

Influence of random errors on the spectrophotometric analysis of multicomponent mixtures

^aJ. ŠUSTEK and ^bO. SCHIESSL

^aResearch Institute of Agrochemical Technology,
810 04 Bratislava

^bState Forest Product Research Institute,
891 99 Bratislava

Received 20 March 1974

Accepted for publication 14 October 1974

The influence of random errors on the spectrophotometric determination of multicomponent mixtures was studied on a three-component model system, the spectrum of which was approximated by the Lorentz functions. The random errors simulated by a computer were normally distributed. The concentrations of the particular components were computed by solving an overdetermined system of linear equations by the method of least squares.

At the present time several ways of the evaluation of the spectrophotometric method for the determination of multicomponent mixtures are used. One of them is based on the solving of an overdetermined system of linear equations by means of the method of least squares [1–3]. It is an advantage of this procedure that it allows to estimate the error of the determination of the particular component concentrations even from a single recording of the sample spectrum if the sample contains only the components considered at the calibration and the number of analytical positions is greater than that of the components analyzed. The overdetermined system can be chosen in three ways [3]:

– Absorbance is measured at a greater number k of wavenumbers (wavelengths) than that of the components being determined n ($k > n$).

– The number of the calibration mixtures m is greater than that of the components being determined n ($m > n$).

– A double overdetermined system ($k > n$, $m > n$).

The algorithm of the calculation using the method of least squares with a double overdetermined system of linear equations is, in the matrix form, illustrated by a block diagram in Fig. 1. The calibration data are computed from the known concentrations of the components in calibration mixtures and the absorbances measured on their spectra. Small random errors of the absorbance measurements, which cause the inconsistency of equations, are distributed among particular equations in such a way that the squares of deviations of the measured absorbances from the calculated absorbances are minimal. The occurrence of systematic measuring errors or blunders in an equation shows itself in the matrix of the residues of the mixture absorbances E_A . Examining the deviations of the measured absorbances from the products of the absorptivities and the mixture component concentrations (*i.e.* from the computed absorbances) it can be found whether there are only small random errors in the system or if the errors appertain to one equation prevalingly [4].

