

CHROM. 14,985

RESOLUTION OF OVERLAPPING CHROMATOGRAPHIC PEAKS WITH AN INTERFERING BACKGROUND

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

Institute of Chemical Process Fundamentals, Czechoslovak Academy of Sciences, Prague (Czechoslovakia)

(Received April 6th, 1982)

SUMMARY

Overlapping chromatographic peaks were separated mathematically by means of a Hewlett-Packard 9825 calculator without digitization equipment and possessing a minimum capacity of the internal memory. It was shown by the numerical comparison of the calculated results with basic and double representation accuracy of the selected points of the chromatogram that the accuracy of the method is most affected by the choice of the mathematical model and/or by the interaction of components. When a gaussian curve was chosen for approximation of a peak the accuracy was better than 1%.

INTRODUCTION

Analytical laboratories can utilize various data processors, including simple calculators. However, the desk-top calculators are not provided with special libraries of programs, and we have therefore started to compile programs of properly working algorithms into particular calculator languages. The programming language HPL is highly effective and we consider it useful to inform users of the Hewlett-Packard HP 9825 about a memory-saving solution for the evaluation of chromatograms or spectrograms with overlapping peaks. We have established how to store data sets using the basic properties of integers and their binary representation¹ in the calculator memory.

Evaluation of chromatograms by using small computers has some limitations, e.g., human factors, the large amount of work involved and suitability of mathematical models. Often the direct determination of the components of a mixture is carried out by means of a microprocessor, which is programmed so that the input is the deviation from the baseline and the time dependence of the quantity followed. The accuracy is acceptable only for cases in which the curves for the elution of individual component are completely separated in time. In other instances the controlling program of the integrator must take into account the values of the first and second derivatives and it has to determine the times of the local extremes and/or inflection points. The method of direct integration does not seem suitable.

Our laboratory is equipped with an HP 9825, which can be used for off-line

evaluation of the chromatograms using a reader/plotter, as described in a previous paper². We have modified our approach for the case in which only a minimum central memory unit is available. The program used needs about half of the standard memory (6844 bytes) and hence the array of experimental points and intermediate results is small for an ordinary solution of the problem. An internal representation of one piece of digital information in the calculator needs 8 bytes of the memory and therefore every point needs 16 bytes. The use of the external cartridge memory wastes time, so we have chosen the other programming means, binary storage of experimental points. The $f_{ti}(D)$ function³ of the HP 9825 calculator makes it possible to transform a decimal number into its binary representation and back using the $itf(B)$ function in the interval -32767 to $+32767$. Thus, we are able to represent any point with an accuracy of $1/(2 \cdot 32767 + 1)$, i.e., $1/65535$. This is more than enough as other members of the measuring chain adversely affect the treated information to much greater extent. This makes it possible to use one memory unit for storage of both coordinates of the point on the elution curve. This is especially suitable with the fast HP 9825 calculator but would not be so convenient with other slower desk-top calculators. The suggested method will first be described using as an example a decimal number.

Let us have 9999 resolving levels of a decimal number. These can be rewritten into the form of a sum:

$$9999 = 9900 + 99 \quad (1)$$

or more generally into

$$C = X \cdot 100 + Y \quad X, Y \in \langle 0, 1, \dots, 99 \rangle$$

The transformation is evident for the storage of X and Y into 2 left positions and 2 right positions, respectively. Hence we have 100 levels for each coordinate. This means that information is transferable within an accuracy of 1%. However, one does not need to transform the information about a particular coordinate with the same precision as that about the other coordinate.

In binary representation it seems suitable to choose 128, 256 or 512 levels. With regard to the preliminary study of the problem we have chosen 256 levels on the X coordinate and 128 levels on the Y coordinate. The precision of reading off a graph manually cannot be better than 1%. As was shown for a decimal number where a shift of the digits in the representation of a number is accomplished by multiplication with an integral power of ten, in binary representation one has to multiply by an integral power of 2. In our program we use arithmetic operation with a decimal number (however, in some modifications of the HP 9825 one can use the function shf)³.

