Journal of Chromatography, 186 (1979) 443–455

CHROM. 12,020

NUMERICAL ANALYSIS OF CHROMATOGRAPHIC–SPECTROMETRIC DATA

PURITY OF PARTIALLY RESOLVED PEAKS AND PATTERN SEPARATION OF THE PURE COMPONENT SPECTRA

JOHN M. HALKET

Institute for Reproductive Biology, Osterstr. 86, 2000 Hamburg 19 (G.F.R.)

SUMMARY

Factor analysis and other matrix manipulation techniques are applied to some data arrays obtained by successive scanning in two chromatographic–spectrometric systems: high-performance liquid chromatography–ultraviolet (HPLC–UV) and gas chromatography–mass spectrometry (GC–MS), the former containing both real and simulated data and the latter simulated data only. In all multi-component cases considered, the substances are assigned identical retention behaviour, or have very nearly the same experimental retention times. In the HPLC system, the method is a generalization of the absorbance-ratio technique for the determination of peak purity where no choice of suitable wavelengths need be made. Factor analysis reveals the number of hidden components and, in a three-component GC–MS example, a computer programme (UNRAVL) is employed to extract the pure component mass spectra.

INTRODUCTION

Factor analysis, in the form of matrix rank analysis (MRA) has found a large number of chemical applications, particularly for the calculation of the number of overlapping components in arrays of fluorescence and absorption spectra[1–7]. A similar technique, principal components analysis (PCA)[8], has also been applied to the analysis of arrays of overlapping absorption spectra[9–12]. Monteiro and Reed[13] used MRA to calculate the number of overlapping components in a mass spectral data array and PCA was applied by Davis *et al.*[14] to the spectra of binary mixtures obtained by gas chromatography–mass spectrometry (GC–MS). Ritter *et al.*[15] also used PCA successfully to calculate the number of components in several mass spectra of mixtures. Both techniques, MRA and PCA were employed by Halket and Reed to analyse mass spectral data arrays obtained by scanning during direct probe evaporation[16,17]. More recently, MRA and PCA have been employed to investigate steroid mass spectra of liquid chromatographic fractions containing up to four components[18], repetitively scanned sterol GC–MS data[19] and also some simulated data matrices obtained by successively scanning UV spectra across a high-performance liquid chromatographic

(HPLC) peak containing up to four components[20]. Other chromatographic applications of factor analysis, but not involving spectra, include a study of retention data[21] and peak shapes[22].

In this paper, factor analysis is applied to several simulated HPLC data arrays containing up to three components having the same retention characteristics but present in different amounts. In addition two actual examples of scanning UV spectra across HPLC peaks after stopping the pump are presented.

Although the determination of the number of spectral components is fairly straightforward, the problem of separating the pure component spectra is more difficult, and no unique solution may exist. In GC–MS, a suitable method might be that of Biller and Biemann[23], provided that the components are sufficiently resolved to allow the detection of maximizing groups of ions. A complementary method of performing this task may be that which was first described by Meyerson[24] for separating the component mass spectra of binary mixtures and later extended by Reed and co-workers to n components[13,25,26]. In this paper, this approach is illustrated by a simple example, the complete separation of the mass spectra of the column background (OV-101) and two components assigned the same retention times in GC–MS.

EXPERIMENTAL

Chemicals and solvents were of analytical-reagent grade, purchased from E. Merck (Darmstadt, G.F.R.).

Simulation of chromatographic systems

The chromatographic profiles used in the simulation study were obtained by injecting various amounts of *p*-cresol (1–20 μg) and recording the eluting peaks (271 nm) on chart paper (10 cm/min). The HPLC system used is described below. The curves thus obtained were superimposed and assigned the various identities and spectra given below.

HPLC–UV spectra simulation

The curves illustrated in Fig. 1A were made to represent an HPLC peak containing three components, a, b and c. These were assigned the identities and UV spectra (230–280 nm) of N-acetyl-*p*-aminophenol, phenacetin and caffeine, respectively, as obtained with a Perkin-Elmer LC55 detector during HPLC[27]. The spectra were manually digitized at 2.5-nm intervals, giving a total of 21 wavelengths. In Fig. 1, 1–9 indicate the positions where the pump might be stopped in order to scan a spectrum. Such a spectrum would then consist of the superimposed spectra of the pure components in the proportions given at that position. The superimposed spectra were calculated as previously described[18] by multiplying a matrix D, containing the peak heights of each of the three components at each of the nine positions, with a matrix X containing the three pure component spectra (21 wavelengths), *i.e.*,

$$M_{9\times21} = D_{9\times3} \cdot X_{3\times21} \quad (1)$$

Thus, M is a 9×21 data matrix containing the three spectra mixed together in the proportions given at each of the nine positions.

