SOME STUDIES
ON
A MASS SPECTROMETRIC MIXTURES PROBLEM

A thesis presented for the degree of

DOCTOR OF PHILOSOPHY

to

The University of Glasgow

by

JOHN McKNIGHT HALKET

August, 1973

To My Parents

CONTENTS

Acknowledgements

The author is indebted to his supervisor, Dr. R.I. Reed, for his unceasing interest and encouragement during the course of this work.

He also wishes to express his gratitude to Mr. A. Maclean for technical assistance; to Mr. E. Kendrick, Dr. M.J. Price, Dr. J.P. Nice, Mr. S.F. Nöel and Dr. R.T.W. Hall of Esso Research Centre, Abingdon, Berks. for providing hospitality and many discussions during three months of industrial experience; to several members of Glasgow University Computing Service for assistance with "de-bugging" many of the sub-routines; to Dr. M.V. Spinnler for a translation of the original mixtures programme; to Mr. J. Hardy and various members of the Chemistry Department Electronics Workshop for technical advice and assistance; to Mr. N.S. Miller of the Department of Naval Architecture for permission to use the d-mac Digitizer; to Dr. J. Haselgrove (Computing Science) and to his laboratory colleagues, Miss M.E.S.F. Silva, Miss E. Jardim and Mr. I. Jardine, for helpful discussions.

His gratitude is also due to the Science Research Council and Esso Petroleum Co. for the provision of a Co-operative Award in Pure Science.

CHAPTER ONE

General Introduction

GENERAL INTRODUCTION

Mass spectrometry has grown within the last fifty years from a relatively simple physical phenomenon to a technique of major importance in analytical chemistry. This rapid progress has been made possible by the invention of new methods and applications together with ready availability of increasingly sophisticated commercial instruments. The analytical requirements of the petroleum industry were responsible for many of these developments, particularly during the Second World War(1). Commercial instruments at that time employed the electron impact method of ionization, normally at 70 electron volt energies. Due to inherent advantages with respect to qualitative and, more importantly, quantitative analysis this method is still the most prevalently used. More will be said about the advantages later in this Chapter. All the mass spectra subsequently referred to will be mass spectra produced by this means, unless stated otherwise. The needs of the petroleum industry, particularly quantitative analysis of mixtures, also favoured the utilisation of computer methods.

A short description will now be given of such analyses together with some other applications involving computer methods, in order to place the present study in perspective.

1. Mixtures Analyses

The large numbers of tedious calculations involved in mixtures analysis provided the impetus for some of the first routine applications of computing methods in mass spectrometry. Such analyses are basically quantitative in that the spectra of the components in the mixture be known. Some requirements for satisfactory analytical work are discussed by Barnard(2), the major ones being:

a) In the pressure ranges normally employed the mass spectra of mixtures are linear superpositions of the individual mass spectra of the components.

b) The fragmentation patterns of the components should not vary significantly during the analysis and nor should their sensitivity coefficients (an explanation of the latter follows).

The above conditions are generally satisfied where:

1. 70 eV. electron impact energies are used. Around this figure the ionization efficiency is nearly independent of the energy(3).
2. Ion-source pressures of less than 10^{-5} Torr are employed in order to minimise ion-molecule collisions which perturb the fragmentation patterns.
3. Ion-source temperatures are kept constant.
4. Instrumental electronics are well adjusted.

There are two major methods of mixtures analysis:

A. Using Pressure Measurements.

B. Using a Calibration Mixture.

A. requires that the sensitivity coefficients, S_{ij}, be obtained for each component, "j", at each of the mass values, "i", used in the analysis. The sensitivity coefficients so defined have dimensions of peak height (or, more accurately, area) per unit pressure, where the partial pressure of each component is required. The values of S_{ij} are obtained by measuring the pressures of individual samples of the components when their mass spectra are being obtained. At least "n" peaks are considered where "n" is the number of components. The product of the sensitivity coefficient, S_{ij}, and the partial pressure, p_j, is the contribution of component, "j", to the height, H_i, of the mass peak, "i", in the spectrum of the mixture. A set of "n", or more, simultaneous linear equations is set up and solved for p_j.

$$\begin{array}{l} S_{11}\cdot p_1 + S_{12}\cdot p_2 + \ldots\ldots + S_{1n}\cdot p_n = H_1 \\ S_{21}\cdot p_1 + S_{22}\cdot p_2 + \ldots\ldots + S_{2n}\cdot p_n = H_2 \\ \vdots \\ S_{n1}\cdot p_1 + S_{n2}\cdot p_2 + \ldots\ldots + S_{nn}\cdot p_n = H_n \end{array} \quad \ldots (1).$$

Some excellent discussions of these quantitative methods, which have remained largely unchanged for twenty-five years, are given by Barnard(4) and the notation used in equations (1) is as used by Reed(5). Matrix notation is useful in representing this system of equations more compactly:

$$Sp = h \qquad \ldots . (2),$$

where the required partial pressures are:

$$p = S^{-1}h$$

The method involves the inversion of matrix S and requires the following information:

(a) Identities of the components.

(b) Mass spectra of the components.

(c) Mass spectrum of the unknown mixture.

(d) Pressure of each pure component during recording of its spectrum to permit calculation of S. *i.e.* the method requires pressure measurements.

Some considerations in solving such systems of equations are given by Barnard and Fox(6).

B.

In order to avoid making pressure measurements, which may be difficult in certain cases, alternative methods have been devised. Johnsen$^{(7)}$ used a two-component calibration mixture, *i.e.* of known composition, to obtain a mass spectrum. The ratio of known uni-component peaks in this spectrum was used to determine the composition of any other mixture of the same two components. The general solution of this standard mixture technique was recently published by Ruth$^{(8)}$. Information (a), (b) and (c), above, is still necessary; (d) is satisfied by obtaining the mass spectrum of a calibration mixture of all the components where their mole fractions are accurately known. The accuracy of the method does not therefore depend on the accuracy of pressure measurements but on the accuracy with which the calibration mixture is made up. Since many situations arise where no pressure measurements are possible, only application of this technique seems capable of solving the problem *e.g.* as when a sample is introduced directly into the ion source by means of a probe.

Equations (1) were written with a view to deriving the partial pressures of the components. The sensitivities for each component differed according to the mass number used. If the equations are rewritten incorporating the mass spectra of each component, given as a percentage of the base-peak (largest peak), then a sensitivity coefficient for each component may be defined. The latter gives the actual abundance of the base-peak in so many units per unit pressure:

$$a_i . S_A . p_A + b_i . S_B . p_B + \ldots\ldots + n_i . S_N . p_N = H_i \qquad \ldots (4)$$

for each mass number, "i".

p_A is the partial pressure of component A, *etc.*

a_i is the abundance at mass "i" in the spectrum of component A expressed as a percentage of the base-peak.

Since these new sensitivities are constant for each component, it is usual to combine the factors $S_j.p_j$ giving X_j, referred to as <u>composites</u>. The equations (4) may be re-written with composites:

$$X_A.a_i + X_B.b_i + \ldots\ldots + X_N.n_i = H_i \qquad \ldots (5)$$

and in matrix notation:

$$xA = h \qquad \ldots (6)$$

The solution, "x" of this equation may be achieved in practice by employing a number of important considerations and time-saving procedures. Some methods are discussed by Barnard(9), Barnard and Fox(6) and Kiser(10). Many factors such as random changes of sensitivity, interference effects, preferential adsorption and desorption, and fractionation can cause disparities in the results obtained. These effects are discussed by Barnard(2) and also in Chapter 3 of this thesis. Consistent results are normally obtained where the mass spectrometer is used regularly. In a properly adjusted and maintained system, calibrations need be performed only every month or so(11).

Such methods of analysis can only be used to determine mixtures containing no extra components although, by examination of residues, the unexpected presence of the latter may be detected.

Related methods utilizing the molecular weight distribution in the mixtures spectrum together with inverse matrices have been used to give so-called "type-analyses" <u>i.e.</u> the amounts of various types of compounds present(12) <u>e.g.</u> aromatics, branched-chain hydrocarbons or sulphur compounds. In one recent case(13) the composition is accounted for in terms of four saturated hydrocarbon types, twelve aromatic hydrocarbon types, three thiopheno types and six unidentified aromatic groups.

High resolution mass spectrometry(14) has contributed significantly to the qualitative aspects of mixtures analysis by recognition of single compounds(15).

Both high resolution and low electron energy has proved to be a powerful combination for qualitative analyses of very complex mixtures e.g. crude oil fractions(16).

2. Gas-Liquid Chromatography/Mass Spectrometry

Some of the potential of electronic computers in mass spectrometry is being shown by large computerized GC-MS systems. Excellent descriptions of some of the major developments are given in Waller(17) and in a recent review by Burlingame and Johanson(18).

3. Spectrum Recognition

i.e. recognizing a mass spectrum as being characteristic of some chemical compound is almost impossible without computer aid. The major difficulties are the large numbers of chemical compounds and the fact that characteristic spectra can vary considerably according to the type of mass spectrometer used. Several methods of approaching the problem have been devised involving comparison of the test spectrum with spectra held within the computer. Use of only the six strongest peaks together with internal computer techniques have greatly reduced the time necessary to effect comparisons(19). Information theory has been employed to compare spectra using the eight largest peaks(20). A review is given by Ridley(21).

Recent work has considered the possibility of identification using binary-coded spectra (i.e. only the positions of masses, presence or absence of peaks) and optional weighting of certain masses with surprisingly good results(22,23). These studies will be shown to have considerable influence on the present work (Chapter 5).

For some analytical work a complete identification is not really necessary, only a determination of the types of molecules present being required. An approach to this problem was given above - "type analysis". A new technique considering the recognition of features peculiar to classes of compounds is given by Crawford and Morrison(24) and extended by Smith(25) in a form suitable for use by a small computer. Artificial intelligence methods (see below) are being applied to the problem with encouraging results(26). The computer programme learns to recognize features in patterns, based on its past experience.

4. Spectrum Interpretation

Recently adapted artificial intelligence techniques have been applied to mass spectra in order to work out the structure or possible structures of the parent compound. The HEURISTIC DENDRAL programme(27) generates possible structures and proceeds to eliminate them on the basis of mass spectral and other information. A simpler method by Crawford and Morrison(28) termed "ab-initio" is less systematic and is at an early stage of development. A strategy is employed which is similar to that of the mass spectroscopist and in the same way it is not always successful! It is potentially more suitable for dealing with large structures than the current DENDRAL methods.

5. Learning Machines

The computer, when used as a learning machine, has been able to predict the molecular formula of a compound given its low resolution mass spectrum(29). The programme is provided with large numbers of low resolution spectra with the corresponding molecular formulae. In this way it can be trained to recognize the latter given a spectrum, the degree of success being related to the degree of training.

6.

Calculation of the most likely molecular weight has been an interesting application of computer methods(30). Given a low resolution spectrum lacking a parent ion the programme calculates the most likely value of the latter based upon the fragments provided.

7.

An important biological application is in the field of protein sequencing(31).

8.

The future of computing is mass spectrometry is assured particularly with the utilization of such systems as described by Waller(17), Burlingame(32), Biemann(33) and Henneberg(34). Needless to say, an important part of the development of mass spectrometry is assured by computer applications. The latter will become of increasing use in the analysis of mass spectra obtained by other means e.g. ion-cyclotron resonance(35), chemical ionization(36), field ionization(37) and, no doubt, a host of other useful techniques, as yet undiscovered.

9. Mixtures Analysis without Prior Knowledge of Components

All previously described methods of mixtures analysis have required either that the components be known or that the computer be programmed with information concerning likely components. The subject of the present study is that of mixtures analysis without prior knowledge of the numbers and types of compounds present. This may be regarded as an extension of the conventional methods outlined in 1. above, and would have been studied in depth some time ago had not GLC methods(38) arrived on the scene. Although such methods are now very efficient some limitations still exist(39-41) e.g. organometallic petroleum additives. Hence some revival of interest is being shown in the fractionating properties of the mass spectrometer itself, particularly using the direct insertion probe(42).

Nearly fifteen years ago Meyerson(43) derived for the first time the spectra of the components of binary mixtures, without prior total separation or knowledge of their identities. Two different mixtures of the compounds are necessary. At least one peak in each pure component spectrum (unknown) must be unique to that spectrum. Such unique peaks are detected by listing peak ratios at each mass in the two spectra and mass numbers chosen where the ratios are highest and lowest; peaks at these masses contain the least contribution from the other component. The ratio of the abundances of a peak unique to one component is found and all the peaks in one mixtures spectrum multiplied by it. Subtraction of the product from the other spectrum yields the mass spectrum of one component. The other may be obtained using the other ratio or by difference. Meyerson succeeded in identifying an unsuspected impurity by this method.

McCormick(44) has used the same technique to separate and identify a mammein homologue of molecular weight 358, present as an impurity, in studies of the compounds derived from Mammeia Americana(45). Fractionation was performed by thin-layer chromatography. The spectrum of a disulphide impurity present in a sulphoxide sample was separated by Laurie(46) using the same method. Fractionation in this case was performed by adjustment of the direct insertion probe to give different spectra at a constant monitor current reading(47).

It was believed, by Monteiro(48), that behind the empirical approach of Meyerson a more general theory might be hidden that could be made to yield the mass spectra of the components of a mixture starting from several different mixtures. The development of such a theory has been made by Monteiro and Reed(49) and this work forms the background of the present study. The method requires that

(a) there be more different mixtures spectra than there are components (at least one more)

and (b) each unknown component spectrum has at least two peaks which are unique <u>i.e.</u> uni-component.

In a few suitable cases condition (b) could be reduced to only one unique peak per component. Effusiometric techniques, which are experimentally exacting, would have to be employed. Further details are given in Chapter 6.

A reading of this work(49) by Monteiro and Reed is essential for a proper understanding of what follows although additional explanations have been attempted at each stage.

The basis of the analysis begins with an extension of equation (6) to include a number of different mixtures of the same components <u>i.e.</u>

<u>one mixture</u> $xA = h$ . . . (6)

<u>several mixtures</u> $XA = M$ . . . (7)

The latter equation may be written out in full for L mixtures of N components A, B, . . . N, the highest individual masses being mwa, mwb, etc. and the highest mass being mw.

$$\underset{\text{COMPOSITIONS}}{\begin{bmatrix} x_{AI} & x_{BI} & \cdots & x_{NI} \\ x_{AII} & x_{BII} & \cdots & x_{NII} \\ \vdots & & & \vdots \\ x_{AL} & x_{BL} & \cdots & x_{NL} \end{bmatrix}} \underset{\substack{\text{COMPONENTS} \\ \text{(rows)}}}{\begin{bmatrix} a_1 & a_2 & \cdots & a_{mwa} \\ b_1 & b_2 & \cdots & b_{mwb} \\ \vdots & & & \vdots \\ n_1 & n_2 & \cdots & n_{mwn} \end{bmatrix}} = \underset{\substack{\text{MIXTURES} \\ \text{(rows)}}}{\begin{bmatrix} h_{1I} & h_{2I} & \cdots & h_{mwI} \\ h_{1II} & h_{2II} & \cdots & h_{mwII} \\ \vdots & & & \vdots \\ h_{1L} & h_{2L} & \cdots & h_{mwL} \end{bmatrix}}$$

The elements of the matrices have the same meanings as before. Spectra are represented by rows and this convention is adopted in all that which follows, including computer manipulations. The columns may be referred to as Mass-Profiles.

A computer programme was used by Monteiro(48) to analyse numerically exact mixtures, calculated in such a way that they satisfied the necessary conditions. The programme was stated to be at an intermediate stage of development, incapable of application to all experimental data since a statistical treatment was not included. This leads to a statement of the current problem and its associated study.

The Problem

Examination of the afore-mentioned computer programme with a view to its experimental application and the development and improvement of the methods used, in order to widen its scope.

Practical Interest

It is probably a fact that most of the samples run in an industrial laboratory are mixtures, owing to the fact that fine separation procedures (*e.g.* preparative GLC techniques or liquid chromatography(50)) can be time-consuming and in many cases unnecessary. A rapid separating technique, e.g. column chromatography with FLORISIL, may be conveniently employed to obtain simple mixtures of up to, say, ten components. Such mixtures may contain high molecular weight oil additives, perhaps with an organometallic content. Examination of these materials is conveniently done by allowing the mixture to distil from the direct insertion probe i.e. utilizing the fractionating properties of the mass spectrometer. The initial spectra consist of the volatile constituents, more heat/time being required to observe those which are less volatile and generally more interesting. In many cases it is possible, by examination of successive spectra, to pick out spectral features from a number of the components. Part of the present work might then be considered with such a technique in mind.

The mixtures spectra referred to will be spectra obtained by fractionation within the mass spectrometer itself, either the gas-inlet system or direct insertion probe, although much of what follows applies to spectra obtained in other ways. e.g. running mixtures obtained by thin-layer chromatography.

The next Chapter includes some explanatory notes and initial attempts at applying the afore-mentioned programme to experimental data.

The method will henceforth be referred to as the "ab-initio" mixtures analysis.

CHAPTER TWO

I - Some Explanations

II - The Algol Computer Programme

I - SOME EXPLANATIONS

It may help to clarify the formation of a mixtures array by considering a simple example e.g. the formation of seven different mixtures of four components each. This is done by mixing four mass spectral vectors, $(A_1\ A_2 \ldots A_8)$, $(B_1\ B_2 \ldots B_8)$ etc., together; each vector has eight elements (peaks).

The mixtures array, M, is formed by performing a simple matrix multiplication:

$$\underset{X}{\begin{bmatrix} X_{a1} & X_{b1} & X_{c1} & X_{d1} \\ X_{a2} & X_{b2} & X_{c2} & X_{d2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ X_{a7} & X_{b7} & X_{c7} & X_{d7} \end{bmatrix}} \underset{A}{\begin{bmatrix} A_1 & A_2 & \ldots & A_8 \\ B_1 & B_2 & \ldots & B_8 \\ C_1 & C_2 & \ldots & C_8 \\ D_1 & D_2 & \ldots & D_8 \end{bmatrix}} = \underset{M}{\begin{bmatrix} M_{11} & M_{12} & \ldots & M_{18} \\ M_{21} & M_{22} & \ldots & M_{28} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ M_{71} & M_{72} & \ldots & M_{78} \end{bmatrix}} \qquad \ldots (8)$$

The elements of X have the same meanings as before.

Element M_{11} consists of

$$X_{a1}.A_1 + X_{b1}.B_1 + X_{c1}.C_1 + X_{d1}.D_1$$

and the element M_{12} is

$$X_{a1}.A_2 + X_{b1}.B_2 + X_{c1}.C_2 + X_{d1}.D_2$$

etc. so that the first mixtures spectrum is formed by mixing the four spectra of matrix A together in relative amounts given by the first row of matrix X. The second mixture, $(M_{21}\ M_{22} \ldots M_{28})$, is formed by mixing according to the second row of X, and so on. The four spectra will then be mixed up linearly within all seven mixtures spectra. If the rows of X are all linearly independent(51) then the rows of M will be different but only four of them will be linearly independent. This means that at least four mixtures spectra would be necessary to construct one of the others through multiplication by suitable factors and addition of results (Jacobi operations)(49,51). It is assumed, of course, that the rows of matrix A are linearly independent, like mass spectra. The rank (49,51,52) of the matrix, M, is then four and this may be determined experimentally by well known methods(53,54).

In the previous chapter equation 7 was used to represent the mixtures array formation:

$$XA = M$$

Since it is desired to derive both A and X from M, it is obvious that M must satisfy certain conditions(49). These have already been mentioned in Chapter 1, but will be repeated here in greater detail:

(1) at least one more row than there are components, where the rows are distinct.

(2) at least two uni-component peaks for each component. In certain circumstances this can be reduced to one per component (see Chapter 6)

(3) for a quantitative analysis the pressure of each mixture is required or must be the same in all cases.

The basis of M satisfying certain conditions must come from the properties of A <u>i.e.</u> the unknown spectra themselves, as X is of a general nature; the fractionation or mixing cannot be expected to satisfy any conditions apart from being random(49). By examination of M the properties of A must reveal themselves. The difficulty arises because an infinite number of solutions exists where no conditions are imposed. A mathematical representation of such a situation may be given as follows:

$$M = XA \qquad \ldots (7)$$

This equation, where A is the desired solution, is equivalent to

$$M = X\ B\ B^{-1}\ A$$

where B is any matrix (of correct dimensions)

<u>i.e.</u> $M = X^1\ A^1$

and A^1 is another solution. Also

$$M = X^{11}\ A^{11} \qquad \text{etc.}$$

It might be helpful at this stage to touch upon the reason why condition (2), above, is necessary. <u>i.e.</u> having to know which peaks are unique to each component. (see Chapter 5 for more details). If X can be fixed in some way then A can be determined uniquely:

$$A = X^{-1}\ M$$

A convenient way of fixing X is by considering the behaviour of a peak unique to each component. These are most readily located if there are at least two(49). The size of a unique peak in each mixture is directly proportional to the amount of component present. Such a proportionality constant is related to the peak sensitivity coefficient and is unknown as the pattern and instrumental sensitivity are unknown. If the unique peak heights U_{a1}, for component A in mixture 1, and U_{a2} etc. multiplied by the unknown constant for the first component, A, are substituted in the first column of X in equation (8), and the same is done for the other components a matrix X_{ku} obtains which is equivalent to X

$$\begin{bmatrix} k_1.U_{a1} & k_2.U_{b1} & k_3.U_{c1} & k_4.U_{d1} \\ k_1.U_{a2} & k_2.U_{b2} & k_3.U_{c2} & k_4.U_{d2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ k_1.U_{a7} & k_2.U_{b7} & k_3.U_{c7} & k_4.U_{d7} \end{bmatrix} \qquad \underline{\text{i.e.}}\ X_{ku}$$

X_{ku} is equivalent to the product:

$$\begin{bmatrix} U_{a1} & U_{b1} & U_{c1} & U_{d1} \\ U_{a2} & U_{b2} & U_{c2} & U_{d2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ U_{a7} & U_{b7} & U_{c7} & U_{d7} \end{bmatrix} \begin{bmatrix} k_1 & & & \\ & k_2 & & 0 \\ & 0 & k_3 & \\ & & & k_4 \end{bmatrix}$$

$$\underline{\text{i.e.}}\ X_u K$$

where X_u is known and K is a diagonal matrix of constants.

On substituting into equation 7

$$M = X_u KA$$

is produced. Only normalised spectra, given by KA, may be obtained in this way:

$$X_u^{-1} M = KA$$

<u>i.e.</u> by inversion of X_u, the matrix of unique peak heights.

A similar result was derived by Monteiro and Reed$^{(49)}$ who gave the equation

$$QM = KA$$

where Q is a matrix of unknowns which on multiplication into M gives the normalised component spectra. Q may be found by equating its unknown elements to known zeros in KA (found by knowing the positions of unique peaks, for example). The latter method is to be preferred because it is more amenable to solution by a least squares method i.e. statistically advantageous. (see Chapter 5 - IV).

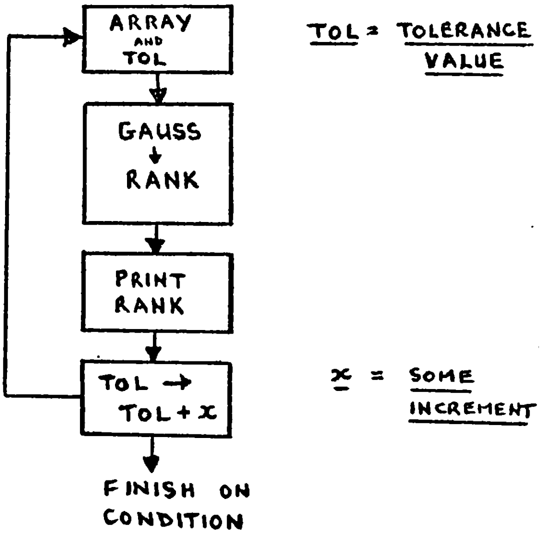
<u>II THE ALGOL COMPUTER PROGRAMME</u> was an attempt to put into practice the principles laid down in the afore-mentioned paper(49). It was stated to be at an intermediate stage of development, incapable of application to experimental data because it lacked a statistical treatment. Monteiro(48) used this programme to solve numerically exact mixtures arrays satisfying the necessary conditions; individual component spectra were obtained. The programming language used was Whetstone Algol(55) punched on paper-tape for use with the English Electric KDF9 computer.

The first stage in the examination of the programme involved processing an experimental array obtained by fractionating a four-component mixture in the gas-inlet system of the A.E.I. MS9 mass spectrometer(56). Experimental details are given in Section II of the next Chapter. Measurements obtained from ultra-violet charts(56) were checked and punched on paper-tape in the required format together with certain arbitrary parameters as outlined in(48). The programme failed to produce the desired result.

Investigation of the failures began with a detailed examination of the programme in order to understand what it was doing. Various sections were found to be redundant, having been included for testing purposes. Other sections which were found to function independently in order were dissected out giving a number of sub-programmes capable of independent testing.

The section which read in data was simplified and a background subtraction sub-programme discarded as being unnecessary for development purposes. In later work it was never re-introduced since considerable variation in background was experienced during experiments.

The first sub-programme to be tested was that used to determine the rank of the array. Gaussian elimination(53,54) was employed and some explanation was given(57). The programme required three parameters, referred to as tolerances, to allow for round-off (see Chapter 4) i.e. in order to decide when a transformed element could be zero. It was decided to investigate the effect of these parameters (which are roughly related) on the rank obtained as this could vary widely. Two of the values were fixed at what was thought to be reasonable levels and the third varied by writing a loop into the programme. This is illustrated by the block diagram:



It was found that the rank decreased as the TOL value was increased and so by varying the latter any reasonable rank could be obtained.

The loop proved to be very slow, taking up to one hour to cover a reasonable range of values.

In view of the above and other experiences a strong case was made for changing the computer system used. Some disadvantages of the Algol system are summarized below:

1. The particular system in use was old and suitable advisory services almost non-existent.
2. Both programme and data were necessarily on punched tapes which proved to be cumbersome, particularly in view of their size. Large amounts of time were required for corrections to be made.
3. Turn-round time was very long as the system involved over-night running only. Runs could easily take several days, or more, if hampered by tape errors or machine break-downs.
4. Actual running times were long compared with other available systems.

In view of possible alternatives two paths were followed simultaneously:

Short Term: Conversion of the programme to a similar ALGOL language on punched cards.

Longer Term: A programmer with experience in both ALGOL and FORTRAN IV languages was charged with converting the original programme. This work is duly acknowledged and was undertaken considering current departmental developments.

The initial delay in real development was considered to be worthwhile and this was eventually shown to be the case.

Some advantages of the changes are:

1. Punched cards allow changes to be made, in both programme and data, reliably in a much shorter time.

2. Development time was reduced by much improved turn-round of work.
3. The IBM 370/155 system on FORTRAN IV allows instant access to valuable scientific sub-routines and statistical packages.
4. Advisory services were much more efficient.

The initial conversion to FORTRAN IV took several months. The conversion to punched-card ALGOL also took several months as the languages were not identical and many problems were encountered.

Eventually FORTRAN IV (IBM 370/155) was used for all programming and the ALGOL system (KDF9) abandoned, except for one case (see APPENDIX A).

The investigation of the rank-determining programme, necessary for the subsequent analysis, will be outlined in Chapter 4, together with details of subsequent improvements.

The next Chapter deals with some factors to be considered when obtaining a suitable mixtures array in the laboratory.

CHAPTER THREE

I - Some Factors relating to Spectra Collection

II - Experimental - Collection of Spectra

III - Spectrum Measurement

IV - Pre-treatment of Data

I Some Factors Relating to Spectra Collection

$$M = XA$$

As discussed in Chapter 2, M is an array of mixtures spectra constructed from pure component spectra A, in relative amounts given by X. It is required to determine both A and X. In "ab-initio" analysis this information must be derived from an examination of array M which is the only experimentally known quantity. Some of the properties of such an array will now be discussed together with some properties to facilitate its analysis. These latter properties will have a strong bearing on experimental method.

Where M contains real spectra (as opposed to calculated mixtures) each element (peak) will deviate from its average value by an amount depending on a number of factors. Assuming that the component spectra are linearly additive and response linear with respect to partial pressure (see Chapter 1, section I) then some of the more important factors are:

(a) scanning speed
(b) amplification level (related to sample pressure)
(c) errors of measurement (human or digitiser)
(d) source temperature changes
(e) tuning of instrumental electronics
(f) impurities and adsorption/desorption
(g) arcing - related to condition of instrument
(h) peak height/area approximation

The problem of interference has been largely overcome in modern instruments(58,48). Limitations in the superposition of mass spectra are discussed by Careri(59).

Each of the above-mentioned factors will be discussed later in more detail. Fluctuations caused by these effects tend to swamp the interesting properties of array M making their detection difficult. The minimization of the fluctuations is all the more important because the subsequent use of numerical methods superimposes round-off errors. The latter are discussed in Chapter 4.

The important point is simply that bad data cannot be compensated for by computer methods i.e. the accuracy of the calculated A matrix can be no better than the M matrix from which it derives. This problem is discussed by Barnard(6) in connection with conventional mixtures computations.

A statistically well-conditioned mixtures array might be defined as conforming to the following rules:

1. The percentage error in any element should be a minimum. Large fluctuations will tend to decrease the important differences between mixtures spectra. (see 3. below).
2. Each mixtures spectrum should be made in such a way that a minimum degree of fractionation or pressure change can occur during its scan. If such conditions are not satisfied the essential linear nature of the mixtures spectra will be destroyed i.e. the recorded spectra will not consist of fixed mass spectra superimposed linearly. A simple method of checking this effect will be given.
3. It is necessary to have at least one more distinct mixtures spectrum than there are components(49).

Where the spectra are experimental they must be statistically distinct. A simple method of investigating this condition will be described.

Some of the afore-mentioned factors causing fluctuations and errors will now be discussed in more detail.

(a) Scanning Speed

This must be adjusted to deal with two situations:

(a) 1. For a fast scanning speed the fluctuation size increases, particularly for the smaller peaks where fewer ions are being collected(40). Depending on the type of instrument used the fluctuations can usually be held to within 5% at reasonably fast speeds(60). In compensation for possible ill-effects of fast scanning, more spectra are obtained. These can be averaged by a small on-line computer system(61) producing a corresponding increase in accuracy.

(a) 2. If scanning speed is too slow to capture the instantaneous condition of the changing system a mixtures spectrum is obtained which is unsuitable for inclusion in M (see condition 2. above)

Scanning speed should therefore be balanced with the rate of fractionation of the mixture, normally carried out in the inlet-system or ion-source of the mass spectrometer. The correct speed to use will depend on a number of factors including sample quantity, type and the temperature and nature of the inlet-system or ion-source. It is best found by trial and error until experience is gained. The rate of fractionation can perhaps be reduced by performing the experiment at a lower temperature. A particularly suitable method might be that described by Bokhoven and Theeuwen(62).

(b) Amplification Level

The peak heights, measured on a chart for small intensity ions, are less reliable because fewer ions are being collected. This effect is increased with scanning speed and amplification level. The latter will generally require some adjustment between scans as the ion-source pressure changes during the experiment. This is lower towards the end of the fractionation as the sample is pumped away (see (g) below).

(c) Errors of Measurement and Transcription

In the case of a good automatic data acquisition system these will merely be the digitiser errors, probably increasing with scanning speed. The present study initially involved measurement of all charts (Honeywell, ultra violet) by hand, in millimetres. Peak heights were noted on the chart itself, multiplied up according to the relevant scale factor and transcribed on to graph paper. This allowed comparison of spectra in order to detect gross errors and missing numbers. Further transcription and card-punching completed the process. This system proved capable of producing:

1. large relative errors in small peaks
2. large errors in scale-multiplied peaks (i.e. conversion between chart scales)
3. gross errors and missed measurements (e.g. failure to scale-up a peak)
4. errors of transcription and punching

It is obvious that large errors can have serious consequences(6). In the absence of automatic on-line facilities it was apparent that a faster and less error-prone system had to be developed. This feeling was reinforced by the need for large amounts of data in any statistical study. The development of a suitable semi-automatic system is given in Section III.

(d) Temperature Effects

In general, mixtures analyses should be run isothermally since changes in ion-source temperature can cause changes in the component spectra. This phenomenon is well known(63-66). The system would again lose its linear characteristics. It is thought that small temperature gradients of the order of ten to twenty Centigrade Degrees, around normal operating temperatures, would be unlikely to alter the patterns significantly (to some extent dependent on the particular substances employed). Such temperature gradients would have useful fractionating properties if used in conjunction with e.g. the G.E.C. - A.E.I. heated direct insertion probe(56). Temperature effects using a similar system have been reported (67), their magnitude depending on the temperature gradients and particular substances employed. An investigation of the fluctuations, mainly due to source temperature changes, present in high resolution spectra has been undertaken(68).

(e) Tuning will be discussed in Section II. See also ref. (83)-repeller potential.

(f) Impurities

These do not necessarily refer to fully recognizable components of the mixture introduced for analysis but could be background impurities increasing in significance as they are desorbed (perhaps preferentially by a component of the mixture) from the inlet and ion-source surfaces. It is obvious that analyses should be conducted only after extensive baking. As this is not always effective or practicable it may be that flooding the system with a desorbing agent such as ammonia or pyrrolidine will suffice. The unexpected presence of desorbed impurities during an experiment was discovered and is described in connection with MIXTURE II in Chapters 4 and 5.

(g) Arcing

This affects only one spectrum at a time and frequency depends on the type of sample, its pressure and the condition of the instrument. It is always easily detected and leads to rejection of the affected spectrum.

(h) Peak Height/Area Approximation

The ion abundance is taken to be proportional to the peak height. This is only possible where the peaks always have the same widths in the base and are of simple geometric shape. Consideration is given to this is Section II (experimental).