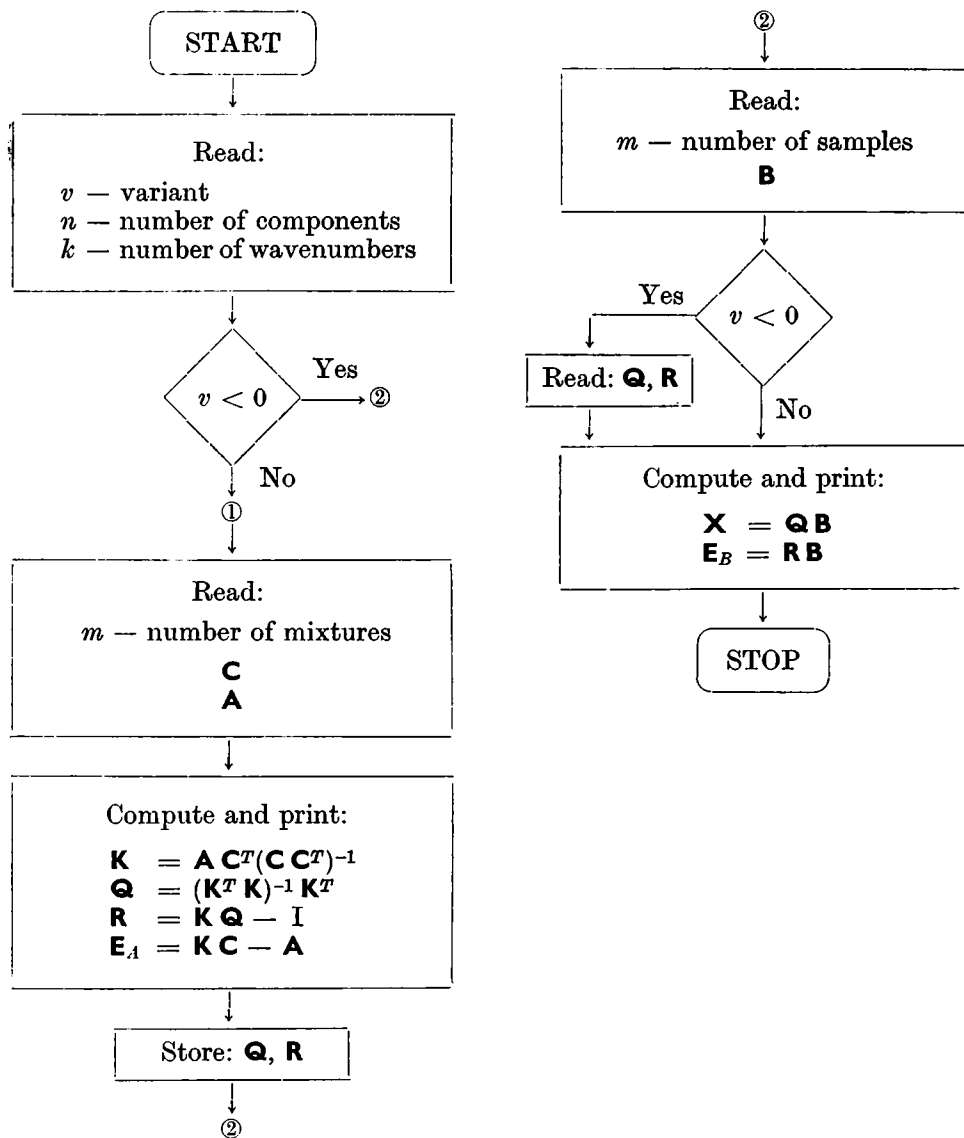


Fig. 1. Block diagram.

Solution of an overdetermined system of linear absorption equations by the method of least squares.

The aim of the present work is to examine the influence of random errors on the precision of the determination of multicomponent compounds. We are not going to deal with the provenance of random errors as their sources are the object of numerous papers [4–8].

Theory

Let

$$\mathbf{C} = \|c_{ji}\|_{nm}$$

be the matrix of the actual concentrations of n components in m mixtures and

$$\mathbf{A} = \|a_{ji}\|_{km}$$

the matrix of the absorbances of those m mixtures at k analytical positions. Supposing the validity of the Lambert—Beer law and of the additivity of the absorbances of the particular mixtures,

$$\mathbf{C} = \mathbf{Q} \cdot \mathbf{A} \quad (1)$$

is valid. Taking into account the inaccuracy of measuring, the deviations from the above-mentioned theoretical relations and the presence of further (unknown) components, instead of eqn (1) a system of equations for computing the matrix \mathbf{X} of the approximate values of the component concentrations in the analyzed samples is obtained

$$\mathbf{X} = \mathbf{Q} \cdot \mathbf{B}. \quad (2)$$

The matrix \mathbf{B} can be considered as the sum of the theoretical absorbance matrix \mathbf{A} and the matrix \mathbf{B}_1 , which involves deviations from the theoretical relations and the measuring errors

$$\mathbf{B} = \mathbf{A} + \mathbf{B}_1. \quad (3)$$

By substituting (3) into (2) we obtain

$$\mathbf{X} = \mathbf{Q} \cdot (\mathbf{A} + \mathbf{B}_1) \quad (4)$$

or

$$\mathbf{X} = \mathbf{C} + \mathbf{Q} \cdot \mathbf{B}_1. \quad (5)$$

The term

$$\mathbf{E}_{X_1} = \mathbf{Q} \cdot \mathbf{B}_1 \quad (6)$$

represents the matrix of deviations of the concentrations found experimentally from the actual concentrations of the considered components.

If the elements of matrix \mathbf{B}_1 are multiplied by constant \varkappa , we obtain the matrix

$$\mathbf{B}_2 = \varkappa \cdot \mathbf{B}_1 \quad (7)$$

and in eqn (5) there will be the term \mathbf{E}_{X_2} instead of $(\mathbf{Q} \cdot \mathbf{B}_1)$, where

$$\mathbf{E}_{X_2} = \mathbf{Q} \cdot \mathbf{B}_2 \quad (8)$$

which, after substitution of \mathbf{B}_2 from (7) and taking into account (6), gives

$$\mathbf{E}_{X_2} = \varkappa \cdot \mathbf{E}_{X_1}. \quad (9)$$

Comparing (7) to (9) we arrive at the conclusion that with the increasing deviations of the absorbances from the theoretically exact values the deviations of the computed concentrations from the actual concentrations of the considered components increase proportionally.