Evaluation of an elution curve is performed in two steps, in two programs. The first program stores the data obtained from a graphically represented curve on magnetic tape and the second program transfers the data into the internal memory of the calculator and calculates the composition of the mixture being analysed.

Data for the first program consist of the following:

- (1) number of peaks, number of read off points and limits of the points;
- (2) identification of the problem (maximum 32 characters);
- (3) coordinates of the points X , Y ($X \in \langle 0, 128 \rangle$; $Y \in \langle 0, 255 \rangle$);

(4) first guesses of the parameters of the model used and their lower and upper limits;

(5) identification number of the data set on the tape.

It is assumed that the tape is marked, e.g., trk 0; mrk 10, 1500. This statement marks on track no. 1 ten sets of size 1500 bytes each.

In the second program the data set is identified by the number of the data set (5). The user modifies only the subroutine "func" where models of individual peaks are defined. In our listing this model is assumed to be gaussian for all peaks of the chromatogram.

$$Y_{ki} = B_{1i} \exp \left\{ -[(X_k - B_{2i})/B_{3i}]^2 \right\} \quad (2)$$

$$i \in \langle 1, j \rangle$$

$$k \in \langle 1, n \rangle$$

where j is number of peaks and n is number of digitized points. Parameter B_{1i} denotes the amplitude of the peak and B_{2i} is the coordinate of the maximum of the peak. B_{3i} is a measure of the dispersion of the peak around the maximum. Evaluation of parameters is a non-linear optimization problem⁴ and therefore the calculation has to start with first guesses of the B parameters. It is easy to guess B_{1i} as it is the magnitude of the corresponding peak. The X position of the maximum of the peak is an excellent first guess for parameter B_{2i} . The only problem is the first guess of parameter B_{3i} , which has to be guessed from the difference in the X -coordinates of the peak maximum and the point where the Y -coordinate is equal to the value $B_{1i} \exp(-1)$, i.e., $0.368 B_{1i}$. Algorithm of the second program leads to values of parameters which by a simple product yield a value proportional to the area of the peak. The total area below one peak can be expressed as

$$A = B_{1i} \int_{-\infty}^{+\infty} \exp[-(X - B_{2i})/B_{3i}]^2 dX \quad (3)$$

Taking into account that

$$\frac{1}{\sqrt{(2\pi)\sigma}} \int_{-\infty}^{+\infty} \exp(-(X - \mu)/\sqrt{2}\sigma)^2 dX = 1 \quad (4)$$

and substituting B_{3i} for $\sqrt{2}\sigma$ into eqn. 4 yields $A = \sqrt{\pi} B_{1i} B_{3i}$.

Assuming that individual components of the mixture are in the same ratio as that of the areas of their corresponding peaks, one can write the very simple equation

$$C_i = \frac{B_{1i} B_{3i}}{\sum_{l=1}^j B_{1l} B_{3l}} \quad (5)$$

EXPERIMENTAL

Materials

m-Xylene, *p*-xylene and ethylbenzene (all not less than 99.5% pure by gas chromatographic (GC) analysis) were obtained from the Institute of Chemical Technology (Prague, Czechoslovakia). Bentone 34 and dinonyl phthalate (for GC) were supplied by Applied Science Labs. (State College, PA, U.S.A.) and BDH (Poole, Great Britain), respectively.

Apparatus

The GC measurements were carried out with a Pye Model 64 heated dual flame-ionization detector programmed chromatograph (Series 104; Pye Unicam, Cambridge, Great Britain). It was equipped with a 0–1 mV recorder (Honeywell Elektronik 194). The column employed was 150 × 0.3 cm I.D. glass tube operating at 104°C. Nitrogen was used as the carrier gas at a flow-rate of approximately 30 ml/min. The samples were introduced with a 10- μ l Hamilton microsyringe. The injection port and detector were maintained at 200 and 150°C, respectively.

Column packing

A mixture of Bentone 34 and dinonyl phthalate (2:1) was used as the stationary phase. Gas-Chrom Q of particle size 80–100 mesh (Applied Science Labs.) served as the support. The packing was prepared in the usual manner by dissolving the individual components in chloroform and slurring with the support. The column packing contained 5% of the mixed stationary phase. The packed column was pre-conditioned at 120°C for 12 h with the carrier gas flowing.