An experimental scheme is then required which will provide a mixtures array satisfying the above requirements i.e. one in which the analyst can have confidence. The particular scheme devised is shown in block diagram form - FIGURE 1. Spectra are examined visually is Checking Scheme I and, if not rejected, more carefully by computer methods in Scheme II.

A. Experimental Methods are outlined in Section II.

B. The spectra were obtained in groups of at least two "identical" spectra by successive scanning (described in Section II). This simple requirement has the immediate advantage that:

a. all gross errors of measurement and transcription can be detected and localised by comparing the spectra of such a group.

b. possible fractionation or pressure change during a scan is detected because then no neighbouring spectra would be "identical" i.e. condition 2. above, is tested.

c. where more than two "identical" spectra can be obtained a statistical bonus is gained in providing some measure of the fluctuations in a particular array.

d. as no automatic measuring facilities were available a preliminary visual comparison of the spectra was possible. The comparison is represented as Checking Scheme I in FIGURE 1 and enables a quick decision to be made as to the suitability of the spectra for time-consuming measurement. In several cases this preliminary check revealed that fractionation was too fast or the scanning speed too slow. This allowed the experiment to be repeated with more success.

C. and D. Measurement and Card-Punching

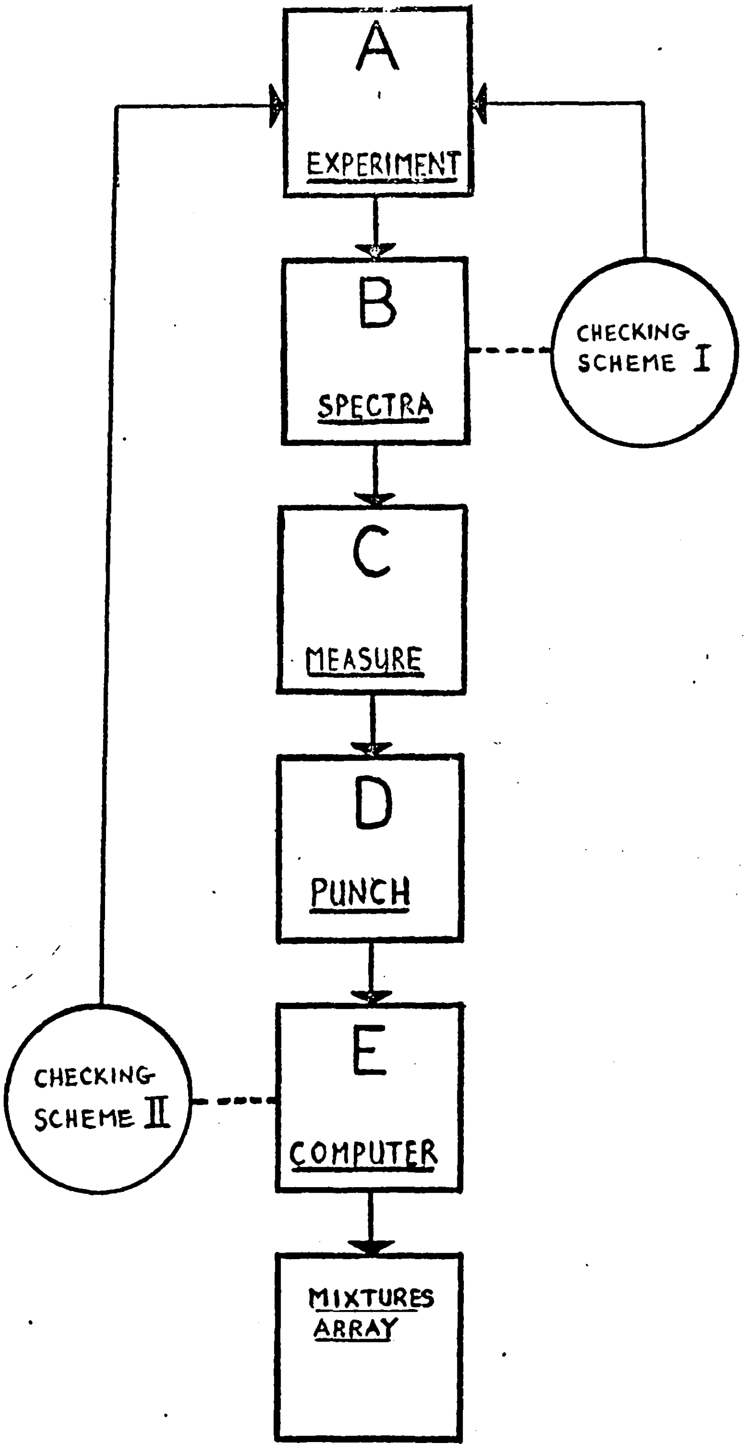
Initially these sections were separate but became a single operation on introduction of the semi-automatic system (Section III of this Chapter).

E.

The existence of groups of identical spectra greatly simplifies screening of the data by computer. As will be shown a computer examination can be very revealing and could be performed at once where on-line facilities exist.

Checking Scheme II (FIGURE 1) involves a further check for gross errors and illustrates the nature of "identical" when considering such groups of spectra. This involves a study of the differences between spectra and introduces the term "Degree of Fractionation" and the use of

FIGURE 1.


A
EXPERIMENT
B
SPECTRA
CHECKING SCHEME I
C
MEASURE
D
PUNCH
E
COMPUTER
CHECKING SCHEME II
MIXTURES ARRAY

Product-Moment Correlation Coefficients (see APPENDIX C). It is obvious that the greater the degree of fractionation of the components, the more distinct will be the mixtures spectra obtained and the more reliable the final results *i.e.* the significance of the inherent fluctuations will be reduced. An equivalent way of saying this is that the elements of the unknown matrix X, as previously defined, should be as diverse as possible. Such considerations are directly related to condition 3, above.

The use of product-moment correlations are well known in scientific applications, particularly psychological ones (see APPENDIX C).

The checks being made in this scheme are then:

a. there are no gross errors within the matrix.

b. each spectrum in a group (as previously defined) actually belongs to it *i.e.* no fractionation or pressure change is occurring.

These tasks are performed by a FORTRAN IV sub-routine called SCREEN, described in APPENDIX B as are all subsequently mentioned sub-routines.

a., above, is accomplished by comparing the spectra in a manner which reveals the position of a serious deviation. For this purpose it is best to have at least three spectra per group.

b., above, is accomplished by computing product-moment correlation coefficients (as described in APPENDICES B and C) of the first mixture spectrum and all others in the matrix. A sub-routine called FRACT is used by SCREEN to compute the correlation matrix of all the spectra. The required coefficients are present in the first row of this matrix. Each coefficient is a statistical measure of the similarity between the first spectrum and each subsequent one. A value of +1.000 means that the spectra are identical or proportional. In practice this might be +0.998 owing to experimental fluctuations. The coefficient values can range down through 0.000 (no correlation) to -1.000 (inverse correlation).

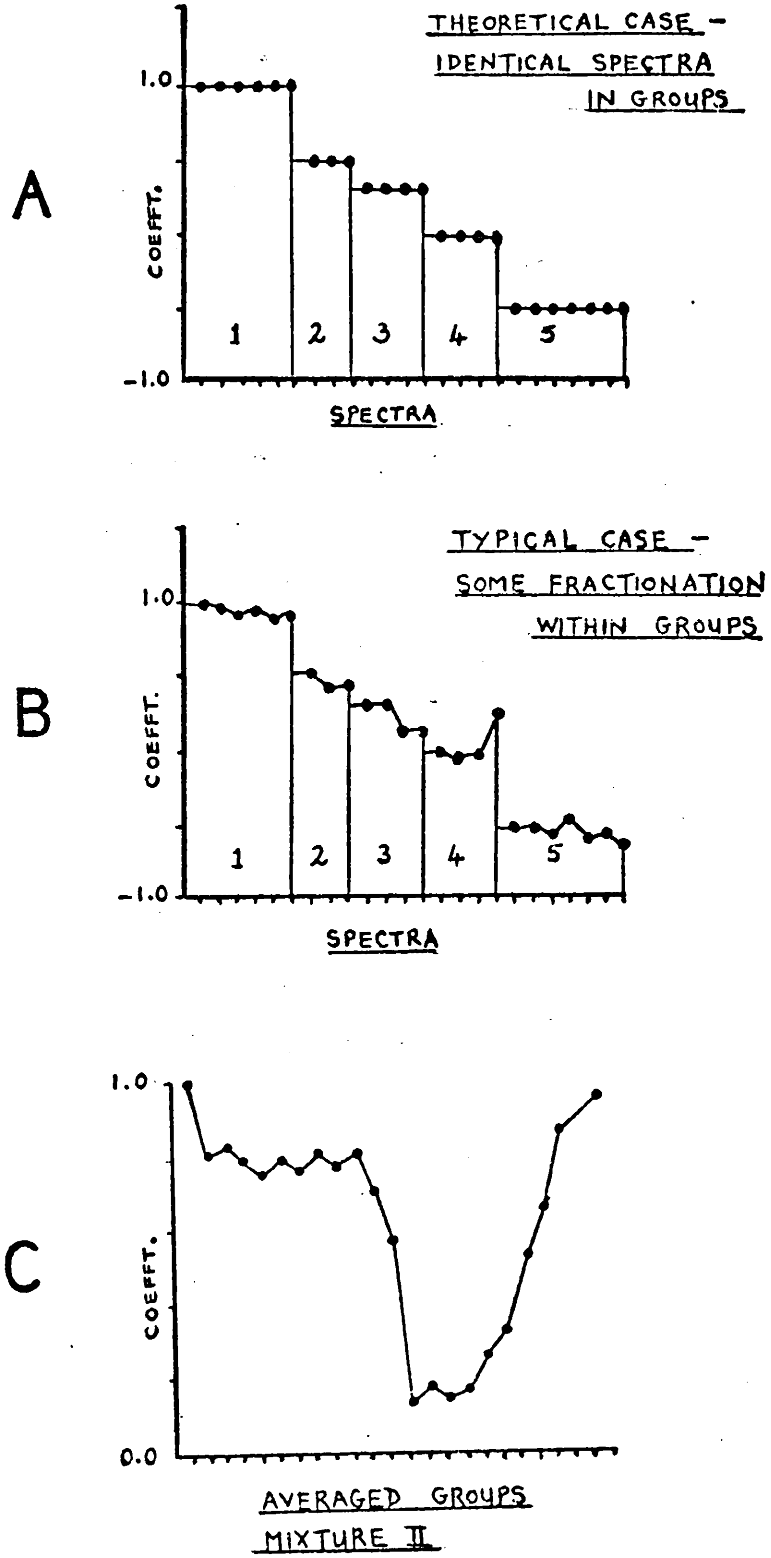
It may be convenient at this time to illustrate by means of a FRACTIONATION DIAGRAM the type of results obtained - FIGURE 2. The diagram is obtained by plotting the correlation coefficients against spectrum number, in experimental order. In the experimental case (B) fractionation is found to be taking place within groups of supposedly identical spectra (as shown in (A)). If this is judged to be slight it may be ignored or the groups re-arranged in some way for subsequent averaging of suitable spectra. The diagram shows that 3 would be better considered as two separate groups, containing two nearly identical spectra each. The final spectrum in 4 would be rejected because it does not correlate well with its immediate neighbours.

The "Degree of Fractionation" may be taken as the smallest correlation coefficient. This value, however, can have no real physical meaning attached to it i.e. it is not translatable into any physical or chemical coefficient and will only be used here for reference purposes. As will be seen it is not necessarily the final coefficient owing to the fact that a later mixture may by chance correlate highly with the first. The coefficients, as used here, do give an overall measure of the degree of similarity between mixtures spectra.

Some preliminary results are listed in Section IV of this Chapter.

FIGURE 2.

FRACTIONATION DIAGRAMS



II Experimental - Collection of Spectra

General Considerations

Tuning of Instruments A.E.I. MS9, MS12 and G.E.C. - A.E.I. MS902S.

The instruments were tuned in the manner recommended by the manufacturers except that particular attention was paid to peak shape. Considering that peak height was to be used as the measure of ion abundance in all cases it was essential to obtain peaks which were flat-topped (or trapezoidal) i.e. gaussian shapes were to be avoided.

Background

Experiments were carried out after baking the instrument for several days. Before the introduction of the mixture in each case background spectra were obtained at a variety of amplification levels in order to have some measure of their significance. Such measures were later found to be inaccurate representations of background levels on sample introduction probably because adsorbed impurities were preferentially desorbed by components in the mixture (see Mixture II, below).

Source Pressure

It was considered desirable for afore-mentioned reasons to maintain low pressures, less than 10^{-5} Torr, in the ion-source.

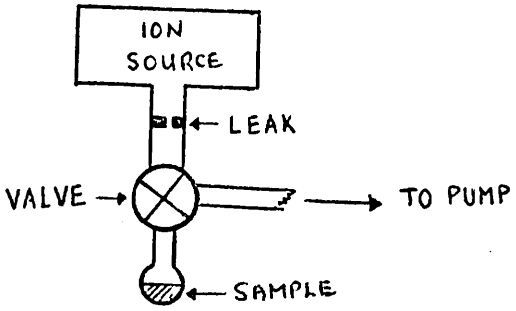
Two types of mixtures experiments were performed using:

(a) The A.E.I. Cold-Inlet System(60).

(b) the G.E.C. - A.E.I. Heated Direct Insertion System(60).

Fractionation Methods

(a) A simplified drawing of the gas-inlet system is shown:



The general principles of the method adopted were as follows:

1. Sample of mixture (ca. 0.1 ml.) placed in bottle as shown.
2. De-gassed with liquid nitrogen in the usual manner.
3. On thawing some mixture vapour was allowed to enter the ion-source via a doser to regulate the amount, an expansion vessel and a leak.
4. Some time (a minute) was allowed for equilibration <u>i.e.</u> constant ion-current monitor reading. Several spectra were taken in succession, at constant monitor reading when the pressure was low enough. The latter was adjusted by pumping some sample, if necessary, from the ion-source. Charts were numbered in groups as defined in Section I.
5. Spectral changes were observed by switching over to the oscilloscope(60). If no change in the pattern was observed after a few minutes some of the mixture was pumped away. If a change was observed some spectra were taken after equlibration.
6. Where no significant change was noticed after several such attempts the system was evacuated and some of the sample in the bottle pumped away. The whole process was then repeated on re-admission until the sample bottle was exhausted.

Several mixtures were obtained in this way:

MIXTURE I		COMPONENTS	
Instrument:	MS9	25%	THIOPHENOL
No. of Spectra:	17	25%	*n* - HEPTANE
No. of Groups:	6	25%	*ISO* - OCTANE
		25%	*p* - XYLENE
MIXTURE II			
Instrument:	MS12	25%	*n* - PROPYL ETHER
No. of Spectra:	133	25%	TOLUENE
No. of Groups:	36	25%	BENZENE
		25%	ETHYL ACETATE
MIXTURE III			
Instrument:	MS12	100%	ETHYL ACETATE
No. of Spectra:	16		
No. of Groups:	1		
MIXTURE IV			
Instrument:	MS12	20%	N,N - DIMETHYLANILINE
No. of Spectra:	35	20%	TOLUENE
No. of Groups:	8	20%	BROMO - ETHANE
		20%	*o* - DICHLOROBENZENE
		20%	THIOPHENOL

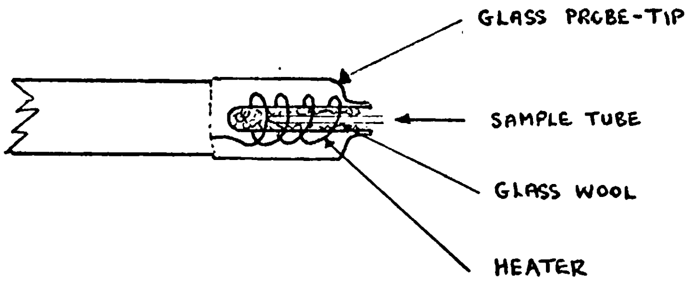
Mixtures II and III were obtained consecutively. Scanning speed was the same in all cases taking approximately ten seconds to cover these mass ranges.

Towards the end of each experiment care was taken to ensure that the amplification level necessary to obtain measurable spectra was not too high i.e. such that the spectra could contain a significant portion of the previously observed background.

(b) Heated Direct Insertion System

The G.E.C. - A.E.I. Heated probe consists of a hollow probe into which cooling gas may be blown. The tip contains a thermocouple and heater surrounding a glass sample cup; the instrument has a working temperature range of -50°C. to +350°C.

The method of placing samples in the cup suggested by G.E.C. - A.E.I.(60) was adopted i.e. dry glass wool was placed inside surrounding a small sample tube made by sealing a capillary tube at one end. The situation is shown in the diagram:



MIXTURE V		COMPONENTS	
Instrument:	MS902S	0.6 m.g.	5α- CHOLESTAN
No. of Spectra:	28	0.8 m.g.	5α- CHOLESTAN-3-ONE
No. of Groups:	15	1.1 m.g.	STIGMASTERYL ACETATE
Scanning Speed:	as before		

The above spectra were obtained according to the following method:

1. 0.6 mg. of the steroid sample were placed evenly in the sample tube using pure diethyl ether.
2. The probe itself was baked in the source for one hour before the experiment.
3. The probe was removed and the source temperature allowed to equilibrate at 95°C. with the electron beam switched on. This temperature was maintained throughout.
4. Background was negligible at reasonable amplification levels.
5. The sample tube was placed in the probe-tip and inserted into the source. The probe temperature remained constant at approximately 35°C. throughout the experiment.
6. Spectra were immediately visible. At constant ion current reading several spectra were taken rapidly, the first being used to adjust chart peak heights. Three spectra were finally obtained after checking visually for fractionation (Checking Scheme I - Section I of this Chapter).
7. Patterns were observed on the oscilloscope between running of groups of spectra at constant ion current. When a new stable pattern was recognizable more spectra were taken.
8. The rate of fractionation was slightly increased where necessary by making slight adjustments in the heater control such that no temperature change was observed.
9. Fifteen groups were obtained. Five were rejected on employment of Checking Scheme I.

MIXTURE VI

A similar technique was employed with an unknown mixture thought to consist of a side-chain cyclohexanone with an acidic impurity i.e. probably a binary mixture. Seventeen spectra were obtained.

Unfortunately the lowest source temperature possible at the time of the experiment was 165°C. thereby causing more rapid fractionation. This could have been offset by cooling the probe in some way but there was insufficient time. The spectra were not rejected but subjected to analysis in order to compare results obtained from Mixture V.

It has not been considered necessary to list copious details of the above experiments. The emphasis has been on allowing the experimental requirements to be as unexacting as possible by employing simple techniques. Several more experiments would be necessary in order to devise a standard procedure e.g. of source and probe temperatures and sample weight. The rate-determining step in the current investigation was that of spectrum measurement and computer acquisition. These topics are discussed in the next Section.

III - Spectrum Measurement

In view of the numbers of spectra to be measured during the course of this work (for statistical reasons) an automatic or, at least, semi-automatic system is to be preferred. This is not only because of the boring and repetitive nature of the work but also the extreme error-proneness of any hand measuring system where numbers must be written down, perhaps multiplied up, and re-copied (in order to punch cards etc.).

A semi-automatic device was available in another department and it was decided to adapt this for mass spectral use. The device is a "pencil-follower"(69) and consists of a glass table beneath which is a sensitive magnetic detector movable in two dimensions by means of a servo-mechanism. A diagram is given in APPENDIX A. The detector follows a magnetic pencil moved on the table and the co-ordinates of its position, to 0.1 mm., continuously fed to a pair of digital voltmeters. On depression of a foot-pedal the co-ordinates are punched on to paper tape.

The pencil was moved by hand over charts held horizontally by means of a specially constructed perspex rectangle.

The paper-tapes, obtained in the manner described in APPENDIX A, are processed by a computer programme and the spectral peak heights punched directly on to cards for storage and subsequent analysis.

IV - Pre-treatment of Data

Application of SCREEN revealed hitherto undetected gross errors and spectra correlating differently from their immediate neighbours. Elimination of such "odd" spectra was accomplished by SELM. Averaging of spectra in re-arranged groups was carried out either by MAV or MSTATS (details in APPENDIX B).

MIXTURE II

Preliminary computer analysis using SCREEN revealed twenty-three human errors (these spectra were measured by hand). Corrections were made by reference to the original charts. A primary data matrix containing 133 spectra in 32 groups with 37 masses in range m/e 103-45 was obtained. The whole matrix was stored on a disk-file(70) for subsequent recall and processing.

Spectral correlation coefficients were obtained (SCREEN) by correlating the first spectrum with all others in four groups since there are more spectra than masses in this case. i.e. the first spectrum was placed first in four sub-matrices of the whole and the four corresponding correlation matrices calculated as discussed in APPENDIX B. The reason for adopting this procedure where there are more spectra than masses is given in APPENDIX C.

The first rows of the correlation matrices are shown in TABLE 1. Nine spectra do not correlate well with their neighbours and these were eliminated. Twenty-three new groups were formed and averaged using MSTATS. It is apparent from TABLE 1 that the fractionation was not a straightforward one in that the final spectrum does not have the smallest coefficient. The effect is ascribed, at least partly, to desorption phenomena (see Chapter 4 and 5-II).

In spite of this twenty-three mixtures spectra were obtained. The fractionation may be represented by the correlation coefficients obtained for these spectra using FRACT:

SPECTRUM	COEFFICIENT
1	1.0000
2	0.8019
3	0.8336
4	0.8056
5	0.7809
6	0.8048
7	0.7883
8	0.8118
9	0.7874
10	0.8044
11	0.6805
12	0.5748
13	0.1512
14	0.1881
15	0.1574
16	0.1858
17	0.2789
18	0.3306
19	0.5105
20	0.6754
21	0.8625
22	0.8114
23	0.9591

or these figures may be plotted and shown in FRACTIONATION DIAGRAM form as is FIGURE 2.C. It is apparent that this diagram allows a measure of the similarity between all mixtures spectra. Such a measure will be an important consideration in choosing suitable spectra for subsequent analysis.

The other mixtures, measured as described in Section III, were similarly treated and some coefficients are listed in TABLES 2, 3 and 4.

In each case the correlation coefficients vary with the particular masses chosen. As could be predicted where the higher masses are used much greater differences in coefficients are observed because there is less likelihood of overlapping *i.e.* fewer components are present at higher masses and so greater changes in patterns are observed. An extreme example of this would be a series of uni-component peaks all unique to different components; as the fractionation proceeds the change in pattern would be a maximum. Examples of this effect are shown for Mixtures IV and V in TABLES 3 and 4 respectively.

MIXTURE I

The results in TABLE 2 indicate that the degree of fractionation obtained was not very good compared with Mixture II for instance.

MIXTURE II

The upward swing in coefficients from spectrum 13 (FIGURE 2.C) coincides with a sudden increase in intensity of ions m/e 87 and 58. This increase is ascribed to an impurity or impurities appearing in the system, perhaps by preferential desorption in the inlet-system or ion-source.

MIXTURE III - TABLE 2

The effect of inherent fluctuations on the coefficients in this case is shown to be very slight. The spectra were obtained using similar conditions to those for Mixtures I, II and IV.

MIXTURE IV

Once again the general shape of the curve (if the coefficients in TABLE 3 were plotted) is as in Mixture II. The increase in coefficients towards the final spectrum could be due to impurities. The effect is partially offset by excluding those peaks due to air. (increased amplification in final spectra).

MIXTURE V - TABLE 4

The coefficients obtained are very encouraging and indicate that the fractionation was under reasonable control with relatively little effort.

MIXTURE VI - TABLE 4

As expected, the coefficients show that fractionation was too rapid for the scanning speed used. However, averaging produced six spectra (see below).

The spectra present in most of the above arrays were grouped according to their correlation coefficients (the re-grouping described in Section I) and the groups averaged giving a final mixtures array in each case:

MIXTURE	NUMBER OF SPECTRA	NUMBER OF MASSES
II	23 (133)	37
IV	8 (35)	88
V	12 (28)	142
VI	6 (13)	71

The values in brackets indicate the numbers of spectra originally present in each array. The m/e values of each peak measured are given in the TABLES.

A sub-routine called MASSES was written to pick out the most significant masses for analysis (see APPENDIX B for details).

TABLE 1

MIXTURE II

CORRELATION COEFFICIENTS

spectrum 1 with:

spectrum	COEFFT.	spectrum	COEFFT.	spectrum	COEFFT.	spectrum	COEFFT.
2	0.9998	35	0.7866	68	0.8032	101	0.1752
3	0.9999	36	0.7923	69	0.8100	102	0.1417
4	0.9970	37	0.7888	70	0.8036	103	0.1452
5	0.9990	38	0.7786	71	0.7890	104	0.1506
6	0.9992	39	0.7818	72	0.8136	105	0.1586
7	0.8093	40	0.7942	73	0.7996	106	0.1637
8	0.7994	41	0.7949	74	0.7996	107	0.1692
9	0.8068	42	0.7941	75	0.6727	108	0.1808
10	0.8078	43	0.7867	76	0.6714	109	0.1836
11	0.8068	44	0.7940	77	0.6757	110	0.1817
12	0.8016	45	0.8008	78	0.6321	111	0.1772
13	0.8126	46	0.8000	79	0.6929	112	0.1790
14	0.8197	47	0.7979	80	0.5758	113	0.2478
15	0.8099	48	0.8047	81	0.5473	114	0.2999
16	0.8355	49	0.8103	82	0.5371	115	0.3231
17	0.8472	50	0.8179	83	0.5558	116	0.3280
18	0.8341	51	0.8136	84	0.5652	117	0.5036
19	0.8397	52	0.8162	85	0.5688	118	0.5089
20	0.7857	53	0.7583	86	0.5738	119	0.5949
21	0.8247	54	0.7720	87	0.5568	120	0.6709
22	0.8119	55	0.7824	88	0.5578	121	0.6723
23	0.8053	56	0.7886	89	0.5092	122	0.7752
24	0.8063	57	0.7859	90	0.5822	123	0.7926
25	0.8169	58	0.7898	91	0.6444	124	0.7948
26	0.7921	59	0.7792	92	0.2005	125	0.8025
27	0.7710	60	0.7834	93	0.1603	126	0.8054
28	0.7866	61	0.7741	94	0.1354	127	0.8650
29	0.8040	62	0.7768	95	0.1320	128	0.8696
30	0.8169	63	0.7900	96	0.1248	129	0.9442
31	0.8080	64	0.7861	97	0.1257	130	0.9551
32	0.7856	65	0.7719	98	0.1961	131	0.9506
33	0.7842	66	0.7820	99	0.1882	132	0.9670
34	0.7818	67	0.7976	100	0.1716	133	0.9608

PEAKS - 103, 102, 93-87, 79-73, 71, 70, 66-55, 53-49, 46, 45

TABLE 2

MIXTURE 1

CORRELATION COEFFICIENTS

spectrum	COEFFT.
1	1.0000
2	0.9996
3	0.8468
4	0.8491
5	0.8435
6	0.8585
7	0.8418
8	0.8378
9	0.8370
10	0.8408
11	0.8259
12	0.8344
13	0.8219
14	0.8265
15	0.8322
16	0.8272

PEAKS - 107-105, 103, 100, 99-97, 92, 91, 85-81, 79-77, 72-65, 63, 62

MIXTURE III

CORRELATION COEFFICIENTS

spectrum	COEFFT.
1	1.0000
2	0.9999
3	0.9999
4	1.0000
5	0.9998
6	0.9999
7	0.9997
8	0.9999
9	0.9999
10	0.9999
11	0.9997
12	0.9998
13	0.9999
14	0.9999
15	0.9999
16	0.9999

TABLE 3

MIXTURE IV CORRELATION COEFFICIENTS

88 PEAKS (INC. AIR)	FIRST 60 PEAKS (EXCL. AIR)	FIRST 35 PEAKS
1.0000	1.0000	1.0000
0.9986	0.9998	0.9999
0.9968	0.9997	0.9998
0.9979	0.9998	0.9997
0.9982	0.9998	0.9999
0.9979	0.9998	0.9998
0.6297	0.4128	0.3855
0.6359	0.4042	0.3763
0.6144	0.4050	0.3782
0.6368	0.4063	0.3785
0.6276	0.4032	0.3766
0.0530	0.0440	-0.0162
0.0379	0.0333	-0.0291
0.0399	0.0204	-0.0435
0.0396	0.0187	-0.0456
0.0289	0.0024	-0.0623
0.0259	0.0011	-0.0656
0.0236	-0.0033	-0.0695
0.0226	-0.0037	-0.0693
0.0196	-0.0560	-0.1246
0.0259	-0.0574	-0.1256
0.0196	-0.0574	-0.1255
0.0234	-0.0579	-0.1258
0.6983	-0.0811	-0.1331
0.7088	-0.0826	-0.1331
0.7073	-0.0810	-0.1311
0.7214	-0.0844	-0.1358
0.3163	-0.0798	-0.1274
0.2953	-0.0816	-0.1290
0.2991	-0.0808	-0.1281
0.3052	-0.0825	-0.1290
0.5249	-0.0801	-0.1308
0.5097	-0.0817	-0.1317
0.5091	-0.0810	-0.1317
0.5362	-0.0832	-0.1340

PEAKS - 149-145, 121, 120, 113-102, 95-89, 87-49, 46, 45.5, 45-34, 32-24

TABLE 4

CORRELATION COEFFICIENTS

MIXTURE V

28 SPECTRA	
142 PEAKS	
SPEC.	COEFFT.
1	1.0000
2	0.9994
3	0.9992
4	0.9750
5	0.9744
6	0.9577
7	0.9699
8	0.9694
9	0.9424
10	0.9413
11	0.9377
12	0.9340
13	0.9118
14	0.9131
15	0.9166
16	0.8949
17	0.8879
18	0.8539
19	0.8514
20	0.8036
21	0.8066
22	0.7358
23	0.7570
24	0.7016
25	0.7246
26	0.6727
27	0.6737
28	0.6692

MIXTURE V

12 AVERAGED GROUPS	
FIRST 25 PEAKS	
SPEC.	COEFFT.
1	1.0000
2	0.9681
3	0.9642
4	0.9221
5	0.9173
6	0.8794
7	0.8373
8	0.7846
9	0.7088
10	0.6237
11	0.6222
12	0.5659

MIXTURE V PEAKS - 394, 387, 386, 372, 371, 358, 357, 364-255, 247-242, 234-227, 220-213, 205-199, 193-185, 179-173, 167-157, 152-145, 139-131, 125-117, 111-105, 97-91, 85-77, 71-65, 57-50, 45-41.

MIXTURE VI

13 SPECTRA	
71 PEAKS	
SPEC.	COEFFT.
1	1.0000
2	0.9741
3	0.8898
4	0.8862
5	0.8728
6	0.8677
7	0.8528
8	0.7496
9	0.7338
10	0.6353
11	0.6248
12	0.5684
13	0.5638

CHAPTER FOUR

Determination of the Number of Significant Components

I - Matrix Rank Analysis

II - Principal Components Analysis

IV - General Conclusions

DETERMINATION OF THE NUMBER OF SIGNIFICANT COMPONENTS

I - Matrix Rank Analysis

An essential part of the "ab-initio" analysis of the mixtures array must necessarily be the determination of the number of components making significant contributions. As discussed in Chapter 2, this is equal to the rank of the matrix. Initial experiences of using the Algol Programme is also given in Chapter 2 together with references to the method used i.e. Gaussian elimination(53,54). This procedure is used to transform certain elements to zero or small numbers which may be taken to be zero. Testing a transformed element against zero is accomplished in the programme by tolerance values which set quite arbitrary limits. Rows of the array undergoing transformation are interchanged where necessary to provide a pivot element greater than zero. The meaning of "pivot element" is given later in this Chapter where Gaussian elimination is more fully explained.

The generally recommended method of rank determination by Gaussian elimination(53,54) involves the interchange of rows and columns (pivoting) to provide the largest possible matrix element as divisor. This process minimises error build-up during the transformation and will be discussed in greater detail.

When the tolerance values are zero it is obvious that (a) round-off errors and (b) experimental fluctuations will raise the rank to a maximum.

(a) Round-off

Consider the matrix

$$\begin{bmatrix} 1 & \frac{1}{3} \\ 2 & \frac{2}{3} \end{bmatrix}$$

The rank is one because one row is a multiple of the other. (In a more complex example it could be some linear combination of any of the other rows). In the computer, however, round-off is present and the matrix might be

$$\begin{bmatrix} 1.000 & 0.333 \\ 2.000 & 0.667 \end{bmatrix}$$

The rank is now two. As shown in this simple example when decimals are rounded off the rank is increased. This is because 0.667 is not exactly twice 0.333. If, however, it is allowed to be twice by means of a suitable tolerance value, the rank will again be one. In this case the necessary value would be greater than or equal to 0.001 i.e. the value obtained by subtracting twice the first row from the second. In a more complex example, where more operations are required to reduce the matrix, the tolerance would have to be greater in order to allows for increased round-off errors. The actual value used will depend on the number of decimals and the number of operations involved.

In the case of the numerically exact mixtures arrays solved by the Algol Programme(48) any small tolerance, sufficient to overcome round-off, enabled the correct rank to be obtained. The situation with experimental data is quite different.

(b) Experimental Fluctuations

If the mixtures array contains these relatively large errors a much greater tolerance value is needed to produce a satisfactory rank. The actual magnitude of this tolerance will vary with the magnitudes of the fluctuations and the number of operations in a particular array. The reason is that the transformed elements, which could be zero, have much greater errors associated with them. This leads to the question of significance when dealing with mixtures spectra.

