

CHROM. 16,644

## INDIRECT METHOD FOR THE RESOLUTION OF OVERLAPPING CHROMATOGRAPHIC PEAKS

M. ROSENBAUM\*, V. HANČIL and R. KOMERS

*Institute of Chemical Process Fundamentals, Czechoslovak Academy of Sciences, 165 02 Prague 6 - Suchbátka (Czechoslovakia)*

(First received November 7th, 1983; revised manuscript received February 6th, 1984)

---

### SUMMARY

The resolution of overlapping peaks depends on the time available for processing the original data. Owing to the high demands on time, complex algorithms cannot be used for the resolution of peaks from a chromatograph used as a control unit of a chemical reactor, where a rapid response is essential. On the other hand, these complex algorithms can be used for analytical purposes when the evaluation time is unimportant. The Gaussian model of a peak and its simplification, enabling a proper description of a peak to be made, are discussed together with an iterative algorithm for the calculation of parameters. The iterative algorithm utilizes the possibility of determining parameters relating to a certain peak separately and so decreases time and memory requirements.

---

### INTRODUCTION

The control of chemical processes using chromatographs is sometimes hindered by problems connected with the physico-chemical separation of the components of the mixture being analysed. Under laboratory conditions there are difficulties resulting from the low accuracy of direct integrating devices, especially when there are overlapping peaks. Such cases require the use of a computer in the measuring sequence or with the controller. In such a way quantitative parameters for controls can be obtained by indirect numerical separation of peaks. A retention curve is usually represented by a time series of digitized deviations from the zero signal of the detector.

Our attempts to resolve indirectly overlapping peaks using mathematical methods<sup>1,2</sup> suggested that the method offers further possibilities. Similar conclusions were drawn by other workers<sup>3-5</sup>. The key factor, however, is the choice of the mathematical model and the choice of the starting parameters. We tested the method on standard mixtures, the composition of which was determined by weighing of the pure components<sup>2</sup>. The problems could be described by a probability peak model of the Gauss type and of other forms of retention curves are approximated by a final series similarly as in ref. 5. The increase in the number of parameters led to ambiguity in

the approximations and subsequently to modification of programs used. We have therefore tried to describe the theoretical considerations involved.

## THEORY

Real retention curves can be approximated by a series of mathematical models characterized by parameter vectors, the only independent variable being time. The separation algorithm should find parameters that ensure a minimum of the chosen objective function. In our case, variance was used as an objective function and was calculated as the sum of the squares of the deviation of the model from experimental values at chosen time intervals. The whole procedure belongs to the class of optimization problems, where it is usually stated that the result depends on the first guess of the parameter vector. Let us first solve the question of how many minima can exhibit an objective function and what the probability that the found estimates are optimal is. We can utilize distinguishing features that characterize a separated peak, namely the continuity of the time function and the possibility of assignment of autonomous subsets of parameters. The set of parameters of the mathematical model can be represented by the vector  $B$ :

$$B = B(B_1, B_2, B_3, \dots, B_i, \dots, B_{(k-1)}, B_k) \quad (1)$$

Assigning groups of components  $B_i$  to individual peaks means transformation of vector  $B$  to matrix  $B$  with elements  $B_{rn}$ , where the row index denotes the serial peak number and the column index serial number of the parameter of an elementary peak:

$$B = \begin{vmatrix} B_{11}, B_{12}, B_{13}, \dots, B_{1m} \\ B_{21}, B_{22}, B_{23}, \dots, B_{2m} \\ \dots\dots\dots \\ B_{n1}, B_{n2}, B_{n3}, \dots, B_{nm} \end{vmatrix} \quad (2)$$

The mathematical model of the retention curve can thus be written as a linear superposition of elementary mathematical models of separated peaks:

$$S = P_1(B_{11}, B_{12}, \dots, B_{1m}, t) + P_2(B_{21}, B_{22}, \dots, B_{2m}, t) + \dots + P_i(N_{i1}, B_{i2}, \dots, B_{im}, t) + \dots + P_n(B_{n1}, B_{n2}, \dots, B_{nm}, t) \quad (3)$$

