2580

Uses of Digital Computers in Theoretical Analytical Chemistry. I. Separation of a Composite Infrared Band into Two Component Peaks

C. ROLAND EDDY and JOHN S. SHOWELL,
Eastern Regional Research Laboratory,[1] Philadelphia, Pennsylvania

Abstract

A procedure has been developed for representing an infrared absorption peak by a general mathematical form, not restricted to Gaussian or Lorentzian shape. A computer program using this procedure will be described for obtaining the individual components from an infrared band consisting of two overlapping peaks. The method is also applicable to other overlapping peaked curves, such as are obtained in chromatography, countercurrent distribution, x-ray diffraction, and other branches of spectroscopy.

Introduction

THERE IS MORE INFORMATION in the typical infrared absorption band than can be expressed by merely its center frequency, its peak absorbance, and its width at half height. The increasing use of higher-resolution instruments on liquid and solid samples reveals that many typical infrared peaks contain underlying substructure consisting of overlapping components. The substructure of a composite band can often reveal subtle molecular effects if the information in the band can be sorted out.

Previous computational methods for resolving overlapping peaks have assumed special curve shapes, usually Gaussian or Lorentzian (1), or have required special-purpose instruments (2). We have developed a procedure for separating two overlapping peaks with a general shape function, suitable for a small general-purpose computer. The computer program is designed to operate on data from a limited region of spectrum containing primarily two overlapping peaks. In favorable cases, the computation can be extended to accommodate a third peak if it is small compared with the other two.

Shape of Component Peak

Several papers have attempted to describe an infrared absorption peak by a mathematical function. An excellent review of the factors involved has been written by Seshadri and Jones (3). Theoretical considerations have led many to adopt Gaussian or Lorentzian shapes, but neither of these adequately fitted the curves we obtained in some of our work (4). After preliminary computational experiments with peaks that appeared single to the eye, we have chosen the following mathematical representation for a general infrared peak:

(a) The function is to be symmetrical about the band center, i.e., a function of x^2, where x is the wavenumber displacement from the band center.

(b) The reciprocal of the absorbance is a polynomial in z $(=x^2)$.

(c) The polynomial must have no maxima, minima, nor inflection points when expressed as a function of z.

[1] E. Utiliz. Res. Dev. Div., ARS, USDA.

643–646 USES OF DIGITAL COMPUTERS IN THEORETICAL ANALYTICAL CHEMISTRY. I. SEPARATION OF A COMPOSITE INFRARED BAND INTO TWO COMPONENT PEAKS, by C. Roland Eddy and John S. Showell

647–648 USES OF DIGITAL COMPUTERS IN THEORETICAL ANALYTICAL CHEMISTRY. II. SIMULATION OF THE PASSAGE OF A COMPOUND THROUGH A COUNTERCURRENT DISTRIBUTION APPARATUS, by C. Roland Eddy and John S. Showell

649–651 VARIABLES IN THE TECHNICON-N-24 METHOD FOR THE DETERMINATION OF SERUM CHOLESTEROL, by Dorothy Roland, Gerald R. Cooper and Robert F. Witter

652–656 A SEMIAUTOMATED PROCEDURE FOR THE DETERMINATION OF SERUM CHOLESTEROL USING THE ABELL-KENDALL METHOD, by Eloise Eavenson, O. T. Grier, J. G. Cisson and R. F. Witter.

These requirements are met by the function

$$A = \frac{1}{\sum_{i=0}^{n} b_i z^i}$$

where $z = (\nu - \nu_0)^2$
ν = wavenumber
ν_0 = wavenumber at band center
A = absorbance of one component

The Lorentzian function is a special case of this when the polynomial is of first degree. The Gaussian function is a special case when the polynomial is an exponential function of z.

Basic Iterative Procedure

The computer program is an iterative procedure in which a theoretical approximation to one of the two peaks is subtracted from the experimental absorbance curve to get the input data for the next curve-fitting step. In the curve-fitting step, the band center is found by fitting a parabola to a narrow range of points near the maximum. A polynomial is then fitted to the reciprocal of absorbance as a function of square of wavenumber displacement from the band center. $R(z) = 1/A(\nu)$ The polynomial fitting is a least-squares procedure subject to the constraint that the polynomial coefficients must be interrelated in such a way that the polynomial has no maxima, minima, nor inflection points within the wavenumber range of the experimental data. A computer program for this type of smooth curve fitting will be published elsewhere. The smoothness constraint is essential to the iteration method of subdividing overlapping peaks. Without the constraint, the iterative process rapidly goes out of control with certain types of infrared data.

