Method for Resolving and Measuring Overlapping Chromatographic Peaks by Use of an On-Line Computer with Limited Storage Capacity

Charles D. Scott, Dennis D. Chilcote, and W. Wilson Pitt, Jr.

A simple routine for distinguishing and measuring the positions and areas of overlapping peaks comprising a complex chromatographic "envelope" has been developed for a small on-line computer. The degree of peak overlap is such that no minima and only one maximum occur within the bounds of the envelope. This routine determines corrected initial estimates of the Gaussian parameters of the peaks. The routine is currently being used, in conjunction with a program devised to analyze data generated by a chromatographic system, specifically to evaluate data from high-resolution analyzers.

Additional Keyphrases  peak stripping  •  physiological fluid analysis  •  column chromatography  •  mathematical peak envelope analysis  •  Taylor series expansion

COLUMN CHROMATOGRAPHIC methods yield much information about many constituents of a sample mixture. This is particularly true of analysis of physiological fluids by high-resolution chromatography; more than 100 constituents may be separated and represented as peaks on a single chromatogram (1-3). The chromatogram is usually a tracing that shows the concentrations of the eluted constituents (or an equivalent property, such as the absorbance of light) as a function of either elution time or elution volume (e.g., Figure 1).

Each separated constituent of the sample mixture can be quantified by determining the area of its corresponding chromatographic peak. When adjacent peaks are discrete, they can be measured simply by numerically integrating the curve or by some similar method; however, when two or more peaks overlap, each peak must be "stripped out" of the combined peak "envelope" before it can be measured (Figure 1).

Manual evaluation of complex chromatograms is tedious and requires training. Alternatively, a small, on-line, high-speed digital computer can be used to evaluate the data more rapidly and to resolve the interfering peaks more accurately.

Several methods have been suggested for separating interfering peaks1; however, one of the most accurate methods to be used on a large off-line computer incorporates nonlinear least-squares techniques. This method is effective if the shape of the chromatographic peak can be approximated by a suitable mathematical model (4). Although such techniques are relatively easy to use on a large off-line computer having an extensive storage capacity in the fast memory, they must be simplified before they can be used on a small on-line computer that evaluates the data from several channels of chromatographic systems.

We have modified the nonlinear least-squares stripping technique, thereby simplifying the mathematical approach to the problem. This technique has been developed and evaluated for an on-line Model PDP-8/I (Digital Equipment Corp., Maynard, Mass.) computer, which is being used to acquire and evaluate the data from several channels of high-resolution chromatographic systems.

The Problem

Consider the specific type of chromatographic system that generates data in the form of a series of absorbance values as a function of time (Figure

---

1). When a small, on-line computer is used to evaluate chromatographic data, each individual data point or small “block” of data must be evaluated as it is generated since the data-storage capacity of the computer is limited. This is particularly important if the small computer is to be used on a time-sharing basis for more than one system.

The well-defined single peak is relatively simple to analyze. Such a peak rises from the baseline, attains a maximum value, and then decreases more or less symmetrically until it reaches the baseline again. Each chromatographic peak has two inflection points that may also aid in its identification.

Interfering peaks (Figure 2) represent a more difficult problem for analysis: first, the fact that the envelope contains two or more peaks must be established; next, some method must be used to isolate or “strip” each of the individual peaks from the envelope. After the peaks have been stripped out, their positions and areas must be determined.

To analyze interfering peaks, the computer must have a data-storage capacity that will accommodate the data for the entire envelope. Thus, for the small on-line computer, the computer technique used to strip out the interfering peaks must be one requiring a minimum of stored data.

Solution

For computer application, we will assume that a peak has begun when the slope of the baseline becomes more positive than some predetermined value; thus, small variations in the baseline will be overlooked (Figure 2). The slope of the peak can be found by numerically calculating the first derivative of absorbance with respect to elution time (5). The inflection points can be defined as the points at which the second derivative changes its sign. The second derivative can also be calculated by numerical methods (5).

The number of pairs of inflection points is assumed to be equivalent to the number of peaks
The elution position of each peak in the envelope (as means of identifying the peak) is the position of the maximum absorbance value of that peak.