For example, for a Gauss-type model the  $i$ th peak can be written as

$$P_i = P_{i1} \exp[-(t - B_{i2})^2/B_{i3}^2] \quad (4)$$

and the subset of parameters has three elements. Similarly, we adopt the idea that the retention curve can be divided into sub-curves  $E_i(t)$ , assuming the validity of

$$E(t_j) = E_1(t_j) + E_2(t_j) + \dots + E_n(t_j) \quad (5)$$

A real objective function in real time can be considered as discrete and non-normalized (by the number of sampling points of the time series):

$$OF = \sum_k [S(t_k) - E(t_k)]^2 \tag{6}$$

Using eqns. 3 and 5, the last relationship can be rewritten as

$$OF = \sum_k [P_1(\dots, t_k) - E_1(t_k)]^2 + \sum_k [P_2(\dots, t_k) - E_2(t_k)]^2 + \dots + \sum_k [P_i(\dots, t_k) - E_i(t_k)]^2 + \dots + \sum_k [P_n(\dots, t_k) - E_n(t_k)]^2 + R \tag{7}$$

The term  $R$  incorporates all mixed members of series expansion 6. Expansion 7 has  $n + 1$  members, which can equal zero, and in this way the objective function can have local minima. By binomial expansion of  $(1 + 1)^j$ , it can be easily shown that

$$\binom{n+1}{1} + \binom{n+1}{2} + \dots + \binom{n+1}{i} + \dots + \binom{n+1}{n+1} = 2^{n+1} - 1 \tag{8}$$

The number of local minima can therefore increase according to

$$\begin{matrix} n = 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ m = 7 & 15 & 31 & 63 & 127 & 255 & 511 & 1023 & 2047 \end{matrix} \tag{9}$$

or, when we do not expect a minimum of the member  $R$ , the sequence  $m$  begins with 3. In any case, the probability of finding a global minimum decrease exponentially with  $n$ .

If we ensure autonomous variations of the parameters of the mathematical model in groups, then the complexity of the model of the peak is not significant. Treatment of subsets of parameters is described in the next section.

Assignment 9 also describes the ambiguity of the results of optimization routines, which is often encountered in practice.

In the design of the mathematical model of separated peaks we started from a probability assumption of the Gauss type:

$$y = y_0 \exp(-x^2/2) \tag{10}$$

which corresponds to physico-chemical interaction of the following component, and other effects will be described by higher derivatives of eqn. 10.

Let us choose in this connection a model of a chosen normalized peak as a differential equation with constant coefficients:

$$a_n y^{(n)} + a_{(n-1)} y^{(n-1)} + \dots + a_1 y^{(1)} + a_0 y = f(x) \tag{11}$$

Linear multipliers  $a_i$  represent functions 10 and their derivation in the function  $f(x)$ . Integration or numerical solution of eqn. 11 will be performed by elementary properties of eqn. 10. The first and the second derivatives can be rewritten into a set of equations:

$$\begin{aligned}
 y &= p_0(x)y \\
 y^{(1)} &= p_1(x)y \\
 y^{(2)} &= p_2(x)y \\
 &\vdots \\
 y^{(k)} &= p_k(x)y \\
 &\vdots
 \end{aligned}
 \tag{12}$$

where the functional multipliers  $p_k(x)$  are linear polynomial of the  $k$ th order for which the following recurring relationship is valid:

$$p_k(x) = p_{(k-1)}^{(1)}(x) - x p_{(k-1)}(x) \tag{13}$$

For example,

$$\begin{aligned}
 p_0(x) &= 1 \\
 p_1(x) &= -x \\
 p_2(x) &= x^2 - 1 \\
 p_3(x) &= -x^3 + 3x \\
 p_4(x) &= x^4 - 6x^2 + 3
 \end{aligned}
 \tag{14}$$

If any equation in the set 12 is integrated, then the primary function of the integral is defined by eqns. 14.

Multiplying assignment 13 by the exponential function 10, represented by  $y$ , we obtain

$$p_k(x)y = [p_{(k-1)}^{(1)}(x) - x p_{(k-1)}(x)]y \tag{15}$$

which is a limited integral of derivative  $y^{(k+1)}$ . It was determined, *e.g.*, by separation of variables:

$$\int_{-\infty}^{\infty} y^{(k+1)} dx = \lim_{a \rightarrow \infty} [p_k(x) y]_{-a}^a \tag{16}$$

If the order of the polynomial  $k$  is even, then the even function on the right-hand side of eqn. 16 is of the form

$$p_k(-a)y(-a) = p_k(a) y(a) \tag{17}$$

and integral 16 converges to zero regardless of the magnitude of its limits. In the case of odd  $k$ , the type of function 16 is given by the polynomial  $p_k(x)$ :

$$p_k(-a) y(-a) = -p_k(a) y(a) \tag{18}$$

and integral 16 can be transformed to

$$\int_{-\infty}^{\infty} y^{(k+1)} dx = \lim_{a \rightarrow \infty} 2[p_k(a) y]_a^0 \quad (19)$$

the resulting value of which need not be defined. However, taking into account the properties of the exponential function  $y$  we are in fact solving extrinsic limits of functions of an exponential class where the right-hand side of eqn. 19 can be transformed to eqn. 20 by expansion of the exponential function from eqn. 10:

$$\int_{-\infty}^{\infty} y^{(k+1)} dx = \lim_{a \rightarrow \infty} \left( \frac{a_k x^k + a_{(k-1)} x^{k-1} + \dots + a_1 x + a_0}{b_0 + b_1 x^2 + b_2 x^4 + \dots + b_h x^{2h} + \dots} \right)_{x=a} \quad (20)$$