The result of each iteration is a theoretical function representing one of the component peaks. This is used as input information in the next iteration for determining the next approximation to the other component peak.

In many infrared spectrophotometers, the noise is roughly independent of transmittance, which means that the uncertainty on the absorbance scale is not independent of the absorbance. Taking reciprocal of the absorbance introduces still further variation in the uncertainty. Theory shows (5) that the change of variable $R = 1/A = -1/\log T$ requires a weighting factor proportional to T^2A^4. This weighting function is used in the computer program.

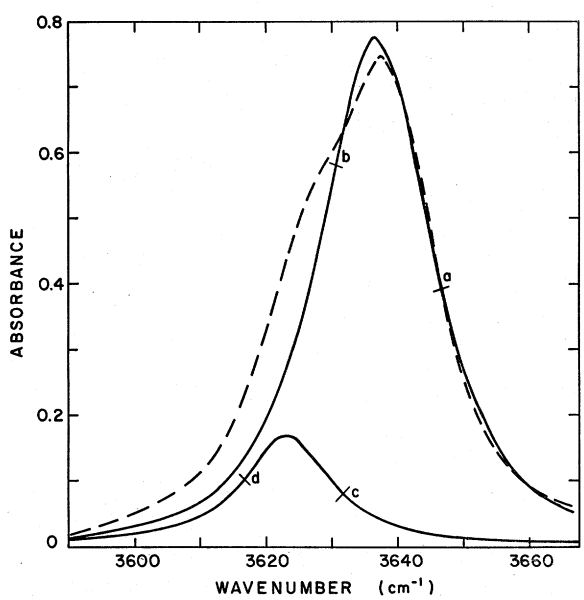

FIG. 1. First iterations in subdividing infrared curve of 1-octadecanol. Dashed curve: Experimental. Solid curves: First approximation to component peaks. See text for details.

Starting the Iterations

To begin the iterations, a portion of the composite curve is selected that, by visual inspection, seems to be dominated by the stronger component with a minimum of interference from the other component. The first iteration uses only this limited region for its curve fitting. The second iteration also uses a limited region where the weaker component is as strong as possible relative to the stronger component, although the contribution by the strong component need not be small since its first theoretical approximation is now available for subtraction. Both limited regions are under human control, are selected by visual inspection, and are entered with the input data cards. After the first iterations, the computer gradually enlarges the limited regions until the entire set of experimental points is finally used in each iteration.

Figure 1 shows an example of the starting procedure. The dashed line is the infrared curve for 1-octadecanol in dilute solution in carbon tetrachloride in the hydroxyl stretching region (4). By visual inspection, the wavenumber range between *a* and *b* is chosen for parabolic fitting to locate the band center of the strong peak. The region from the high wavenumber end of the curve to point *b* is chosen as the region dominated by the strong component and is used for reciprocal curve fitting during the first iteration. The upper solid curve shows the theoretical function produced by the first iteration. This function is subtracted from the dashed curve. On the difference curve, the region *c* to *d* is chosen for locating the band center, and the region *c* to the low wavenumber end chosen for fitting the weaker peak. The lower solid curve shows the theoretical function produced by the second iteration. The sum of these two theoretical curves is only a rough first approximation to the experimental curve. As the iterations continue, the fit improves. The procedure converges in 80 iterations to the two theoretical functions shown in Figure 2. The standard error of estimate for this case is then 0.63 on the percent-transmittance scale.

Stopping the Iterations

At the end of each iteration, the computer synthesizes the complete curve from the two theoretical components and calculates the sum of squares of deviations from the experimental data, on the scale of transmittance. The sum of squares gradually decreases as the fit of the theoretical function improves. When the sum of squares reaches a minimum and begins to rise again, the computer saves the best pair of theoretical peaks, performs 20 more iterations, then stops if the sum of squares has not fallen below its best again.

As the computation ends, it prints out a table showing the theoretical component peaks, the calculated composite, and a comparison with the original experimental points.