Significance

An experimental mixture may contain several major components and a number of minor ones in which the analyst has less interest. It is easy to envisage a mixtures array made up of four component spectra in similar relative amounts. If other arrays are visualised having smaller and smaller relative contributions from one of the components, a stage is eventually reached where difficulty is experienced in choosing between ranks three and four. At this stage the contribution of the fourth component is on the threshold of significance. The tolerance values used will therefore have to be chosen with this in mind. As will be shown such a problem need not be a serious one.

Having decided to change computer systems work was begun on the rank analysis of mixtures arrays. The rank determining sub-routine available with the I.B.M. Scientific Library(71) was utilised for this purpose. Its use involved reading in an array together with a suitable factor which, on multiplication by the largest element, produced a tolerance. The array, M, is effectively transformed according to

$$M'_{ij} = M_{ij} - \frac{M_{i1}.M_{1j}}{M_{11}}$$

where M_{11} is the largest element pivoted to the leading position by suitable interchange of rows and columns. The largest element in M' is compared with the tolerance and if it turns out to be effectively zero the rank is one. If not it is pivoted to become M'_{11} and the elimination repeated to give M''_{ij} etc., until the rank is found.

The immediate advantages of this system over the previous method were:

(1) speed of operation

and (2) complete pivoting was used.

The disadvantage was the inaccessibility of the sub-routine for development purposes e.g. the inclusion of further tolerances as in the Algol Programme. The problem was not considered insurmountable at this stage and "looping" studies were begun (see Chapter 2 - II). It was hoped to relate empirically the tolerances necessary to give correct ranks for a number of known arrays. Progress was held up by the need to obtain suitable arrays by previously described methods.

Work was then channelled into two areas:

(a) obtaining and measuring mixtures arrays for testing purposes.

(b) writing a new programme in FORTRAN IV readily capable of modification.

Coinciding with this work discussions(72) at Esso Research Centre (Abingdon, Berkshire) revealed relatively long-established applications of matrix rank analysis(73). These were mainly in the field of biochemistry for arrays of absorbtion data analogous to the mass spectral arrays considered here. Two particular methods of a similar nature seemed to be of particular relevance to the current problem. A combination of these techniques was then incorporated into the new programme mentioned in (b), above.

The first is due to Wallace and Katz(74). The Gaussian elimination is performed as previously described i.e.

$$M'_{ij} = M_{ij} - \frac{M_{i1} . M_{1j}}{M_{11}}$$

No single tolerance is incorporated but a companion matrix, S, is constructed the elements of which are the individually estimated errors in M.

As M is reduced by the elimination process S is transformed according to the error propagation equation$^{(75)}$:

$$S'_{ij} = \sqrt{S_{ij}{}^2 + S_{ij}{}^2 . \left[\frac{M_{i1}}{M_{11}}\right]^2 + S_{i1}{}^2 . \left[\frac{M_{ij}}{M_{11}}\right]^2 + S_{11}{}^2 \left[\frac{M_{i1} . M_{1j}}{M_{11}}\right]^2} \; .$$

The procedure is the same as that previously described except that the largest element in M$'$ is compared with its corresponding error in S$'$, perhaps multiplied by a suitable constant. The reason for this suggested extra criterion is not explained but may help to take account of inaccuracies in the original error estimates. In a study by Ainsworth$^{(75)}$ there was some doubt as to the existence of a third component as a transformed pivot element was not much greater than its error. Perhaps this is not surprising in view of the illustration of "significance" given earlier.

The possibility of including an extra criterion was kept in mind when applying the method to mass spectral data with its relatively large and varying errors.

The second method, due to Katakis$^{(76)}$, is similar to the above but has a statistical advantage. Instead of comparing only the pivot element with its transformed error he suggests comparing all the elements in the transformed data matrix with their corresponding errors. In this way a percentage of elements may be allowed greater than their errors.

It is apparent from the error propagation equation that the value of any element in S' cannot exceed twice the value of the largest element in S and will probably be much less. This is only true where the matrix has been pivoted i.e. M_{11} is always greater than M_{i1} etc. If the matrix is not pivoted the errors can become very large indeed, particularly where the divisors (leading elements) are small.

The remaining problem in application of the technique is the construction of the error matrix itself. Here several possibilities exist depending on the nature of the particular data:

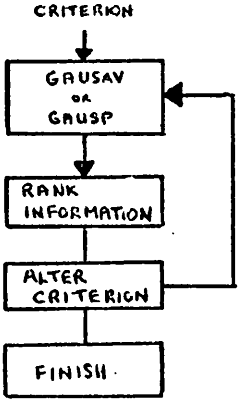
1. Where large spectral groups are obtained from the fractionation, the averaged deviations from the mean (for each peak) are given by sub-routine MSTATS. These values can become the elements of S corresponding to the elements of M, the averaged data matrix. This would be suitable where good on-line facilities are available.
2. Where fewer spectra are obtained in each group (e.g. two or three) it is convenient to form the error matrix elements as some fixed percentage of the corresponding averaged data matrix elements e.g. 5%.

Another possibility borne in mind was that of defining the percentage by obtaining an experimental array with one component. This could be run immediately after the mixture has been pumped away. (Mixture III, described in Chapter 3 - II, is such an array). The percentage error required to give a rank of one could be used to choose the value for the mixtures analysis. However, the difference in numbers of operations required to give the results in both cases would have to be taken into account.

The sub-routines used to determine rank in both cases 1. and 2. are described in APPENDIX B. They are called GAUSAV and GAUSP respectively.

The effect of some extra criterion on the rank obtained is built into both sub-routines thereby giving the degree of sensitivity of results to changes in the errors used. The elements of the error matrices in both cases are altered by some constant factor and the determination repeated. A diagrammatic illustration is given:

FIGURE 3.



The programme lists "rank statistics" i.e.

(a) percentage of transformed elements greater than errors

and (b) pivot element and its error

after each elimination step. The original data of Wallace and Katz(74) was processed and yielded the same transformed matrix elements and errors.

Results are given in TABLES 5 - 10 for the mixtures obtained as outlined in Chapter 3 - II.

MIXTURE I (TABLE 5)

After three eliminations at an error of 5% the absolute value of the largest element (-32.3) is less than its error but a substantial percentage (16.3) of values are greater than their corresponding errors.

After four eliminations the largest element (23.8) is not much less than its error (29.5).

The columns showing the corresponding errors and percentages at 3% and 7% give some indication of the error sensitivity of results.

The most probable rank, based on these figures, is three, possibly four. The difficulty is probably because the spectra are all fairly similar - a low degree of fractionation i.e. perhaps the compositions covered by these mixtures are biased in favour of three components where one has either been pumped too quickly or too slowly.

MIXTURE II (TABLE 6)

Results using both GAUSP and GAUSAV on the averaged array are given (errors in B. were produced by sub-routine MSTATS).

In both cases the absolute value of the largest element drops sharply after four eliminations. This most probably indicates the presence of four major components. After five eliminations the largest element is approximately equal to its error indicating a fifth component of lesser significance. The presence of a sixth minor component is also indicated but this could be due to the size of the array and some inaccuracy in the error transformation equation.

TABLE 6-B yields a similar result using arbitrary criteria within the ranges covered. Smaller percentages are obtained because the errors happen to be larger. Bigger "percentage greater" values are also caused by taking the percentage of each peak as the error in GAUSP; many small peaks will be greater than their tiny errors after transformation i.e. the actual percentage error in small peaks may easily be 50%. In this respect GAUSAV which uses the actual error in each peak is better but large numbers of spectra are necessary.

The fifth component (or fifth and sixth) was thought to be present mainly in the final ten spectra corresponding to sudden increases in the intensities of ions at m/e 87, 75 and 58. In TABLE 7 rank analyses for

A, the first nine spectra and B, the final ten spectra are given. No significant difference is observed.

In TABLE 8 results from the elimination of the whole array of 124 spectra are given. As expected the figures are not very good although the sudden decrease in size of the largest element is still present after four eliminations.

MIXTURE III (TABLE 8)

The rank is clearly one. Large "percentage greater" values after one elimination are caused by the method of initial error calculation (see MIXTURE II, above).

MIXTURE IV (TABLE 9)

A. (INCLUDING AIR) The rank is six, possibly seven.

B. (EXCLUDING AIR) The rank is five, possibly six.

MIXTURE V (TABLE 10)

The indicated rank is three in both cases.

MIXTURE VI (TABLE 5)

The rank indicated by GAUSP is clearly two.

Some Conclusions and Comments

1. The rank determination is best where few spectra are used i.e. where small numbers of operations are necessary e.g. compare rank analysis on 124 spectra in TABLE 8.
2. Results obtained for each mixture are clearly satisfactory and not very sensitive to changes in the initial error values used. The usefulness of the "percentage greater" columns (Katakis) would be increased where the errors in the peaks are estimated individually e.g. by MSTATS where sufficient numbers of spectra are available.

An improvement could probably be made by using sub-routine MASSES to remove the very small mass-profiles from consideration. An alternative improvement would be to increase the percent error in the smaller peaks.

3. In both GAUSP and GAUSAV the elimination process is repeated each time a new percentage error or criterion is used (see FIGURE 3). In future work this need be performed once and only the new errors transformed (see APPENDIX B).

TABLE 5

MIXTURE I   16 SPECTRA, 28 MASSES, 4 COMPONENTS

STEP	LARGEST ELEMENT	ERROR AT:			% > ERROR AT:		
		3%	5%	7%	3%	5%	7%
0	1080.0	32.4	54.0	75.6	100.0	100.0	100.0
1	302.3	10.1	16.8	23.5	80.2	72.6	63.5
2	71.2	23.6	39.4	55.1	54.7	40.4	29.4
3	-32.3	52.1	86.8	121.5	33.8	20.6	12.9
4	-23.8	17.7	29.5	41.3	30.2	16.3	9.4

MIXTURE VI   6 SPECTRA, 71 MASSES, 2 COMPONENTS

STEP	LARGEST ELEMENT	ERROR AT:			% > ERROR AT:		
		3%	5%	7%	3%	5%	7%
0	10542.4	316.3	527.1	737.9	100.0	100.0	100.0
1	1642.2	52.1	86.8	121.4	99.1	99.1	98.9
2	-116.1	69.5	115.8	162.1	66.7	48.6	29.3
3	-33.2	34.1	56.8	79.5	43.6	22.1	13.2

TABLE 6

A. MIXTURE II 23 SPECTRA, 37 MASSES GAUSP

STEP	LARGEST ELEMENT	ERROR AT:			% > ERROR AT:		
		3%	5%	7%	3%	5%	7%
0	424.9	12.8	21.3	29.7	100.0	100.0	100.0
1	291.1	8.9	14.9	20.8	93.4	90.3	87.0
2	-217.1	12.0	20.0	28.0	84.4	76.5	69.0
3	149.3	5.6	9.4	13.2	64.7	51.0	41.6
4	27.2	6.4	10.6	14.9	53.1	36.4	27.1
5	12.5	4.3	7.2	10.1	47.9	28.3	18.8
6	-8.7	6.8	11.4	15.9	36.2	17.5	10.4
7	6.3	5.4	9.0	12.6	31.5	15.2	9.8

B. MIXTURE II 23 SPECTRA, 37 MASSES GAUSD

(AVERAGED DEVIATIONS AS ERRORS)

STEP	LARGEST ELEMENT	ERROR, TIMES			% > ERROR AT:		
		2	6	8	x2	x6	x8
0	424.9	7.8	23.4	31.2	93.3	86.8	83.5
1	291.1	13.2	39.5	52.7	77.3	52.3	46.0
2	-217.1	16.2	48.5	64.7	60.0	29.8	23.7
3	149.3	5.8	17.3	23.0	33.2	12.2	8.2
4	27.2	8.0	24.0	32.0	19.1	1.4	0.3
5	12.5	11.5	34.4	45.9	13.9	0.5	0.0
6	-8.7	34.1	102.4	-	8.0	0.2	-
7	6.3	17.8	53.5	-	4.8	0.0	-

TABLE 7

A. MIXTURE II FIRST 9 SPECTRA, 37 MASSES

STEP	LARGEST ELEMENT	ERROR AT:			% > ERROR AT:		
		3%	5%	7%	3%	5%	7%
0	424.9	12.8	21.3	29.7	100.0	100.0	100.0
1	-217.9	12.0	20.0	28.0	88.5	85.1	80.6
2	186.3	5.7	9.5	13.4	60.8	47.8	42.0
3	149.3	5.6	9.4	13.2	55.4	40.2	32.8
4	27.3	6.4	10.6	14.9	33.3	15.8	10.3
5	-4.9	11.1	18.5	25.8	15.6	3.9	1.6
6	-4.9	4.9	8.2	11.4	5.4	0.0	0.0

B. MIXTURE II LAST 10 SPECTRA, 37 MASSES

STEP	LARGEST ELEMENT	ERROR AT:			% > ERROR AT:		
		3%	5%	7%	3%	5%	7%
0	297.3	8.9	14.9	20.8	100.0	100.0	100.0
1	67.4	2.2	3.6	5.1	86.1	78.4	72.5
2	25.5	1.5	2.5	3.5	73.2	61.1	54.3
3	22.7	3.4	5.7	8.0	66.0	48.7	34.9
4	-6.7	3.3	5.4	7.6	61.1	39.4	26.3
5	-5.8	2.8	4.6	6.4	38.1	23.1	10.6
6	-2.4	2.1	3.5	4.9	34.7	12.9	9.7
7	3.5	2.5	4.2	5.8	27.8	12.2	8.9

See FRACTIONATION DIAGRAM in FIG. 2(C)

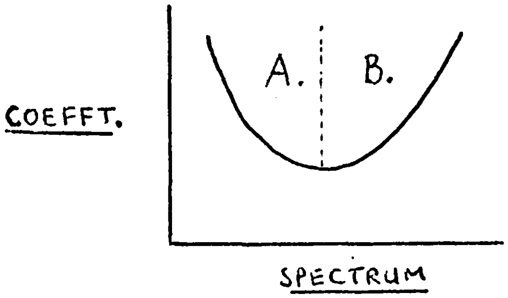


TABLE 8

MIXTURE II 124 SPECTRA, 37 MASSES

STEP	LARGEST ELEMENT	ERROR AT:			% > ERROR AT:		
		3%	5%	7%	3%	5%	7%
0	428.5	12.8	21.4	30.0	100.0	100.0	100.0
1	283.1	8.7	14.5	20.3	93.0	88.3	83.6
2	-224.9	11.8	19.7	27.6	86.4	78.2	71.5
3	147.8	5.6	9.4	13.2	63.6	48.9	38.6
4	24.6	1.5	2.5	3.5	53.6	36.7	26.3
5	-22.6	5.7	9.5	13.3	46.4	32.3	22.8
6	17.9	6.8	11.4	15.9	40.8	25.5	17.0
7	-10.5	2.9	4.8	6.7	32.2	18.8	12.2
8	-10.6	2.5	4.2	5.8	31.5	17.9	11.3
9	-9.1	1.6	2.7	3.7	27.5	13.9	8.5
10	-9.0	2.0	3.3	4.6	27.3	13.6	7.5
11	-8.2	3.4	5.7	8.0	29.0	14.3	7.6

MIXTURE III 16 SPECTRA, 27 MASSES, 1 COMPONENT

STEP	LARGEST ELEMENT	ERROR AT:			% > ERROR AT:		
		3%	5%	7%	3%	5%	7%
0	2900.3	87.0	145.0	203.0	100.0	100.0	100.0
1	-65.1	60.7	101.1	141.6	33.6	24.4	18.5
2	29.8	27.3	45.4	63.6	23.7	14.3	7.7
3	18.6	34.9	58.2	81.5	13.8	5.4	1.9
4	-19.1	31.8	53.0	-	8.7	3.3	-
5	16.8	51.4	85.7	-	1.7	0.0	-

TABLE 9

MIXTURE IV

A. 8 SPECTRA, 88 MASSES (INCLUDING AIR PEAKS)

STEP	LARGEST ELEMENT	ERROR AT:			% >ERROR AT:		
		5%	7%	9%	5%	7%	9%
0	13234.7	661.7	926.4	1191.1	100.0	100.0	100.0
1	8090.8	406.5	569.1	731.7	98.0	96.9	96.4
2	6193.7	354.5	496.3	638.2	90.1	85.9	81.4
3	3016.6	236.9	331.8	426.6	86.4	80.5	76.2
4	1098.8	176.5	247.2	317.8	60.7	50.0	42.3
5	-534.4	156.2	218.7	281.1	37.8	26.9	20.9
6	275.9	216.9	303.7	390.5	19.5	8.5	4.9
7	-53.3	210.2	294.3	378.4	4.9	1.2	0.0

B. 8 SPECTRA, 60 MASSES (EXCLUDING AIR PEAKS)

STEP	LARGEST ELEMENT	ERROR AT:			% >ERROR AT:		
		5%	7%	9%	5%	7%	9%
0	8127.4	406.4	568.9	731.5	100.0	100.0	100.0
1	6898.6	348.5	487.9	627.3	89.3	86.4	82.8
2	3160.1	236.9	331.8	426.6	84.8	80.7	74.4
3	2243.2	112.5	157.5	202.5	49.5	38.2	29.5
4	-406.6	146.6	205.2	263.9	43.8	33.0	24.1
5	273.6	210.4	294.5	378.7	29.1	18.8	13.3
6	84.4	105.2	147.2	189.3	18.5	14.8	13.0
7	-58.6	226.9	317.7	408.1	0.0	0.0	0.0

TABLE 10

MIXTURE V

A. 12 SPECTRA, 142 MASSES

STEP	LARGEST PEAK	ERROR AT:			% > ERROR AT:		
		3%	5%	7%	3%	5%	7%
0	6762.7	202.8	338.1	473.4	100.0	100.0	100.0
1	1717.8	54.9	91.4	127.9	97.5	95.3	93.4
2	995.3	64.3	107.1	149.9	91.1	84.4	77.9
3	-107.6	102.7	171.2	239.6	47.8	25.3	14.2
4	-104.4	123.7	206.1	288.5	22.3	7.1	2.2
5	-76.0	95.7	159.5	223.3	10.2	2.3	0.2
6	-72.6	139.1	231.8	324.5	2.1	0.2	0.0
7	50.3	235.2	391.9	-	0.9	0.0	-

B. 12 SPECTRA, 60 HIGHEST MASSES

STEP	LARGEST PEAK	ERROR AT:			% > ERROR AT:		
		3%	5%	7%	3%	5%	7%
0	6762.7	202.9	338.1	473.4	100.0	100.0	100.0
1	1627.6	52.4	87.3	122.3	98.2	96.3	94.3
2	334.1	131.4	219.0	306.6	57.2	37.9	23.6
3	-84.8	85.4	142.3	199.2	37.0	12.3	3.1
4	-99.7	140.4	233.9	327.6	11.8	2.0	0.4
5	-44.6	69.9	116.4	162.9	2.6	0.5	0.3
6	48.6	155.3	258.9	362.4	0.9	0.0	0.0
7	-52.3	173.8	-	-	0.4	-	-

II - Principal Components Analysis

This is a long established technique having been extensively employed in the analysis of psychological data(76). Recent years have seen its application to a number of spectroscopic(78) and chromatographic(79) problems. A short explanation of the method and its terminology is included in APPENDIX C. The number of factors needed to account for the VARIANCE(76) in the mixtures data is estimated. This is done by forming a correlation matrix and computing its EIGENVALUES(78). The method requires fewer operations than Gaussian elimination and so error build-up is much reduced. As with Gaussian elimination no gross errors can be tolerated as they would introduce more variance to be accounted for i.e. extra components would be found.

The criterion problem still exists because it is necessary to decide on the number of significant eigenvalues i.e. above some pre-set tolerance value.

A certain amount of the variance in each array is accounted for by the inherent errors in the data. As the levels of these vary between experiments so will the eigenvalue threshold of significant components.

Eigenvalues are computed by sub-routine PCA which is explained in APPENDIX B.

Results obtained for each mixture are given in TABLES 11 and 12. If values of logarithms are calculated as indicated it would seem that the number of components is given by the positive values in each case (see (78)).

MIXTURE I (TABLE 11)

The fourth factor accounting for the variance has a negative logarithm. The rank is then probably three i.e. in agreement with the rank analysis (see MIXTURE IV).

MIXTURE II (TABLE 11)

Four major components are indicated by the logarithms. The sixth component gives a negative value. The effect of processing a large array is similar to that in rank analysis.

MIXTURE III (TABLE 11) ONE COMPONENT

The rank is clearly one. The eigenvalue 0.002 gives an indication of the significance of the inherent fluctuations in the spectra. This second factor and subsequent factors <0.001 account for the "error variance".

MIXTURE IV (TABLE 12)

The effect of including air peaks is as before i.e. the rank increases by one. Results for eight spectra (averaged groups) indicate that a first negative logarithm of the value shown could represent a significant factor; otherwise the rank could be four. Where all 35 spectra are considered the rank is five.

MIXTURE V (TABLE 12)

The rank is three where 142 masses are used. Consideration of the 60 highest masses reduces the apparent rank i.e. one of the components appears to have less significance in this range.

MIXTURE VI (TABLE 12)

The number of components is two.

No serious differences are observed in results obtained from both Principal Components Analysis and Rank Analysis in each case (where constant criteria are used).

It appears, from the few examples tested and the criteria used, that the method of Principal Components tends to produce a minimum rank whereas the method of rank analysis tends to be more generous. A combination of both might prove useful.

TABLE 11

EIGENVALUES →

MIXTURE I 16 SPECTRA, 28 MASSES

NUMBER	λ	$\ln(\lambda . 10^2)$
1	16.258	7.394
2	0.604	4.101
3	0.119	2.477
4	0.007	-0.357
5	0.002	-1.609
6	0.002	-1.609
:	<0.001	-

MIXTURE II

124 SPECTRA, 37 MASSES

NUMBER	λ	$\ln(\lambda . 10^2)$
1	20.797	7.640
2	9.783	6.886
3	4.600	6.131
4	0.819	4.406
5	0.307	3.425
6	0.188	2.934
7	0.105	2.349
8	0.070	1.947
9	0.059	1.770
10	0.046	1.526
:	:	:

23 SPECTRA, 37 MASSES

NUMBER	λ	$\ln(\lambda . 10^2)$
1	16.062	7.382
2	5.496	6.309
3	0.804	4.388
4	0.606	4.105
5	0.162	0.482
6	0.005	-0.635
7	0.004	-1.050
8	0.003	-1.347
9	0.001	-2.207
.	<0.001	-

MIXTURE III 16 SPECTRA, 27 MASSES ——— ONE COMPONENT

NUMBER	λ	$\ln(\lambda . 10^2)$
1	15.997	7.378
2	0.002	-1.609
3	<0.001	-
.		

TABLE 12

MIXTURE IV

8 SPECTRA, 88 MASSES
(INCLUDING AIR PEAKS)

NUMBER	λ	$\ln(\lambda.10^2)$
1	5.137	6.242
2	1.771	5.177
3	0.824	4.416
4	0.195	2.970
5	0.068	1.917
6	0.004	-0.916
7	0.001	-2.303

8 SPECTRA, 60 HIGHEST MASSES
(EXCLUDING AIR PEAKS)

NUMBER	λ	$\ln(\lambda.10^2)$
1	5.436	6.298
2	1.473	4.968
3	0.750	4.317
4	0.335	3.513
5	0.004	-0.821
6	0.001	-2.056
7	<0.001	-

35 SPECTRA, 88 MASSES

NUMBER	λ	$\ln(\ .10^2)$
1	21.379	7.668
2	8.597	6.757
3	3.716	5.918
4	0.984	4.589
5	0.285	3.350
6	0.018	0.588
7	0.005	-0.617

35 SPECTRA, 60 HIGHEST PEAKS

NUMBER	λ	$\ln(\lambda.10^2)$
1	21.921	7.693
2	7.883	6.670
3	3.617	5.891
4	1.543	5.039
5	0.018	0.621
6	0.006	-0.462
7	0.004	-0.821

MIXTURE V

12 SPECTRA, 142 MASSES

NUMBER	λ	$\ln(\lambda.10^2)$
1	11.292	7.029
2	0.664	4.196
3	0.036	1.273
4	0.002	-1.514
5	0.002	-1.609
6	0.001	-2.079
7	<0.001	

12 SPECTRA, 60 HIGHEST PEAKS

NUMBER	λ	$\ln(\lambda.10^2)$
1	11.252	7.026
2	0.736	4.298
3	0.0088	-0.131
4	0.0016	-1.858

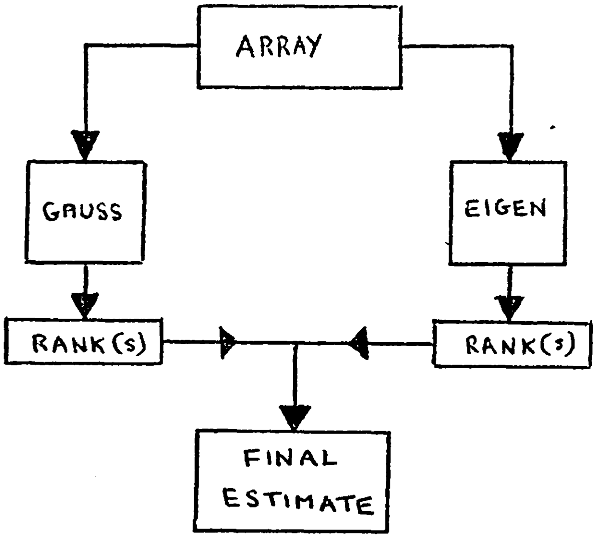
MIXTURE VI

6 SPECTRA, 71 MASSES

NUMBER	λ	$\ln(\lambda.10^2)$
1	12.031	7.093
2	0.958	4.563
3	0.008	-0.198
4	<0.001	

III - GENERAL CONCLUSIONS

Results indicate that a statistical package has been produced which allows the determination of the number of significant components in a suitable array of mixtures spectra. The package is shown in the diagram;



It may be that once a uniform experimental method is established, such analyses may be carried out more simply by employing a fixed and reliable criterion giving the most likely rank. Such a system would facilitate the operation of a fully automatic mixtures analysis programme.

The next Chapter considers some methods of detecting the components contributing to each column of the mixtures array.

CHAPTER FIVE

Pattern Separation Techniques

I - Determination of the Components present in Mixtures Peaks

II - Mass Profile Correlation Analysis

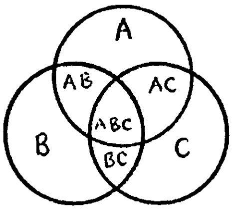
III - Region Peak Analysis

IV - Spectrum Derivation and Refinement

PATTERN SEPARATION TECHNIQUES

I - Determination of the Components present in Mixtures Peaks

The next stage in the analysis according to Monteiro and Reed(49) is a determination of the components contributing to each peak *i.e.* each column in the mixtures array as previously defined. This process will be referred to as *peak analysis* . At this point the determination takes little account of the relative amounts of the components but endeavours to say whether they are present or absent in any column. The maximum number of possibilities for any mixture may be illustrated by a Venn Diagram *e.g.* for a three-component mixture :



i.e. seven distinct mixtures peaks are possible. For a more complex mixture the maximum number increases dramatically,being

$$\sum_{M=1}^{M=N} \frac{N!}{M!\,(N - M)!}$$

for an N-component mixture. This equation is simply a summation performed on the well-known equation giving the number of combinations of items taken M at a time(80).

The method of peak analysis proposed by Monteiro and Reed(49) requires the existence of

(a) at least two uni-component peaks for each component *i.e.* at least two columns of the mixtures array should be unique to each component.

or (b) one uni-component peak per component where an effusiometric method(81) is applicable (see Chapter 6-I).

Once a uni-component peak (column of the array) has been identified for each component further peak analysis is possible

The reason for (a), above, is to allow the detection of the unique peaks as the relevant columns are proportional. The method of analysing other peaks involves the formation of suitable sub-matrices of the array and evaluation of their determinants(49). Before discussing how this is done it is convenient to discuss the reasons for determining which columns are uni-component and carrying out further peak analysis.

A simple illustration of the equations which may be set up to solve a five-component mixture are shown below. The meanings of the arrays are given in (49) and also on page 17 of this thesis

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 \\ \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \gamma_5 \\ \delta_1 & \delta_2 & \delta_3 & \delta_4 & \delta_5 \\ \phi_1 & \phi_2 & \phi_3 & \phi_4 & \phi_5 \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} & m_{13} & \cdots \\ m_{21} & m_{22} & m_{23} & \cdots \\ m_{31} & m_{32} & m_{33} & \cdots \\ m_{41} & m_{42} & m_{43} & \cdots \\ m_{51} & m_{52} & m_{53} & \cdots \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots \\ b_1 & b_2 & b_3 & \cdots \\ c_1 & c_2 & c_3 & \cdots \\ d_1 & d_2 & d_3 & \cdots \\ e_1 & e_2 & e_3 & \cdots \end{bmatrix}$$

or

$$QM' = A'$$

Q is a matrix of unknowns which will give the required normalised component spectra on multiplying into M'. The latter is a data matrix (the only known quantity) whose rows are five distinct mixtures spectra.

A' is a matrix containing the five normalised component spectra.

The important consequence of analysing the columns of M' (M) is that the positions of zeros in A' are immediately known. Every zero found in the first row of A' may be used to form an equation given by the relevant sum of products in Q and M . At least four such equations are necessary to determine α_1, α_2, α_3, α_4 and α_5 (since these are equated to zero). Once the constants have been determined the first row of A' (first component spectrum) is known. The minimum conditions (properties of A' and M') required for the determination are :

(a) at least four (or one less than the rank) of the elements in the first row of A' are zero, and

(b) all other components contribute to this <u>set</u> of peaks <u>i.e.</u> not necessarily to each and every member. This is to ensure that the system of simultaneous equations has maximum rank, there being four unknowns in this case.

These requirements are described in a similar form by Thurstone(82) and by Ainsworth(73).

Part of A' might have the form :

$$\begin{array}{cc}
\begin{array}{ccccccccccccccccccc}
 & & & & & & & \downarrow & & \downarrow & & \downarrow & \downarrow & & & & & & \\
\end{array} & \\
\left[\begin{array}{ccccccccccccccccccc}
\cdot & \cdot & \cdot & \cdot & + & + & + & 0 & + & 0 & + & 0 & 0 & + & + & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & + & + & + & + & + & 0 & + & + & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 & 0 & 0 & + & + & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & + & + & + & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 & + & + & 0 & + & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot
\end{array}\right] & \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array}
\end{array}$$

"+", any number greater than zero.
"0", zero.
".", any number.

The components are A, B, C, D and E. The conditions necessary to determine the first row are satisfied (at indicated masses).

Four equations can now be set up in α_1, α_2, α_3, α_4, and α_5 may be set to -1 since the equations are set to zero. In this way the spectrum of A may be derived.

The simplest way in which both conditions (a) and (b) may be satisfied for all components is to have identified one uni-component peak per component. This may again be illustrated by part of A' :

$$\begin{bmatrix}
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & + & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & + & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 & + & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 & 0 & + & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 & 0 & 0 & + & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot
\end{bmatrix}$$

where four equations satisfying (a) and (b) are possible for each of the five components.

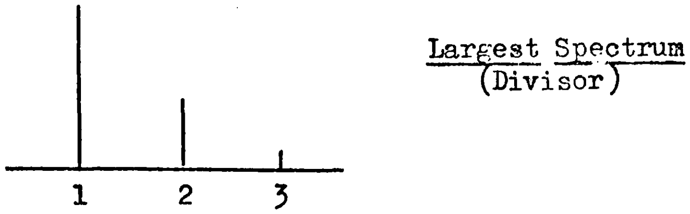
It is easily seen that the important consideration is not necessarily the fact that a peak is uni-component but that both conditions can be shown to be satisfied.

The afore-mentioned Algol programme(48) relies on the detection of proportional columns to locate unique peaks so at least two for each component are necessary. In this present case five such groups would have to be detected or the programme would fail . A short account of the numerical method used will now be given in order to compare it with other possibilities.

The largest mixtures spectrum is placed first and all other spectra divided by it. Where the divisor is too small by comparison with some tolerance the peak is not considered i.e. cannot be analysed. The first column of quotients is then subtracted from all other columns. When a column of zeros is obtained then another mass of the same type has been located i.e. the original columns were proportional. Such detected masses are then eliminated from further consideration as they have been identified as uni-component. The process is begun again by comparing the next available mass (column) with every other remaining mass. Such subtractions are continued until all possibilities have been exhausted and N groups of uni-component peaks found where N is the number of components.

The method was found to work well with calculated data but was unsuitable for use with real data. Some disadvantages of the method are

(1) The use of division to compare the profiles may make analysis of some masses difficult or impossible. Since fractionation is taking place the biggest number at a certain mass may not be in the biggest spectrum, which is placed first as divisor. This difficulty is illustrated by :



The divisor at mass 3 is very small and would lead to large differences being obtained on subtraction of quotients. This mass may be unique to some component but might not be detected as unique , although :

(1) other peaks unique to the same component exist

and 2) there might be significant contributions at this mass number as fractionation proceeds e.g. this component might be relatively involatile taking longer to appear significantly in M.

(2) When a peak is located as a member of a uni-component group it is eliminated from further consideration i.e. it is not possible to compare the uni-component mass profiles within a group. The particular masses being eliminated are regulated by means of a single

arbitrary tolerance value which may be difficult to define experimentally, particularly in view of (1) , above. Considerable difficulty was experienced in detecting the groups of Mixture II by this method.