Taking into account the first  $h$  members from an unlimited expansion in the denominator, where  $2h$  is greater than  $k$ , the limit of integral 16 is also zero. It is evident from this procedure that the only non-zero integral is assigned to the nominal function 10 and equals  $\sqrt{2\pi}y_0$ . If we are able by a numerical procedure to identify the nominal function  $y$  from the total function  $F(x)$ , then its integral represents the area of the following peak. In real conditions the formal distortion of a peak is then explained by the effects of process dynamics and is described by higher derivatives of the model used. This is valid for deviations of peak heights, retention times, component interactions and others. Non-linear regression analysis of statistical data sets leads to the estimation of at least three parameters for one peak. Other parameters describing distortions by higher derivatives and zero drifts increase the time of analysis and this also increase the period of the control or measuring cycle. The suggested method of indirect identification of overlapping peaks by a mathematical model requires  $k + 3$  parameters per peak. The value of  $k$  is the order of the differential equation representing the mathematical model of the peak and the number of the nuisance parameter for its quantification.

#### TREATMENT OF PARAMETERS

Effects of acceleration and retardation forces as components of the probability description of a separated peak are incorporated into a physico-chemical interpretation of the model used (eqn. 11) and are proportional to the first and second time changes of the response defining the curve:

$$\sum_{i=1}^n (a_{0i} p_{0i} + a_{1i} p_{1i} + a_{2i} p_{2i}) y_i \doteq E(x) \quad (21)$$

Using the equation

$$x_i = (t - B_{i2})/B_{i3} \quad (22)$$

we can transform independently variable coordinates of the  $i$ th peak to real time. In

parameter  $B_{i3}$  the effect of constant  $\sqrt{2}$  is incorporated. From eqn. 21 it thus follows that

$$\sum_{i=1}^n \left[ \frac{a_{i2}}{B_{i3}^2} (t - B_{i2})^2 - \frac{a_{i1}}{B_{i3}} (t - B_{i2}) + (a_{0i} - a_{2i}) \right] - B_{i1} \exp[-(t - B_{i2})^2/B_{i3}^2] \doteq E(t) \quad (23)$$

The value of  $n$  determines the number of peaks in a time expansion of studied elution curve  $E(t)$ . A value of unity can be assigned to  $a_{0i}$  without any loss of generality. The area of a Gauss-type peak is proportional to  $B_{i1}B_{i3}$ , the proportionality constant being  $\sqrt{\pi}$ . Using the procedure shown in relationship 2 the multipliers  $a_{1i}$ ,  $a_{2i}$  are called  $B_{i4}$  and  $B_{i5}$ . The number of columns of the matrix  $B$  is fixed at 5 and the number of rows equals to number of overlapping peaks. Non-linear regression methods, in which some of the parameters are in mutual products or quotients, exhibit convergence problems. However, in our case we have left the model in the form of eqn. 23 as we can easily guess likely values of  $B_{4i}$  and  $B_{5i}$ . Let us now suppose that the optimization algorithm has found a local minimum of the objective function  $OF$ . We can suggest a suitable identification subroutine that can check whether the minimum is global in a given parameter space.

Let us denote the second column of matrix  $B$  as a critical one. We shall investigate the parametric sensitivity of objective function 6 to variation of nearly independent parameters  $B_{i2}$ , which define the relative tops of Gauss-type peaks on the time axis. Let us choose the sequence

$$\{b\} = \{\dots, B_{i2} - 2\Delta B_{i2}, B_{i2} - \Delta B_{i2}, B_{i2}, B_{i2} + \Delta B_{i2}, B_{i2} + 2\Delta B_{i2}, \dots\} \quad (24)$$

where  $\Delta B_{i2}$  is a suitably chosen step, e.g.,  $B_{i3}/10$ .