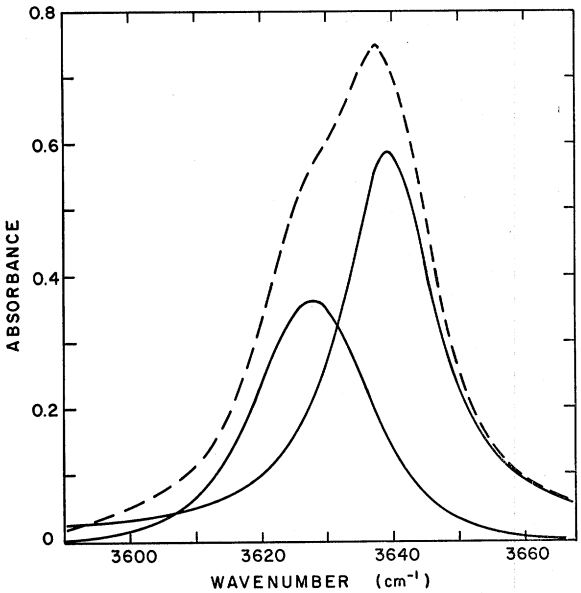

FIG. 2. Final result of subdivision of infrared curve of 1-octadecanol. Dashed curve: Experimental. Solid curves: Component peaks.

Bands with Three Components

Visual inspection of either the original curve or the final printout from the program sometimes suggests the presence of a third peak along with the two main components. If this peak is not strong, it can be included in the resolution by the following procedure.

The program contains provision for omitting a selected portion of the curve. The data are first computed by the two-peak program with the region of the third peak excluded. The sum of the two theoretical peaks is then subtracted from the absorbance of the complete experimental curve to get the data for the third peak. The computer then fits a theoretical function to the third peak alone, followed by a comparison of the experimental curve with a composite of all three theoretical curves. Visual inspection determines whether this process requires further iteration.

In principle, a fourth peak might be similarly obtained, provided that it is weak and does not overlap the third peak.

Work is in progress on automating the third peak, but so far the intelligent interaction between man and machine has been more successful than complete automation of three peaks.

Figure 3 shows an example of a three-component curve fitted by this procedure.

Representation of Tails

In principle, no fitted curve should be considered valid beyond the range of the experimental points. However, constraints in defining a mathematical function do establish a trend and set limits on the behavior of a curve outside the experimental range. Polynomial curve fitting gives numerical values for a limited number of coefficients. For our infrared peaks, fifth degree polynomials have been adequate. The routine for eliminating maxima, minima, and inflection points only applies up to a specified distance from the band center. Beyond that point the empirical polynomial may violate the constraints. It is assumed that the true polynomial describing the infrared peak contains higher degrees that are negligible within the experimental wavenumber range but become appreciable as z becomes large, in such a way that the constraints would always be obeyed. However, the coefficients of these higher degrees are not determinable from the experimental data. Hence the empirical polynomial cannot be used beyond the experimental wavenumber range. A different function must be used for representing the tails of the curve.

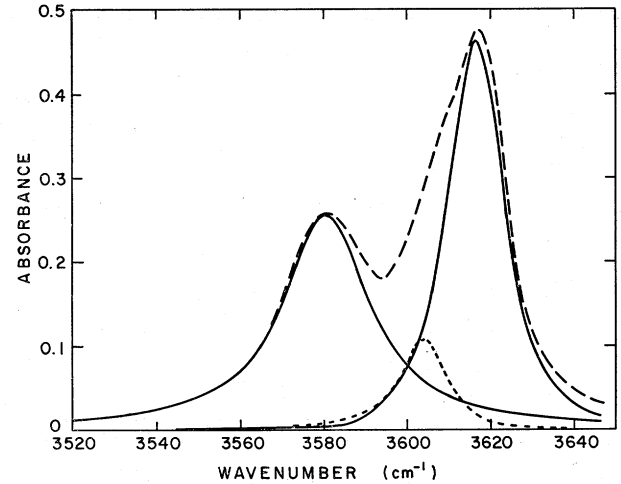

FIG. 3. Subdivision of infrared curve of 6-phenyl-3-ethylhex-5-en-3-ol. Upper, dashed curve: Experimental. Lower three curves: Theoretical component peaks.

The constraints imply that the first and second derivative of the polynomial must never change sign. The first derivative must always be positive in order for the absorbance to be peaked. The second derivative could have either sign. Figure 4 shows the two types of trend. If the second derivative is positive in the experimental range, the lowest value it can take outside the range is zero. If it is negative inside the range, the highest it can be outside the range is zero. Zero is the limiting value in either case, corresponding to the Lorentzian function. Thus a straight line can be drawn tangent to the empirical polynomial at the end of its valid range, to limit