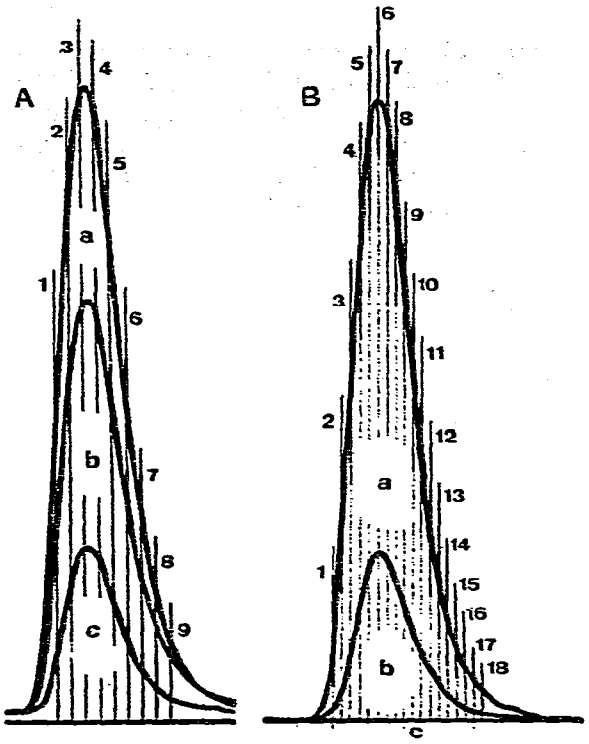

Fig. 1. Overlapping chromatographic profiles for simulation of: (A) HPLC–UV system—(a) N-acetyl-*p*-aminophenol, (b) phenacetin and (c) caffeine; nine positions are indicated where spectra are recorded during stopped flow; (B) GC–MS system—(a) *n*-hexane, (b) vinyl chloride and (c) OV-101; eighteen mass spectra were recorded by repetitive scanning.

Similarly, one- and two-component data matrices were constructed by leaving out the relevant chromatographic profiles and their corresponding spectra, *e.g.*, for two components:

$$M_{9\times21} = D_{9\times2} \cdot X_{2\times21} \tag{2}$$

GC–MS simulation

The curves used to simulate a GC–MS peak are shown in Fig. 1B. The three components, a, b and c, were assigned the identities and mass spectra of *n*-hexane, vinyl chloride[28] and the major mass spectral peaks of the column background, OV-101[29], respectively. The data matrix was constructed as described above:

$$M_{18\times36} = D_{18\times3} \cdot X_{3\times36} \tag{3}$$

Thus, M in this instance is an array of 18 superimposed mass spectra containing the three component spectra (36 mass positions) mixed together in the proportions determined by the respective chromatographic profiles.

High-performance liquid chromatography

A Varian 8500 instrument was employed, equipped with a reversed-phase column (octadecylsilane on LiChrosorb, MicroPak CH-10, 25 cm × 2 mm I.D.). Cresol samples of 20 μg were injected and peaks eluted with 20% methanol in water. The flow-rate was 50 ml/min at a pressure of 180 atm. Spectra were recorded with a Variscan detector in the range 290–230 nm at 10 nm/min. The peak obtained upon injection of 20 μg of *m*-cresol, recorded at 271 nm, is shown in Fig. 2.

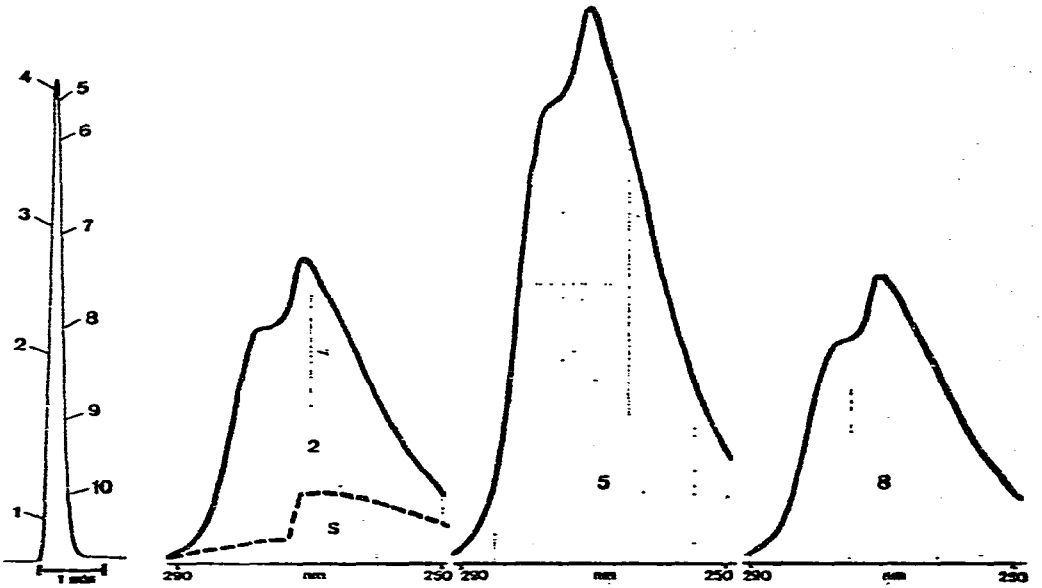

Fig. 2. HPLC peak containing 20 μg of *m*-cresol together with ten positions where UV spectra were recorded during stopped flow. Three partial spectra (290–250 nm) are shown (2, 5 and 8) together with the corresponding solvent spectrum,(s).