(3) Only the information from the first mass in each group of uni-component masses is retained and utilized in the subsequent analysis. The programme appears to use the first uni-component peak heights themselves and the corresponding columns of the data matrix to form a matrix like X_u as described on page 16 (Chapter II) i.e. the numerical method used differs from that method described by Monteiro and Reed but is really equivalent to it. The present observation is then the limited amount of information used to determine A and the fact that it is the first unique mass in each group which is taken. In many cases the latter will be the smaller isotope peak in a parent ion group. The use of such relatively small numbers could lead to numerical problems as discussed by Barnard and Fox(6) particularly in view of the experimental fluctuations present.

(4) No other peaks are analysed. This would provide more zeros in each row of A leading to a larger number of equations. Solution of the latter by a suitable least-squares method(84) would produce a more reliable result.

A more practical and flexible approach was then sought in order to remove or reduce some of the above disadvantages. Methods of tackling each of these problems will be discussed and illustrated by experimental results.

The division method of proportional column detection was replaced by a method involving the subtraction of mass profiles normalised to the same sum. In this way all masses could be tested i.e.

the first column was subtracted from all others and the differ-

-ences subjected to some analysis :

1. each was compared with some tolerance value related to the size of the data fluctuations and the differences in sizes of the original profiles.
2. a facility was provided to allow a certain number of differences to be greater than this tolerance.
3. the average absolute deviation for each difference was obtained.

In the case of calculated data all differences would be zero where proportionality exists. In experimental cases both experimental errors and errors produced during the normalisation are introduced. The disadvantage of the method is then where proportional mass profiles have very different sizes i.e. a small peak and a large peak due to the same component.

Columns which were particularly small before normalisation will produce larger deviations from zero on subtraction and so should require larger tolerance values. This allowance was not made in view of the fact that less significance should be given to numbers which were originally quite small and relatively inaccurate.

Difference values obtained for masses 103 and 102 (both unique to n-propyl ether) in Mixture II are given in Table 13 . The second set of values given are for masses 92 and 91 (unique to toluene). Differences for masses 102 and 92 are provided for comparison.

The values were calculated using the sub-routine FILTER. Once a peak is found to be potentially uni-component it is not eliminated from further comparisons. An overall picture of the detection is then obtained. In this way the analyst is able to exercise his judgement in deciding whether a peak is really uni-component or not, perhaps by comparing values with those for peaks which are obviously unique. Thus (2) is satisfied.

Only peaks satisfying certain criteria are stored for printing out and subsequent use. The difficulty of defining these criteria was obvious during examination of peaks in Mixture II. However, employment of generous criteria gave large numbers of possible uni-component peaks i.e. peaks obviously not unique were excluded from consideration. Examination of the set of possible unique peaks allowed the most likely to be chosen.

On a statistical basis the greater the degree of fractionation and the larger the number of distinct spectra the more efficient will be the detection of groups of unique peaks. If these conditions are not well satisfied then many more columns might appear to be proportional considering the fluctuations present in the data.

In attempting to overcome such difficulties (i.e. of defining a tolerance value and comparing it with a number of differences) a new and more direct approach to unique peak detection was successfully investigated. This is discussed in the next Section.

TABLE 13

	DIFFERENCES		
MASSES :	103 with 102	92 with 91	102 with 92
	-9.7	0.5	42.5
	3.6	- 2.0	- 131.4
	2.1	1.0	- 62.8
	21.8	- 0.0	-113.7
	26.5	0.6	- 117.8
	- 3.7	0.7	- 67.3
	6.8	- 1.4	- 64.2
	7.6	1.2	- 47.6
	9.6	-1.0	- 46.0
	4.8	4.1	- 7.5
	21.7	- 7.3	64.4
	- 1.7	1.0	82.9
	- 7.7	- 0.0	65.5
	0.8	0.8	85.1
	-7.1	- 0.0	57.2
	- 11.9	- 0.1	56.6
	-11.8	- 0.4	27.6
	- 3.0	1.6	38.8
	- 9.5	- 0.9	32.4
	- 6.0	- 2.1	31.2
	- 11.7	0.5	25.9
	- 13.1	2.7	25.1
	- 8.4	0.6	23.0
absolute average differences over 23 spectra in each case	9.15	1.32	57.24
	UNIQUE but one profile originally much bigger than other	UNIQUE	NOT UNIQUE

II - Mass Profile Correlation Analysis

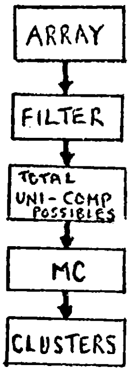
The previous section described methods used to detect proportional columns in the mixtures array. Whereas in Chapter 3 the rows of the mixtures array were correlated this new method of detecting uni-component peaks involves correlating the columns. At least two uni-component peaks per component are still required but the method will be shown to be of greater potential than was first thought.

It is obvious that the correlation coefficients (see Appendix C) of mass profiles unique to the same component will be very high, 1.00000 in numerically exact cases. In experimental situations this could be 0.990 depending upon the size of the particular fluctuations present. Thus, by forming a correlation matrix of the masses and searching for very high values all clusters of uni-component peaks may be identified in one single operation. The degree to which a profile may be considered as having one major component is given simply by the correlation coefficient.

Part of the Mass Profile Correlation Matrix for Mixture II is given in Table 14

Such an analysis is best accomplished where there are more spectra than masses. Thus the method would be particularly valuable where a fast and accurate data acquisition system is available. Large numbers of distinct spectra are, of course, to be preferred for any method of analysis where large experimental errors are involved. By means of a simple technique the method will be shown to be practicable in cases where there are more masses than spectra so this need not necessarily be a serious limitation.

The sub-routine which performs the analysis is called MC . The afore-mentioned sub-routine FILTER may be used initially to exclude all peaks which are obviously multi-component (or uni-component where only one exists). This is done by providing a tolerance large enough to allow some multi-component peaks to be included. Mass correlation analysis may then be applied to the set obtained. Thus FILTER provides a convenient filtering mechanism to avoid peaks being considered by MC which are clearly not proportional to any of the others. The mechanism may be represented diagrammatically thus :



It is reasonable to suppose that any method of pattern separation using all available information will give better results. If equations are formed using all the zeros found in the uni-component peak detection better solutions should be obtained. Least-squares methods of dealing with such systems of equations would be particularly useful. If this is not done the system will tend to be more ill-conditioned i.e. a small amount of numerical information is being used to derive a large amount (the component spectra). This will be discussed in Section IV where spectrum derivation is considered.

Further peak analysis was considered desirable, if not essential, when processing experimental data in order to provide positions of further zero elements in the unknown spectra and hence more simultaneous equations . More zero elements will also help to overcome the ill-conditioning by providing more absolutely determined spectral features.

Such peak analysis, if extensive, could be extremely valuable for identification purposes in its own right. This has been shown by recent work(22,23) involving binary-coded spectra and has to some extent shifted the emphasis of the present study away from a numerical calculation of relative peak heights.

Methods of peak analysis involving determinants are suggested by Monteiro and Reed(49) and by Ainsworth(73), the latter for use with equivalent absorbtion spectral data. They both involve forming suitable sub-matrices of the unique columns of the mixtures array in various combinations with unknown multi-component columns, taken one at a time. If the determinant of such a sub-matrix is found to be zero then the rank of the sub-matrix is less than its smallest dimension. This method is not readily programmable in a general sense and would require the calculation of prohibitively large numbers of determinants. The latter could be readily found by means of the error compensating sub-routines GAUSAV or GAUSP developed in the previous Chapter. The real problem is then the formation of suitable sub-matrices where large numbers of possible combinations present themselves.

A simpler solution to the problem presented itself in utilisation of the large quantity of hitherto unused information present in the mass correlation matrix. A mass profile containing a large contribution from <u>e.g.</u> component B will correlate highly with a mass profile which is known to be unique to B. If the correlation coefficients of a mass profile under test with each unique mass profile be listed (they are all present in the matrix) it is possible to accomplish much valuable peak analysis with very little effort. So far only large correlations have been mentioned. It is obvious that a small correlation means that the mass under test is less likely to contain that particular component as one or two with which it correlates highly.

The listing for a five-component calculated mixture is shown in Table 15 with the corresponding pure-component spectra in Table 16. The listing was produced by sub-routine CDG.

The particular coefficient obtained will depend on the relative amounts of the components present and the particular correlation coefficients of their unique mass profiles. In this case the latter are shown at the foot of Table 15

It appears that in any mixed mass profile a small or negative correlation with a uni-component profile need not necessarily mean an absence of that component. It could mean a relatively small contribution compared with the contribution from another component; or a smaller contribution from the latter where it correlates badly with the uni-component profile. This consideration may be illustrated by reference to mass 19 (Table 15) which has a relatively small contribution from component A compared with component B; its correlation with a profile unique to A is - 0.13 (yet it contains A) and with B it is 0.97.

In Chapter 6-IIIthe use of such correlation coefficients in peak analysis is illustrated by a numerical example.

Part of the COMPONENT DIAGRAM for experimental Mixture II is given in Table 17 . The corresponding spectra are shown in FIGS. 4 and 5. It was clear from the first four columns that ions having m/e 87, 75, 66, 64, 58 and 57 have large contributions from other components, probably impurities i.e. their mass profiles correlate well with each other but badly with the four pure component profiles. The impurity "uni-component peak" was taken as m/e 58 and all correlations included in the fifth column (see note on weighting at end of Appendix C). This result was borne out by the rank analyses described in Chapter 4.

The construction of such a table, taken from the correlation matrix assumes that there are more spectra than masses for best results. By means of a simple device this condition can always be satisfied:

Fewer Spectra than Masses

A mass profile known to be unique to the first component, A, may be correlated with all other masses in groups containing fewer masses than the number of spectra available e.g. if there are three times more masses than the number of spectra, "n", then the correlations may be:

first unique peak with masses 1 to n
first unique peak with masses n+ 1 to 2n
first unique peak with masses 2n+ 1 to 3n

In this way the first column of the COMPONENT DIAGRAM may be produced.

The second column is formed by correlating a mass profile, unique to the second component, with all others in similarly sized groups i.e. less than "n" . The first mass to be considered this time will be that where the significant contribution from the second component appears. This mass may be determined by a more sophisticated application of matrix rank analysis as described in the next section.

The remainder of the COMPONENT DIAGRAM is obtained in the same way.

Once the diagram is complete all small coefficients will indicate small or zero contributions from the relevant components. This is not unfortunate because interest really lies in the identification of zero contributions. Doubts may be cast on a small coefficient by the other coefficients in that row of the diagram which might show a large correlation with another component. Further qualification is obtained by the correlations between the unique mass profiles themselves. Experience gained by forming diagrams for numbers of known mixtures will improve an understanding of the limitations of the method enabling spectral features to be guessed reliably.

TABLE 14

Mass Profile Correlation Matrix - Mixture II

	103	102	93	92	91	90	89	88	87	79	78	77	76	75	74	73	71
103	1.00	0.99	0.44	0.42	0.42	0.46	0.71	0.87	-0.20	0.82	0.82	0.82	0.81	-0.30	0.87	0.99	0.91
102		1.00	0.42	0.44	0.42	0.44	0.70	0.87	-0.25	0.82	0.83	0.83	0.81	-0.30	0.87	0.99	0.91
93			1.00	0.99	0.99	0.99	0.91	0.08	0.21	0.12	0.10	0.17	0.27	0.15	0.49	0.40	0.30
92				1.00	0.99	0.99	0.91	0.08	0.20	0.16	0.10	0.16	0.27	0.14	0.49	0.41	0.30
91					1.00	0.99	0.90	0.07	0.20	0.10	0.10	0.16	0.26	0.14	0.48	0.41	0.29
90						1.00	0.92	0.11	0.21	0.13	0.12	0.18	0.29	0.15	0.51	0.43	0.34
89							1.00	0.46	0.13	0.44	0.43	0.48	0.57	0.06	0.75	0.69	0.62
88								1.00	-0.10	0.90	0.91	0.89	0.90	-0.18	0.86	0.89	0.92
87									1.00	-0.35	-0.36	-0.32	-0.06	0.99	0.07	-0.20	-0.18
79										1.00	0.99	0.99	0.93	-0.40	0.82	0.83	0.87
78											1.00	0.99	0.93	-0.40	0.82	0.85	0.87
77												1.00	0.95	-0.37	0.84	0.84	0.87
76													1.00	-0.01	0.91	0.83	0.88
75														1.00	0.01	-0.30	-0.25
74															1.00	0.89	0.88
73																1.00	0.92

TABLE 15

CORRELATION COEFFICIENTS

	MASS	RELATIVE CONTRIBUTIONS OF THE 5 COMPONENTS IN CALCULATED MIXTURES SPECTRA				
		A	B	C	D	E
1	1,00	1,00000	0,00064	0,53909	-0,70015	-0,31722
2	2,00	1,00000	0,00064	0,53909	-0,70015	-0,31722
3	3,00	1,00000	0,00064	0,53909	-0,70015	-0,31722
4	4,00	1,00000	0,00064	0,53909	-0,70015	-0,31722
5	5,00	1,00000	0,00064	0,53909	-0,70015	-0,31722
6	6,00	1,00000	0,00064	0,53909	-0,70015	-0,31722
7	7,00	0,00064	1,00000	-0,62670	0,48260	0,84552
8	8,00	0,53948	0,84235	-0,23715	0,02901	0,54097
9	9,00	0,04953	0,99880	-0,59958	0,44777	0,82897
10	10,00	0,00064	1,00000	-0,62670	0,48260	0,84552
11	11,00	0,68973	0,72451	-0,08220	-0,13316	0,39356
12	12,00	0,53909	-0,62670	1,00000	-0,90611	-0,82860
13	13,00	0,57024	-0,56631	0,99717	-0,91177	-0,79476
14	14,00	0,79147	-0,42732	0,94040	-0,93607	-0,71170
15	15,00	-0,70015	0,48260	-0,90611	1,00000	0,84104
16	16,00	-0,70015	0,48260	-0,90611	1,00000	0,84104
17	17,00	0,50242	-0,63937	0,99751	-0,87398	-0,81148
18	18,00	1,00000	0,00064	0,53909	-0,70015	-0,31722
19	19,00	-0,01366	0,97239	-0,71712	0,60175	0,91993
20	20,00	0,53909	-0,62670	1,00000	-0,90611	-0,82860
21	21,00	0,08512	0,99253	-0,52693	0,38446	0,79243
22	22,00	-0,56315	0,65944	-0,90972	0,97221	0,94431
23	23,00	-0,31722	0,84552	-0,82860	0,84104	1,00000
24	24,00	-0,31722	0,84552	-0,82860	0,84104	1,00000
25	25,00	0,00064	1,00000	-0,62670	0,48260	0,84552
26	26,00	-0,29966	0,85259	-0,80247	0,82385	0,99892
27	27,00	-0,47896	0,74186	-0,89336	0,93523	0,97804
28	28,00	-0,62885	0,61721	-0,92910	0,98698	0,91084
29	29,00	0,56447	-0,61418	0,99954	-0,91333	-0,82354
30	30,00	0,91352	0,25302	0,36038	-0,43038	0,02743
31	31,00	0,65554	-0,56179	0,98943	-0,93298	-0,79755
32	32,00	0,89020	0,45577	0,18901	-0,39437	0,11048
33	33,00	-0,30811	0,65948	-0,87127	0,89477	0,93144
34	34,00	-0,70015	0,48260	-0,90611	1,00000	0,84104
35	35,00	-0,67693	0,51890	-0,89817	0,99810	0,86797
36	36,00	-0,31722	0,84552	-0,82860	0,84104	1,00000

		SELF CORRELATIONS - INTER-UNIQUE				
2	2,00	1,00000	0,00064	0,53909	-0,70015	-0,31722
7	7,00	0,00064	1,00000	-0,62670	0,48260	0,84552
12	12,00	0,53909	-0,62670	1,00000	-0,90611	-0,82860
15	15,00	-0,70015	0,48260	-0,90611	1,00000	0,84104
24	24,00	-0,31722	0,84552	-0,82860	0,84104	1,00000

TABLE 16

MASS	COMPONENT SPECTRA				
	A	B	C	D	E
1	13.00	0.0	0.0	0.0	0.0
2	111.00	0.0	0.0	0.0	0.0
3	2.30	0.0	0.0	0.0	0.0
4	1.00	0.0	0.0	0.0	0.0
5	19.00	0.0	0.0	0.0	0.0
6	83.00	0.0	0.0	0.0	0.0
7	0.0	44.00	0.0	0.0	0.0
8	14.00	53.00	0.0	0.0	0.0
9	2.00	99.00	0.0	0.0	0.0
10	0.0	32.00	0.0	0.0	0.0
11	77.00	196.00	0.0	0.0	0.0
12	0.0	0.0	88.00	0.0	0.0
13	0.0	14.00	77.00	0.0	0.0
14	47.00	18.00	119.00	0.0	0.0
15	0.0	0.0	0.0	457.00	0.0
16	0.0	0.0	0.0	106.00	0.0
17	0.0	0.0	140.00	27.00	0.0
18	26.00	0.0	0.0	0.0	0.0
19	24.00	149.00	0.0	58.00	0.0
20	0.0	0.0	16.00	0.0	0.0
21	0.0	333.00	24.00	0.0	0.0
22	0.0	0.0	0.0	199.00	114.00
23	0.0	0.0	0.0	0.0	28.00
24	0.0	0.0	0.0	0.0	94.63
25	0.0	65.30	0.0	0.0	0.0
26	0.0	6.00	7.00	5.00	69.00
27	0.0	6.00	0.0	60.00	73.00
28	0.0	27.00	0.0	88.00	0.0
29	2.00	0.0	66.00	0.0	0.0
30	33.00	0.0	22.00	21.00	20.00
31	9.00	0.0	57.00	0.0	0.0
32	81.00	91.00	0.0	0.0	5.00
33	39.00	0.0	0.0	99.50	4.00
34	0.0	0.0	0.0	41.20	0.0
35	0.0	2.00	4.00	59.00	4.00
36	0.0	0.0	0.0	0.0	4.00

TABLE 17

CORRELATION COEFFICIENTS

	MASS	RELATIVE CONTRIBUTIONS OF THE 5 COMPONENTS IN FIRST 90 GOOD SPECTRA				
		n-PROPYL ETHER	TOLUENE.	BENZENE	ETHYL ACETATE	IMPURITIES
1	103,00	0,98870	0,47170	0,76064	0,86640	-0,09460
2	102.00	1,00000	0,46593	0.76277	0,87319	-0,10592
3	93,00	0,48387	0.99450	0,06577	0.07181	0.12499
4	92.00	0,46593	1.00000	0,04513	0,05335	0.14496
5	91,00	0.45159	0,99857	0,03066	0,03678	0,14430
6	90.00	0.50271	0,98848	0,06654	0,09715	0,15809
7	89,00	0,78524	0.85137	0.44193	0,49824	0,12149
8	88,00	0,85414	0.02805	0,90138	0,99391	-0,02796
9	87,00	-0,07521	0,18511	-0.19989	-0.04511	0,99464
10	79,00	0,75155	0.03140	0.98636	0,90448	-0,20056
11	78,00	0.76277	0,04513	1,00000	0.91280	-0,20405
12	77.00	0,75782	0.08493	0.99230	0,89858	-0.18439
13	76,00	0.77717	0,13047	0.98107	0.91168	-0,04842
14	75,00	-0,21487	0,11890	-0.28254	-0.16497	0,99095
15	74.00	0,90363	0,39590	0.86174	0,88687	0,08460
16	73,00	0,99743	0,42609	0,78729	0.89711	-0.10100
17	71,00	0,89251	0,14815	0.84850	0.94435	-0,08904
18	70,00	0,83967	-0.00440	0.91609	0.99682	-0,06212
19	66,00	-0,12552	0,41804	-0,33427	-0,22657	0,92695
20	65,00	0,42745	0.99268	-0,00381	0,01233	0,19730
21	64,00	-0,22661	0,20810	-0,35505	-0.23286	0,97672
22	63,00	0,69693	0,78525	0,41377	0,42831	0.12751
23	62,00	0,78737	0,77432	0,48023	0,54439	0,16741
24	61.00	0,87319	0,05335	0,91280	1.00000	-0.05838
25	60,00	0,81761	0,13237	0.79052	0,93204	0,28444
26	59.00	0,91933	0,47261	0.65175	0.82185	0,26891
27	58,00	-0,10592	0.14496	-0,20405	-0.05838	1.00000
28	57,00	-0,23793	0,11534	-0,33024	-0,19954	0,98894
29	56,00	0.34173	0,32733	0,17480	0.34745	0,88192
30	55,00	0,99383	0.44553	0,77040	0,88415	-0,07675
31	53,00	0,76607	0,58870	0,67348	0.67254	0,29132
32	52,00	0,77693	0,11035	0,98934	0,89575	-0,19552
33	51,00	0,83928	0.33334	0.91515	0.84241	-0,12841
34	50,00	0,80935	0,24678	0.91015	0,84491	-0,13972
35	49,00	0,33772	0,21578	0,35485	0,41901	0,75469
36	46,00	0,88551	0,62427	0,66723	0.76172	0,24338
37	45,00	0,88352	0,07604	0,89032	0,99363	-0,01343

		SELF CORRELATIONS - INTER-UNIQUE				
2	102,00	1,00000	0.46593	0,76277	0,87319	-0.10592
4	92,00	0,46593	1,00000	0,04513	0.05335	0,14496
11	78,00	0,76277	0,04513	1,00000	0,91280	-0,20405
24	61,00	0,87319	0.05335	0,91280	1,00000	-0,05838
27	58,00	-0,10592	0,14496	-0,20405	-0,05838	1,00000

III - Region Peak Analysis

The chances are in any mixture most components will have different molecular ion m/e values or different m/e values of first significant appearance. If a rank analysis is performed on sub-matrices, of a mixtures array M, formed by including increasing numbers of mass profiles down the mass range it is possible to detect the masses at which significant contributions of successive components first occur. This is illustrated in Table 18 by a three-component example.

It is thus possible to determine large numbers of zeros in the unknown component spectra as a direct result of such an analysis. As the GAUSP or GAUSAV sub-routines are ideally suited for use with such variable sizes of sub-matrices they were incorporated into a region peak analysis sub-routine called RPA. This gives a statistical table for each sub-matrix and can give the mass positions where the rank has increased.

The use of this technique also reduces the number of masses to be analysed but the reduction need not be a significant one. This depends on the positions of first significant ions for each component.

If the difference in numbers of mass profiles between the first component and the last is at least one less than the number of components an interesting result obtains i.e.

> the spectrum of the lowest molecular weight component may be derived directly without any knowledge of unique peaks or zeros.

e.g. in Table 18 there are six zeros between the highest mass and the first appearance of the third component. Thus the third component spectrum may be derived directly because the necessary conditions are satisfied <u>i.e.</u> known zeros in the first six masses and all other components contributing to them. An illustration of such a determination will be given in Chapter 6-III.

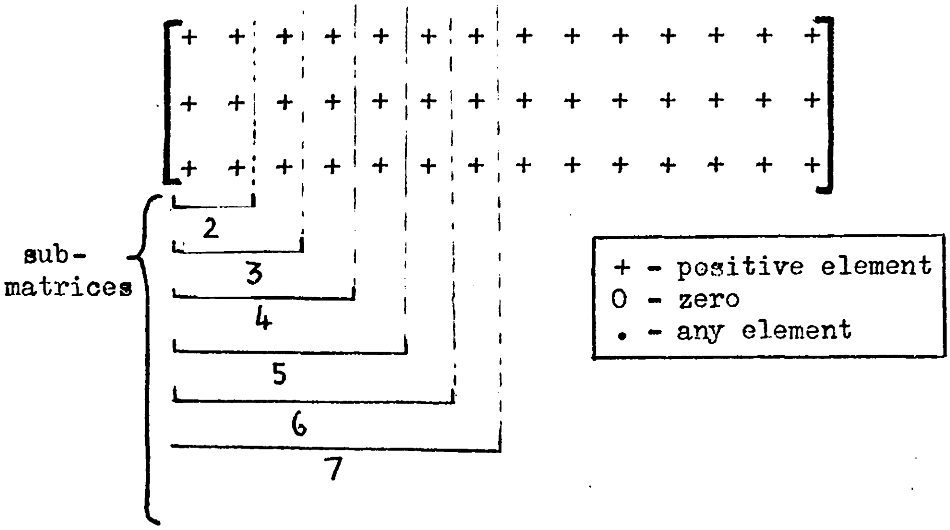
The zeros found by this simple technique may also be used to check and refine spectra derived by other methods.

TABLE 18

UNKNOWN COMPONENT SPECTRA

$$\begin{array}{c} A \\ B \\ C \end{array} \begin{bmatrix} + & + & + & + & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & + & + & + & + & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & + & + & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

CORRESPONDING MIXTURES SPECTRA



+ - positive element
0 - zero
. - any element

RESULTS OF RANK DETERMINATIONS

submatrix	rank
1	1
2	1
3	1
4	2
5	2
6	2
7	3
8	3
.	3
.	3

IV - Spectrum Derivation and Refinement

Derivation

The Algol programme(48) method will not be described in detail, suffice to say that

(a) the peak analysis and spectrum derivation sections are tied together.

(b) only a little information, the first unique peak in each component group, is used.

(c) N mixtures spectra are chosen during the rank determining step as being the first N spectra which are distinct and not necessarily the most distinct spectra in the array. The more different are the spectra used to form the equations the more significant will be the results. It is clear that the N most distinct spectra (patterns) should be chosen.

The derivation method adopted here incorporates the following :

(1) choice of the most distinct spectra.

(2) utilization of all known zero elements to form equations.

A sub-routine called UNIQS is provided with all clusters of unique peaks and uses a least-squares method to solve all the equations it constructs.

A similar sub-routine called ZEROS is provided with the positions of all zeros in the unknown spectrum where these are known from uni-component peaks or other sources.

Both sub-routines use a method which calculates the matrix Q , as previously described, and produce normalised component spectra. An example of the use of ZEROS is given in Chapter 6-III. Full details of both sub-routines are to be found in Appendix B.

Any spectra derived as a result of either method should roughly agree with the pattern of correlation coefficients in the corresponding column of the COMPONENT DIAGRAM (Section II). This can therefore serve as a semi-quantitative check on any spectra produced.

The most distinct mixtures spectra are recognised as having the most diverse correlation coefficients given by the sub-routine FRACT (see Chapter 3).

Refinement

Refining methods are based on subjecting the derived spectra to some check. The spectra may be mixed together in proportions dictated by their unique peak sizes to form an array M''. The difference between this and the original matrix, M', gives a residue matrix R.

$$R = M' - M''$$

An examination of R at each mass number gives an indication of the errors present. These may be considerable, anyway, if only one unique peak is used to re-combine each derived spectrum; several should be tried, if possible. The sub-routine RESIDU will produce R for any set of spectra derived and also gives the sum of the deviations at each mass number. Any method of refinement will involve minimisation of the residues. Some refining was carried out by a sub-routine REFINU which is given the masses present in each cluster of unique peaks and alters the spectra accordingly i.e. where spurious numbers appear in positions which should contain zeros.

The particular mixtures spectra used (M') are shown in Table 19 and the component spectra derived using the following unique peaks :

from	masses	no. of equations
n-propyl ether	102,73	4
toluene	55,92	4
ethyl acetate	61	5
benzene	78	5

Only these mass profiles were judged to be suitable for inclusion in the simultaneous equations by virtue of their sizes.

The matrix Q derived from the equations is shown in Table 20 together with the derived spectra. Table 21 shows the spectra after application of REFINU *i.e.* introduction of zeros given by unique peak positions (including unique peaks having profiles considered numerically too small for inclusion in UNIQS).

The matrix R given by RESIDU for the refined spectra, using masses 102, 92, 78 and 61, are given in Table 22. The differences between this case and the residues obtained using the original derived spectra are slight except at the unique mass positions. (original residues are not shown). The right hand column lists the absolute sums of the deviations for each mass. The fact that most of the deviations are negative illustrates the defiency in using only one unique peak per component in RESIDU. Perhaps the deviations could be reduced by:

(a) trying a variety of unique peaks in RESIDU.
(b) altering the unique peak heights relative to all others.
(c) a different choice of N mixtures spectra originally, where N is the rank of the array.

Pure component spectra were obtained for the four major components under almost identical conditions and are shown beside the calculated spectra in FIGS. 4 and 5.

It was found that the best results were obtained where only the larger unique mass profiles were used *i.e.* where the inherent fluctuations have less significance. This will also be considered to some extent in the next Chapter where methods of analysing "more difficult" mixtures arrays are considered.

TABLE 19

mass	mixtures spectra			
103,0	1,63	4.87	7,67	2,36
102.0	14,25	59,53	87.60	24,32
93.0	2.78	2.37	9.47	4.57
92,0	30.82	26,07	121,03	56,19
91,0	50,80	43.03	194,13	92,62
90,0	1,87	1.83	6,50	3,23
89,0	3.02	4,00	10.77	4.53
88,0	5,90	17.27	13,70	3,62
87,0	0,90	2.30	2,97	1.53
79,0	9,63	10,87	9.10	2,58
78.0	136,60	150.80	129.30	31,01
77.0	29,60	28,50	28.33	7.74
76.0	7,72	8,60	7.93	2.58
75,0	3.55	4,37	5.60	2,81
74,0	9.23	12,07	18.17	5,76
73,0	49.83	191.70	262,43	72,62
71.0	1.83	4,80	4,67	1.54
70,0	17,67	51,60	37.47	8,77
66.0	1.55	1.40	3,80	2.21
65.0	7,82	6.77	28.03	13.74
64.0	1.73	1,67	4.93	2,74
63,0	10,27	11.20	26.00	10,94
62.0	4,13	7.33	11,27	5,17
61.0	30,32	89,00	73.10	17.32
60.0	2.60	7,10	6.10	1,99
59,0	8.77	34,17	46,07	14.12
58.0	2.27	5,60	6.60	2.70
57,0	3,07	8,10	16.33	7,39
56,0	1.50	3.03	3.93	1.59
55.0	11.48	41.50	57.20	16,29
53.0	2,57	2.73	4.10	1.89
52,0	31,72	34.43	31,77	9.68
51,0	32.10	35,17	42,83	14,92
50.0	27,50	31,57	34,67	11.82
49,0	4,75	5.40	5,97	2.28
46,0	2,43	3.83	6,07	2,56
45,0	35,00	106,07	87.30	23,41

TABLE 20

MATRIX Q

$$\begin{bmatrix} -0.01 & -0.23 & 0.53 & -1.00 \\ -0.00 & -0.08 & 0.34 & -1.00 \\ 0.14 & -0.37 & 0.52 & -1.00 \\ 0.51 & -0.32 & 0.41 & -1.00 \end{bmatrix}$$

mass	component spectra			
103.0	2.61	0.57	0.43	0.20
102.0	38.28	0.90	9.11	0.21
93.0	0.76	5.11	1.17	0.02
92.0	6.16	57.13	9.94	2.22
91.0	3.86	100.00	6.50	1.36
90.0	1.24	3.88	2.28	0.43
89.0	0.87	4.02	0.04	0.37
88.0	2.54	1.35	16.66	0.99
87.0	2.54	2.33	5.85	1.32
79.0	2.23	1.34	3.94	5.99
78.0	0.05	0.02	0.32	100.00
77.0	0.62	1.88	5.57	22.63
76.0	2.51	2.00	4.31	4.31
75.0	4.54	4.15	8.26	0.17
74.0	4.34	2.08	4.30	5.89
73.0	100.00	0.02	2.39	0.09
71.0	1.13	1.18	5.17	0.47
70.0	6.50	1.16	47.60	1.49
66.0	2.71	3.39	4.40	0.68
65.0	3.04	15.78	5.05	0.87
64.0	2.70	3.95	4.56	0.81
63.0	0.14	10.11	1.20	3.24
62.0	4.85	6.38	11.80	1.69
61.0	0.03	0.08	64.48	0.13
60.0	2.29	1.66	8.75	0.90
59.0	10.22	4.56	13.17	3.47
58.0	2.73	3.02	8.34	1.36
57.0	3.42	8.26	12.11	3.79
56.0	1.19	1.66	3.72	0.38
55.0	19.82	1.42	2.70	0.14
53.0	1.96	2.38	3.31	0.55
52.0	6.94	5.94	11.05	19.65
51.0	4.83	11.07	8.92	17.98
50.0	6.34	8.94	12.67	14.67
49.0	2.24	2.31	4.05	2.02
46.0	1.38	2.68	3.91	0.08
45.0	13.04	8.62	100.00	7.23

TABLE 21

mass	refined component spectra			
103.0	2.61	0.0	0.0	0.0
102.0	38.28	0.0	0.0	0.0
93.0	0.0	5.11	0.0	0.0
92.0	0.0	57.13	0.0	0.0
91.0	0.0	100.00	0.0	0.0
90.0	1.24	3.88	2.28	0.43
89.0	0.87	4.02	0.04	0.37
88.0	0.0	0.0	16.66	0.0
87.0	2.54	2.33	5.85	1.32
79.0	0.0	0.0	0.0	5.99
78.0	0.0	0.0	0.0	100.00
77.0	0.62	1.88	5.57	22.63
76.0	2.51	2.00	4.31	4.31
75.0	4.54	4.15	8.26	0.17
74.0	4.34	2.08	4.30	5.89
73.0	100.00	0.0	0.0	0.0
71.0	1.13	1.18	5.17	0.47
70.0	6.50	1.16	47.60	1.49
66.0	2.71	3.39	4.40	0.68
65.0	0.0	15.78	0.0	0.0
64.0	2.70	3.95	4.56	0.81
63.0	0.14	10.11	1.20	3.24
62.0	4.85	6.38	11.80	1.69
61.0	0.0	0.0	64.48	0.0
60.0	2.29	1.66	8.75	0.90
59.0	10.22	4.56	13.17	3.47
58.0	2.73	3.02	8.34	1.36
57.0	3.42	8.26	12.11	3.79
56.0	1.19	1.66	3.72	0.38
55.0	19.82	0.0	0.0	0.0
53.0	1.96	2.38	3.31	0.55
52.0	6.94	5.94	11.05	19.65
51.0	4.83	11.07	8.92	17.98
50.0	6.34	8.94	12.67	14.67
49.0	2.24	2.31	4.05	2.02
46.0	1.38	2.68	3.91	0.08
45.0	13.04	8.62	100.00	7.23

TABLE 22

mass	deviations				sum
103,0	0,66	0,81	1.70	0.70	3,87
102,0	0.00	0.00	0.00	0.00	0.00
93,0	0.03	0,04	-1.36	-0.46	1,88
92,0	0.00	0.00	0,00	0,00	0,00
91,0	-3,14	-2.59	-17,73	-5.73	29.20
90,0	-2,35	-5.66	-7.69	-2,11	17,82
89,0	0,00	0,20	-0.26	-0,10	0,57
88.0	-1.93	-5.72	-5.18	-0.85	13.69
87,0	-5,85	-12,76	-16.11	-4.35	39.07
79.0	1,46	1.84	1.36	0,72	5,37
78,0	0,00	0.00	0.00	0,00	0,00
77.0	-5,17	-15,13	-12,64	-3.01	35.95
76.0	-2,21	-8,66	-12,50	-3.48	26,86
75,0	-4,50	-16,24	-23.16	-6,43	50.33
74,0	-3,57	-10.45	-8.67	-2,03	24.72
73,0	12,61	36,17	33.58	9.08	91.43
71,0	-2.30	-5,34	-6.89	-1.87	16.40
70,0	-9,79	-26.98	-35,75	-9.75	82.27
66.0	-4.28	-11.45	-15.44	-4.23	35,41
65,0	-0,69	-0.43	-5,39	-1,77	8,29
64.0	-4,66	-11.86	-15.84	-4.33	36.69
63,0	-0.23	-0.17	-1.29	-0.42	2,10
62.0	-8,97	-21.95	-28.91	-7.89	67.72
61,0	0.00	0.00	0.00	0.00	0,00
60,0	-4,50	-10,67	-13.76	-3.73	32,65
59,0	-8,43	-7,22	-6.41	-1.47	23,53
58,0	-6,16	-13,59	-17.27	-4.67	41.69
57,0	-13,54	-23.43	-27,62	-7.34	71,92
56,0	-2,10	-5.27	-7,00	-1,91	16.28
55,0	4,10	10,67	11.83	3,69	30.29
53,0	-1,75	-6.78	-9.87	-2,75	21.15
52,0	-6,11	-23.95	-34.63	-9.64	74,33
51.0	-4,42	-16.82	-25,02	-7,00	53.26
50.0	-5.68	-21,98	-32,12	-8,96	68,74
49.0	-1,99	-7,78	-11,27	-3,14	24,17
46,0	-1,47	-5.05	-7,31	-2,03	15,87
45,0	-31,39	-67,27	-83,51	-22.45	204,42

FIGURE 4.