The envelope is analyzed by fitting the curve with the number of individual peaks corresponding to the number of pairs of inflection points detected within the envelope. In this procedure, individual peaks must first be represented by a mathematical model. Many can be adequately described as Gaussian or modified-Gaussian curves. The initial estimates of the parameters of the assumed Gaussian peaks comprising the envelope (e.g., peak position, mean value, and standard deviation) are obtained from the envelope. Corrections to these parameters are calculated, and the area and position of the peaks in the envelope are then computed. A brief mathematical description of the method follows.

If the peaks in the chromatogram are assumed to be Gaussian in shape, then the envelope can be described by

\[
A = \sum_{j=1}^{n} \bar{A}_j \exp \left[ -\left( T - \bar{T}_j \right)^2 / 2\sigma_j^2 \right] \tag{1}
\]

where \( A \) is the total absorbance value at \( T \); \( \bar{A}_j \), the maximum absorbance of individual peak \( j \); \( T \), the elution time; \( \bar{T}_j \), the value of \( T \) at \( \bar{A}_j \); \( \sigma_j \), the standard deviation of the Gaussian peak \( j \); and \( n \), the number of interfering peaks.

Equation 1 can be rewritten as

\[
A = A(T; \beta_1, \beta_2, \ldots, \beta_k) \tag{2}
\]

where \( \beta \) denotes the parameters \( \bar{A}_j, \bar{T}_j, \) and \( \sigma_j \).

The experimental data points are denoted as

\[
(\bar{A}_i, \bar{T}_i); \; i = 1, \ldots, k \tag{3}
\]

where \( k \) is the number of data points needed to analyze each envelope and equals the number of parameters to be corrected.

Let \( A_i^* \) denote the predicted value (from the mathematical model) obtained for \( A \) when, for the \( i \)th data point, the value for the independent variable is substituted into Equation 2. The actual estimates of \( \beta_1, \beta_2, \ldots, \beta_k \) will be denoted by \( b_1, b_2, \ldots, b_k \). Thus,

\[
A_i^* = A(\bar{T}_i; \beta_1, \beta_2, \ldots, \beta_k) \tag{4}
\]

For the conventional least-squares analysis of the problem, the unknown parameters are varied until

\[
\phi = \sum_{i=1}^{k} [\hat{A}_i - A_i^*]^2 \tag{5}
\]

is a minimum (\( \theta \)). Since the function \( \hat{A} \) in Equation 2 is nonlinear in the parameters, Equation 4 is expanded in a Taylor series about the unknown parameters. The Taylor series is truncated after the linear terms (see Equation 8) and substituted into Equation 5. Corrections to the initial choices of the parameters are found by setting the partial derivative of \( \phi \) with respect to the corrections to the parameters equal to zero. New estimates for the parameters are found, and the process is repeated until the corrections become negligible.

Since the initial estimates of the parameters must be accurate to ensure convergence, more sophisticated approaches have been devised \( (7, 8) \). However, for the purpose of envelope analysis by an on-line computer with limited storage ability, a simple, but efficient, approach that uses a minimal amount of fast memory is mandatory.

The simplest approach is to dispense with the iteration procedure in the least-squares analysis and to assume that the initial estimate of the parameters are sufficiently accurate. However, determination of peak areas by this method gave unacceptably large errors.

The next approach is to use the truncated Taylor series expansion of Equation 4 to obtain corrections to the initial estimates of the parameters. This method produced satisfactory results, and is currently being used in our laboratory. For this simplified analysis, the assumption is made that

\[
\hat{A}_i = A_{i*} \tag{6}
\]

where

\[
A_{i*} = A_i^* + \Delta A_i^* \tag{7}
\]

The quantity \( \Delta A_i^* \), a correction to \( A_i^* \), is the collection of linear terms from a Taylor series expansion of \( A \). Therefore,

\[
\hat{A}_i = A_i^* + \sum_{j=1}^{k} \Delta \beta_j \frac{\partial A_i}{\partial \beta_j}; \; i = 1, \ldots, k \tag{8}
\]

where \( A_i^* \) and the partial derivatives \( \partial A / \partial \beta_j \) are evaluated at \( \bar{T}_i \) and \( b_j \), the initial estimates for \( \beta_j \). The quantity \( \Delta \beta_j \) is the correction to the initial estimate of \( \beta_j \).