At each of the ten positions indicated, the pump was stopped and spectra were recorded on chart paper. Partial spectra (290–250 nm) 2, 5 and 8 are shown for comparison. The corresponding solvent spectrum is also shown, superimposed on spectrum 2. The spectra were then manually digitized at 2.5-nm intervals, giving extinction values at 25 different wavelengths, *i.e.*, a 10 × 25 data matrix was obtained.

The corresponding peak and sample spectra for a mixture of 10 μg of *m*-cresol and 10 μg of *p*-cresol are shown in Fig. 3. Changes in the spectral pattern are seen to occur in the centre and latter half of the peak.

Matrix rank analysis

The data matrices were transformed by the process of gaussian elimination with pivoting[30] and the transformed pivot elements (the largest elements in the matrix) compared with their corresponding propagated errors. The number of elimination steps required to reduce this pivot element (absolute value) to a value less than its error determines the number of components[5,17]. In the present simulated data matrices,

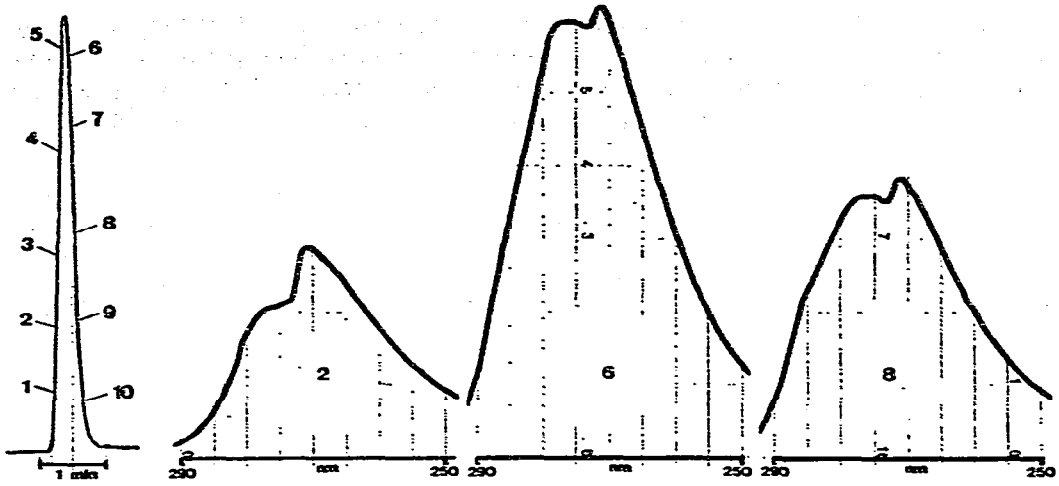

Fig. 3. Badly resolved HPLC peak containing 10 μg of *m*-cresol + 10 μg of *p*-cresol, together with three of the scanned spectra.

the error was taken to be zero. The error level necessary for experimental data was determined empirically.

Principal component analysis

The computer programme forms a covariance matrix from the data matrix and calculates the number of eigenvalues required to account for the variance. This number is equal to the number of underlying components[31]. At this stage, no error comparison routine has been incorporated as these may be unreliable[32] and a simpler approach is adopted. The smaller eigenvalues then represent the system noise and experimental error.

Computer programmes and data preparation

The present studies were performed off-line using a Telefunken TR440 computer (University of Hamburg). Data were punched on to cards and sub-routines for the calculation of covariance matrices and eigenvalues were kindly provided by Dr. W. Rehpenning (University of Hamburg).

RESULTS AND DISCUSSION

HPLC–UV simulation

The computer results obtained for MRA and PCA of five different combinations of N-acetyl-*p*-aminophenol, phenacetin and caffeine are given in Table I.

In this simple example, free of experimental errors, the determination of the numbers of overlapping component spectra is successful in each instance, using both techniques. Logarithms of these eigenvalues are shown later in Fig. 4, where the results are compared with those obtained from experimental data matrices. If the chromatographic profiles of any of the components had been exactly overlapping then they would not be distinguished by the programme, *i.e.*, the spectra of the com-

binations would be proportional and only one component would be detected. In the case of experimental data, errors due to digitization and wavelength reproducibility may obscure such small differences in the composite spectra where the chromatographic profiles nearly overlap or the spectra of the components are very similar.

TABLE I

RESULTS OF MATRIX RANK ANALYSIS (MRA) AND PRINCIPAL COMPONENTS ANALYSIS (PCA) OF SIMULATED HPLC–UV SPECTRAL DATA MATRICES CONTAINING UP TO THREE COMPONENTS HAVING SIMILAR RETENTION BEHAVIOUR

Matrix	*Components**	*MRA: pivot elements***				*PCA: eigenvalues* ($\times 10^{-2}$)*****			
		0	*1*	*2*	*3*	*1*	*2*	*3*	*4*
(a)	a	11502.600	0.000	—	—	*503219.5*	0.000	—	
(b)	a + b	20708.400	11.907	0.000	—	662352.7	*0.900*	0.000	—
(c)	b + c	10135.750	−174.663	0.000	—	274430.1	*35.910*	0.000	—
(d)	a + c	12113.100	−186.472	0.000	—	409854.8	*20.640*	0.000	—
(e)	a + b + c	21318.900	−189.527	9.569	0.000	492607.7	20.735	*0.289*	0.000

* a = N-Acetyl-*p*-aminophenol; b = phenacetin; c = caffeine.