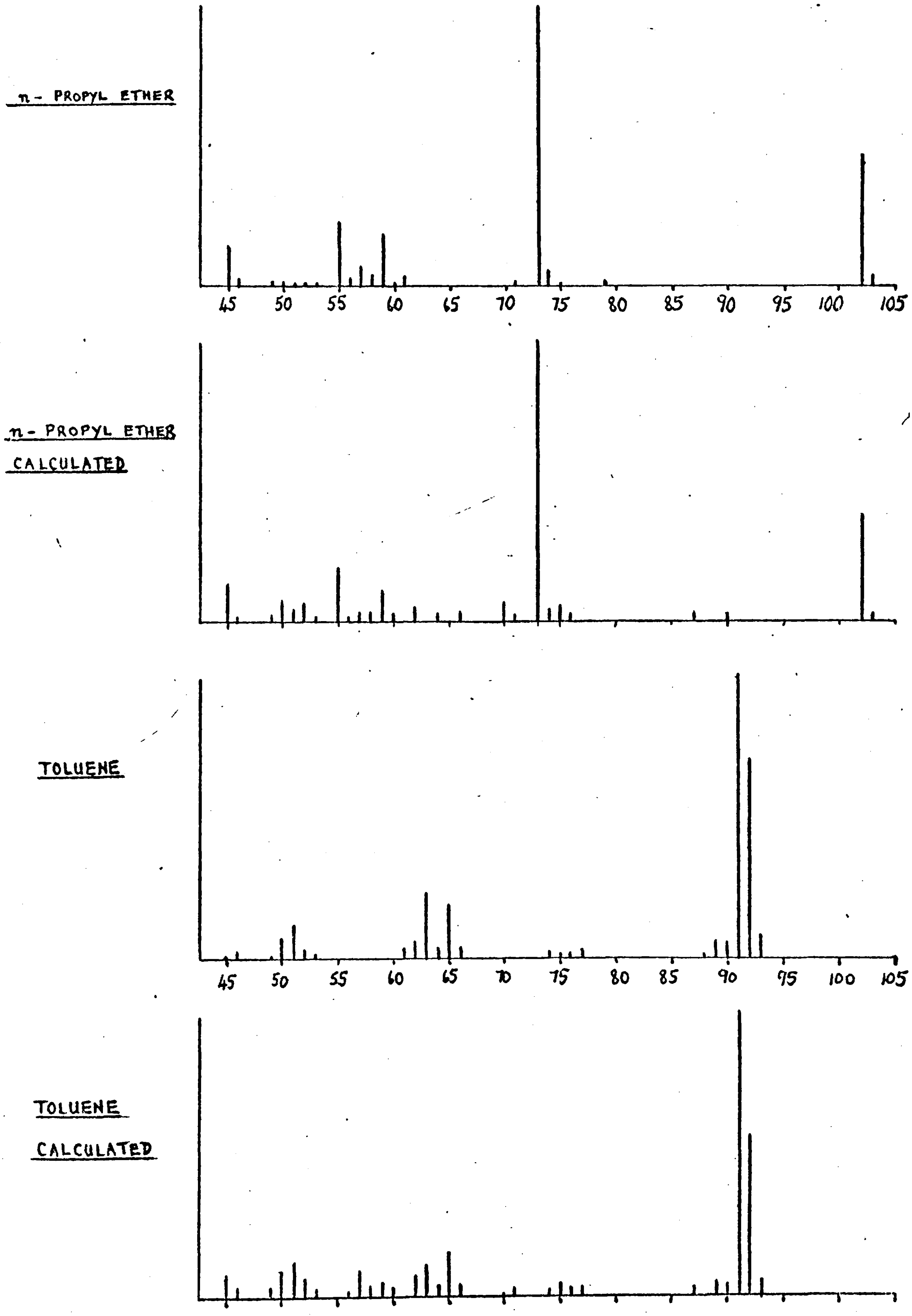
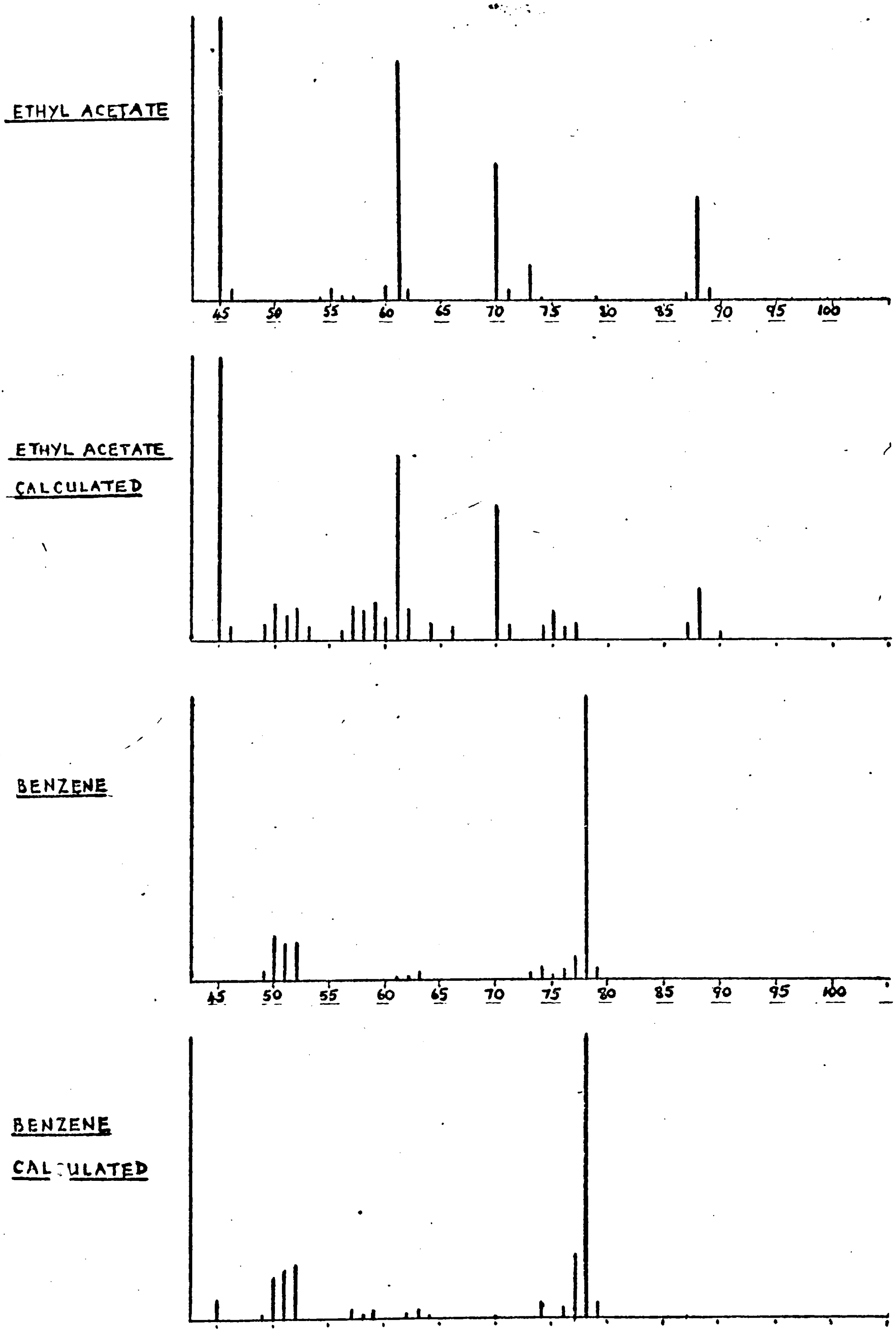

n- PROPYL ETHER
45
50
55
60
65
70
75
80
85
90
95
100
105
n- PROPYL ETHER
CALCULATED
TOLUENE
45
50
55
60
65
70
75
80
85
90
95
100
105
TOLUENE
CALCULATED

FIGURE 5.



ETHYL ACETATE
45
50
55
60
65
70
75
80
85
90
95
100
ETHYL ACETATE CALCULATED
BENZENE
45
50
55
60
65
70
75
80
85
90
95
100
BENZENE CALCULATED

CHAPTER SIX

I - Extension of Mixtures Analysis Theory

II - Notes on Quantitative Analysis

III - Example of the Application of the UNRAVL Procedure

I - Extension of Mixtures Analysis Theory.

As discussed by Monteiro and Reed(49) effusiometry(81) could provide a method, albeit a difficult one, of detecting a peak unique to a component where only one such peak exists. A logarithmic plot of its abundance with time will yield a straight line where a leak has been used in the inlet - system. The method is therefore only applicable where the components are sufficiently volatile to allow their entry into the ion-source via a leak. It is also apparent that if there is only one peak unique to a particular component there might just as easily be none (apart, perhaps, from the highest mass peak).

A more complex example of this type of approach has been given by Grigsby and Cole(85) where graphs of peak height vs. time were plotted and fitted to a polynomial by the involved least-squares technique of Sillén(86). This was done for materials distilled from the direct insertion probe and its application to the problem, where a leak could not be used, was considered. It was claimed that the components present in some peaks could be identified but this was not their aim. However, such a method would still be experimentally exacting and the necessary computer programming difficult and time consuming.

Both of the above methods would still require at least one unique peak for each component. The simplest method at present(49) requires at least two unique peaks per component for a direct analysis.

It was thought by Reed(87) that the problem should be capable of a unique solution where there are two peaks unique to one component and one peak unique to each of the others.

One can imagine situations in which one component has several unique peaks, another component has one unique peak and some other component none at all. A method is then required which is not experimentally exacting , uses all the information it has derived and yields a complete or partial solution when it can proceed no further.

The following is a simple illustration of the development of such a method. A generally applicable system is devised which greatly reduces the current restrictions. For statistical reasons the method functions more reliably where there are large numbers of spectra. This is true of any method of mixtures analysis where the data contains inherent fluctuations.

As discussed in Chapters 2 and 5 , unique peak detection may be considered as a convenient method of detecting zeros in the unknown spectra and ensuring that condition (b) of Chapter 2 is satisfied. Any method of reducing the present restrictions must necessarily involve an alternative method of doing this. One possible solution to this problem presented itself in Chapter 5-II where mass profile correlations were introduced.

The first column of a COMPONENT DIAGRAM can be obtained without reference to groups of unique peaks if it is assumed that the highest mass peak in the mixture is unique to the first component. This is done by calculating correlation coefficients of the first mass profile (highest mass or unique peak where known) with all others. Similarly any other column of the COMPONENT DIAGRAM is found by correlating another unique profile with all other profiles. The tendency of a component to be absent from a particular mass is given by a small coefficient, subject to certain qualifications (see Chapter 5-II).

In order to simplify the description and development it will be assumed that the highest mass in the mixture is unique to one of the components, A, the others being B, C, and D , all in an array, M, of rank four. No peaks have been detected as unique to any of these components <u>i.e.</u> there could be one peak unique to any of them but these cannot be detected since at least two are required.

In order to calculate the spectrum of A it is necessary to:

(a) detect as many mass numbers as possible to which A does not contribute (at least one less than the number of components), and

(b) be sure that all other components contribute to this set of masses.

If sufficient numbers of zeros are detected then the chances are that condition (b) is satisfied. This latter condition is necessary to ensure that the system of simultaneous equations has maximum rank, as previously discussed.

The positions of zeros in the spectrum of A may be found in two ways:

(1) As before, utilizing peaks found unique to other components (does not apply in this particular case).

(2) Using information derived from the COMPONENT DIAGRAM. In this particular case the first column is known.

As discussed in Chapter 5-IV the zeros detected in this way may be used by sub-routine ZEROS to derive the spectrum of A. Since the positions of many of the zeros may have been guesses it is best that several combinations be used in order to check the accuracy of the result i.e. if the derived spectrum changes significantly on leaving out a zero or several zeros then a wrong choice has been made. The process may be repeated until a reliable spectrum is obtained i.e. one formed using several different sets. Further details are given in Section III. The sub-routine TRIAL produces spectra for any number of combinations of such zeros.

Once the spectrum of the first component is derived it may be subtracted from each mixtures spectrum in the array M by taking account of its unique peak sizes, in this case the first peak. This may lead to error build-up as in the subtraction technique of Kiser(10) in conventional mixtures analysis. Where several unique peaks are available the accuracy of the subtraction procedure may be improved.

If U_a is a matrix whose rows are the spectra of A as they appear in M then a matrix M' is obtained on subtraction:

$$M' = M - U_a$$

The rows of U_a are obtained by considering the size of the unique peak in each row of M. The matrix M' has rank one less than M and it is apparent that all peaks previously containing A contain it no longer. This important result means that mixtures peaks which originally contained A in combination with one other component now contain only the latter <u>i.e.</u> are now uni-component.

The minimum condition necessary to derive the second component, B, is that at least two of the columns of M originally contained contributions from both A and B only. On subtraction two peaks unique to B would be left and detected by the "perfect" correlation of their mass profiles (see Chapter 5-II and sub-routine MC).

Once a peak unique to B is found the absence of this component in all other peaks may be indicated, as before, by correlating the mass profiles. Zeros thus found may be used to derive a consistent spectrum of B. Formation of a matrix U_b by considering the unique peaks in M', and its subtraction would yield a new mixtures array M'' :

$$M'' = M' - U_b$$

M'' has rank two less than M. Peaks which previously contained <u>B and C</u> or <u>A,B and C</u> would now be unique to component C and could be readily identified by mass profile correlation provided there were at least two of them.

The spectrum of C may now be derived and subtraction from M'' should yield the spectrum of D, the remaining component:

$$M''' = M'' - U_c$$

i.e. M''' contains the spectra of D as they appear in M.

The above development represents the worst possible case and in practice many more unique peaks would probably come to light very early in the process, before error build-up could seriously affect the results. As soon as unique peaks are detected more zeros become known and also the relevant columns of the COMPONENT DIAGRAM may be constructed. As discussed in Chapter 5-II the more columns of this which are available the more accurate will be the initial guesses as to the positions of zeros. Further zeros may be detected by the Region Peak Analysis as described in Chapter 5-III.

The method is then capable of application in any situation between the following limits:

(a) several unique peaks per component, and

(b) one peak unique to the first component if this of highest mass.
OR
two peaks unique to the first component

AND

two peaks containing a first component and second component.
two peaks containing a first, second and third component.
two peaks containing a first, second, third and fourth component.
etc.

The accuracy of a determination would decrease between (a) and (b) but can always be checked in some way:

(1) using sub-routine RESIDU to recombine spectra utilizing several unique peaks, perhaps.

(2) using various combinations of zeros to examine the stability of the derived spectrum (see Section III, sub-routine TRIAL).

(3) observing the values calculated in a spectrum which should be zero.

(4) **comparing** the spectral pattern with the corresponding column of the COMPONENT DIAGRAM (see Section III).

The method described above has been partially programmed and an example using a calculated five-component mixture containing only one single unique peak will be described in Section III. The system of sub-routines has been named UNRAVL .

The next section contains a few notes on quantitative analysis illustrated by the above example.

II - Notes on Quantitative Analysis

An approximate quantitative distribution of the spectra in M may be obtained by considering the matrices U_a , U_b , U_c and $U_d \equiv M'''$. If ΣU_a is the sum of all the peaks in U_a and $\sum m$ is the sum of all peaks in M then the percentage contribution of the spectrum of A to M, by total ion current, is given by

$$\frac{\sum U_a}{\sum m} \cdot 100$$

The percentage of component N is similarly given by

$$\frac{\sum U_n}{\sum m} \cdot 100$$

Results may be checked by summation as the sum of the ion currents of all the components as they appear in M should equal the sum of the ion currents in M, <u>i.e.</u>

$$\sum m = \sum U_a + \sum U_b + \sum U_c + \sum U_d$$

The method does not, of course, apply only to UNRAVL but may be used where the component spectra and unique peaks are known.

A proper quantitative analysis of the mixture may be made by knowing the sensitivities of the components.

III - Application of the UNRAVL procedure to solve a five-component system of mixtures in which there is one single peak unique to only one of the component spectra.

The method is applicable to any case between the limits discussed in Section I of this Chapter.

Five pure-component spectra, A - E, as illustrated by the columns in Table 23A were mixed together in different proportions to give forty distinct mixtures spectra, five of which are shown in Table 23B. All forty mixtures spectra were stored on a disk-file and were numerically exact i.e. no errors were associated with them.

A mass profile correlation matrix was formed using sub-routine MC and no groups of at least two proportional profiles detected (such profiles would have correlation coefficients of exactly 1.00000 since the mixtures were calculated). The first mass profile was then assumed to be unique to the first component (highest mass), A , and its correlations with all other masses examined (Table 23C). The most negative and smallest values were taken to represent those masses least likely to contain any contribution from A i.e. detection of the most likely positions of zero elements in the spectrum of A. These mass numbers were re-arranged in order of likelihood of containing no A, i.e.

MASS	COEFFT.
9.0	- 0.17903
4.0	- 0.13083
6.0	- 0.12714
18.0	- 0.11924
5.0	- 0.08600
10.0	- 0.05407
15.0	0.01660
21.0	0.03069
14.0	0.07436
23.0	0.12671
.	.
.	.
etc.	

Two of the masses listed above i.e. "5.0" and "14.0" have contributions from A but this is not known from the coefficients. If all the masses contained no A and all such zero positions ($\geqslant 4$) were used in various combinations to form equations and hence spectra of A, all such spectra would be identical - provided that the rank of each system of equations is four, as discussed in Chapter 5-I, condition (b).

If a peak containing some contribution from A is used to form an equation in such a combination a deviant, erroneous spectrum will be produced.

A sub-routine called TRIAL was written to produce spectra for any desired number of combinations of zeros guessed from the correlation coefficients. Five representative mixtures spectra were chosen (in the case of experimental data FRACT would be employed to pick the best spectra as previously discussed). Some of the spectra derived for the above example are listed in Table 24 together with the particular masses used in each case. From the whole set tried it was apparent that masses "5.0" and "14.0" were "bad" i.e. combinations containing them produced spectra deviating greatly from the majority. The sub-routine ZEROS, similar to the afore-mentioned UNIQS, is called by TRIAL and uses the zeros to derive a spectrum. A warning is given when the system of equations has rank less than the number of unknowns e.g. combination of "4.0", "6.0", "18.0" and "23.0" had rank three since no contribution from E is present (Table 24). In later work employment of GAUSAV or GAUSP is recommended.

A spectrum produced by TRIAL can be easily checked because the pattern of peaks should be roughly similar to the pattern of correlation coefficients (in this case Table 23C). It is seen that the deviant spectra are very different.

Once a consistent spectrum is found the corresponding zero positions, in this case at masses "9.0", "4.0", "6.0" and "18.0", are fed to a subroutine called UNRAVL. In dealing with experimental data any number greater than four could be used to give a least-squares solution. TRIAL and UNRAVL can accept any number.

UNRAVL uses a peak unique to this component to subtract the derived spectrum from all forty mixtures spectra. A matrix U is formed as described in Section I and subtracted from the original mixtures array, M, forming M'. The existence of large negative peaks at various masses in a real situation would indicate errors and allow adjustment of relevant peak heights in the derived spectrum. Estimates of errors may also be made by performing UNRAVL using various unique peaks where these are known.

The first five mixtures spectra in M' are shown in Table 25A. The first mass profile is, as expected, reduced to zero and the rank of the mixtures array is now four. UNRAVL also writes M' on to the same disk-file, replacing M, and a new mass profile correlation analysis is carried out.

Examination of the correlation matrix revealed two clusters of unique peaks i.e. unique to two components designated B and C. Peaks present in each cluster were:

B - "2.0" , "3.0" , "20.0"
C - "14.0" , "17.0"

The relevant correlation coefficients for these masses are given in Tables 25B and 25C .

The masses in these Tables were re-arranged in order of least likelihood of containing their respective components:

B

MASS	COEFFT.
10.0	- 0.06539
16.0	- 0.06122
9.0	- 0.05311
*17.0	- 0.04804
*14.0	- 0.04804
5.0	- 0.01626
etc.	

C

MASS	COEFFT.
* 2.0	- 0.04804
* 3.0	- 0.04804
* 20.0	- 0.04804
12.0	0.01771
7.0	0.02040
13.0	0.02882
etc.	

The starred masses indicate that the information was also available from the cluster of peaks found unique to the other component. This illustrates the fact that as the analysis proceeds any new information can be used to improve its accuracy.

Since the rank of M' is four only three suitable equations are necessary to derive the spectrum of B or C.

Derivation of B and C Spectra

Positions of zeros indicated by the above Tables were used by TRIAL to find consistent spectra, as before.

The most consistent spectrum of B is given in Table 26A and compares well with the coefficients in Table 25B , as a check. Once again, suitable spectra (four, this time) were chosen at random although this would be best done by FRACT. Note was taken of the masses giving this derived spectrum in TRIAL.

The spectrum of C can be obtained in two ways:

1/ derived immediately from M' in the same way as the spectrum of B, or

2/ the spectrum of B may be taken from M' giving M'' (using UNRAVL). New clusters of peaks unique to C would be found on application of MC. In the present case clusters would also be found for components D and E i.e. enough information would be available to derive spectra of C , D and E directly from M'' in the usual way.

In the present case of numerically exact mixtures the results of 1/ and 2/ would be almost identical so 1/ will be described as an illustration of the procedure in a less favoured example.

The whole procedure is illustrated by the block diagram in FIG. 6 and can be accomplished by the sub-routines listed at the end of APPENDIX B. Both 1/ and 2/ , above, would be carried out i.e. a mass correlation matrix is listed after subtraction of each component. The system is then quite flexible allowing decisions to be made at each step.

The spectrum of C derived is given in Table 26B and compares well with the coefficients in Table 25C. Note was made of the masses used to derive this spectrum.

Sub-routine UNRAVL was given this mass information for components B and C, derived each spectrum in turn and subtracted them from M' giving an array M''' of rank two. In this case UNRAVL is activated twice by the parameter "NREDS" being set equal to two. During the earlier subtraction of A this parameter was set to one.

Part of M''' is shown in Table 26C . Mass correlation analysis revealed two clusters of uni-component profiles. One was due to component D and the other to E. The masses were :

D	E
6.0	10.0
7.0	11.0
8.0	13.0
12.0	16.0
19.0	24.0
23.0	

These masses are equivalent to those which would be found as the highest and lowest ratios in Meyerson's method of solving binary mixtures(43). Actual coefficients are listed in Tables 27A and 27B. M''' then contains forty binary mixtures spectra. They may be separated by Meyerson's method or by an equivalent UNRAVL method using two spectra. The latter has the advantage that all the uni-component peaks are used at the same time to give an averaged solution. Two spectra were chosen from M'''. The zeros in the spectrum of D are at masses "10.0", "11.0", "13.0", "16.0" and "24.0" . Similarly the zeros in the spectrum of E are those at masses unique to D. Submission of both sets of masses to UNRAVL (NREDS = 2) produced the spectra of D and E shown in Tables 27C and 27D , respectively.

TRIAL and UNRAVL as described in APPENDIX B and FIG. 6 can be applied to any situation where fewer than N clusters of uni-component peaks are detected, the rank of the array being N. The limitations are discussed in Section I.

The unravelling process at each stage in the above example is illustrated by the peak compositions in Table 28.

Determination of the spectrum of E immediately from M

If the first eight masses are introduced into TRIAL as having zeros in the spectrum of E the latter is immediately derived and may be used to check the spectrum of E as derived during the last stage of UNRAVL. The zeros in this case may be obtained by employment of sub-routine RPA as discussed in Chapter 5-III. A spectrum of E derived from M, by chance, is the ninth column in Table 24 . It is identical to the spectrum given by the first eight zeros and may be compared with the spectrum in Table 27D i.e. the spectrum given by UNRAVL. The slight differences may be accounted for by :

(1) round-off (see Chapter 4-I)
(2) errors introduced by the subtraction process.
(3) perturbations in the derived spectra caused by the rank of the system of simultaneous equations being affected by the chance presence of a relatively minor component i.e. at the particular masses considered one of the components may be present to a lesser extent than the others. This reinforces the desirability of using as many masses as possible when forming the equations. Such perturbations are noticeable on comparison of spectra from TRIAL.

Application of UNRAVL may prove useful in separating the superimposed spectra of pyrolysis products, where the existence of two unique peaks per component is less likely. This would also apply to species being formed thermally in the ion-source.

FIGURE 6.

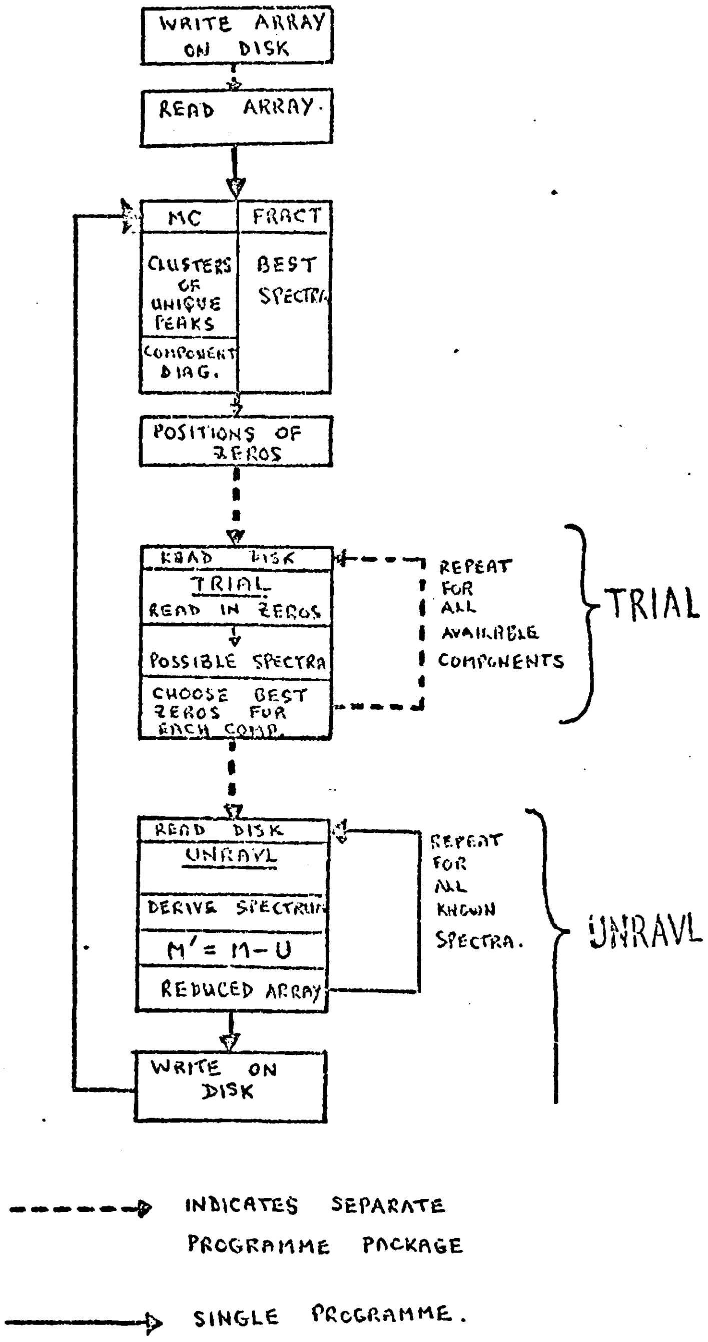

WRITE ARRAY ON DISK
READ ARRAY.
MC
FRACT
CLUSTERS OF UNIQUE PEAKS
COMPONENT DIAG.
BEST SPECTRA
POSITIONS OF ZEROS
READ DISK
TRIAL
READ IN ZEROS
POSSIBLE SPECTRA
CHOOSE BEST ZEROS FOR EACH COMP.
REPEAT FOR ALL AVAILABLE COMPONENTS
TRIAL
READ DISK
UNRAVL
DERIVE SPECTRUM
M′ = M − U
REDUCED ARRAY
REPEAT FOR ALL KNOWN SPECTRA.
UNRAVL
WRITE ON DISK
INDICATES SEPARATE PROGRAMME PACKAGE
SINGLE PROGRAMME.

TABLE 23

A.

A	B	C	D	E	MASSES
411.000	0.000	0.000	0.000	0.000	1.00
14.000	253.000	0.000	0.000	0.000	2.00
65.000	22.000	0.000	0.000	0.000	3.00
0.000	14.000	77.000	0.000	0.000	4.00
50.000	18.000	560.000	0.000	0.000	5.00
0.000	0.000	140.000	27.000	0.000	6.00
0.000	111.000	0.000	840.000	0.000	7.00
140.000	333.000	0.000	840.000	0.000	8.00
0.000	0.000	666.000	7.000	99.000	9.00
0.000	0.000	90.000	0.000	860.000	10.00
0.000	560.000	65.000	0.000	25.000	11.00
100.000	22.000	0.000	50.000	0.000	12.00
0.000	19.000	0.000	0.000	16.000	13.00
20.000	0.000	77.000	0.000	0.000	14.00
0.000	45.000	190.000	99.000	20.000	15.00
70.000	0.000	0.000	0.000	165.000	16.00
90.000	0.000	200.000	0.000	0.000	17.00
0.000	20.000	90.000	0.000	0.000	18.00
120.000	0.000	0.000	55.000	0.000	19.00
580.000	100.000	0.000	0.000	0.000	20.00
0.000	0.000	0.000	25.000	100.000	21.00
0.000	155.000	23.000	99.000	50.000	22.00
0.000	0.000	80.000	110.000	0.000	23.00
77.000	23.000	0.000	0.000	111.000	24.00
190.000	460.000	80.000	0.000	0.000	25.00

B.

250.710	423.330	41.100	295.920	427.440
71.790	67.550	16.580	63.792	115.735
45.150	71.570	7.820	51.471	76.398
80.777	3.371	0.848	34.542	8.309
597.016	58.416	6.136	269.421	78.910
143.204	6.454	3.049	74.140	24.368
111.750	199.710	101.076	544.365	649.189
252.650	390.530	128.396	692.296	883.567
749.515	26.435	41.909	357.590	60.163
788.902	184.974	356.651	733.500	278.368
225.542	123.327	43.974	165.788	234.232
71.500	118.120	16.940	107.671	148.798
17.747	7.422	7.775	16.994	12.718
89.477	21.031	2.008	45.970	23.510
228.080	35.594	22.141	165.034	102.363
176.729	107.492	75.425	184.050	125.600
255.620	93.820	9.020	146.800	100.640
95.324	4.704	1.209	41.146	11.166
78.700	135.150	18.182	120.500	164.400
378.800	618.400	64.000	438.830	643.190
83.730	26.700	44.280	96.500	50.000
112.348	64.194	41.165	144.217	150.074
91.288	23.548	12.372	101.000	82.016
142.885	107.949	55.112	150.233	124.798
311.188	292.748	46.608	267.258	384.370

C.