To any element of sequence 24 is assigned a value of the objective function  ${}^jOF_i$ :

$$\{OF_i\} = \{\dots, {}^{-2}OF_i, {}^{-1}OF_i, {}^0OF_i, {}^1OF_i, {}^2OF_i, \dots\} \quad (25)$$

We can expect that the monotonous sequence 25 will correspond to the situation when our algorithm did not fit the  $i$ th peak into the neighbourhood of the relative minimum  $OF$ . If there is a value of  ${}^0OF_i$  in the sequence that is very small, then we can expect that the model of the  $i$ th peak will have parameters that are close to the optimum. A decision algorithm is chosen so that it examines signs of differences of elements 25. If a sign change is recorded, then the  $i$ th row is not subjected to optimization

$$B = \begin{vmatrix} B_{11}, B_{12}, B_{13}, B_{14}, B_{15} \\ B_{21}, B_{22}, B_{23}, B_{24}, B_{25} \\ \vdots \\ C_{i1}, C_{i2}, C_{i3}, C_{i4}, C_{i5} \\ \vdots \\ B_{n1}, B_{n2}, B_{n3}, B_{n4}, B_{n5} \end{vmatrix} \quad (26)$$

If all rows of matrix  $B$  are fixed, then the calculation is terminated, or matrix  $C$  is considered as a new guess of the parameters and iteration proceeds.

The procedure outlined can be realized by elementary routines, but it is suitable to incorporate a strategy ensuring termination of the calculation at the moment when the accuracy of the result is better than 1%.

In order to estimate the computer time required, we consider that in the most favourite case the transformation of matrix  $B$  to  $C$  needs  $n$  cycles. The modified program SECOND from an earlier paper<sup>2</sup> is shown in Fig. 1. To demonstrate the properties of the mathematical model used for a separated peak, a Gauss-type peak and its first and second derivatives are shown in Fig. 2.

It is evident that the initial guesses of  $B_{i4}$  and  $B_{i5}$  must take into account their effect on the expected superposed time dependence of the  $i$ th component  $E_i(t)$ . They were not greater than  $0.1B_{i1}$ . Non-Gaussian peaks should be corrected much more strongly. As an example of the suggested method we have performed a numerical separation of the mixture described previously<sup>2</sup>, which exhibited the greatest devia-

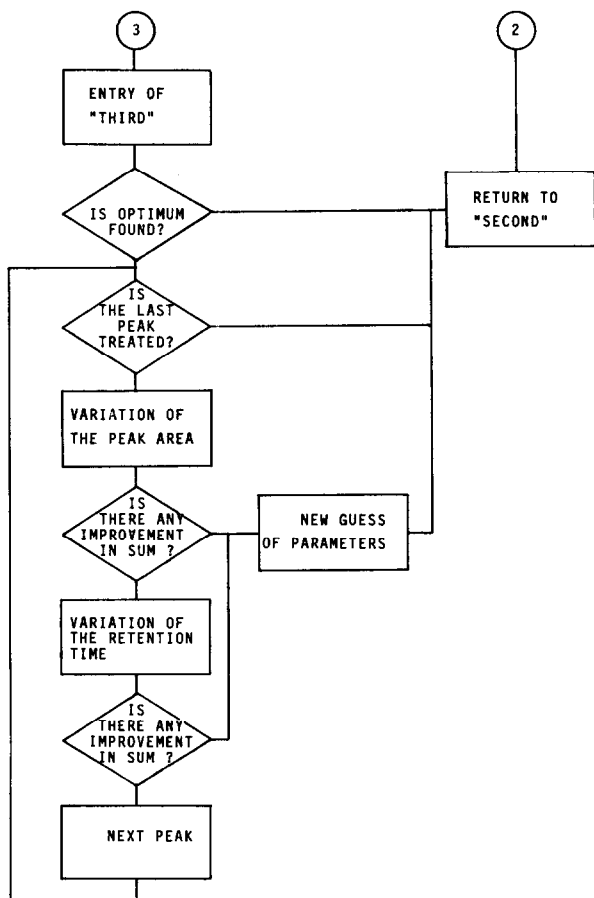


Fig. 1. Modified program SECOND.