** Position of zero element determines number of components.

*** Position of eigenvalue in italics determines the number of components.

HPLC–UV experimental data

The extinction values measured for the first ten wavelengths of the data matrix, M, for the one-component peak (Fig. 2) are reproduced at the top of Table II and underneath are given the corresponding pivoted matrices obtained after one and two gaussian eliminations, respectively. At the same time an error matrix, S, was transformed[6] so that each element of the transformed data matrix could be compared with its error. The results of MRA on both one-(f) and two-component(h) data matrices are given in Table III, which also gives the corresponding values obtained after subtracting the solvent spectrum in each case, *i.e.*, (g) and (i), respectively.

The initial error value incorporated (0.9 unit) was necessarily very large in order to reduce the matrix to the known rank. This has also been observed by Wallace and Katz[6], who found it necessary to use inflated errors during their rank analyses of absorption spectra. At the present stage, it appears that the only really reliable way to study such examples is to calibrate the particular system being used with results on pure substances and known test mixtures[19].

The PCA results for the same data matrices, (f)–(i), are plotted in logarithmic form in Fig. 4 and compared with the logarithms of the eigenvalues (λ) of the simulated data matrices (a)–(e) given in Table I. If the simple criterion is adopted that the negative values represent noise and experimental error, then the matrix (f), *i.e.*, *m*-cresol + solvent, shows up two components, one being the solvent contribution. On subtracting the solvent spectrum, matrix (g) is obtained, having only one positive eigenvalue. In the case of matrix (h), having two cresols + solvent, the result is less satisfactory, the third eigenvalue being slightly negative. On removal of the solvent, matrix (i), its third eigenvalue decreases further, showing more clearly the presence of two components. It should also be taken into account in this instance that the spectra of the cresols are very similar, that they have very nearly the same retention times and

TABLE II

PARTIAL ONE-COMPONENT SPECTRAL DATA MATRIX OBTAINED BY SCANNING DURING STOPPED FLOW IN HPLC, TOGETHER WITH THE CORRESPONDING MATRICES AFTER ONE AND TWO GAUSSIAN ELIMINATION STEPS WITH PIVOTING

Pivot elements (largest absolute elements) are given in italics. These are compared with their error tolerances in Table III.

Wavelength (nm)	*Extinction*									
290.0	0.25	0.45	0.55	0.50	0.50	0.50	0.48	0.40	0.30	0.25
287.5	0.48	0.95	1.20	1.20	1.20	1.20	0.90	0.80	0.60	0.49
285.0	0.85	1.90	2.95	3.00	3.10	3.00	2.15	1.55	1.10	0.80
282.5	1.55	4.30	7.60	8.00	8.00	7.00	5.00	3.60	3.40	1.40
280.0	2.40	7.05	11.40	13.15	13.05	11.25	8.60	6.10	3.85	2.15
277.5	2.88	8.08	12.86	16.10	15.70	13.28	10.20	7.50	4.55	2.58
275.0	3.00	8.35	13.30	16.50	16.05	13.65	10.50	7.70	4.67	2.70
272.5	4.15	9.75	15.35	18.50	18.00	15.65	12.10	9.00	6.00	3.50
270.0	4.62	10.25	15.15	*18.90*	18.15	15.65	12.65	9.80	6.50	4.33
267.5	4.30	9.05	13.30	16.80	16.50	13.80	11.20	8.80	6.00	4.10
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
After one	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
step	0.0	*−2.38*	1.39	−2.14	−0.64	0.62	−1.03	−1.81	−1.84	−1.49
	0.0	0.11	0.54	0.11	0.21	0.51	0.14	−0.01	0.06	0.27
	0.0	−0.43	1.18	−0.40	0.31	0.37	−0.35	−0.54	0.64	−0.03
	0.0	−0.86	0.85	−0.81	0.42	0.36	−0.20	−0.71	−0.67	−0.08
	0.0	−1.10	−0.04	−1.05	0.23	−0.05	−0.57	−0.84	−0.98	−0.65
	0.0	−1.08	0.07	−1.03	0.20	−0.01	−0.54	−0.85	−1.00	−0.59
	0.0	−0.73	0.52	−0.37	0.23	0.33	−0.28	−0.59	−0.36	−0.28
	0.0	0.13	0.14	0.12	0.02	0.08	0.14	0.14	0.12	0.17
	0.0	0.25	−0.16	0.19	0.36	−0.11	−0.01	0.08	0.22	−0.06
	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.
After two	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
steps	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	*0.98*	−0.01	0.43	0.26	−0.16	−0.21	0.93	0.23
	0.0	0.0	−0.01	0.01	0.18	0.54	0.09	−0.09	0.61	0.20
	0.0	0.0	−0.00	−0.03	0.65	0.13	0.17	−0.06	0.35	0.46
	0.0	0.0	−0.12	−0.05	0.53	−0.34	−0.09	−0.00	−0.69	0.04
	0.0	0.0	−0.16	−0.05	0.49	−0.29	−0.07	−0.03	−0.56	0.08
	0.0	0.0	0.21	0.29	0.43	0.13	0.03	−0.03	0.08	0.18
	0.0	0.0	0.02	0.01	−0.01	0.12	0.08	0.03	0.22	0.09
	0.0	0.0	0.02	−0.03	0.29	−0.04	−0.15	−0.10	−0.01	−0.21
	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.