Calibration mixtures

For the determination of ethylbenzene (EB), *p*-xylene (PX) and *m*-xylene (MX), calibration mixtures A, B and C were prepared with EB:PX:MX mass proportions of 1:2:1, 1:4.5:4.5 and 3.5:3.5:1, respectively. Therefore, the following amounts of individual compounds were weighed and mixed: 0.09326 g of EB, 0.18293 g of PX and 0.08250 g of MX (A); 0.02673 g of EB, 0.12908 g of PX and 0.12119 g of MX (B); and 0.13657 g of EB, 0.14589 g of PX and 0.04131 g of MX (C).

RESULTS

To evaluate the statistical validity of the technique in combination with the method of data representation one could start from the assumption that the composition of the model mixtures as relative percentages is determined by weighing to 2 decimal places (Table I). This means that we were able to judge the approximations resulting from the programs. Using the relationship experimental concentration = \hat{f} (calculated concentration) and by linear regression, the correlation coefficient (*R*) and the slope of the regression line were evaluated. Three compositions of the mixtures (see Table I) were considered as one statistical set. Coordinates of the points of the curves (see Fig. 1) were introduced in millimetres (set 1) or in divisions of the chart used (sets 2 and 3). Sets 1 and 3 were stored so that one coordinate was in one byte; with set 2, two bytes per coordinate were used. Numerical results of the

TABLE I
COMPOSITIONS OF THE MIXTURES USED

Mixture	Component (%)		
	Ethylbenzene	p-Xylene	m-Xylene
A	26.00	51.00	23.00
B	9.65	46.60	43.75
C	42.18	45.06	12.76

TABLE II
STATISTICAL CHARACTERISTICS OF THE TREATED DATA SETS

m = Slope of the regression line; q = intercept; R = regression coefficient; Δ = difference between the actual and the calculated amounts of a compound; σ = standard deviation.

Set	Sample	Composition calculated (%)	Δ	m	q	σ	R
1	A	25.95	0.05	1.0086	-0.2859	0.56478	0.99916
		50.85	0.15				
		23.20	-0.20				
	B	9.81	-0.16				
		45.68	0.92				
		44.51	-0.76				
	C	42.68	-0.51				
		45.50	-0.44				
		11.81	0.95				
2	A	25.41	0.59	1.0551	-1.886	1.3608	0.99754
		51.24	-0.24				
		23.35	-0.35				
	B	9.12	0.53				
		45.94	0.66				
		44.94	-1.19				
	C	42.27	-0.09				
		47.66	-2.60				
		10.07	2.69				
3	A	25.10	0.90	1.0204	-0.6799	0.50686	0.99963
		51.68	-0.68				
		23.22	-0.22				
	B	9.01	0.64				
		46.81	-0.21				
		44.18	-0.43				
	C	41.76	0.42				
		45.47	-0.42				
		12.76	0.00				