1.00000
0.32376
0.95771
-0.13083
-0.08600
-0.12714
0.37499
0.51974
-0.17903
-0.05407
0.25137
0.93198
0.19129
0.07436
0.01660
0.35485
0.25535
-0.11924
0.95925
0.98748
0.03069
0.31856
0.12671
0.56990
0.55467

TABLE 24

1	2	3	4	5	6	7	8
70.86	50.68	70.86	70.86	70.86	23.53	70.86	70.86
2.40	20.33	2.42	2.41	2.42	21.49	2.42	2.41
11.21	9.63	11.21	11.21	11.21	5.52	11.21	11.21
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8.63	0.00	8.62	8.62	8.62	3.99	8.62	8.62
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
0.04	66.41	0.00	0.00	0.00	73.83	0.00	0.00
24.09	100.00	24.14	24.14	24.14	100.00	24.14	24.13
0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.05	71.95	0.00	0.00	0.00	80.00	0.00	0.01
0.02	42.44	0.01	0.00	0.01	47.19	0.00	0.00
17.24	17.42	17.24	17.24	17.24	11.38	17.24	17.24
0.00	2.76	0.00	0.00	0.00	3.07	0.00	0.00
3.45	1.44	3.45	3.45	3.45	0.00	3.45	3.45
0.01	9.33	0.00	0.00	0.00	10.38	0.00	0.00
12.06	22.67	12.07	12.07	12.07	19.61	12.07	12.07
15.52	8.42	15.52	15.52	15.52	2.18	15.52	15.52
0.00	0.27	0.00	0.00	0.00	0.30	0.00	0.00
20.69	18.61	20.69	20.69	20.69	11.11	20.69	20.69
100.00	78.87	100.00	100.00	100.00	41.38	100.00	100.00
0.01	10.24	0.00	0.00	0.00	11.38	0.00	0.00
0.01	22.21	0.00	0.00	0.00	24.69	0.00	0.00
0.00	6.56	0.00	0.00	0.00	7.29	0.00	0.00
13.27	20.63	13.28	13.28	13.28	16.79	13.28	13.27
32.74	56.19	32.76	32.76	32.76	47.30	32.76	32.76

9	10	11	12
0.01	42.49	70.86	70.86
0.01	1.45	2.40	2.41
0.00	6.72	11.21	11.21
0.00	0.00	0.00	0.00
0.00	5.17	8.63	8.62
0.0	0.00	0.00	0.00
0.03	0.03	0.04	0.00
0.03	14.48	24.09	24.14
11.51	11.51	0.01	0.00
100.00	100.00	0.04	0.00
2.92	2.91	0.02	0.00
0.00	10.34	17.24	17.24
1.86	1.86	0.00	0.00
0.00	2.07	3.45	3.45
2.33	2.33	0.00	0.00
19.19	26.42	12.08	12.07
0.00	9.31	15.52	15.52
0.00	0.00	0.00	0.00
0.00	12.41	20.69	20.69
0.01	59.97	100.00	100.00
11.63	11.63	0.00	0.00
5.82	5.81	0.01	0.00
0.00	0.00	0.00	0.00
12.91	20.87	13.28	13.28
0.01	19.64	32.74	32.76

Spectrum	Masses used
1	9 4 6 18
2	9 4 6 5
3	9 4 6 10
4	9 4 6 15
5	9 4 6 21
6	9 4 6 14
7	9 4 6 23
8	9 4 6 13
9	4 6 18 14
10	4 6 18 23
11	4 6 18 13
12	9 4 6 18 10 15 21 23 and 13

TABLE 25

A.

1.0	0.00	0.00	0.00	0.00	0.00
2.0	63,29	91.12	306.19	86.03	2.55
3.0	5.50	7.92	26.63	7.48	0.22
4.0	80.78	12.76	17.89	5.04	43.68
5.0	566.50	62.63	28.64	8.13	316.79
6.0	143,20	14.58	29.53	4.28	86.72
7.0	111.60	56.59	999.25	155.30	236.21
8.0	167.44	136.89	1268.44	230.87	238.66
9.0	749.51	67.93	19.36	84.90	378.63
10.0	788.73	17.43	35.21	708.49	51.80
11.0	225.46	208.27	679.26	211.20	42.32
12.0	10.51	8.93	78.14	14.48	14.23
13.0	17.74	6.99	23.62	19,63	0.21
14.0	77.27	7.72	0.94	0.28	43.53
15.0	228.06	37,41	159.52	46.31	135.61
16.0	134.06	1,69	6.66	135,89	0.22
17.0	200.71	20.05	2.45	0.72	113.08
18.0	95.32	16.23	25.31	7,12	51.09
19.0	5.51	1.11	56.67	7.70	15.41
20.0	25.01	36.02	121,02	34.00	1.01
21.0	83.71	1.47	29.71	85.84	7.10
22.0	112.30	60.53	291.72	107.81	42.30
23.0	91.27	10.21	114.26	15.68	76.02
24.0	95.94	9.42	32.32	99.24	0.38
25.0	195.35	173.70	557.69	156.70	49.88

B. **C.**

MASS 2,00		MASS 14,00
0,20628	1,0	-0,21717
1,00000	2,0	-0,04804
1,00000	3,0	-0,04805
0,13082	4,0	0,98398
-0,01626	5,0	0,99949
0,00072	6,0	0,98967
0,47308	7,0	0,03985
0,67021	8,0	0,02040
-0,05311	9,0	0,98956
-0,06539	10,0	0,20075
0,99156	11,0	0,07388
0,69380	12,0	0,01771
0,74420	13,0	0,02882
-0,04804	14,0	1,00000
0,27155	15,0	0,88867
-0,06122	16,0	0,09923
-0,04804	17,0	1,00000
0,16948	18,0	0,97625
0,33377	19,0	0,05077
1,00000	20,0	-0,04804
0,00173	21,0	0,10791
0,88593	22,0	0,12432
0,20237	23,0	0,71137
0,14238	24,0	0,08865
0,98467	25,0	0,12689

TABLE 26

A.

0,0
45,21
3,93
2,50
3,22
0,06
17,97
57,76
0,00
0,00
100,00
3,83
3,39
0,00
7,82
0,03
0,00
3,57
0,11
17,87
0,06
27,46
0,24
4,13
82,19

B.

0,0
0,00
0,00
11,56
84,08
21,02
0,00
0,00
100,00
13,52
9,76
0,00
0,00
11,56
28,53
0,00
30,03
13,51
0,00
0,00
0,00
3,45
12,01
0,00
12,01

C.

1,0	0,0	0,0	0,0	0,0
2,0	-0,00	0,00	0,00	0,00
3,0	-0,00	0,00	-0,00	0,00
4,0	0,00	0,00	0,00	-0,00
5,0	-0,00	0,00	0,00	0,00
6,0	2,63	0,43	27,44	3,67
7,0	86,41	20,36	877,51	121,10
8,0	86,58	20,47	877,24	120,95
9,0	81,13	1,15	11,21	82,51
10,0	698,40	8,41	34,11	708,16
11,0	20,24	0,19	1,14	20,66
12,0	5,15	1,22	52,22	7,20
13,0	12,99	0,16	0,64	13,18
14,0	-0,00	-0,00	-0,00	-0,00
15,0	26,43	2,60	104,21	30,74
16,0	134,02	1,64	6,49	135,84
17,0	0,00	0,00	0,00	0,00
18,0	-0,00	0,00	0,00	0,00
19,0	5,35	0,88	55,89	7,48
20,0	-0,00	-0,00	0,00	0,00
21,0	83,63	1,36	29,33	85,74
22,0	50,78	2,88	105,44	55,46
23,0	10,65	1,70	111,65	14,94
24,0	90,16	1,10	4,36	91,38
25,0	-0,00	0,00	0,00	0,00

TABLE 27

A. MASS 6.00		B. MASS 10,00
0,0	1,00	0,0
0,14591	2,00	0,07837
-0,06544	3,00	0,24879
0,22161	4,00	-0,07736
0,24512	5,00	-0,29681
1,00000	6,00	-0,04109
0,99957	7,00	-0,04255
0,99959	8,00	-0,04251
0,01268	9,00	0,99855
-0,04109	10,00	1,00000
-0,04036	11,00	0,99989
0,99959	12,00	-0,04252
-0,04103	13,00	1,00000
-0,19052	14,00	-0,10980
0,96581	15,00	0,21774
-0,04113	16,00	1,00000
-0,02120	17,00	0,38412
0,20805	18,00	-0,15989
1,00000	19,00	-0,04108
-0,04410	20,00	-0,09314
0,14569	21,00	0,98251
0,82652	22,00	0,52789
1,00000	23,00	-0,04103
-0,04114	24,00	1,00000
-0,07173	25,00	0,20197

C.		D.
0.0	1,00	0.0
0,00	2,00	0,00
0,00	3,00	0,00
0,0	4,00	0.0
0,00	5,00	0,00
2,98	6,00	0.15
99,21	7,00	0,22
100.00	8,00	0,34
0,94	9,00	11,43
0,03	10,00	100,00
0,42	11,00	3.21
5,95	12,00	0,02
0,01	13,00	1.87
0.0	14,00	0.0
11,70	15,00	2.35
0,17	16,00	19.06
0,0	17,00	0,0
0,0	18,00	0,0
6,06	19,00	0,31
0,0	20,00	0,0
2,71	21,00	11,79
11,58	22,00	5,92
11,99	23,00	0,71
0,11	24,00	12,82
0,00	25,00	0,00

TABLE 28

Peak Compositions at Each Stage of the Unravelling Procedure

Mass	M	M'	M''	M'''	M''''
1	A	0	0	0	0
2	AB	B	0	0	0
3	AB	B	0	0	0
4	BC	BC	C	0	0
5	ABC	BC	C	0	0
6	CD	CD	CD	D	0
7	BD	BD	D	D	0
8	ABD	BD	D	D	0
9	CDE	CDE	CDE	DE	E
10	CE	CE	CE	E	E
11	BCE	BCE	CE	E	E
12	ABD	BD	D	D	0
13	BE	BE	E	E	E
14	AC	C	C	0	0
15	BCDE	BCDE	CDE	DE	E
16	AE	E	E	E	E
17	AC	C	C	0	0
18	BC	BC	C	0	0
19	AD	D	D	D	0
20	AB	B	0	0	0
21	DE	DE.	DE	DE	E
22	BCDE	BCDE	CDE	DE	E
23	CD	CD	CD	D	0
24	ABE	BE	E	E	E
25	ABC	BC	C	0	0

APPENDIX A - Spectrum Measurement Details

APPENDIX B - Computer Programmes

APPENDIX C - Statistical Notes

APPENDIX A

The method of preparing tapes from the d-mac "pencil-follower" will be illustrated by an example in which three simple spectra of six masses each are measured. The third mass in case has been measured on the second galvanometer scale (x10) on ultra-violet sensitive paper and the fifth mass on the third scale (x30).

A print-out of the tape (with explanations) is given below together with the corresponding computer programme.

The programme ignores all "x" co-ordinates though use could be made of these to fix mass numbers. Repetition of base-line digitisations is made each time to allow termination of the process when desired (removal of the perspex rectangle).

In practice charts were prepared beforehand by ringing those groups of peaks to be digitised. During digitisation several carriage returns were introduced (via the key-board) after each group. The patterns produced on printing out tapes (flexowriter) were used to check for missing numbers and extra digitisations.

The programme listed below is written in Egdon Algol for use with the English Electric KDF9 computer which has a tape-reading facility. Base-line "y" co-ordinates are subtracted from relevant peak co-ordinates to yield peak heights which are scaled up as necessary. A FORTRAN IV sub-routine called CARD is then called up to print out the spectra and punch them on to cards in any desired format.

A diagrammatic plan of the "pencil-follower" is given in FIG. A.

	MEANING	D = DIGITISATION
£DATA	CONTROL CHARACTERS	
1000;	EXPERIMENT NUMBER	
10;	FIRST SCALE FACTOR	
30;	SECOND SCALE FACTOR	
1;	SPECTRUM NUMBER	
1600;1000;	D ON FIRST BASE-LINE	X;Y;
1600;800;	D ON SECOND BASE-LINE	
1600;200;	D ON THIRD BASE-LINE	
2000;1543;	D - TOP OF FIRST PEAK	
1800;1250;	D - " " SECOND "	
-1;1600;930;	-1; SIGNIFIES D ON SECOND BASELINE	
1400;2000;	D	
-10;1200;360;	-10; SIGNIFIES D ON THIRD BASE-LINE	
1000;1100;	D	
9999;	SIGNIFIES END OF A SPECTRUM	
2;	SPECTRUM NUMBER	
1600;999;	ETC.	
1600;799;		
1600;199;		
2000;2000;		
1800;1260;		
-1;1600;370;		
1400;1100;		
-10;1200;270;		
1000;1275;		
9999;		
3;	SPECTRUM NUMBER	
1600;1000;	ETC.	
1600;800;		
1600;200;		
2000;1760;		
1800;1290;		
-1;1600;960;		
1400;2010;		
-10;1200;360;		
1000;1143;		
999999;	SIGNIFIES END (FINAL SPECTRUM)	
£CARD+++++++++++++	CONTROL CHARACTERS	

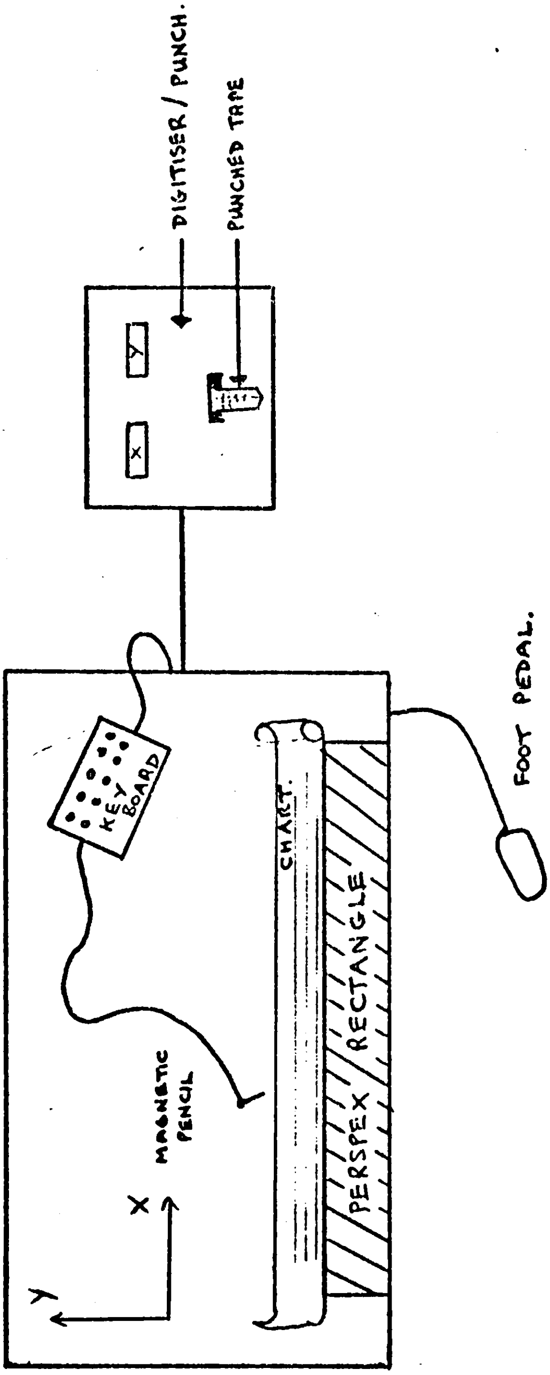
```
'BEGIN'
'REAL'
F,
Y,Y3,
YB,
F1,F2,Y1,Y2,DIS,X;
'INTEGER'
FM,
LABEL,
   I,J,K,NEXP,FMT;
'INTEGER' 'ARRAY'
 SLAB(1'.'20);
'ARRAY'
 MASS(1'.'150),
 PEAK(1'.'20,1'.'150);
FM=LAYOUT('('NDDD.DD')');
FMT=LAYOUT('('4S-NDDDD.DDD')');
WRITET(70,'('READ TEST TAPE 3 SPECTRA AND BASELINES')');
LABEL=READ(20);
OUTPUT(70,LABEL);
F1=READ(20);
OUTPUT(70,F1);
F2=READ(20);
OUTPUT(70,F2);
 WRITET(70,'(''('CC')'')');
I=1;
L1'.'
  SLAB(I)=READ(20);
OUTPUT(70,SLAB(I));
DIS=READ(20);
Y1=READ(20);
OUTPUT(70,Y1);
DIS=READ(20);
Y2=READ(20);
OUTPUT(70,Y2);
DIS=READ(20);
Y3=READ(20);
OUTPUT(70,Y3);
WRITET(70,'(''('CCC')'')');
WRITET(70,'(''('CCC')'')');
WRITE(70,LAYOUT('('NDDD')'),SLAB(I));
 WRITET(70,'(''('CC')'')');
'FOR'  J=1  'STEP' 1 'UNTIL'  150 'DO'
'BEGIN'
X=READ(20);
   OUTPUT(70,X);
'IF' ABS(X-999999) 'LT' 0.01 'THEN'
'GOTO' L5;
'IF' ABS(X-9999) 'LT' 0.01 'THEN'
'BEGIN'
I=I+1;
'GOTO' L1;
'END';
'IF' X 'LT' 0.0 'AND' X 'GT' -2.0 'THEN'
'BEGIN'
YB=Y2;
F=F1;   'GOTO' L4;
'END';
'IF' X 'LT' -5.0 'AND' X 'GT' -12.0 'THEN'
'BEGIN'
```

CONT'D
OVERLEAF

```
YB=Y3;
F=F2;     'GOTO' L4;
'END';
Y=READ(20);
    OUTPUT(70,Y);
PEAK(I,J)=(Y-Y1);
 OUTPUT(70,PEAK(I,J));
'GOTO' L3;
L4'.'
DIS=READ(20);
Y=READ(20);
    OUTPUT(70,Y);
PEAK(I,J)=F*(Y-YB);
 OUTPUT(70,PEAK(I,J));
L3'.'
'END';
L5'.'
NEXP=I;
OUTPUT(70,NEXP);
K=J-1;
    OUTPUT(70,K);
WRITET(70,'(''('CC')''')');
CARD(PEAK,NEXP,K);
'END'
```

```
EGTRAN COMPILER          MARK NO. 302          DATE    25/07/73
      SUBROUTINE CARD(X,N,M)
      DIMENSION X(20,150)
   10 FORMAT(10F7.1)
   11 FORMAT(10F10.2)
      DO 100 I=1,N
  100 PUNCH 10,(X(I,J),J=1,M)
      DO 200 J=1,M
  200 PRINT 11,(X(I,J),I=1,N)
      RETURN
      END
```

FIG. A.


DIGITISER / PUNCH.
PUNCHED TAPE
Y
X
KEY BOARD
CHART.
PERSPEX RECTANGLE
FOOT PEDAL.
MAGNETIC PENCIL
X
Y

APPENDIX B.

Although considerable time was spent writing and developing the programmes it cannot be claimed that the methods used are the best with regard to saving of time and storage space.

It has not been the purpose of this work to produce a single computer programme but rather a number of useful sub-routines which can be conveniently joined by the user depending on his particular needs. The present system allows the analyst to interact with the data and bring his judgement to bear on the problem. Several examples of calling programmes are given.

In one case use was made of COMMON statements to reduce storage requirements (rank analysis on Mixture II). It is obvious that greater use could be made of such statements in future.

The programmes were not written by a specialist and the particular style used is very simple. Anyone with a knowledge of FORTRAN IV should experience little difficulty in following them. It cannot be claimed that the best possible numerical methods have been used but the system is now a very flexible one.

Array Dimensions. In each case the array dimensions as written at the beginning of each programme should be equal to or greater than the actual dimensions used. It is important to ensure that corresponding arrays in a series of sub-routines called by one programme have identical dimensions.

In the sub-routines FRACT, MC, CDG, and PCA each of which call up an I.B.M. Library sub-routine, CORRE, the dimension statements must contain exact dimensions unless the sub-routine ARRAY is also used. An example of the use of ARRAY in variable dimensioning is given in sub-routines UNIQS and ZEROS. Details of the storage methods used are given in the I.B.M. manual(71). CORRE computes means, standard deviations and product-moment correlation coefficients.