that the solvent spectrum is relatively insignificant. On this basis, plots such as Fig. 4 could give a fair indication of the purity of HPLC and other chromatographic peaks. This approach is then a generalization of the absorbance-ratio method of investigating peak purity[33], the latter being a particularly important consideration in quantitative work or where sample size is severely limited. In this instance, no decision has to be made concerning the two or three wavelengths to be chosen for ratioing so that the technique can be readily automated.

TABLE III

MATRIX RANK ANALYSIS OF SOME HPLC–UV SPECTRAL DATA MATRICES

Error levels in the transformed elements are given in parentheses.

Matrix	*Contents*	*Pivot elements and errors*				*Number of components*
		Step 0	*Step 1*	*Step 2*	*Step 3*	
(f)	*m*-Cresol	18.900	−2.380	0.985	—	2
		(0.900)	(1.152)	(1.320)	—	
(g)	*m*-Cresol less solvent	17.400	−1.172	—	—	1
		(0.900)	(1.176)	—	—	
(h)	*m*-Cresol, *p*-cresol	15.300	3.539	1.349	0.929	3
		(0.900)	(1.261)	(1.225)	(2.251)	
(i)	*m*-Cresol, *p*-cresol less solvent	15.300	2.530	1.331	—	2
		(0.900)	(1.403)	(1.436)	—	

These powerful techniques, MRA and PCA, are potentially useful for determining the purity of chromatographic peaks. The methods require that the profiles of the components do not exactly overlap within the peak, that they have significantly different spectral patterns and, if n components are present, then at least $n + 1$ spectra must be recorded containing the component spectra in different proportions. No prior knowledge of the pure component spectral patterns is required and the techniques are not dependent on peak shape analysis and the detection of maxima, so that, *e.g.*, filter changes can be made during spectrum recording, as in the present experiments, or the signal amplification can be altered between scans.

Extraction of the pure component spectra from the raw data matrix

Calculation of the totally unknown component spectra is difficult and may be

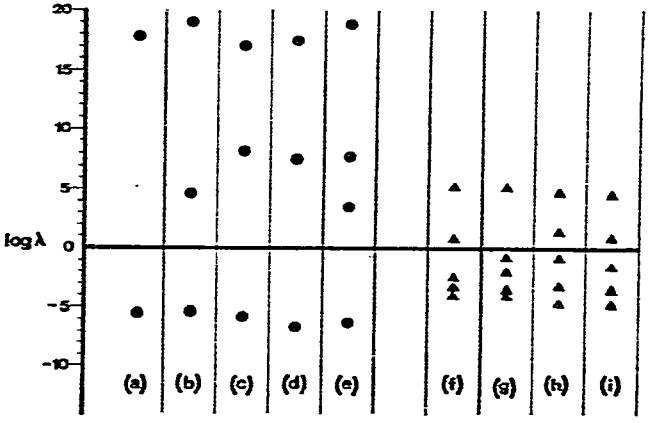

Fig. 4. Plots of logarithms of eigenvalues (λ) of simulated, (a)–(e), and experimental, (f)–(i), HPLC–UV data matrices: (a) N-acetyl-*p*-aminophenol; (b) N-acetyl-*p*-aminophenol + phenacetin; (c) phenacetin + caffeine; (d) N-acetyl-*p*-aminophenol + caffeine; (e) all three compounds; (f) *m*-cresol; (g) *m*-cresol less solvent spectrum; (h) *m*-cresol + *p*-cresol; (i) *m*-cresol + *p*-cresol less solvent spectrum.

impossible. The following investigation illustrates a possible approach to this problem. The data matrix of overlapping spectra, M (the only experimental information), is the product of the matrices D, the amounts and retention data of the components within the peak, and X, the component spectral patterns. When nothing is known about D or X, an infinite number of possible solutions for X exist. However, factor analysis, as described above, provides the number of spectra in X so that the dimensions of D and X become known. The further pattern separation of the component spectra has been studied by Ainsworth[5] and Lawton and Sylvestre[11] for binary absorption spectra, and by Meyerson[24] for binary mass spectra. The latter technique has been extended to n components by Reed and co-workers[13,25], resulting in the UNRAVL programme[26]. Such techniques depend on the detection of wavelengths or m/e values to which individual component spectra do not contribute or contribute uniquely. Such situations are more commonly encountered with mass spectra than with absorption spectra. The application of part of the UNRAVL programme will be illustrated by a simple example of simulated GC–MS data, constructed as described under Experimental.