Study of the influence of random errors

For the study of the influence of random errors we chose a multicomponent model system in which one absorption band of the Lorentzian type appertains to each component [9]

$$A_j(\nu) = \frac{A_{0j}}{1 + \frac{4(\nu - \nu_{0j})^2}{\Delta\nu_{1/2}^2(j)}}. \quad (10)$$

We supposed the validity of the absorption spectra additivity from which it ensues that the total absorbance is the sum of the absorbances of the particular components in a chosen analytical position (wavenumber ν)

$$A(\nu) = \sum_{j=1}^n A_j(\nu). \quad (11)$$

We also supposed the validity of the Lambert–Beer law, *i.e.* the constant A_{0j} in (10) is proportional to the concentration of the j th component.

For the chosen matrix of concentrations \mathbf{C} , by means of eqns (10) and (11), we computed the relevant absorbance matrix \mathbf{A}_0 , to which we added the random errors. Their distribution was given by the Gaussian frequency. We chose a set of errors, from which the particular values were taken, as follows.

We chose the standard deviation σ of the Gaussian error distribution and an interval of the absorbance deviations ($-t \cdot \sigma$, $+t \cdot \sigma$) from which the errors were taken. This interval was divided into r equal intervals (β_p, β_{p+1}) and we calculated the integrals over those intervals

$$F_p = \int_{\beta_p}^{\beta_{p+1}} G \, d\beta, \quad (12)$$

where G is the Gauss function.

Furthermore, we divided the particular intervals (β_p, β_{p+1}) equally so as to obtain a number N_p of fine dividing intervals, approximately proportional to the area F_p

$$N_p \doteq k \cdot F_p, \quad (13)$$

where k is the proportionality coefficient*.

In the interval (β_p, β_{p+1}) there are N_p points $\beta_{p1}, \beta_{p2}, \dots, \beta_{pN_p}$, which are the limits of the intervals of fine dividing and represent the set of selected errors. At the selection, the left side limit of the interval (β_p, β_{p+1}) was always left out. The total number of points N , *i.e.* the number of the selected values of the errors was

$$N = \sum_{p=1}^r N_p. \quad (14)$$

The selected errors were at random inserted into the error field, from where they were gradually added to the values $A(\nu)$, computed according to eqn (11).

* The approximation of eqn (13) lies in the rounding up of the real number to an integer.

Model system

A three-component model system was chosen as an example. In the calculation of the calibration data, the matrix **C** represented the set of 30 calibration mixtures and the concentrations of the particular components were within the limits of the interval 0–1.00. For 30 wavenumbers the absorbance matrix **A**₀ was computed by means of eqn (10), where

$$A_{0j} = k_{0j} \cdot c_j. \quad (15)$$

The parameters used are given in Table 1. According to the procedure described above, the random errors were added to the matrix **A**₀ computed in this way and

Table 1

Parameters for calculating the model spectra

<i>j</i>	ν_{0j} [cm ⁻¹]	g_{0j}	k_{0j}	$\Delta\nu_{1/2}(j)$ [cm ⁻¹]	c_j
1	5	5	0.4	3	0–1.0
2	13	13	0.8	4	
3	25	25	0.8	5	

so the matrix **A** was obtained ($N = 800$, $r = 154$, $\sigma = 0.01$, and $t = 3$). The model spectrum of one of the mixtures used is shown in Fig. 2. The calibration data were calculated from the matrices **C** and **A** by using the block diagram in Fig. 1.

Analogously to the calculation of the matrix **A**₀, the column matrix **B**₀ of the sample absorbances (chosen concentrations of the particular components in the sample: $c_1 = c_2 = c_3 = 0.5$) was calculated. By adding ten different random errors to the elements of the column matrix **B**₀ the set of absorbances representing the

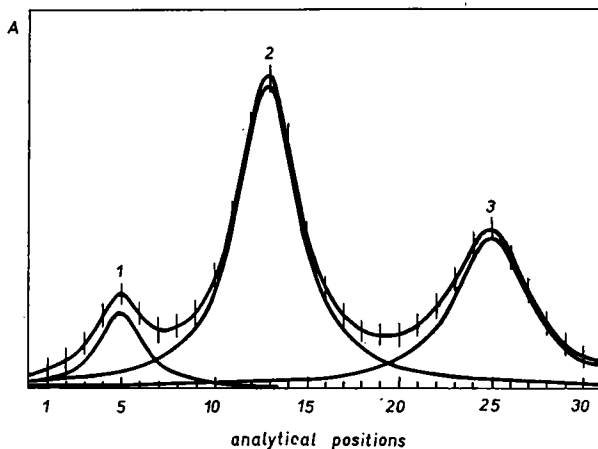


Fig. 2. Model spectrum of a calibration mixture.

matrix **B** was obtained. In this way we approximated 10 repeated measurements of one sample. The set of statistical data was thus obtained.