List of Sub-routines

Name	Reference	Page No.
PRA	general	123
PRAG	general	123
SELM	general	124
SELECT (MASSES)	general	125
example: Use of SCREEN	Chapter 3	126
SCREEN	"	127,128
FRACT	general	129,130
DATA	general	130
DIFF	Chapter 3	131,132
MAV	"	133
MSTATS	"	134,135
example: Use of GAUSP	Chapter 4-I	136
GAUSP	"	137
GAUSS	"	138,139
ELIM	"	140
PIVOT	"	141
GAUSAV	"	142
PCA	Chapter 4-II	143
FILTER	Chapter 5-I	144,145
MC	Chapter 5-II	146
CDG	"	147,148
RPA	Chapter 5-III	149
example: Use of UNIQS	Chapter 5-IV	150
data listing: Mixture II	"	151
UNIQS	"	152,153
ARRAY (double precision)	"	154
RESIDU	"	155
REFINU	"	156
example: Use of TRIAL	Chapter 6-III	157
TRIAL	"	158
ZEROS	"	159,160
SEL	general	160
example: Use of UNRAVL	Chapter 6-III	161
UNRAVL	"	162

```
      SUBROUTINE PRA(A,NEXP,NOM)
      DIMENSION A(28,142)
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C               INPUT - MATRIX A, NEXP ROWS, NOM COLUMNS
C               SUBROUTINE PRINTS A
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
    1 FORMAT(12F10.3,//)
    2 FORMAT(12F10.3)
    3 FORMAT(/////)
    4 FORMAT(//)
      JL=0
      JT=0
      JH=0
  100 JT=JT+1
      JL=JH+1
      JH=JT*12
      JH=MIN0(JH,NOM)
      PRINT 4
      DO 200 I=1,NEXP
      PRINT 2,(A(I,J),J=JL,JH)
  200 CONTINUE
      IF(JH.LT.NOM)GO TO 100
      PRINT 3
      RETURN
      END
```

```
      SUBROUTINE PRAG(A,NEXP,NOM,M,N)
      DIMENSION A(28,142)
C               INPUT-
C               MATRIX A HAVING NEXP ROWS AND NOM COLUMNS
C               OUTPUT-
C               ROWS M TO N ARE PRINTED OUT
C
    1 FORMAT(12F10.3,//)
    2 FORMAT(12F10.3)
    3 FORMAT(/////)
    4 FORMAT(//)
      JL=0
      JT=0
      JH=0
  100 JT=JT+1
      JL=JH+1
      JH=JT*12
      JH=MIN0(JH,NOM)
      PRINT 4
      DO 200 I=M,N
      PRINT 2,(A(I,J),J=JL,JH)
  200 CONTINUE
      IF(JH.LT.NOM)GO TO 100
      PRINT 3
      RETURN
      END
```

```
      SUBROUTINE SELM(A,NOS,NOM,K)
C
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C          PROGRAM ELIMINATES CERTAIN SPECTRA FROM DATA MATRIX
C          GIVING CONTRACTED ARRAY.
C
C          INPUT - A, OF NOS ROWS AND NOM COLUMNS.
C          READS IN FORM(I4) INTEGERS WHICH ARE ROW NUMBERS OF
C          UNDESIRED SPECTRA.
C          FINAL CARD CONTAINS 0.
C          OUTPUT - CONTRACTED ARRAY A WITH K ROWS.
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      DIMENSION A(133,37)
      DOUBLE PRECISION A
    1 FORMAT(I4)
      K=1
      L=0
      READ 1,IL
      DO 100 I=1,NOS
      K=I-L
      IF(I.EQ.IL)GOTO 120
      DO 110 J=1,NOM
  110 A(K,J)=A(I,J)
      GOTO 100
  120 CONTINUE
      L=L+1
      READ 1,IL
  100 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE SELECT(A,NOM,NOC)
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C          THIS PROGRAM FORMS A CONTRACTED ARRAY 'A' BY
C          CHOICE OF CERTAIN ROWS (SPECTRA) FROM 'A',
C
C          INPUT -
C          A, INITIAL DATA MATRIX,
C          NOM, NO. OF COLUMNS OF A
C          NOC, NO. OF ROWS SELECTED,
C          READS IN -
C          KVS, VECTOR CONTAINING NOC INTEGERS GIVING ROW
C          NUMBERS SELECTED. FORMAT(12I4)
C
C          SUBROUTINE 'MASSES' IS SIMILAR BUT COLUMNS ARE
C          SELECTED.
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      DIMENSION A(23,37),KVS(10)
      DOUBLE PRECISION A
    3 FORMAT(12I4)
      READ 3,(KVS(I),I=1,NOC)
      DO 100 I=1,NOC
      DO 100 J=1,NOM
      K=KVS(I)
  100 A(I,J)=A(K,J)
      RETURN
      END
```

```
      DIMENSION A(28,142),AT(142,28),KVG(28)
C
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C          EXAMPLE OF USE OF 'SCREEN'.
C          REF.     MIXTURE V.
C
C          PROGRAM READS IN -
C          NOS, NUMBER OF ROWS IN A(SPECTRA).
C          NOM, NUMBER OF PEAKS (COLUMNS OF A).
C          NOG, NUMBER OF GROUPS OF "IDENTICAL " SPECTRA.
C          KVG, VECTOR CONTAINING ROW NUMBERS OF FINAL
C                  SPECTRA IN EACH GROUP.
C          A, MIXTURES SPECTRA.
C          A IS PRINTED OUT, TRANSPOSED AND PASSED
C                  TO 'SCREEN'.
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
    1 FORMAT(I4)
    2 FORMAT(10F7.1)
    3 FORMAT(12I4)
      READ 1,NOS
      READ 1,NOM
      READ 1,NOG
      READ 3,(KVG(I),I=1,NOG)
      DO 100 I=1,NOS
      READ 2,(A(I,J),J=1,NOM)
  100 CONTINUE
      CALL PRAG(A,NOS,NOM,1,NOS)
C
C          TRANSPOSE MIXTURES ARRAY.
C
      DO 150 I=1,NOS
      DO 150 J=1,NOM
  150 AT(J,I)=A(I,J)
      CALL SCREEN(A,AT,NOS,NOM,KVG,1)
      CALL EXIT
      END
```

```
      SUBROUTINE SCREEN(A,AT,NOS,NOM,KVG,ITI)
      DIMENSION A(28,142),AM(28,142),AT(142,28),FC(28)
      DIMENSION KVG(28),TIC(28)
C
C-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
C             PURPOSE-
C             THIS PROGRAM PROCESSES MIXTURES SPECTRA FROM
C             A MASS SPECTROMETER. THE SPECTRA ARE IN
C             THE FORM OF GROUPS OF "IDENTICAL" SPECTRA
C             WHICH ARE EXAMINED TO REVEAL
C                    POSSIBLE GROSS ERRORS.
C             INPUT-
C             A IS MIXTURES ARRAY.
C             AT IS TRANSPOSE OF MIXTURES ARRAY.
C             NOS IS NUMBER OF SPECTRA I.E. ROWS OF A.
C             NOM IS NUMBER OF PEAKS I.E. COLUMNS OF A.
C             KVG IS A VECTOR CONTAINING ROW NUMBERS OF ENDS
C             OF EACH EXPTL. GROUP.
C             IF ITI=1 A FULL P/O OF EACH EXPTL. GROUP
C                     IS GIVEN.
C             IF ITI=0 A FULL P/O IS GIVEN ONLY WHERE GROSS
C                ERRORS ARE DETECTED.
C             OUTPUT-
C             CORRELATION COEFFICIENTS OF ALL SPECTRA.
C             LOCATIONS OF GROSS ERRORS IN SPECTRAL GROUPS.
C                 SPECTRA IN EACH GROUP ARE NORMALISED TO
C             THE SUM OF THE LARGEST ION CURRENT.
C             MATRIX AN THUS FORMED IS PASSED TO 'DIFF'.
C             OTHER SUBROUTINES REQUIRED-
C             PRAG,FRACT,DIFF,DATA
C-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
C
    1 FORMAT(I4)
    5 FORMAT(1X,'FULL P/O EVERY GROUP',///)
    6 FORMAT(1X,'P/O SUSPECT GROUPS ONLY',///)
      CALL FRACT(AT,NOM,NOS,KVG,FC)
    7 FORMAT(1X,'SUBMATRIX IN ORIGINAL FORM',//)
      IF(ITI.EQ.0)GOTO 51
      PRINT 5
      GOTO 55
   51 PRINT 6
   55 CONTINUE
      K=1
      M=1
   50 N=KVG(K)
      NUMGP=N-M+1
C
C        FINDS LARGEST SPECTRUM IN GROUP BY SUM.
C
      BIG=0.0
      DO 130 I=1,NUMGP
      L=M+I-1
      TIC(I)=0.0
      DO 140 J=1,NOM
      TIC(I)=TIC(I)+A(L,J)
  140 CONTINUE
```

```
            IF(TIC(I).GT.BIG)BIG=TIC(I)
        130 CONTINUE
      C
      C        NORMALISATION OF OTHER MEMBERS TO THIS SUM,
      C
            DO 150 I=1,NUMGP
            L=M+I-1
            DO 150 J=1,NOM
        150 AN(I,J)=A(L,J)*BIG/TIC(I)
            CALL DIFF(AN,NUMGP,NOM,FC,M,ITI,ITO)
            IF((ITO.EQ.0).AND.(ITI.EQ.0))GOTO 170
            PRINT 7
            CALL PRAG(A,NUMGP,NOM,M,N)
        170 IF(N.EQ.NOS)GOTO 500
            M=N+1
            K=K+1
            GOTO 50
        500 RETURN
            END
```

```
      SUBROUTINE FRACT(X,N,M,KVG,FC)
      DIMENSION X(142,28)
      DIMENSION B(200),FC(200),S(200),T(200),XBAR(200),KVG(200)
      DIMENSION V(8000),R(8000)
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C             INPUT-
C             X IS THE TRANSPOSE OF THE CONVENTIONAL
C             DATA MATRIX. THERE MUST BE MORE MASSES
C             THAN SPECTRA I.E. N>=M. DIMENSIONS OF X
C             IN STATEMENT MUST BE (N,M) UNLESS 'ARRAY'
C             IS USED.
C             N IS THE NUMBER OF MASSES.
C             M IS THE NUMBER OF SPECTRA,
C             KVG CONTAINS ROW NUMBERS OF FINAL
C             SPECTRA IN EACH GROUP.
C
C             OUTPUT-
C             FC CONTAINS CORRELATION COEFFICIENTS OF
C             FIRST SPECTRUM WITH ALL OTHERS IN ORIGINAL
C             ARRAY, GROUPED ACCORDING TO EXPYL. GROUPS,
C
C             SUBROUTINES DATA AND PRAG ARE REQUIRED.
C             CORRE IS AN I.B.M. SCIENTIFIC SUBROUTINE
C             IT CALCULATES THE PRODUCT-MOMENT CORRELATION
C                     COEFFICIENTS.
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
    1 FORMAT(I4)
    2 FORMAT(1X,I4,2X,'SPECTRA')
    3 FORMAT(1X,I4,2X,'MASSES',//)
    4 FORMAT(///)
    5 FORMAT(1X,I4,3X,F12.5)
    6 FORMAT(1X,'SPEC',3X,'FRACTN.COEFFICIENT',//)
      PRINT 2,M
      PRINT 3,N
      IO=1
      CALL CORRE(N,M,IO,X,XBAR,S,V,R,FC,B,T)
      I=1
      DO 110 J=1,M
      IF(I-J)102,104,104
  102 L=I+(J*J-J)/2
      GOTO 110
  104 L=J+(I*I-I)/2
  110 FC(J)=R(L)
      PRINT 6
      K=1
      KV=KVG(K)
      DO 100 I=1,M
      WRITE (6,5) I,FC(I)
      IF(I.NE.KV)GOTO 100
      IF(KV.EQ.M)GOTO 100
      PRINT 4
      K=K+1
      KV=KVG(K)
```

```
AN IV G LEVEL  20                 FRACT               DATE = 73200

          100 CONTINUE
              RETURN
              END
```

```
AN IV G LEVEL  20                 DATA                DATE = 73200

              SUBROUTINE DATA
        C
        C          THIS DUMMY SUBROUTINE IS USED
        C          WITH FRACT,MC AND PCA   (CDG)
        C
              RETURN
              END
```

```
      SUBROUTINE DIFF(AN,NUMGP,NOM,FC,M,ITI,ITO)
      DIMENSION AN(28,142),D(28,142),FC(28),DEV(28)
C
C-----------------------------------------------------------------
C          INPUT-
C          AN,NORMALISED SPECTRA IN GROUP BEGINNING
C                 WITH SPECTRUM M.
C          NUMGP,NO. OF SPECTRA IN GROUP.
C          NOM, NO. OF MASSES (COLUMNS)
C          FC, VECTOR CONTAINING SPEC. CORRLN. COEFFTS.
C          M, FIRST ROW IN GROUP IS SPEC.M OF ORIGINAL
C                 DATA MATRIX.
C
C          OUTPUT-
C
C          POSITIONS  OF SUSPECT DEVIATIONS ON SUB-
C          TRACTION OF NORMALISED SPECTRA (AND O/P
C          OF DIFFERENCES IF ITO OR ITI=1).
C          ITO BECOMES 1 IF SUSPECT DEVIATIONS ARE FOUND.
C
C          OTHER SUBROUTINES NECESSARY-  PRAG.
C
C      ADJUSTMENT OF ERROR TOLERANCES - SEE COMMENTS, BELOW
C-----------------------------------------------------------------
C
    1 FORMAT(1X,'GROUP BEGINNING SPECTRUM ',I4,///)
    2 FORMAT(1X,'SUSPECT DEVIATION...SPEC NO. ',I4,3X,'MASS ',I4)
    3 FORMAT(1X,///,'NO SUSPECT SPECTRA',///)
    4 FORMAT(1X,////,1X,'GROUP DIFFERENCE MATRIX',///)
    5 FORMAT(1X,'SUM DEVIATIONS    FRACTN. COEFFT.',//)
    6 FORMAT(1X,F12.5,4X,F12.5)
    7 FORMAT(I4)
    8 FORMAT(////)
    9 FORMAT('1')
      PRINT 9
      PRINT 1,M
      CALL PRAG(AN,NUMGP,NOM,1,NUMGP)
      ITO=0
C
C        SUBTRACTION OF FIRST SPECTRUM IN EACH NORMALISED
C        GROUP TO GIVE DIFFERENCE ARRAY 'D'.
C
C
C        CALCULATION OF % DEVIATION FROM MEAN.
C
C
C      IF % > 20 AND SPECTRAL ELEMENTS > 100.0 THEN ERROR .
C          THESE VALUES WILL DEPEND ON THE PARTICULAR UNITS
C          USED TO MEASURE SPECTRA
C
      DO 120 I=1,NUMGP
      DEV(I)=0.0
      DO 100 J=1,NOM
      D(I,J)=AN(1,J)-AN(I,J)
      D1=D(I,J)
      D2=ABS(D1)
```

```
      DEV(I)=DEV(I)+D2
      D(I,J)=(D2/AN(I,J))*100.0
      PER=D(I,J)
      IF(PER.LT.20.0)GOTO 100
      K1=AN(1,J)
      K2=AN(I,J)
      IF((K1.LT.100.0).OR.(K2.LT.100.0))GOTO 100
      ITO=1
      PRINT 2,I,J
  100 CONTINUE
  120 CONTINUE
      PRINT 8
      PRINT 5
      DO 150 I=1,NUMGP
      NO=I+M-1
      PRINT 6,DEV(I),FC(NO)
  150 CONTINUE
      IF(ITO.EQ.1)GOTO 140
      PRINT 3
  140 CONTINUE
      IF((ITI.EQ.0).AND.(ITO.EQ.0))GOTO 130
      PRINT 4
      CALL PRAG(D,NUMGP,NOM,1,NUMGP)
  130 RETURN
      END
```

```
N IV G LEVEL  20                  MAV                  DATE = 73200
      SUBROUTINE MAV(A,AV,NOS,NOM,NOG)
      DIMENSION A(28,142),AV(12,142),KV(30)
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C           THIS PROGRAM AVERAGES SPECTRA PRESENT IN GROUPS.
C
C           A IS THE FULL MIXTURES ARRAY HAVING NOS ROWS,
C           NOM COLUMNS AND NOG SPECTRAL GROUPS
C           THE VECTOR KV IS READ IN CONTAINING ROW NUMBERS
C           OF THE FINAL SPECTRA IN EACH GROUP, I.E. NOG
C           INTEGERS IN ALL.
C
C           AV IS THE AVERAGED DATA MATRIX FORMED.
C           VALUES IN FIRST AND FINAL COLUMNS OF AV ARE
C           PRINTED AS A CHECK.
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
    2 FORMAT(2F10.2)
    4 FORMAT(12I4)
      READ 4,(KV(I),I=1,NOG)
      PRINT 4,(KV(I),I=1,NOG)
      M=1
      K=1
   50 N=KV(K)
      NUMGP=N-M+1
      DO 130 J=1,NOM
      SUM=0.0
      DO 140 I=M,N
      SUM=SUM+A(I,J)
  140 CONTINUE
      AV(K,J)=SUM/NUMGP
  130 CONTINUE
      IF(KV(K).EQ.NOS)GOTO 500
      M=KV(K)+1
      K=K+1
      GOTO 50
  500 CONTINUE
      DO 150 K1=1,K
      PRINT 2,AV(K1,1),AV(K1,NOM)
  150 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE MSTATS(A,AV,AVDEV,NOS,NOM,NOG)
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C          THIS PROGRAM AVERAGES GROUPS OF 'IDENTICAL'
C          SPECTRA IN MATRIX A GIVING AV,
C          INPUT A,
C          NOM - NUMBER OF SPECTRA ROWS) IN A,
C          NOM - NUMBER OF MASSES (COLS) IN A AND AV,
C          NOG - NUMBER OF GROUPS (ROWS OF AV),
C          KVG, A VECTOR CONTAINING ROW NUMBERS OF FINAL
C          SPECTRA IN EACH GROUP OF A, IS READ IN FORM(12I4)
C          IF THERE ARE < E.G. 4 SPECTRA/GROUP MAV SHOULD
C          BE USED.
C          OUTPUT
C          AVDEV IS THE AVERAGED DEVIATION FROM THE MEAN
C          VALUE OF EACH PEAK.
C          % ERRORS ARE PRINTED AND AVERAGE % ERROR.
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      DIMENSION A(10,8),AV(10,8),AVDEV(10,8),AN(10,8)
      DIMENSION PER(10,8),TIC(10),DEV(10),KVG(10)
      K=1
      M=1
      L=1
    1 FORMAT(I4)
    3 FORMAT(12I4)
    4 FORMAT(1X,'AVERAGED ARRAY',//)
    5 FORMAT(1X,'AVERAGE DEVIATIONS FROM MEAN',//)
    6 FORMAT(1X,'NORMALISED GROUPS',//)
    7 FORMAT(1X,'% ERROR IN EACH PEAK',//)
   13 FORMAT(1X,'AVERAGE  % DEVIATION = ',F6.2)
      READ 3,(KVG(I),I=1,NOG)
      PRINT 3,(KVG(I),I=1,NOG)
      PRINT 6
   50 N=KVG(K)
      NUMGP=N-M+1
      BIG=0.0
      DO 130 I=M,N
      TIC(I)=0.0
      DO 140 J=1,NOM
      TIC(I)=TIC(I)+A(I,J)
  140 CONTINUE
      IF(TIC(I).GT.BIG)BIG=TIC(I)
  130 CONTINUE
      DO 150 I=M,N
      DO 150 J=1,NOM
      AN(I,J)=A(I,J)*BIG/TIC(I)
  150 CONTINUE
      CALL PRAG(AN,NUMGP,NOM,M,N)
      DO 160 J=1,NOM
      SUM=0.0
      DO 170 I=M,N
      SUM=SUM+AN(I,J)
  170 CONTINUE
      AVJ=SUM/NUMGP
      SUMDEV=0.0
```

```
      DO 180 I=M,N
      DIFF=AVJ-AM(I,J)
      DEV(I)=ABS(DIFF)
      SUMDEV=SUMDEV+DEV(I)
  180 CONTINUE
      AVDEV(K,J)=SUMDEV/NUMGP
      AV(K,J)=AVJ
      PER(K,J)=(100.0*AVDEV(K,J))/AV(K,J)
  160 CONTINUE
      IF(N.EQ.NOS)GOTO 500
      L=1
      M=N+1
      K=K+1
      GOTO 50
  500 CONTINUE
      PRINT 4
      CALL PRAG(AV,NOG,NOM,1,NOG)
      PRINT 5
      CALL PRAG(AVDEV,NOG,NOM,1,NOG)
      PRINT 7
      CALL PRAG(PER,NOG,NOM,1,NOG)
      DO 200 I=1,NOG
      DO 200 J=1,NOM
  200 SUM=SUM+PER(I,J)
      AVPER=SUM/(NOG*NOM)
      PRINT 13,AVPER
      RETURN
      END
```

```
N IV G LEVEL  20                  MAIN                DATE = 73200

            DIMENSION A(28,142),AV(12,142)
      C-------------------------------------------------------------------
      C
      C
      C             EXAMPLE OF USE OF GAUSP WITH MIXTURE V.
      C
      C             INPUT-
      C             A, ARRAY OF MIXTURES SPECTRA HAVING-
      C             NEXP ROWS (SPECTRA)
      C             NOM COLUMNS (MASSES)
      C             NOG GROUPS OF "IDENTICAL" SPECTRA.
      C
      C             SPECTRAL GROUPS ARE AVERAGED BY MAV GIVING AV.
      C             RANK ANALYSIS IS PERFORMED ON AV.
      C-------------------------------------------------------------------
      C
          1 FORMAT(I4)
          2 FORMAT(10F7.1)
            READ 1,NEXP
            READ 1,NOM
            READ 1,NOG
            DO 100 I=1,NEXP
        100 READ 2,(A(I,J),J=1,NOM)
            CALL PRA(A,NEXP,NOM)
            CALL MAV(A,AV,NEXP,NOM,NOG)
            CALL GAUSP(AV,NOG,NOM,0,2,0,1.0,10.0)
            CALL EXIT
            END
```

```
      C
            SUBROUTINE GAUSP(AV,NOG,NOM,IPER,PEROR,STEP,FPEROR)
            DIMENSION AV(12,142),AG(12,142),S(12,142)
      C
      C-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      C              THIS PROGRAM DETERMINES THE NUMBER OF SIGNIFICANT
      C              COMPONENTS PRESENT IN AN ARRAY OF MIXTURES MASS
      C              SPECTRA,
      C
      C              GAUSP IS THE RANK DETERMINING SUBROUTINE WHICH
      C              TAKES THE ERRORS IN THE DATA MATRIX TO BE
      C              A PERCENTAGE ,
      C
      C              AV IS THE  ARRAY OF MIXTURES SPECTRA HAVING
      C              NOG ROWS AND NOM COLUMNS,
      C              IPER IS THE NUMBER OF TRANSFORMED ELEMENTS
      C              ALLOWED GREATER THAN THEIR ERRORS-FOR EXACT RANK
      C              DETERMINATION,  FOR OUTPUT OF FULL RANK STATISTICS
      C              IPER SHOULD BE ZERO,
      C              IPER IS AN INTEGER,
      C              PEROR IS THE INITIAL PERCENTAGE ERROR TAKEN,
      C              STEP IS THE INCREASE IN % EACH TIME,
      C              FPEROR IS THE  FINAL PERCENT ERROR TO BE TRIED,
      C-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      C
          8 FORMAT(1X,'PERCENTAGE ERROR USED =',F6,2,'%',//)
         40 FORMAT(//////)
         50 DO 140 I=1,NOG
            DO 140 J=1,NOM
            AG(I,J)=AV(I,J)
            S(I,J)=AG(I,J)*PEROR*0,01
        140 CONTINUE
            PRINT 40
            PRINT 8,PEROR
            CALL GAUSS(AG,S,IPER,NOG,NOM)
            PEROR=PEROR+STEP
            IF(PEROR,GT,FPEROR)GOTO 34
            GOTO 50
         34 RETURN
            END
```

```
      SUBROUTINE GAUSS(AG,S,IPER,IFNOG,NOM)
      DIMENSION AG(12,142),S(12,142)
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C             THIS PROGRAM IS PART OF THE RANK DETERMINING
C             SUBROUTINE. SUCCESSIVE GAUSSIAN ELIMINATIONS
C             ARE PERFORMED ON MATRIX AG AND RANK STATISTICS
C             PRINTED OUT.
C
C             INPUT-
C             AG, DATA MATRIX OF IFNOG ROWS AND NOM COLUMNS.
C             S, ERROR MATRIX OF AG.
C             IPER, SEE GAUSP OR GAUSAV
C             L  DETERMINES THE NUMBER OF ELIMINATIONS.
C
C             OTHER SUBROUTINES REQUIRED-
C             PIVOT,ELIM
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
    1 FORMAT(I4)
    2 FORMAT(F9.3)
    3 FORMAT(2F9.3)
  108 FORMAT(1X,'RANK=',I4)
  109 FORMAT(5X,I4,8X,F5.1,8X,F9.3,2X,F9.3)
  110 FORMAT(1X,'PERCENT ALLOWED IN RESIDUE = ',I4,///)
  113 FORMAT(1X,'NO. ELIMS',4X,'PERCENT',8X,'PIVOT EL',6X,
     1'CORRES. ERROR',//)
      PRINT 110,IPER
      FIPER=FLOAT(IPER)
      L=1
      IX=0
      PRINT 113
   30 IF((L.GT.IFNOG).OR.(L.GT.NOM))GOTO 306
      CALL PIVOT(AG,S,IFNOG,NOM,L)
      IY=0
      DO 303 J=L,NOM
      DO 303 I=L,IFNOG
      A=AG(I,J)
      AB=ABS(A)
  303 IF(AB.GT.S(I,J))IY=IY+1
      NOE=(IFNOG-IX)*(NOM-IX)
      PERG=(FLOAT(IY)/FLOAT(NOE))*100.0
      GOTO 307
  306 CONTINUE
      PERG=0.0
      S(L,L)=0.0
      AG(L,L)=0.0
  307 IF(PERG.GT.FIPER)GO TO 111
      IRANK=IX
      PRINT 109,IX,PERG,AG(L,L),S(L,L)
      PRINT 108,IRANK
      GOTO 31
  111 PRINT 109,IX,PERG,AG(L,L),S(L,L)
      IF((L.GT.IFNOG).OR.(L.GT.NOM))GOTO 31
      CALL ELIM(AG,S,IFNOG,NOM,L)
      IX=IX+1
```

```
N IV G LEVEL  20                     GAUSS                    DATE = 73200
          L=L+1
          GO TO 30
       31 RETURN
          END
```

```
N IV G LEVEL  20                  ELIM                DATE = 73200

      SUBROUTINE ELIM(A,S,NEXP,NOM,L)
      DIMENSION A(12,142),S(12,142),AA(12,142),SS(12,142)
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C          THIS PROGRAM IS PART OF THE RANK DETERMINING
C          SUBROUTINE,
C
C          GAUSSIAN ELIMINATION IS PERFORMED ON A AND ITS
C          ERROR MATRIX S IS TRANSFORMED AT THE SAME TIME,
C
C          ON OUTPUT A IS THE REDUCED ARRAY AND S CONTAINS
C           ITS ERRORS,
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
      C=A(L,L)
      DO 40 I=L,NEXP
      DO 40 J=L,NOM
      AA(I,J)=A(I,J)-A(I,L)*A(L,J)/C
      S5=S(I,J)*S(I,J)
      S1=(A(I,L)/C)*(A(I,L)/C)
      S2=(A(L,J)/C)*(A(L,J)/C)
      S3=S(I,L)*S(I,L)
      S4=S(L,J)*S(L,J)
      SS(I,J)=SQRT(S5+S4*S1+S3*S2+S5*S1*S2)
   40 CONTINUE
      DO 50 I=L,NEXP
      DO 50 J=L,NOM
      A(I,J)=AA(I,J)
   50 S(I,J)=SS(I,J)
      RETURN
      END
```

```
N IV G LEVEL  20                   PIVOT                DATE = 73200

       SUBROUTINE PIVOT(A,S,NEXP,NOM,L)
       DIMENSION A(12,142),S(12,142)
C
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C            THIS PROGRAM IS PART OF THE RANK DETERMINING
C            SUBROUTINE.
C            A IS INPUT AND ITS LARGEST ELEMENT PIVOTED TO
C            THE LEADING POSITION BY INTERCHANGE OF ROWS AND
C            COLUMNS. THE CORRESPONDING ROWS AND COLUMNS OF S,
C            THE ERROR MATRIX OF A, ARE INTERCHANGED AT
C            THE SAME TIME.
C
C                ON OUTPUT A AND S ARE THE PIVOTED MATRICES.
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      AMAX=0.0
      JMAX=1
      IMAX=1
      DO 10 I=L,NEXP
      DO 10 J=L,NOM
      AC=A(I,J)
      AB=ABS(AC)
      AM=ABS(AMAX)
      IF(AB.LE.AM)GO TO 10
      AMAX=A(I,J)
      IMAX=I
      JMAX=J
   10 CONTINUE
      DO 19 I=L,NEXP
      B=A(I,L)
      A(I,L)=A(I,JMAX)
      A(I,JMAX)=B
      T=S(I,L)
      S(I,L)=S(I,JMAX)
   19 S(I,JMAX)=T
      DO 30 J=L,NOM
      B=A(L,J)
      A(L,J)=A(IMAX,J)
      A(IMAX,J)=B
      T=S(L,J)
      S(L,J)=S(IMAX,J)
   30 S(IMAX,J)=T
      RETURN
      END
```

```
V G LEVEL  20                          GAUSAV              DATE = 73207

          SUBROUTINE GAUSAV(AV,AVDEV,NOG,NOM,IPER,CRIT,STEP,FCRIT)
          DIMENSION AV(23,37),AG(23,37),S(23,37),AVDEV(23,37)
      C-------------------------------------------------------------------
      C              SAME AS GAUSP EXCEPT THAT ERROR IN S IS OBTAINED
      C              FROM  'MSTATS' AND MULTIPLIED BY CRIT
      C-------------------------------------------------------------------
      C
        8 FORMAT(1X,'ERROR TIMES ',F6.2,//)
       40 FORMAT(//////)
       50 DO 140 I=1,NOG
          DO 140 J=1,NOM
          AG(I,J)=AV(I,J)
          S(I,J)=AVDEV(I,J)*CRIT
      140 CONTINUE
          PRINT 40
          PRINT 8,CRIT
          CALL GAUSS(AG,S,IPER,NOG,NOM)
          CRIT=CRIT+STEP
          IF(CRIT.GT.FCRIT)GOTO 34
          GOTO 50
       34 RETURN
          END
```

```
IV G LEVEL  20                       PCA                    DATE = 73204            0:
        SUBROUTINE PCA(X,M,N,CON)
        DIMENSION X(25,12)
        DIMENSION B(200),D(200),S(200),T(200),XBAR(200)
        DIMENSION V(8000)
        DIMENSION R(8000)
        DIMENSION TV(8000)
C
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C                THIS PROGRAM PERFORMS A PRINCIPAL COMPONENTS
C                ANALYSIS ON MATRIX X,
C
C                INPUT - X IS ARRAY TO BE ANALYZED, DIMENSION IS N BY M
C                WHERE N>=M, DIM IN STATEMENT MUST BE (N,M) UNLESS
C                'ARRAY' IS USED (SEE IBM SSP MANUAL FOR ALL
C                DIMENSION DETAILS), MEANS ,STANDARD DEVIATIONS
C                AND ALL EIGENVALUES > CON ARE PRINTED OUT.
C
C                DUMMY SUBROUTINE 'DATA' ALSO REQUIRED,
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      1 FORMAT(1X,'PRINCIPAL COMPONENTS ANALYSIS',//)
      2 FORMAT(6H0MEANS/(8F15.5))
      3 FORMAT(20H0STANDARD DEVIATIONS/(8F15.5))
      4 FORMAT(25H0CORRELATION COEFFICIENTS)
      5 FORMAT(4H0ROWI3/(10F12.5))
      6 FORMAT(1H0/12H EIGENVALUES/(10F12.5))
      7 FORMAT(37H0CUMULATIVE PERCENTAGE OF EIGENVALUES/(10F12.5))
      8 FORMAT(1X,I4,'VARIABLES')
      9 FORMAT(1X,I4,'OBSERVATIONS')
        PRINT 1
        PRINT 8,M
        PRINT 9,N
        IO=1
        CALL CORRE(N,M,IO,X,XBAR,S,V,R,D,B,T)
        WRITE (6,2)  (XBAR(J),J=1,M)
        WRITE (6,3) (S(J),J=1,M)
        WRITE (6,4)
        DO 120 I=1,M
        DO 110 J=1,M
        IF(I-J)102,104,104
    102 L=I+(J*J-J)/2
        GOTO 110
    104 L=J+(I*I-I)/2
    110 D(J)=R(L)
    120 WRITE (6,5) I,(D(J),J=1,M)
        MV=0
        CALL EIGEN(R,V,M,MV)
        CALL TRACE(M,R,CON,K,D)
        DO 130 I=1,K
        L=I+(I*I-I)/2
    130 S(I)=R(L)
        WRITE(6,6) (S(J),J=1,K)
        WRITE (6,7) (D(J),J=1,K)
        RETURN
        END
```

```
      SUBROUTINE FILTER(AV,IFNOG,NOM,AMASS,TOLL,NGT)
      DIMENSION AV(23,37),AN(23,37),DDD(400),SIGMA(37)
      DIMENSION DD(23,400),AM1(400),AM2(400),AMASS(37),SUM(400)
C
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C               THIS PROGRAM DETECTS AND STORES THE COLUMNS OF A
C               MASS SPECTRAL MIXTURES ARRAY WHICH ARE PROPORTIONAL.
C
C               MASS PROFILES (COLUMNS OF AV) ARE NORMALISED
C               TO THE SAME SUM AND EACH SUBTRACTED FROM EVERY
C               OTHER. ALL PAIRS OF MASSES HAVING ONLY (NGT)
C               DIFFERENCES > (TOLL) ARE STORED IN VECTORS AM1 AND
C               AM2, DIFFERENCES AND THE AVERAGES OF ABS. DIFFS.
C               ARE PRINTED OUT FOR THESE PROFILES ONLY.
C
C               IFNOG IS THE NUMBER OF SPECTRA IN AV (ROWS).
C               NOM IS THE NUMBER OF MASSES IN AV (COLUMNS).
C               AMASS IS A VECTOR CONTAINING THE MASS NUMBERS.
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
C
      INTEGER XY,FNOD
    1 FORMAT(I4)
    2 FORMAT(F10.2)
    3 FORMAT(3F10.2)
    5 FORMAT(1X,F5.1,1X,F5.1)
    6 FORMAT(///////)
    7 FORMAT(12F8.1)
    9 FORMAT(1X,F5.1,1X,F5.1,6X,F10.2)
  180 FORMAT(1X,'MASS 1   ',2X,'MASS 2    ',6X,'DIFFERENCES',//)
  170 FORMAT(1X,F9.3,2X,F9.3,2X,12F8.3)
      SIGMAX=0.0
      DO 100 J=1,NOM
      SIGMA(J)=0.0
      DO 110 I=1,IFNOG
      SIGMA(J)=SIGMA(J)+AV(I,J)
  110 CONTINUE
      IF(SIGMA(J).GT.SIGMAX)SIGMAX=SIGMA(J)
  100 CONTINUE
      DO 120 J=1,NOM
      DO 120 I=1,IFNOG
      AN(I,J)=AV(I,J)*SIGMAX/SIGMA(J)
  120 CONTINUE
      PRINT 6
      PRINT 180
      PRINT 6
      XY=1
      JOF=NOM-1
      DO 130 JO=1,JOF
      JIN=JO+1
      DO 140 JI=JIN,NOM
      SUM(XY)=0.0
      IC=0
      DO 220 I=1,IFNOG
      DDD(I)=AN(I,JI)-AN(I,JO)
      D5=DDD(I)
```

```
      D6=ABS(D5)
      IF(D6.GT.TOLL)IC=IC+1
  220 CONTINUE
      IF(IC-NGT)150,150,140
  150 AM1(XY)=AMASS(JO)
      AM2(XY)=AMASS(JI)
      DO 160 I=1,IFNOG
      DD(I,XY)=DDD(I)
      A=DD(I,XY)
      AB=ABS(A)
      SUM(XY)=SUM(XY)+AB
  160 CONTINUE
      SUM(XY)=SUM(XY)/IFNOG
      PRINT 5,AM1(XY),AM2(XY)
      PRINT 7,(DD(I,XY),I=1,IFNOG)
      PRINT 2,SUM(XY)
      XY=XY+1
  140 CONTINUE
      PRINT 6
  130 CONTINUE
      PRINT 6
      PRINT 1,NOPR
      NOPR=XY-1
      PRINT 180
      DO 190 XY=1,NOPR
      PRINT 9,AM1(XY),AM2(XY),SUM(XY)
  190 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE MC(X,M,N,BMASS)
      DIMENSION X(90,37)
      DIMENSION BMASS(37)
      DIMENSION B(100),D(100),S(100),T(100),XBAR(100)
      DIMENSION V(8000)
      DIMENSION R(8000)
C
C-------------------------------------------------------------------
C          THIS PROGRAM INTER-CORRELATES THE MASS PROFILES
C          IN MIXTURES ARRAY X.
C
C          INPUT - X IS THE CONVENTIONAL MIXTURES ARRAY,
C          N - NUMBER OF ROWS (SPECTRA),
C          M - NUMBER OF COLUMNS (MASSES),
C          N >=M
C          DIM STATEMENT (N,M) UNLESS SSP 'ARRAY' USED
C          BMASS - A VECTOR CONTAINING THE MASS NUMBERS USED,
C          DUMMY SUBROUTINE 'DATA' ALSO REQUIRED,
C
C          THE CORRELATION MATRIX OF MASS PROFILES IS PRINTED,
C-------------------------------------------------------------------
C
    4 FORMAT(25H0CORRELATION COEFFICIENTS)
    5 FORMAT(1X,'MASS',F8.2/,(10F12.5))
      IO=1
      CALL CORRE(N,M,IO,X,XBAR,S,V,R,D,B,T)
      WRITE (6,4)
      DO 120 I=1,M
      DO 110 J=1,M
      IF(I-J)102,104,104
  102 L=I+(J*J-J)/2
      GOTO 110
  104 L=J+(I*I-I)/2
  110 D(J)=R(L)
  120 WRITE (6,5) BMASS(I) ,(D(J),J=1,M )
      RETURN
      END
```

```
      SUBROUTINE CDG(X,M,N,NOC,BMASS)
C
C-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
C
C          THIS PROGRAM PRODUCES A COMPONENT DIAGRAM FROM
C          MIXTURES ARRAY, X.
C
C          INPUT - X OF DIMENSION (N,M) WHERE M IS THE
C          NUMBER OF COLUMNS (MASSES) AND N THE NUMBER OF
C          MIXTURES SPECTRA (ROWS).
C          NOC - NUMBER OF COMPONENTS.
C          BMASS - VECTOR CONTAINING MASSES.
C          CDG READS IN NOC INTEGERS IN FORMAT(12I4) GIVING
C          COLUMN NUMBERS OF ONE MASS UNIQUE TO EACH COMPONENT.
C          DUMMY SUBROUTINE 'DATA' REQUIRED.
C-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
C
      DIMENSION X(90,37),BMASS(37)
      DIMENSION XAB(50,50),IU(10)
      DIMENSION B(100),D(100),S(100),T(100),XBAR(100)
      DIMENSION V(8000)
      DIMENSION R(8000)
    3 FORMAT(12I4)
    4 FORMAT(17X,'MASS',7X,'RELATIVE CONTRIBUTIONS OF THE',
     1I2,1X,'COMPONENTS')
    5 FORMAT(1X,'MASS',F8.2/,(10F12.5))
    6 FORMAT(9X,I4,F10.2,2X,5F10.5)
    7 FORMAT(////)
    8 FORMAT(28X,'SELF CORRELATIONS - INTER-UNIQUE',//)
    9 FORMAT(28X,'CORRELATION COEFFICIENTS',//)
   10 FORMAT(28X,'IN FIRST 90 GOOD SPECTRA',//)
      READ 3,(IU(I),I=1,NOC)
      IO=1
      CALL CORRE(N,M,IO,X,XBAR,S,V,R,D,B,T)
      PRINT 9
      PRINT 4,NOC
      PRINT 10
      K=1
      I1=1
      KV=IU(K)
      DO 120 I=1,M
      IF(I.NE.KV)GOTO 120
      DO 110 J=1,M
      IF(I-J)102,104,104
  102 L=I+(J*J-J)/2
      GOTO 110
  104 L=J+(I*I-I)/2
  110 XAB(J,I1)=R(L)
      I1=I1+1
      K=K+1
      IF(K.GT.NOC)GOTO 150
      KV=IU(K)
  120 CONTINUE
  150 CONTINUE
      DO 160 J=1,M
      PRINT 6,J,BMASS(J),(XAB(J,I1),I1=1,NOC)
```

```
  160 CONTINUE
      PRINT 7
      PRINT 8
      DO 170 K=1,NOC
      J=IU(K)
      PRINT 6,J,BMASS(J),(XAB(J,I1),I1=1,NOC)
  170 CONTINUE
      RETURN
      END
```

```
V G LEVEL  20                  RPA                  DATE = 73215           03/
      SUBROUTINE RPA(AV,NEXP,NOM,IPER)
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C           THIS PROGRAM FORMS MIXTURES ARRAY SUBMATRICES
C           CONTAINING INCREASING NOS, OF COLUMNS (MASSES)
C           STARTING AT THE HIGH MASS END,   RANK ANALYSES ARE
C           CARRIED OUT ON EACH,
C
C           INPUT - AV, MIXTURES ARRAY,
C           NEXP , NO, OF SPECTRA,
C           NOM, NO, OF MASSES TO BE CONSIDERED,
C           IPER, % OF VALUES ALLOWED > ERRORS,
C
C           MASSES AT WHICH COMPONENTS BEGIN TO CONTRIBUTE
C           SIGNIFICANTLY ARE INDICATED,
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      DIMENSION AV(12,142)
    1 FORMAT('1')
   10 FORMAT(1X,I4,1X,'MASSES')
      J1=2
   91 PRINT 1
      PRINT 10,J1
      PEROR=4.0
      CALL GAUSP(AV,NEXP,J1,IPER,PEROR,1,0,6,0)
      IF(J1.EQ,NOM)GOTO 34
      J1=J1+1
      GOTO 91
   34 RETURN
      END
```

```
IV G LEVEL  20                        MAIN                    DATE = 73208              00/
       DIMENSION A(133,37),AV(23,37),AMASS(50)
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C       EXAMPLE OF PATTERN SEPARATION
C       DERIVATION OF MAJOR COMPONENT SPECTRA - MIXTURE II.
C
C       INPUT -
C       NEXP, NO. ROWS (SPECTRA) IN INITIAL ARRAY.
C       NOM, NO. OF COLUMNS (MASSES) IN INITIAL MATRIX.
C       NOG,NO. OF GROUPS OF 'IDENTICAL' SPECTRA FOR AVERAGING.
C       NOC, NO. OF COMPONENTS.
C           MASSES IN AMASS AND THE PEAKS IN 'A' ARE READ FROM
C           DISK-FILE ,UNIT 10
C           SELM READS IN ROWS TO BE ELIMINATED, K IS FINAL NO.
C           MAV AVERAGES REMAINING SPECTRA.
C           'SELECT' CHOOSES NOC OF THESE FORMING ARRAY AV
C           WHICH IS OUTPUT.
C           FINALLY UNIQS PRODUCES COMPONENT SPECTRA.
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
      DOUBLE PRECISION A,AV
    1 FORMAT(I4)
    2 FORMAT(10F7.1)
    3 FORMAT(F6.2)
    4 FORMAT(F5.1)
    5 FORMAT('1')
    7 FORMAT(////////)
    8 FORMAT(12X,F7.1,6X,6F8.2)
      READ 1,NEXP
      PRINT 1,NEXP
      READ 1,NOM
      PRINT 1,NOM
      READ 1,NOG
      PRINT 1,NOG
      READ 1,NOC
      PRINT 1,NOC
      DO 200 J=1,NOM
      READ (10,4) AMASS(J)
      READ (10,2) (A(I,J),I=1,NEXP)
  200 CONTINUE
      CALL SELM(A,NEXP,NOM,K)
      PRINT 7
      CALL MAV(A,AV,K,NOM,NOG)
      CALL SELECT(AV,NOM,NOC)
      PRINT 5
      PRINT 7
      DO 100 J=1,NOM
  100 PRINT 8,AMASS(J),(AV(I,J),I=1,NOC)
      CALL UNIQS(AV,NOM,NOC,AMASS)
      CALL EXIT
      END
```

LISTING OF DATA CARDS

Pattern Separation Example (previous page)

```
133
 37
 23
  4
 20
 21
 78
 79
 89
 90
 91
107
119
  0
  6  15  19  23  26  29  46  50  64  72  75  84
 90  94  99 104 106 108 110 112 117 119 124
  1   5  11  12
  7
  2  16  30   4   5  24  11
  3   5   6   7
  2   4  24  11
 12
  1   2  16  30   3   4   5  20   8  24  10  11
  4   8  10  12
  2   4  24  11
```