Examination of the raw data matrix

The 18 (spectra) × 36 (masses) data matrix was first analysed by factor analysis using the methods described above, and three independent components were unambiguously determined. Thus, the dimensions of the unknown D and X matrices are known, *i.e.*, as in eqn. 3. The object was then to calculate X.

The next step was to subject the data matrix to a correlation analysis where the rows (spectra) were correlated giving an 18 × 18 correlation matrix (i), and the columns were correlated giving a 36 × 36 matrix (ii), containing the correlation coefficients of the 36 mass profiles. Such a procedure has been shown to provide useful information about the fractionation taking place across the chromatographic peak itself[19], matrix (i) and the relationships between the components present within the peak[25], matrix (ii). Selected columns of the correlation matrix (ii) for the present data matrix are shown in Table IV.

The columns contain the correlation coefficients of the profiles of the ions of m/e 207, 86 and 62, respectively, with all other ions. It is apparent from the first column that m/e 207 correlates highly with masses 191, 177, 147, 133, 96, 83 and 73, as they are non-overlapping background peaks. By examination of the whole correlation matrix in this way, it was possible to extract three clusters of highly correlating mass profiles, indicating uniqueness to each of the three unknown components. Because these groups of peaks are unique to each of the components, the corresponding columns (masses) of the matrices M and the unknown X are also unique to each component. Once the uni-component peaks for each component have been determined, the procedure is basically the same as previously described[13,25]. In this instance, simultaneous equations are formed from relevant parts of eqn. 4:

$$Q_{3\times3} \cdot M'_{3\times36} = X_{3\times36} \qquad (4)$$

Q is an unknown matrix (to be determined) which operates on a matrix M' to yield the required component spectra, X. M' is formed by selecting three (as there are three

TABLE IV

SOME SELECTED COLUMNS OF THE MASS PROFILE CORRELATION MATRIX OF A GC–MS DATA SET

Correlations of the selected masses with all others are listed. Clusters are given in italics.

m/e	*207*	*86*	*62*
207	*1.0000*	−0.5714	−0.4903
191	*1.0000*	−0.5714	−0.4903
177	*1.0000*	−0.5714	−0.4903
147	*1.0000*	−0.5714	−0.4903
133	*1.0000*	−0.5714	−0.4903
96	*1.0000*	−0.5714	−0.4903
87	−0.5714	*1.0000*	0.9919
86	−0.5714	*1.0000*	0.9919
85	−0.5714	*1.0000*	0.9919
84	−0.5714	*1.0000*	0.9919
83	*1.0000*	−0.5714	−0.4903
73	*1.0000*	−0.5714	−0.4903
72	−0.5714	*1.0000*	0.9919
71	−0.5714	*1.0000*	0.9920
70	−0.5714	*1.0000*	0.9919
69	−0.5714	*1.0000*	0.9919
65	−0.4903	0.9919	*1.0000*
64	−0.4903	0.9919	*1.0000*
63	−0.4903	0.9919	*1.0000*
62	−0.4903	0.9919	*1.0000*
61	−0.4903	0.9919	*1.0000*
60	−0.4903	0.9919	*1.0000*
59	−0.3871	0.9742	0.9934
58	−0.5714	*1.0000*	0.9919
57	−0.5713	*1.0000*	0.9920
56	−0.5714	*1.0000*	0.9919
55	−0.5707	*1.0000*	0.9920
54	−0.5714	*1.0000*	0.9919
53	−0.5714	*1.0000*	0.9919
50	−0.4903	0.9919	*1.0000*
49	−0.4903	0.9919	*1.0000*
48	−0.4903	0.9919	*1.0000*
47	−0.4903	0.9919	*1.0000*
44	−0.5649	0.9999	0.9931
43	−0.5713	*1.0000*	0.9920
42	−0.5714	*1.0000*	0.9919

components) representative spectra of mixtures from M, the complete data matrix. Eqn. 4 may be written more fully:

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} & M_{13} \ldots . & M_{1m} \\ M_{21} & M_{22} \ldots \ldots \ldots & M_{2m} \\ M_{31} & M_{32} \ldots \ldots \ldots & M_{3m} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & X_{13} \ldots . & X_{1m} \\ X_{21} & X_{22} \ldots \ldots \ldots & X_{2m} \\ X_{31} & X_{32} \ldots \ldots \ldots & X_{3m} \end{bmatrix} \quad (5)$$

An example of a possible simultaneous equation in this instance is eqn. 6:

$$\alpha_1 M_{11} + \alpha_2 M_{21} + \alpha_3 M_{31} = X_{11} \quad (6)$$

Three such equations as eqn. 6 would be necessary to determine a_1, a_2 and a_3, and hence the matrix X, via eqn. 5. If, in eqns. 5 and 6, X_{11} and X_{21} are zero because the first column of X is unique to another component, *e.g.*, where X_{31} is non-zero, *i.e.*, is a unique peak, then eqn. 6 becomes

$$a_1 \; M_{11} + a_2 \; M_{21} + a_3 \; M_{31} = 0 \tag{7}$$

or

$$a_1 \; M_{11} + a_2 \; M_{21} = M_{31} \tag{8}$$

TABLE V

MATRIX OF SELECTED MULTI-COMPONENT MASS SPECTRA IN GC–MS (M'). THE ACTUAL PURE COMPONENT SPECTRA (X) AND THE SPECTRA CALCULATED BY THE UNRAVL PROGRAMME