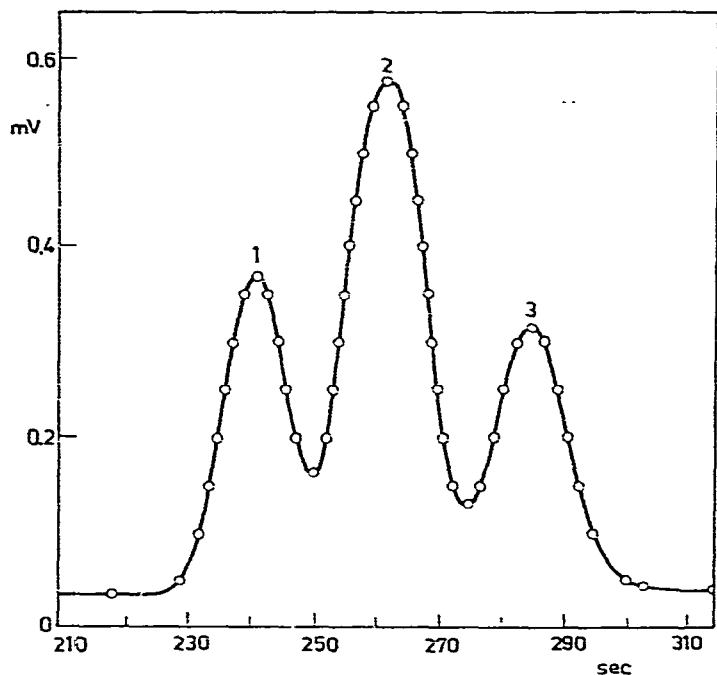


Fig. 1. Chromatogram of mixture A. Peaks: 1 = ethylbenzene (26%, w/w); 2 = *p*-xylene (51%, w/w); 3 = *m*-xylene (23%, w/w).

indirect analysis with statistical parameters are given in Table II. By comparing statistical parameters of sets 2 and 3 in Table II we can find that the two bytes per coordinate representation can provide even worse results than that for one byte per coordinate. This means that the source of the errors need not be the representation but the selection of the model of the peak, the drift of the baseline or inaccuracy during data digitization. When the calculator with a plotter is used, the last source of errors is minimized. This is valid under the assumption that the analogue output from the chromatograph is better than 0.5%. Drift of the baseline will be dealt in a separate paper.

It was found that the basic significance of the statistical validity of the indirect evaluation of the chromatograms is the selection of the models of the peaks and their proper use in the subroutine "func", *i.e.*, in the only part which is modified by the user. In our case a probabilistic model of the normal distribution was used for each peak. The standard errors of sets 1 and 3 are 0.56 and 0.51, respectively, and that of set 2 is 1.36. The two bytes per coordinate representation of the points was thus found to be adequate. The statistical parameters of set 2 indicate some reservations about the method.

CONCLUSION

The method of indirect evaluation of elution curves with overlapping peaks does not adversely affect the accuracy of the overall evaluation sequence and is acceptable from the technical point of view. Its algorithm can be programmed on a standard version of the HP 9825 desk-top calculator without any peripherals. One peak can be represented by 15 points. The maximum error in the determination of the concentration of a particular component is about 1%. The computer time required depends on the first guesses; in our examples it was about 5 min. The programs in the


```

0: "SECOND":
1: "CHROMATOGRAM RESOLUTION FOR 5 PEAKS"-
2: dse "RESOLUTION FOR 5 PEAKS" :sta
3: are "NMB.of PEAKS":K:are:0:3K+2+K:3:up:K:10
4: tsk 0:1:1: 4:sta 14:3:MM:dim: N:10:1:K:K#1:20:2:K:H
5: dse BK1,MC2,K
6: dim KC61,IC31,FC51,MC31
7: are "NMB.of FILE?":A:1:are: A,N:1:2:H:BC:1:MC:1
8: KCN+2)+H+H:A+IC31
9: dim: AK:K+21,CIK,K:1-21:PCW1,PCW
10: .0001-S:1.00001+FC41
11: .01+FC31:10+FC11:1-P:3+IC21:1.001-FC31
12: are "SUMMA DIF12"
13: 1+K+G:1+G+H:KH+KC31:IC31)+K+K:4:3:FC41)+K+K:1
14: are "PROGRAM"
15: "SOLVE":0+KC61
16: for I=1 to K
17: 1+IC51+KC61
18: next I
19: are "A":if KC6K=31-S+MC1:0:are "A"
20: "B":are "B":if KC6K=2+MC1:1:are "B"
21: "C":1+IC11
22: "C":if IC21:0:are "C"
23: for I=1 to K
24: IC31)+I+J:BC11)+PC11:K:4:1+L-1
25: 1.01:abs(BC11)+.01-PC11
26: next I
27: are "D"
28: "D":if FC41,FC51 and IC21:1:are "D"
29: for L=1 to K
30: CL-1:H+C
31: for J=1 to K
32: FC11)+I+A:BC11)+PC11
33: next J
34: L+IC31)+R:L+K:4)+P
35: PC11)+D:1+abs(PC11)+abs(PC11)+abs(PC11)+D
36: .0010+D
37: FC11)+D-B
38: if B'=MC2,L) or S.=MC1)+L:0:are "E"
39: FC11)+D+FC11)+D+D:are "E"
40: "E":FC11)+D+FC11
41: "F":0+G:K:31)+A
42: for I=1 to H
43: all 'func'(I,A,B
44: next I
45: for J=1 to K
46: C+D-A
47: C+FC11)+L:J:1:3-PC11
48: next J
49: "H":next L
50: "I":for I=1 to K
51: CI-1)+H+C:G+AC1,G1
52: for J=1 to N
53: C+J+B
54: AC1,G1)+PC11)+VC11)+AC1)+G1
55: next J
56: for J=1 to K
57: G+AC1,J:1:1)+H+A
58: for L=1 to H
59: C+L+B:N+L+A
60: AC1,J:1)+PC11)+PC11)+AC1)+J