(each row is a card)

```
      SUBROUTINE UNIQS(AMIX,NOM,NOC,AMASS)
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C      THIS PROGRAM USES INFORMATION ABOUT POSITIONS OF UNIQUE
C      PEAKS (ANY NUMBER PER COMPONENT) TO DERIVE PURE COMP.
C      SPECTRA BY A LEAST SQUARES METHOD,
C
C      INPUT - NOC IS NUMBER OF COMPS.,NOM IS NUMBER OF MASSES,
C      AMIX CONTAINS NOC SPECTRA CHOSEN FROM ALL SPECTRA AS
C      HAVING THE MOST DIFFERENT PATTERNS.(FROM 'FRACT').
C      AMIX IS PROBABLY FORMED BY SUBROUTINE 'SELECT'
C      AMASS CONTAINS MASSES USED.
C      READS IN - NM, TOTAL NUMBER UNIQUE PEAKS ,
C      KVM - VECTOR HAVING NM INTEGERS WHICH ARE COLUMN NUMBERS
C      OF UNIQUE PEAKS GROUPED ACCORDING TO COMPONENTS.
C     'IN' - VECTOR CONTAINING NOC INTEGERS WHICH ARE POSITIONS
C      OF FINAL MASSES OF GROUPS IN KVM (NOT IN AMIX), FINAL
C      INTEGER IN VECTOR 'IN' IS NM.
C
C      OUTPUT - Q IS DERIVED 'Q' MATRIX.
C     'AN' CONTAINS DERIVED SPECTRA BEFORE REFINING.
C     'AN' AFTER APPLICATION OF 'REFINU' CONTAINS REFINED SPECTRA.
C      RESIDUES ARE GIVEN BY 'RESIDU',
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      DIMENSION AMIX(23,37),A(50,50),S(50,50),KVM(50),IN(50)
      DIMENSION AMASS(50),Q(10,10),DERV(23,37),D(50),Y(50)
      DIMENSION AN(23,37),IPIV(50),AUX(50)
      DOUBLE PRECISION AMIX,A,D,Y,Q,DERV,AN,S,AUX
    1 FORMAT(I4)
    2 FORMAT(F6.1)
    3 FORMAT(8F9.3)
    4 FORMAT(////////)
    5 FORMAT(12I4)
    7 FORMAT(12X,F7.1,6X,6F8.2)
    8 FORMAT(12X,4F8.2)
    9 FORMAT(12X,'DLLSQ ERROR PARAMETER =',I4,//)
   12 FORMAT('1')
      PRINT 1,NOC
      READ 1,NM
      PRINT 1,NM
      READ 5,(KVM(I),I=1,NM)
      PRINT 5,(KVM(I),I=1,NM)
      READ 5,(IN(I),I=1,NOC)
      PRINT 5,(IN(I),I=1,NOC)
      PRINT 12
      L=1
      N1=1
      N2=IN(L)
      NV=NOC-1
      DO 200 IY=1,NOC
      JX=0
      NE=NM-N2+N1-1
      DO 110 J=1,NM
      IF((J.GE.N1).AND.(J.LE.N2))GOTO 110
      JX=JX+1
```

```
      DO 100 I=1,NV
      K=KVM(J)
      A(JX,I)=AMIX(I,K)
  100 CONTINUE
      D(JX)=AMIX(NOC,K)
  110 CONTINUE
      CALL ARRAY(2,NE,NV,50,50,S,A)
      CALL DLLSQ(S,D,NE,NV,1,Y,IPIV,1.E-5,IER,AUX)
      PRINT 9,IER
      DO 120 J=1,NV
      Q(IY,J)=Y(J)
  120 Q(IY,NOC)=-1.0
      N1=N2+1
      L=L+1
  200 N2=IN(L)
      PRINT 12
      DO 250 IY=1,NOC
  250 PRINT 8,(Q(IY,J),J=1,NOC)
      DO 160 IY=1,NOC
      DO 160 J=1,NOM
      DERV(IY,J)=0.0
      DO 160 I=1,NOC
  160 DERV(IY,J)=DERV(IY,J)+Q(IY,I)*AMIX(I,J)
      DO 170 IY=1,NOC
      ABIG=0.0
      DO 180 J=1,NOM
      DE=DERV(IY,J)
      ABDE=ABS(DE)
  180 IF(ABDE.GT.ABIG)ABIG=ABDE
      DO 170 J=1,NOM
      DE=DERV(IY,J)
      ABDE=ABS(DE)
  170 AN(IY,J)=ABDE*100/ABIG
      PRINT 4
      PRINT 4
      DO 230 J=1,NOM
  230 PRINT 7,AMASS(J),(AN(I,J),I=1,NOC)
      CALL RESIDU(AMIX,AN,NOC,NOM,AMASS)
      CALL REFINU(AN,NOC)
      DO 240 J=1,NOM
  240 PRINT 7,AMASS(J),(AN(I,J),I=1,NOC)
      CALL RESIDU(AMIX,AN,NOC,NOM,AMASS)
      RETURN
      END
```

```
      SUBROUTINE ARRAY(MODE,I,J,N,M,S,D)
C
C-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
C             DOUBLE PRECISION VERSION OF IBM, VARIABLE
C             DIMENSIONING SUBROUTINE (SEE IBM -SSP MANUAL)      ref.(71)
C-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
C
      DOUBLE PRECISION S,D
      DIMENSION S(1),D(1)
      NI=N-I
      IF(MODE-1)100,100,120
  100 IJ=I*J+1
      NM=N*J+1
      DO 110 K=1,J
      NM=NM-NI
      DO 110 L=1,I
      IJ=IJ-1
      NM=NM-1
  110 D(NM)=S(IJ)
      GOTO 140
  120 IJ=0
      NM=0
      DO 130 K=1,J
      DO 125 L=1,I
      IJ=IJ+1
      NM=NM+1
  125 S(IJ)=D(NM)
      NM=NM+NI
  130 CONTINUE
  140 RETURN
      END
```

```
IV G LEVEL  20                    RESIDU               DATE = 73208            07/45
```

```
      SUBROUTINE RESIDU(AV,AN,NOC,NOM,AMASS)
C
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C             THIS PROGRAM CALCULATES RESIDUES FOR DERIVED
C             SPECTRA USING ONE PEAK UNIQUE TO EACH.
C
C             INPUT - AV - ORIGINAL MIXTURES SPECTRA USED TO
C             DERIVE AN WHERE AN CONTAINS DERIVED SPECTRA.
C             NOC IS NUMBER OF COMPONENTS, AMASS CONTAINS MASSES,
C             - READS IN NOC INTEGERS IN 'IU' WHICH ARE COLUMN
C              NUMBERS OF ONE PEAK UNIQUE TO EACH COMPONENT.
C
C             PRINTED OUT - RESIDUE MATRIX AND SUM OF DEVIATIONS
C             FOR EACH MASS.
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      DIMENSION AV(23,37),AN(23,37),DEV(50),QI(10,10),CALCMX(23,37)
      DIMENSION IU(10),AMASS(50)
      DIMENSION DIFF(23,37)
      DOUBLE PRECISION AV,AN,DEV,QI,CALCMX,DIFF
    1 FORMAT(12I4)
    2 FORMAT('1')
    6 FORMAT(////////)
   12 FORMAT(12X,F7.1,3X,4F8.2,3X,F8.2)
      PRINT 2
      READ 1,(IU(I),I=1,NOC)
      PRINT 1,(IU(I),I=1,NOC)
      DO 300 IMX=1,NOC
      DO 300 I=1,NOC
      K=IU(I)
  300 QI(IMX,I)=AV(IMX,K)/AN(I,K)
      DO 310 IMX=1,NOC
      DO 310 J=1,NOM
      CALCMX(IMX,J)=0.0
      DO 310 I=1,NOC
  310 CALCMX(IMX,J)=CALCMX(IMX,J)+QI(IMX,I)*AN(I,J)
      DO 320 J=1,NOM
      DEV(J)=0.0
      DO 320 I=1,NOC
      DIFF(I,J)=AV(I,J)- CALCMX(I,J)
      AB=DIFF(I,J)
      ABC=ABS(AB)
  320 DEV(J)=DEV(J)+ABC
      PRINT 6
      PRINT 6
      DO 330 J=1,NOM
  330 PRINT 12,AMASS(J),(DIFF(I,J),I=1,NOC),DEV(J)
      RETURN
      END
```

```
      SUBROUTINE REFINU(AN,NOC)
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C          THIS PROGRAM REFINES THE DERIVED SPECTRA IN ARRAY 'AN'.
C
C           INPUT - AN MATRIX HAVING NOC ROWS DERIVED BY
C          'UNIQS' OR 'ZEROS'.
C          READS IN -
C          NU, TOTAL NO. OF UNIQUE PEAKS KNOWN.
C          VECTOR KVU - NU INTEGERS CONTAINING COLUMN NOS.
C          OF UNIQUE PEAKS IN MATRIX AN.THESE ARE IN GROUPS
C          ACCORDING TO THE COMPONENTS.
C          VECTOR IG  CONTAINS NOC INTEGERS GIVING POSITIONS IN
C          KVU OF  FINAL NUMBERS IN THE NOC GROUPS. (THE
C          FINAL INTEGER IN IG WILL BE NU).
C
C          OUTPUT - AN CONTAINS REFINED SPECTRA, NOC ROWS.
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      DIMENSION AN(23,37),KVU(30),IG(30)
      DOUBLE PRECISION AN
    1 FORMAT(I4)
    2 FORMAT(12I4)
    3 FORMAT('1')
      PRINT 3
      READ 1,NU
      PRINT 1,NU
      READ 2,(KVU(I),I=1,NU)
      PRINT 2,(KVU(I),I=1,NU)
      READ 2,(IG(I),I=1,NOC)
      PRINT 2,(IG(I),I=1,NOC)
      L=1
      M=1
      N=IG(L)
      DO 410 I=1,NOC
      DO 420 IO=1,NOC
      IF(IO.EQ.I)GOTO 420
      DO 420 J=M,N
      K=KVU(J)
      AN(IO,K)=0.0
  420 CONTINUE
      IF(L.EQ.NOC)GOTO 410
      L=L+1
      M=N+1
      N=IG(L)
  410 CONTINUE
      RETURN
      END
```

```
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
C          PART OF 'UNRAVL' PROCEDURE,
C
C          EXAMPLE OF 'TRIAL' AS IN CHAPTER 6-III,
C
C          READS IN -
C          A, FULL MIXTURES ARRAY FROM DISK '11',
C          NOS, ROW NUMBER,
C          NOM, COLUMN NUMBER,
C          NOC, RANK OF A,
C
C          SMIX CONTAINS NOC SPECTRA SELECTED FROM A,
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      DIMENSION A(40,25),SMIX(5,25),AMASS(25)
      DOUBLE PRECISION A,SMIX
    1 FORMAT(I4)
      READ 1,NOS
      PRINT 1,NOS
      READ 1,NOM
      PRINT 1,NOM
      READ 1,NOC
      PRINT 1,NOC
    3 FORMAT(F7.2)
      READ 3,(AMASS(I),I=1,NOM)
    2 FORMAT(8F9.3)
      READ (11,2) ((A(I,J),I=1,NOS),J=1,NOM)
      CALL PRA(A,NOS,NOM)
      CALL SEL(A,SMIX,NOM,NOC)
    4 FORMAT(1X,F10.2,6X,5F10.3)
      DO 100 J=1,NOM
  100 PRINT 4,AMASS(J),(SMIX(I,J),I=1,NOC)
      CALL TRIAL(SMIX,NOC,NOM)
      CALL EXIT
      END
```

```
      SUBROUTINE TRIAL(AMIX,NOC,NOM)
C
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C        THIS PROGRAM PRODUCES TRIAL SPECTRA FROM GUESSED
C        POSITIONS OF ZEROS,
C
C        AMIX IS A 'NOC' BY 'NOM' MIXTURES ARRAY
C        NOC AND NOM HAVE USUAL MEANINGS
C        READS IN - NZTOT, TOTAL NO, OF ZEROS IN ALL
C        COMBINATIONS TO BE TRIED.
C        NTRLS - NO. OF GROUPS OF GUESSED ZEROS.
C        KVZ - KEY VECTOR CONTAINING COLUMN NOS. (MASSES)
C        WHICH HAVE ZERO CONTRIBUTIONS FROM THE DESIRED
C        SPECTRUM (GIVEN BY 'MC'). THERE ARE NZTOT INTEGERS
C        ORDERED ACCORDING TO THE GROUPS TO BE TRIED,
C        NTR - KEY VECTOR CONTAINING POSITIONS IN KVZ OF
C        FINAL ZERO IN EACH GROUP (NTRLS INTEGERS),
C
C        'ZEROS' IS CALLED FOR EACH GROUP OF ZEROS AND ONE
C        SPECTRUM DERIVED FOR EACH,
C        FINALLY UP TO 18 SPECTRA ARE PRINTED OUT,
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
      DIMENSION AMIX(5,25),ATR(50,25),DS(25)
      DIMENSION KVZ(90),NTR(25)
      DOUBLE PRECISION AMIX,ATR,DS
   12 FORMAT(F10.2)
    1 FORMAT(I4)
    2 FORMAT(12I4)
    3 FORMAT(18F7.2)
      READ 1,NZTOT
      PRINT 1,NZTOT
      READ 1,NTRLS
      PRINT 1,NTRLS
      READ 2,(KVZ(I),I=1,NZTOT)
      PRINT 2,(KVZ(I),I=1,NZTOT)
      READ 2,(NTR(I),I=1,NTRLS)
      PRINT 2,(NTR(I),I=1,NTRLS)
      M=1
      K=1
   50 N=NTR(K)
      PRINT 2,(KVZ(I),I=M,N)
      CALL ZEROS(AMIX,NOC,NOM,M,N,DS,KVZ)
      DO 200 J=1,NOM
  200 ATR(K,J)=DS(J)
      IF(K.EQ.NTRLS)GOTO 500
      M=N+1
      K=K+1
      GOTO 50
  500 CONTINUE
      DO 100 J=1,NOM
  100 PRINT 3,(ATR(I,J),I=1,NTRLS)
      RETURN
      END
```

```
      SUBROUTINE ZEROS(AMIX,NOC,NOM,M,N,DS,KVZ)
C
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C          THIS PROGRAM FORMS SIMULTANEOUS EQUATIONS FROM
C          A MIXTURES ARRAY ACCORDING TO POSITIONS OF ZERO
C          CONTRIBUTIONS FROM THE SPECTRUM TO BE DERIVED. THESE
C          POSITIONS ARE CONTAINED IN KVZ BETWEEN ELEMENTS M
C          AND N. M AND N ARE GIVEN BY 'TRIAL'.
C            SOLUTIONS OF THE EQUATIONS GIVE THE UNKNOWN SPECTRUM
C          ON MULTIPLYING INTO THE MIXTURES ARRAY.
C
C           DS FINALLY CONTAINS THE DERIVED SPECTRUM.
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      DIMENSION AMIX(5,25),DS(25)
      DIMENSION AA(15,15),D(10),Y(10)
      DIMENSION AUX(10),S(15,15),KVZ(90)
      DIMENSION IPIV(15)
      DOUBLE PRECISION AMIX,AA,D,Y,DS
      DOUBLE PRECISION AUX,S
    1 FORMAT(I4)
      NV=NOC-1
      NZ=N-M+1
C
C          NV IS THE NO. OF VARIABLES.
C          NZ IS THE NO. OF ZEROS IN THE GROUP (=NO. OF EQUNS.)
C
C
C          AA IS FORMED FROM COLUMNS OF MIXTURES ARRAY HAVING
C          ZERO CONTRIBUTIONS FROM THE SPECTRUM TO BE DERIVED.
C          AA CONTAINS COEFFTS.
C          D CONTAINS R.H.S.
C
      DO 110 J=M,N
      K=KVZ(J)
      J1=J-M+1
      DO 100 I=1,NV
      AA(J1,I)=AMIX(I,K)
  100 CONTINUE
      D(J1)=AMIX(NOC,K)
  110 CONTINUE
C
C          'ARRAY' ALTERS STORAGE MODE OF AA GIVING S.
C          DLLSQ GIVES A LEAST SQUARES SOLUTION.
C
      CALL ARRAY(2,NZ,NV,15,15,S,AA)
      CALL DLLSQ(S,D,NZ,NV,1,Y,IPIV,1.E-9,IER,AUX)
C
C          1.B.M. ERROR PARAMETER, IER, IS PRINTED
C
      PRINT 1,IER
      Y(NOC)=-1.0
      DO 160 J=1,NOM
      DS(J)=0.0
      DO 170 I=1,NOC
      DS(J)=DS(J)+Y(I)*AMIX(I,J)
```

```
AN IV G LEVEL  20                    ZEROS                DATE = 73213        1

          170 CONTINUE
          160 CONTINUE
              ABIG=0,0
              DO 190 J=1,NOM
              DE=DS(J)
              ADE=ABS(DE)
              IF(ADE.GT.ABIG)ABIG=ADE
          190 CONTINUE
        C
        C          NORMALISATION OF SPECTRA,
        C
              DO 200 J=1,NOM
              DE=DS(J)
              ADE=ABS(DE)
              ADE=ADE*100,0/ABIG
              DS(J)=ADE
          200 CONTINUE
              RETURN
              END
```

```
 AN IV G LEVEL  20                    SEL                  DATE = 73213

                 SUBROUTINE SEL(A,SMIX,NOM,NOC)
           C
           C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
           C           AS SELECT BUT NEW ARRAY FORMED
           C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
           C
                 DIMENSION A(40,25),SMIX(5,25)
                 DIMENSION KVS(10)
                 DOUBLE PRECISION A,SMIX
               3 FORMAT(12I4)
               1 FORMAT(I4)
                 READ 3,(KVS(I),I=1,NOC)
                 DO 100 I=1,NOC
                 DO 100 J=1,NOM
                 K=KVS(I)
             100 SMIX(I,J)=A(K,J)
                 RETURN
                 END
```

```
 IV G LEVEL  20                    MAIN                DATE = 73215            18

      C
      C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
      C          PART OF 'UNRAVL' PROCEDURE,
      C
      C          EXAMPLE OF 'UNRAVL' AS DESCRIBED IN CHAPTER 6-III,
      C
      C          READS IN -
      C          NOS, NO. OF SPECTRA IN FULL MIXTURES ARRAY,
      C          NOM, NO. OF MASSES (COLUMNS) IN ARRAY,
      C          NOC, RANK OF ARRAY,
      C          NREDS, NO. OF SPECTRA TO BE FORMED AND SUBTRACTED
      C          BY UNRAVL. UNRAVL WILL BE ACTIVATED THIS NO. OF TIMES,
      C          AMASS, ARRAY CONTAINING MASS NUMBERS,
      C          A, FULL MIXTURES ARRAY FROM DISK '11',
      C
      C          SMIX CONTAINS NOC SPECTRA SELECTED FROM A, THIS NO,
      C          GOES DOWN AS THE MIXTURES ARRAY IS SIMPLIFIED,
      C
      C          MASS PROFILE CORRELATION ANALYSIS IS PERFORMED ON
      C          EVERY NEW MIXTURES ARRAY .
      C
            DIMENSION A(40,25),SMIX(5,25),AMASS(25)
            DOUBLE PRECISION A,SMIX
          1 FORMAT(I4)
          2 FORMAT(8F9.3)
          3 FORMAT(F7.2)
            READ 1,NOS
            PRINT 1,NOS
            READ 1,NOM
            PRINT 1,NOM
            READ 1,NOC
            PRINT 1,NOC
            READ 1,NREDS
            PRINT 1,NREDS
            READ 3,(AMASS(I),I=1,NOM)
            READ (11,2) ((A(I,J),I=1,NOS),J=1,NOM)
            CALL PRA(A,NOS,NOM)
            CALL SELECT(A,SMIX,NOM,NOC)
          4 FORMAT(1X,F10.2,6X,5F10.3)
            DO 100 J=1,NOM
        100 PRINT 4,AMASS(J),(SMIX(I,J),I=1,NOC)
            DO 200 IX=1,NREDS
        200 CALL UNRAVL(A,SMIX,NOS,NOM,NOC,AMASS)
            CALL PRA(A,NOS,NOM)
            CALL MC(A,NOM,NOS,AMASS)
            CALL EXIT
            END
```

```
      SUBROUTINE UNRAVL(A,AMIX,NOS,NOM,NOC,AMASS)
C
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C        THIS PROGRAM PRODUCES A SPECTRUM FROM POSITIONS OF
C        ZEROS,  AS FOUND BY 'TRIAL'. THIS SPECTRUM IS THEN
C        SUBTRACTED OUT OF THE MIXTURES ARRAY GIVING AN ARRAY
C        OF LOWER RANK.
C
C        A IS THE WHOLE MIXTURES ARRAY.
C        AMIX IS THE SAME AS DESCRIBED IN 'TRIAL'.
C
C        READS IN -
C        NZ ,  NUMBER OF ZEROS USED.
C        KVZ , KEY VECTOR CONTAINING COLUMN NOS. (MASSES)
C        WHICH HAVE ZERO CONTRIBUTIONS FROM THE DESIRED
C        SPECTRUM.  I.E.  NZ  INTEGERS.
C        JU,  COLUMN NO. OF PEAK UNIQUE TO THE DERIVED SPECTRUM.
C
C        A MASS PROFILE CORRELATION ANALYSIS IS CARRIED OUT
C        ON THE NEW MIXTURES ARRAY.
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      DIMENSION A(40,25),AMIX(5,25),DS(25),U(25)
      DIMENSION AMASS(25),KVZ(50),B(25,25)
      DOUBLE PRECISION A,AMIX,DS,U,B
    1 FORMAT(I4)
    2 FORMAT(12I4)
    3 FORMAT('1')
    4 FORMAT(12X,F7.1,3X,6F8.2)
      READ 1,NZ
      READ 2,(KVZ(I),I=1,NZ)
      CALL ZEROS(AMIX,NOC,NOM,1,NZ,DS,KVZ)
      READ 1,JU
      BB=DS(JU)
      DO 220 I=1,NOS
      AX=A(I,JU)
      XK=AX/BB
      DO 220 J=1,NOM
      U(J)=DS(J)*XK
  220 A(I,J)=A(I,J)-U(J)
      PRINT 3
      CALL PRA(A,NOS,NOM)
      PRINT 3
      DO 230 J=1,NOM
  230 PRINT 4,AMASS(J),(A(I,J),I=1,5)
      CALL MC(A,NOM,NOS,AMASS)
      RETURN
      END
```

APPENDIX C.

Mean and Variance

In order to summarise a number of measurements (*e.g.* successively scanned mass spectra) it is convenient to calculate their mean (or average) and their variance. The latter is calculated by expressing each measurement as a deviation from the mean.

The mean is given by $\sum \frac{X}{n}$, where X is each measurement and "n" the number of measurements.

The variance is given by $\sum \frac{x^2}{n}$, where "x" is the deviation of each measurement from the mean.

The mean is a measure of central tendency. The variance is a measure of scatter. These two values are useful in summarising the nature of a distribution of measurements *e.g.* mass spectral peaks.

Correlation gives a measure of the tendency of two things to vary together *i.e.* to be associated or correlated. This is sometimes given by calculation of the Product-Moment Correlation Coefficient , "r" :

$$r = \frac{\sum \frac{x.y}{n}}{\sqrt{(\text{variance})_x} . \sqrt{(\text{variance})_y}}$$

where "x" and "y" are the deviations of each.

The product-moment correlation coefficient is by far the most widely used estimator of the degree of association or correlation. It varies in value from "-1" (perfect inverse relation) through zero (no relation) to "+1" (perfect positive relation). Some interpretations and limitations of the coefficient are given by Guilford(88).

The analysis of variance like most statistical procedures assumes that sampling is random hence the necessity to acquire large numbers of spectra for analysis.

Correlation matrix is a symmetric matrix containing the correlation coefficients of each series of measurements with every other. It therefore has dimensions equal to the number of series being correlated. The elements in the main diagonal are all "1.0" since these are self correlation coefficients.

Principal Components Analysis allows, by examination of the correlation matrix, an estimation of the minimum number of factors accounting for the variance in the data. There is only one possible set of factors for any correlation matrix ; the method therefore gives a unique solution. The factors found in an array of mixtures spectra bear no relation to the pure component spectra apart from their number. This number may be estimated by calculating the eigenvalues of the correlation matrix and deciding how many are significant. An I.B.M. Scientific Sub-routine called EIGEN may be used(71) as described in APPENDIX B (sub-routine PCA). The particular technique used is the diagonalization method originated by Jacobi and adapted by Von Neumann for large computers.(89)

Weighting

When the correlation matrix is calculated equal weighting is given to all the variables no matter how relatively small some of them may be i.e. some of the smaller variables are given too much weight. Perhaps this explains the sensitivity of the mass profile correlation matrix for Mixture II to the impurity peaks (see Table 17). According to the rank analysis these formed a relatively minor component. This effect is discussed by Hope(90). In future work it would be better to work with correlations related to the relative amounts as well as the patterns.

REFERENCES

1. Washburn,H.W, Wiley,H.F. and S.M. Rock, *Industr. Engng. Chem. Anal. Ed.*, *15* (1943) 541.

2. Barnard,G.P., "Modern Mass Spectrometry", The Institute of Physics", London (1953) 204.

3. Tate,J.T., Smith,P.T. and A.L. Vaughan, *Phys. Rev.*, *48* (1935) 525.

4. ref. 2, p. 192.

5. Reed,R.I., "Applications of Mass Spectrometry to Organic Chemistry", Academic Press, London (1966) Chapter 6.

6. Barnard,G.P. and L. Fox, "Conference on Applied Mass Spectrometry", Institute of Petroleum, London (1953) 269.

7. Johnsen,S.E.J., *Anal. Chem.*, *19* (1947) 305.

8. Ruth,J.M., *Anal. Chem.*, *40* (1968) 747.

9. ref. 2, p. 214 ; see also ref. 5.

10. Kiser, R.W., "Introduction to Mass Spectrometry and its Applications", Prentice-Hall, N.J. (1965) 233.

11. Quayle,A., in "Recent Topics in Mass Spectrometry", (R.I. Reed,Ed.) Gordon and Breach, New York (1971) 268.

    Edwards,T.R., in "Advances in Mass Spectrometry", (A. Quayle,Ed.) *5*, Institute of Petroleum, London (1971) 319.

12. Snyder,L.R., Howard,H.E. and W.C. Ferguson, *Anal. Chem.*, *35* (1963) 1676.

13. Robinson,C.J., *Anal. Chem.*, *43* (1969) 1548.

14. Beynon,J.H., "Advan. Mass Spectry.", *1* , Pergamon Press, London (1959) 328.

15. Biemann,K., P. Bommer and D.M. Desiderio, Tetrahedron Letters (1954) 1725 ; Biemann,K., P. Bommer, D.M. Desiderio and W.J. McMurray, "Advan. Mass Spectry.", *3* ,Inst. of Petroleum, London (1966) 639.

16. Johnson,B.H. and T. Aczel, *Anal Chem.*, *39* (1967) 682.

17. Waller,G.R., (Ed.), "Biochemical Applications of Mass Spectrometry", Wiley-Interscience (1972).

18. Burlingame,A.L. and G.A. Johanson, *Anal. Chem.*, *44* (1972) 339R.

19. Crawford,L.R. and J.D. Morrison, *Anal. Chem.*, *40* (1968) 1464.

20. Farbman, Mrs. S., Silva, Miss M.E.S.F. and R.I. Reed, *J. Mass Spectry. and Ion Phys.*, IN PRESS.

21. Ridley, R.G., in ref. 17.

22. Robertson,D.H., Cavagnaro,J., J.B. Holz and C. Merritt,Jr., "Twentieth Annual Conference on Mass Spectrometry and Allied Topics", 1972 , p. 359.

23. Grotch,S.L., Anal. Chem., 42 (1970) 1214.

24. Crawford,L.R. and J.D. Morrison, Anal. Chem., 40 (1968) 1469.

25. Smith,D.H., Anal. Chem., 44 (1972) 536.

26. Jurs,P.C., Kowalski,B.R. and T.L. Isenhour, Anal. Chem., 41 (1969) 695; Tunnicliff,D.D. and P.A. Wadsworth, Anal. Chem., 45 (1973) 12.

27. Buchs,A., Duffield,A.M., Schroll,G.,Djerassi,C., Delfine,A.B., Buchanan,B.G., Sutherland,G.L., Feigenbaum,E.A. and J. Lederberg, J. Amer. Chem. Soc. 92 (1970) 6831.

28. Crawford,L.R. and J.D. Morrison, Anal. Chem., 43 (1971) 1790.

29. Jurs,P.C., Kowalski,B.R. and T.L. Isenhour, Anal. Chem., 41 (1969) 21.

30. Jardine,A., Silva, Miss M.E.S.F. and R.I. Reed, Org. Mass Spec., 7 (1973) 601.

31. Perrine,C., Cannon,E., Cavender,C., Farrwell,T. and R.E. Lovins, "Advan. Mass Spectry.", 5 , Institute of Petroleum, London (1971) 728.

32. Smith,D.H., Olsen,R.W., Walls,F.C. and A.L. Burlingame, Anal. Chem., 43 (1971) 1796.

33. Hites,R.A. and K. Biemann, "Advan. Mass Spectry.", Inst. of Petroleum, London, 4 (1968) 37.

34. Ziegler,E., Henneberg,D. and G. Schomburg, Anal. Chem., 42 (9) (1970) 51A.

35. Baldeschwieler,J.D., Science, 159 (1968) 263.

36. Field,F.H., "Advan. Mass Spectry.", Inst. Petroleum, London (1968) 645.

37. Beckey,H.D., "Field Ionization Mass Spectrometry", Pergamon Press, 1971.

38. James,A.T. and A.J.P. Martin, Biochem. J., 50 (1952) 679.

39. Bowen,H.C., Clayton,E., Shields,D.H. and H.M. Stanier, "Advan. Mass Spectry.", Inst. Petroleum, London, 4 ,(1968) 257.

40. Halliday,J.S., ibid. , 239.

41. Sweeley,C.C., Elliot,W.H., Fries,I. and R. Ryhage, Anal. Chem., 38 (1966) 1549.

42. Snedden,W. and R.B. Parker, Anal. Chem., 43 (1971) 1651; Majer,J.R. and M.J.A. Reade, Chem. Commun., (1970) 58; Jenkins,A.E. and J.R. Majer, Talanta, 14 (1967) 777.

43. Meyerson,S., Anal. Chem., 31 (1959) 174.

44. McCormick,A., Ph.D. Thesis, Glasgow (1966).

45. Crombie,L., Games,D.E. and A. McCormick, Tet. Letters, (1966) 145.

46. Laurie,W.A., Ph.D. Thesis, Glasgow,(1969) 69.

47. Laurie,W.A., *ibid.*, 78.

48. Monteiro,L.F., Ph.D. Thesis, Glasgow (1969).

49. Monteiro,L.F. and R.I. Reed, J. Mass Spectry. and Ion Phys., 2 (1969) 265.

50. Kirkland,J.J., (Ed.) "Modern Practice of Liquid Chromatography", Wiley-Interscience, N.Y., 1971.

51. Margenau,H. and G.M. Murphy, "The Mathematics of Physics and Chemistry" Van Nostrand, New York (1943) Chapter 10.

52. Faddeeva,V.N., "Computational Methods of Linear Algebra", Dover , New York, (1959) 33.

53. Faddeeva,V.N., *ibid.* 63.

54. Pennington, R.H., "Introductory Computer Methods and Numerical Analysis", 2nd. Ed., Macmillan, London (1971) 339, 374.

55. Naur,P., (Ed.) "Revised Report on the Algorithmic Language, ALGOL 60", The Computer Journal (1963).

56. MS9 Mass Spectrometer Manual, A.E.I., Manchester.

57. ref. 48, p.119.

58. ref. 2, p. 65; ref. 83, p. 431.

59. Careri,G., Il Nuovo Cimento, 9, N.2 (1952) 128.

60. MS902S Mass Spectrometer Manual, G.E.C.-A.E.I., Manchester.

61. Smith,D.H., Olsen,R.W., Walls,F.C. and A.L. Burlingame, Anal. Chem. 43 (1971) 1796.

62. Bokhoven,C. and H.J. Theeuwen, "Advan. Mass Spectry.",1 (1959) 222.

63. Fox,R.E. and J.A. Hipple, J.Chem. Phys., 15 (1947) 208.

64. Cassuto,A., "Advan. Mass Spectry.", 2 (1963) 296.

65. Monteiro,L.F., Ph.D. Thesis, Glasgow (1969).

66. Halket,J.McK., B.Sc. Thesis, Glasgow (1970).

67. Haddon,W.F., Chait,E.M. and F.W. McLafferty, Anal. Chem., 38 (1966) 1968.

68. Brown,W.F. and I.P. Fisher, Talanta, 19 (1972) 798.

69. d-mac Pencil Follower - Type pf.10,000 Mark 1B, d-mac Ltd., High Wycombe.

70. "370/155 User's Guide", Edinburgh Regional Computing Centre, July 1972.

71. Programmers Manual, Programme Number 360A-CM-03x, System 360 Scientific Subroutine Package Version III, (August, 1970)

72. Kendrick,E., personal communication.

73. Ainsworth,S., P.S.G. Bulletin No. 20, July 1972, 611.

74. Wallace,R., and S.M. Katz, J. Phys. Chem., 68 (1964) 3890.

75. Ainsworth,S. and W.S.W. Bingham, Biochim. Biophys. Acta., 160 (1968) 10.

76. Katakis,D., Anal. Chem., 37 (1965) 876.

77. Hope,K., "Elementary Statistics- a Workbook", Pergamon Press, Oxford (1967).

Anderson,T.W., "Introduction to Multi-variate Analysis", John Wiley and Sons, New York, N.Y. (1958).
Anderson,T.W., Ann. Math. Statist., 34 (1963) 122.

78. Lawton,W.H. and E.A. Sylvestre, Technometrics, 13 (1971) 617.
Ohta,N., Anal. Chem., 45 (1973) 553.

79. Macnaughtan Jr.,D., Rogers,L.B. and G. Wernimont, Anal. Chem., 44 (1972) 1421.

80. ref. 51, p. 415.

81. Eden,J., Burr,B.E. and A.W. Pratt, Anal Chem., 23 (1951) 1735.

82. Thurstone,L.L., "The Vectors of Mind", Univ. of Chicago Press (1935)

83. Beynon,J.H., "Mass Spectrometry and its Applications to Organic Chemistry", Elsevier,Amsterdam, (1960) 105.

84. Golub,G., "Numerical Method for Solving Linear Least Squares Problems", Numerische Mathematik, 7 (1965) 206.

85. Grigsby,R.D., Hansen,C.O., Mannering,D.G., Fox,W.G. and R.H. Cole, Anal. Chem. 43 (1971) 1135.

86. Sillén,L.G., Acta. Chem. Scand., 18 (1964) 1085.

87. Reed,R.I. in "Proceedings of the International Conference on Mass Spectroscopy, Kyoto", University of Tokyo Press (1970) 1217.

88. Guilford,J.P. "Fundamental Statistics in Psychology and Education", 3rd. Edn. McGraw-Hill (1956).

89. Ralston,A. and H.S. Wilf (Eds.) "Mathematical Methods for Digital Computers", John Wiley & Sons, New York (1962) Chapter 7.

90. ref. 77 (Hope,K.) 49.