Spectra X and calculated, are given in normalized form.

m/e	M'			X			*Calculated spectra*		
207	101.0	103.0	106.0	100.0	0.0	0.0	100.0	0.0	0.0
191	10.1	10.3	10.6	10.0	0.0	0.0	10.0	0.0	0.0
177	20.2	20.6	21.2	20.0	0.0	0.0	20.0	0.0	0.0
147	10.1	10.3	10.6	10.0	0.0	0.0	10.0	0.0	0.0
133	12.1	12.4	12.7	12.0	0.0	0.0	12.0	0.0	0.0
96	11.1	11.3	11.7	11.0	0.0	0.0	11.0	0.0	0.0
87	19.0	10.6	2.4	0.0	1.0	0.0	0.0	1.0	0.0
86	294.5	164.3	37.2	0.0	15.5	0.0	0.0	15.5	0.0
85	7.6	4.2	1.0	0.0	0.4	0.0	0.0	0.4	0.0
84	1.9	1.1	0.2	0 0	0.1	0.0	0.0	0.1	0.0
83	16.2	16.5	17.0	16.0	0.0	0.0	16.0	0.0	0.0
73	45.5	46.4	47.7	45.0	0.0	0.0	45.0	0.0	0.0
72	5.7	3.2	0.7	0.0	0.3	0.0	0.0	0.3	0.0
71	96.0	54.0	13.1	1.0	5.0	0.0	1.0	5.0	0.0
70	13.3	7.4	1.7	0.0	0.7	0.0	0.0	0.7	0.0
69	5.7	3.2	0.7	0.0	0.3	0.0	0.0	0.3	0.0
65	3.5	2.1	0.4	0.0	0.0	0.7	0.0	0.0	0.7
64	153.5	91.5	15.5	0.0	0.0	31.0	0.0	0.0	31.0
63	4.0	2.4	0.4	0.0	0.0	0.8	0.0	0.0	0.8
62	495.0	295.0	50.0	0.0	0.0	100.0	0.0	0.0	100.0
61	44.1	26.3	4.5	0.0	0.0	8.9	0.0	0.0	8.9
60	5.0	3.0	0.5	0.0	0.0	1.0	0.0	0.0	1.0
59	4.5	4.5	3.3	3.0	0.0	0.3	3.0	0.0	0.3
58	7.6	4.2	1.0	0.0	0.4	0.0	0.0	0.4	0.0
57	1905.1	1065.2	245.3	5.0	100.0	0.0	5.0	100.0	0.0
56	860.7	480.2	108.7	0.0	45.3	0.0	0.0	45.3	0.0
55	127.4	72.0	19.0	2.0	6.6	0.0	2.0	6.6	0.0
54	11.4	6.4	1.4	0.0	0.6	0.0	0.0	0.6	0.0
53	32.3	18.0	4.1	0.0	1.7	0.0	0.0	1.7	0.0
50	4.0	2.4	0.4	0.0	0.0	0.8	0.0	0.0	0.8
49	9.4	5.6	1.0	0.0	0.0	1.9	0.0	0.0	1.9
48	11.4	6.8	1.2	0.0	0.0	2.3	0.0	0.0	2.3
47	26.7	15.9	2.7	0.0	0.0	5.4	0.0	0.0	5.4
44	56.8	33.0	9.4	2.5	2.7	0.6	2.5	2.7	0.6
43	1545.3	863.3	197.5	2.5	81.2	0.0	2.5	81.2	0.0
42	777.1	433.5	98.2	0.0	40.9	0.0	0.0	40.9	0.0

where a_3 may be set equal to -1, as the set of such equations is homogeneous. Now, only two such equations are necessary to determine a_1 and a_2 and therefore the first row of matrix Q and hence the first row of X. If one uni-component peak can be detected for each of the three components, then two equations such as eqn. 8 can immediately be set up for each component and hence Q and X determined directly[13]. This task is performed by the computer programme, and the derived spectra are printed out. In the present experiment, masses 207, 86 and 62 were fed to UNRAVL and the following Q matrix was calculated:

$$Q = \begin{bmatrix} 0.50 & -0.66 & -1.00 \\ -1.23 & 2.24 & -1.00 \\ -0.99 & 2.00 & -1.00 \end{bmatrix} \quad (9)$$

The matrix M' used in this instance is given in Table V together with the actual matrix X and the calculated spectra.

In this simple example, free of experimental error and noise, the calculated spectra are the same as the component spectra hidden in M. The programme has worked in practice where a four-component mixture, fractionated within the mass spectrometer, was analysed[25]. The same principles apply to n components, provided that at least one uni-component peak can be detected per component. Under certain circumstances, this condition may be relaxed[26].