Satisfactory results were obtained when only the initial estimates of \( \bar{A}_j \) and \( \bar{T}_j \) were corrected. Thus, in this discussion, \( k = 2n \); and the values of \( A \) and \( T \) at each inflection point in the envelope represent a convenient choice of experimental points.
The initial choices of the parameters $\bar{A}_j$ and $\bar{T}_j$ are

$$\bar{A}_j^0 = (A_{2j} + A_{2j-1})/2, \text{ where } j = 1, \ldots, n$$ (9)

and

$$\bar{T}_j^0 = (T_{2j} + T_{2j-1})/2, \text{ where } j = 1, \ldots, n$$

A careful examination of the chromatogram in Figure 1 leads to another simplifying assumption. Since near-neighbor peaks have almost identical standard deviations, it can be reasonably assumed that each peak in a given envelope has the same $\sigma$. Therefore, the choice of $\sigma$ is

$$\sigma = (T_{2M} - T_{2M-1})/2$$ (10)

where $2M - 1$ and $2M$ are the data points (i.e., the inflection points) on each side of the envelope maximum.

The corrections to $\bar{A}_j^0$ and $\bar{T}_j^0$ are found by solving the set of linear, simultaneous, algebraic equations in Equation 8 by a process of elimination, an efficient method for a small on-line computer (9). Thus, the position of peak $j$ in the envelope is $T_j = T_j^0 + \Delta T_j$; and the area, $S$, of the peak is $S = 2.5066 (A_j^0 + \Delta A_j) \sigma$.

**On-Line Computer**

Several different arrangements could be used for the on-line analyses of data from two or more chromatographs by a small computer. The computer we used (PDP-8/I) has an 8K-12 bit core memory and a 32K-12 bit disk storage. It is also equipped with an ASR 33 Teletype for input-output. A computer program has been written
for acquiring and analyzing the data from one chromatographic channel. A flow diagram of the computer program is shown in Figure 3.

The analysis begins by sequentially searching for a positive first derivative that is greater than some predetermined value (Point 3, Figure 3). When such a derivative is found, the position and the peak index are recorded as the beginning of the peak or envelope. Next, the second derivative at each data point is sequentially examined until its sign changes. At this point, an inflection is found, and its position and absorbance are recorded (Point 5, Figure 3). An examination of the second derivative is continued; another change of sign indicates the presence of a second inflection point and a chromatographic peak (Points 6 and 7, Figure 3). The computer continues to examine the data for additional inflection points (back to Point 4, Figure 3). During the analysis, each new absorbance value is compared to the preceding value until a maximum value is found. The occurrence of a maximum indicates that the peak or envelope is descending toward the baseline. The program begins examining the data for a minimum. Once the minimum is found, the position and area of the peak or envelope are determined.

The area of each individual peak is determined by integration by use of the "trapezoidal rule" (10). The position of the peak is the value, shown on the time axis, corresponding to the maximum absorbance value. If interfering peaks are present, the spectral stripping routine (Figure 4) is used to strip out the individual peaks and determine their areas and positions. Peak positions and areas are recorded, and the evaluation of the chromatogram continues until all of the data have been examined.

Operation with small blocks of data and rejection of each block immediately after evaluation are necessary if optimum use is to be made of the limited memory and computational capability of the small computer, especially if several chromatographic analyses are to be evaluated simultaneously.

The described computer program discards the raw data immediately after evaluation. Only five data points are needed in the fast memory at any given time. However, a data buffer of about 100 points is needed in the auxiliary disk memory in order to use the stripping routine. We believe that it will be possible to use this type of system for at least two automated high-resolution an-

**Fig. 4. Flow diagram for the spectral stripping routine**
alyzers (similar to those developed at ORNL) that are currently used to identify and quantify the molecular constituents of body fluids (1–3).

Results

The computer program just described can analyze complex envelopes containing as many as three Gaussian peaks, with an error of less than 5% (based on peak areas).

Table 1 shows the computer analysis of four typical envelopes. Five hundred data points were calculated for each envelope. The program analyzed the synthetic data and printed out the position of the maximum and the associated inflection points. It also supplied the position of the peaks and the areas, which were found by the stripping routine when interfering peaks were present.

The on-line computer is now being used in our laboratory to analyze one channel of a high-resolution chromatograph.

This research was sponsored by the NIGMS, NIH.

References