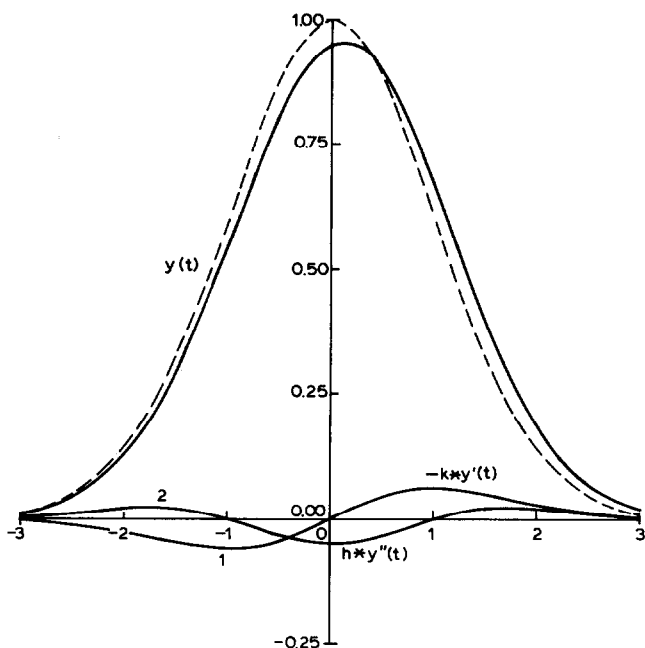


Fig. 2. Gauss-type peak with two components of its derivatives. —, Experimental peak; - - - - - , fitted peak. 1, Negative component of the first-order derivative; 2, positive component of the second-order derivative. Coordinates are expressed on a relative scale.

tions from the expected values found by weighing. The comparison in Table I shows a significant improvement. The elution curve is shown in Fig. 3.

## CONCLUSION

Utilization of numerical methods for resolving overlapping peaks is advantageous when physico-chemical procedures fail or are too expensive.

The choice of the mathematical model of a separated peak, which should be

TABLE I

COMPARISON OF THEORETICAL AND CALCULATED DATA FOR THE COMPOSITION OF THE ANALYSED MIXTURE

Component	Composition (%)				
	Theoretical	Calculated*			
		A	$\Delta$	B	$\Delta$
Ethylbenzene	26.00	25.10	0.9	25.93	0.07
<i>p</i> -Xylene	51.00	51.68	-0.68	51.05	-0.05
<i>m</i> -Xylene	23.00	23.22	-0.22	23.02	-0.02

\* A, Results obtained in ref. 2 (mixture A, set 3); B, results obtained in this study;  $\Delta$ , difference between the actual and calculated amounts of a component.



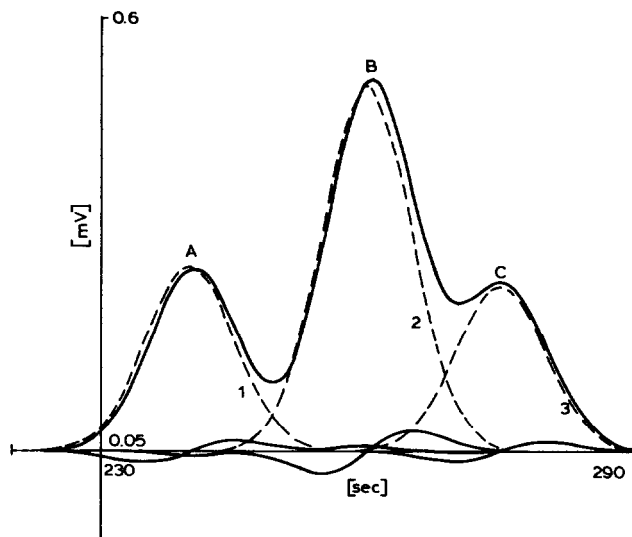


Fig. 3. Chromatogram of mixture A from ref. 2. Peaks: A = ethylbenzene (26.00%, w/w); B = *p*-xylene (51.00%, w/w); C = *m*-xylene (23.00%, w/w). For other comments, see Fig. 2.

similar in type to the time expansion of the retention curve is important. Starting guesses of parameters in the optimization algorithm have to be determined carefully, taking into account that neither the height nor the location of a peak is equal to coordinates of corresponding points on the retention curve, but they are distorted by time dependences of neighbouring peaks.

Special attention must be paid to the choice of the iteration algorithm. For analytical purposes we do not need to take into account the time needed for one numerical separation and the algorithm given previously<sup>2</sup> is suitable. However, when a chromatograph is incorporated into the control loop of a technological process the computer time is crucial, and simpler algorithms with a rapid response which utilize relations among model parameters must be used. We doubt if a general routine exists and it seems wiser to develop special routines than to construct a large general routine. Any routine, however, should ensure an accuracy of 1%.

#### REFERENCES

- 1 M. Rosenbaum, V. Hančil and R. Komers, *J. Chromatogr.*, 191 (1980) 157.
- 2 M. Rosenbaum, V. Hančil and R. Komers, *J. Chromatogr.*, 246 (1982) 1.
- 3 F. Dondi, A. Betti, G. Blo and C. Bigli, *Anal. Chem.*, 53 (1981) 496.
- 4 F. J. Knorr, H. R. Thorsheim and J. M. Harris, *Anal. Chem.*, 53 (1981) 821.
- 5 Y. Mori, *J. Chromatogr.*, 70 (1972) 31.