From Eq. (A11), the unknown coefficients of \( B_k \) and \( C_k \) \((k=1, 2, \ldots, 2m)\) can be determined from the equations
\[
\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_{N-1} \\
x_{N} \\
x_{N-2} \\
\vdots \\
x_{N-2m-1}
\end{bmatrix}
= \begin{bmatrix}
B_1 \\
B_2 \\
\vdots \\
B_{2m}
\end{bmatrix}
\]
\[
\begin{bmatrix}
C_1 \\
C_2 \\
\vdots \\
C_{2m}
\end{bmatrix}
= \begin{bmatrix}
x_1 & 0 & \cdots & 0 \\
x_2 & x_1 & \cdots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
x_{2m-1} & x_{2m-2} & \cdots & x_0
\end{bmatrix}
\times \begin{bmatrix}
B_1 \\
B_2 \\
\vdots \\
B_{2m}
\end{bmatrix}. \tag{A12}
\]

In Eq. (A12), the number of sampling points \( N \) is usually much larger than \( 2m \). Therefore, the conventional linear least-squares method is used to calculate \( B_k \)'s. Then, \( C_k \)'s are given by Eq. (A13).

Our final goal is to calculate values of \( Q_i (i=1, 2, \ldots, 2m) \) and then to estimate the values of \( \gamma_i, \omega_i, \) and \( A_i \). The values of \( Q_i \) are given by solving the \( 2m \)-th order polynomial equation
\[
Q_i^{2m} - B_1 Q_i^{2m-1} - B_2 Q_i^{2m-2} - \cdots - B_{2m-1} Q_i - B_{2m} = 0. \tag{A14}
\]
This equation is derived from Eq. (A7) by substituting \( z = Q_i \). Substituting the value of \( Q_i \) into Eq. (A3), we can obtain \( \omega_i \) and \( \gamma_i \) from the equations
\[
\omega_i = \tan^{-1} \left( \frac{\text{Im}[Q_i]}{\text{Re}[Q_i]} \right), \tag{A15}
\]
\[
\gamma_i = \ln \left( \frac{\text{Re}[Q_i]}{\cos \left\{ \tan^{-1} \left( \frac{\text{Im}[Q_i]}{\text{Re}[Q_i]} \right) \right\}} \right). \tag{A16}
\]
where \( \text{Re}[Q_i] \) and \( \text{Im}[Q_i] \) represents for the real and imaginary part of \( Q_i \), respectively. Substituting \( z = Q_i \) into Eq. (A8), we can derive the value of \( A_i \) from the equation
\[
A_i = \frac{2 \sum_{k=1}^{2m} C_k Q_i^{-(k-1)}}{\prod_{j=1}^{2m} (1 - Q_j Q_i^{-1})}. \tag{A17}
\]
In the course of calculations, we obtain the same values of \( \omega_i, \gamma_i, \) and \( A_i \) twice because we distinguished \( Q_i \) from \( Q_i^* \) in Eq. (A5). Finally, the Lorentzian parameters of the \( i \)-th component can be determined uniquely from Eq. (2).