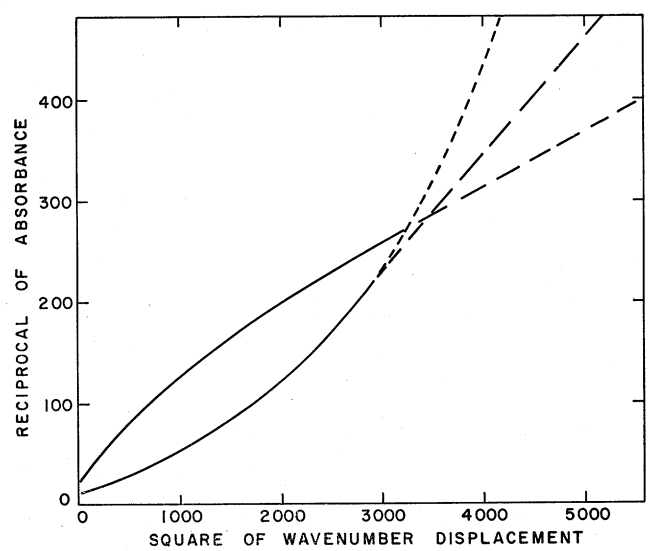

FIG. 4. Representative trends of reciprocal of absorbance as a function of square of wavelength displacement. Solid curve: Polynomial fitted to experimental data. Long dashes: Straight line drawn tangent to solid curve at farthest experimental point. Short dashes: Exponential function ($C = e^{kz}$) tangent to solid curve.

the possible extrapolated curves, as shown by the dashed lines in Figure 4. Outside the valid range, the higher-degree polynomial must always lie above the Lorentzian straight line if the second derivative is positive in the experimental range, or it must lie below the Lorentzian line if the second derivative is negative in the experimental range. Translated back to absorbance, the tails of the infrared curve in the former case must always have lower absorbance than the tangent Lorentzian curve; the tails in the latter case must always have higher absorbance than the tangent Lorentzian curve.

It is well known that the Lorentzian function exaggerates the tail area of infrared peaks of liquid and solid systems. Thus, if the second derivative of the reciprocal polynomial is negative in the experimental range, the curve would be expected to be neither higher nor lower than the Lorentzian line outside the experimental range. For those infrared peaks whose polynomials have negative second derivatives, we have therefore chosen to represent the curve outside the experimental range by a Lorentzian function whose reciprocal is tangent to the empirical polynomial at the highest experimental value of z, the square of the wavenumber displacement from the band center. Only a small number of our component peaks (4) were in this category.

The majority of our components had reciprocal polynomials with positive second derivatives, thus allowing some latitude in choice of function outside the experimental range. Of the possibilities, we have chosen the exponential function, since it contains few parameters and is certain to have positive first and second derivatives. The parameters are chosen to make the exponential function tangent to the empirical polynomial at the maximum z used in setting the constraints on the empirical polynomial, represented by the dotted curve in Figure 4. This corresponds to a tail of Gaussian shape.

The tail parts of the theoretical curves are considered less valid than the central parts of the curves. However, they are useful in estimating the total area under an absorbance curve provided that the tail area is a small fraction of the total. If the tail area is found to be large, it suggests reexamining the curve, to see for example whether a baseline correction is indicated.

Other Applications

Although this program was designed originally for finding subtle molecular effects in the infrared bands of single compounds, it can also be used for analytical applications of mixtures of compounds. Furthermore, it could be applied to other types of curves having overlapping peaks, such as ultraviolet spectra, chromatographic curves, mass spectra, countercurrent distribution, ultracentrifugation, electrophoresis, or x-ray diffraction.

The special curve-shape function used in this program is not essential to the iteration procedure. Other shapes, including unsymmetrical ones, could equally well be used for the component peaks. All that is required is a specified mathematical function suitable for automatic programming, with or without constraints or conditions.

ACKNOWLEDGMENT

K. A. Tabler, Linda E. Preiss, J. W. Tukey, J. N. Boyd, and Elizabeth A. Eddy provided assistance and advice in various phases of this project.

REFERENCES

1. Stone, H., J. Opt. Soc. Am. *52*, 998 (1962).

2. Noble, F. W., J. E. Hayes, Jr., and M. Eden, Proc. I.R.E. *47*, 1952 (1959); Anonymous, Chem. Eng. News *43*, No. 46, 50 (Nov. 15, 1965).

3. Seshadri, K. S., and R. N. Jones, Spectrochim. Acta *19*, 1013 (1963).

4. Eddy, C. R., J. S. Showell and T. E. Zell, JAOCS *45*, 92 (1963).

5. Deming, W. E., "Statistical Adjustment of Data," John Wiley & Sons, 1943, p 21–22, 37–45.