CONCLUSION

Factor analysis shows potential for determining the content and purity of chromatographic–spectrometric systems, provided that the spectra are linearly additive, *e.g.*, HPLC–UV by scanning during the stopped flow.

A further step in such matrix manipulation techniques allows the pattern separation of the pure component spectra to be carried out, provided that certain simple conditions are satisfied.

It is apparent that in the case of real data, experimental errors and instrumental noise will make the application of these techniques more difficult, especially when dealing with very badly resolved peaks. Development of UNRAVL is continuing, including an investigation of the effects of simulated experimental errors, and its adaptation for more efficient handling of real data.

ACKNOWLEDGEMENTS

The author is grateful to Dr. R. I. Reed (University of Glasgow) and Prof. Dr. G. Lehnert (Zentralinstitut für Arbeitsmedizin, Hamburg) for their encouragement, and to Dr. W. Rehpenning (University of Hamburg) for providing some of the computer programmes used. Part of this work was carried out with the financial assistance of the Deutsche Forschungsgemeinschaft (SFB 34) and the Bundesministerium für Forschung und Technologie.

REFERENCES

1 J. C. Sternberg, H. S. Stillo and R. H. Schwendemann, *Anal. Chem.*, 32 (1960) 84.
2 R. M. Wallace, *J. Phys. Chem.*, 64 (1960) 899.
3 G. Weber, *Nature (London)*, 190 (1961) 27.
4 S. Ainsworth, *J. Phys. Chem.*, 65 (1961) 1968.
5 S. Ainsworth, *J. Phys. Chem.*, 67 (1963) 1613.
6 R. M. Wallace and S. M. Katz, *J. Phys. Chem.*, 68 (1964) 3890.
7 D. Katakis, *Anal. Chem.*, 37 (1965) 876.
8 H. Hotelling, *J. Educ. Psychol.*, 24 (1933) 417.
9 J. J. Kankare, *Anal. Chem.*, 42 (1970) 1322.
10 Z. Z. Hugus, Jr., and A. A. El-Awady, *J. Phys. Chem.*, 75 (1971) 2954.
11 W. H. Lawton and E. A. Sylvestre, *Technometrics*, 13 (1971) 617.
12 N. Ohta, *Anal. Chem.*, 45 (1973) 553.
13 L. Fraser Monteiro and R. I. Reed, *Int. J. Mass Spectrom. Ion Phys.*, 2 (1969) 265.
14 J. E. Davis, A. Shepard, N. Stanford and L. B. Rogers, *Anal. Chem.*, 46 (1974) 821.
15 G. L. Ritter, S. R. Lowry, T. L. Isenhour and C. L. Wilkins, *Anal. Chem.*, 48 (1976) 591.
16 J. M. Halket, *Ph.D. Thesis*, University of Glasgow, 1973.
17 J. M. Halket and R. I. Reed, *Org. Mass Spectrom.*, 10 (1975) 808.
18 J. M. Halket, in A. Frigerio and L. Renoz (Editors), *Recent Developments in Chromatography and Electrophoresis*, Elsevier, Amsterdam, Oxford, New York, 1979, p. 327.
19 J. M. Halket, *J. Chromatogr.*, 175 (1979) 229.
20 J. M. Halket, *J. High Resolut. Chromatogr. Chromatogr. Commun.*, 2 (1979) 197.
21 R. B. Selzer and D. G. Howery, *J. Chromatogr.*, 115 (1975) 139.
22 D. Macnaughten, Jr., L. B. Rogers and G. Wernimont, *Anal. Chem.*, 44 (1972) 1421.
23 J. E. Biller and K. Biemann, *Anal. Lett.*, 7 (1974) 515.
24 S. Meyerson, *Anal. Chem.*, 31 (1959) 174.
25 J. M. Halket and R. I. Reed, *Org. Mass Spectrom.*, 10 (1975) 370.
26 J. M. Halket and R. I. Reed, *Org. Mass Spectrom.*, 11 (1976) 881.
27 H. Hein and R. Jöster, *Angew. Chromatogr.*, 28 (1977) 7.
28 E. Stenhagen, S. Abrahamsson and F. W. McLafferty (Editors), *Atlas of Mass Spectral Data*, Vol. I, Interscience, New York, 1969, pp. 43 and 149.
29 M. Spiteller and G. Spiteller, *Massenspektrensammlung von Lösungsmitteln, Verunreinigungen, Säulenbelegmaterialien und einfachen aliphatischen Verbindungen*, Springer, Vienna, 1973.
30 R. H. Pennington, *Introductory Computer Methods and Numerical Analysis*, Macmillan, London, 2nd ed., 1971, pp. 339 and 374.
31 J. L. Simmonds, *J. Opt. Soc. Amer.*, 53 (1963) 968.
32 E. R. Malinowski, *Anal. Chem.*, 49 (1977) 606 and 612.
33 R. Yost, J. Stoveken and W. MacLean, *J. Chromatogr.*, 134 (1977) 73.