```



```

61: next L
62: next J
63: if ACI, I1 > 1e-20; sto "K"
64: "J": for J=1 to G
65: 0 ← ACI, J1
66: next J
67: 1 ← ACI, I1
68: "K": next I
69: 0 ← F
70: for I=1 to K
71: F ← ACI, G1 + 2 ← F
72: next I
73: for I=1 to K
74: (ACI, I1) ← ACI, H1
75: next I
76: for I=1 to K
77: ACI, G1 ← ACI, H1 + ACI, G1
78: for J=1 to K
79: ACI, J1 ← (ACI, H1) + ACI, J1 + ACI, H1 ← ACI, J1
80: next J
81: next I
82: FC21 ← FC11 + R; sto "L"
83: "L": FC11 ← R ← P
84: for I=1 to K
85: for J=1 to G
86: ACI, J1 ← CC1, J1
87: next J
88: R ← CC1, I1 ← CC1, I1
89: next I
90: for I=1 to K
91: 1 ← I ← C
92: for J=C to G
93: CC1, J1 ← CC1, I1 ← CC1, J1
94: next J
95: for J=1 to K
96: if I=J; sto "M"
97: for L=C to G
98: CCJ, L1 ← CC1, L1 ← CCJ, I1 ← CCJ, L1
99: next L
100: "M": next J; next I
101: 0 ← S ← T
102: for I=1 to K
103: CC1, G1 ← ACI, H1 ← CC1, H1
104: KI31 ← I ← A
105: BC11 ← CC1, H1 ← B; if S=ND3, I1; MD2 ← I1 ← B
106: MD1, I1 ← P(A1); if F(S) B ← P(A1)
107: T ← CC1, H1 ← ACI, G1 ← ACI, H1 ← T
108: S ← CC1, H1 ← CC1, H1 ← 1 ← P(A1) ← SC11 ← CC1, H1
109: next I
110: T ← (SF) ← T; 0 ← S; if T < 0; 2 ← S; -T ← T
111: if T > 1; 1 ← T
112: 0 ← KI21; rad; if T < -1; -1 ← T
113: if T > -1 and T < 0; 1 ← KI21
114: if T < 0; abs(T) ← T
115: if T = 0; sto "N"
116: sin(π((1-T+2)/T) ← T; if KI21 = 1; π - T ← T
117: sto "P"
118: "N": π/2 ← T
119: "P": 180/T ← π ← T; if S=0; 120 - T ← T
120: "Q": KI31 ← A; KI51 ← B
121: for I=1 to N
122: call 'func'(I, A, B

```


REFERENCES

- 1 J. E. Kemeny, J. L. Snell and G. L. Thompson, in *Introduction to Finite Mathematics*, Prentice Hall, Englewood Cliffs, NJ, 1966.
- 2 M. Rosenbaum, V. Hančil and R. Komers, *J. Chromatogr.*, 191 (1980) 157.
- 3 *Hewlett-Packard Manual 9825 Desktop Computer Operating and Programming*, Hewlett-Packard, Fort Collins, CO, 1976.
- 4 J. K. Seinfeld and L. Lapidus, in N. R. Amundsen (Editor), *Mathematical Methods in Chemical Engineering*, Vol. 3, Prentice Hall, Englewood Cliffs, NJ, 1974, p. 343.