To be able to evaluate statistically the influence of the introduced errors, we repeated the whole procedure described ten times. The sets of data obtained in this way were evaluated independently. From all the obtained values one larger set was made up and evaluated as a whole.

Statistical evaluation of results

Multiplying the matrix **B** by matrix **Q** using eqn (2) gave the matrix of the component concentrations in the sample. The means \bar{x}_j of ten determinations ($m = 10$) of each component are given in Table 2. As a criterion for evaluating the precision of the determination of the particular components, the values of the estimates of the standard deviations

$$s_{c_j} = \sqrt{\frac{\sum_{i=1}^m (c_j - x_{ji})^2}{m - 1}} \quad (16)$$

and

$$s_{\bar{x}_j} = \sqrt{\frac{\sum_{i=1}^m (\bar{x}_j - x_{ji})^2}{m - 1}} \quad (17)$$

are given in Table 2, where c_j is the actual concentration of the particular components in the sample.

The 95% confidence interval was given by

$$\bar{x}_j \pm t_{\alpha/2, \Phi} \cdot \frac{s_{\bar{x}_j}}{\sqrt{m}}, \quad (18)$$

where $\Phi = m - 1$. For $m = 10$ the value $t_{0.025, 9} = 2.262$.

The mathematico-statistical characteristics for the whole set ($m = 100$) were calculated analogously to the evaluation of the particular set of ten determinations ($m = 10$). At the calculation of the 95% confidence interval the value $t_{0.025, 99} = 1.984$ was used.

The estimates of standard deviations of the absorbance calculated according to eqn

$$s = \sqrt{\frac{\sum_{g=1}^k e_g^2}{k - n}} \quad (19)$$

are not given, as they are in good accordance with the chosen standard deviation $\sigma = 0.01$. In eqn (19), e_g are elements of the matrices of the absorbance residues **E_A** and **E_B** respectively (see block diagram).

Table 2

Statistical evaluation of concentration determinations

<i>j</i>	<i>c_j</i>	Parameter varied	Particular sets of data (<i>m</i> = 10)										Set of data (<i>m</i> = 100)
			1	2	3	4	5	6	7	8	9	10	
1	0.500	\bar{x}_1	0.532	0.520	0.514	0.519	0.506	0.508	0.514	0.505	0.499	0.495	0.511
		$\bar{x}_1 - t$	0.511	0.499	0.497	0.499	0.492	0.487	0.493	0.485	0.481	0.476	0.495
		$\bar{x}_1 + t$	0.553	0.541	0.531	0.539	0.520	0.529	0.535	0.525	0.517	0.514	0.527
		s_{c_1}	0.045	0.036	0.028	0.034	0.020	0.029	0.033	0.028	0.025	0.027	0.030
		$s_{\bar{x}_1}$	0.029	0.030	0.024	0.028	0.019	0.028	0.029	0.028	0.025	0.027	0.026
2	0.500	\bar{x}_2	0.497	0.497	0.498	0.495	0.506	0.495	0.498	0.498	0.499	0.495	0.498
		$\bar{x}_2 - t$	0.490	0.492	0.494	0.489	0.501	0.490	0.492	0.494	0.494	0.489	0.494
		$\bar{x}_2 + t$	0.504	0.502	0.502	0.501	0.511	0.500	0.504	0.502	0.504	0.501	0.502
		s_{c_2}	0.010	0.007	0.007	0.010	0.009	0.008	0.009	0.007	0.007	0.009	0.008
		$s_{\bar{x}_2}$	0.010	0.007	0.006	0.008	0.007	0.007	0.009	0.006	0.007	0.008	0.007
3	0.500	\bar{x}_3	0.502	0.498	0.500	0.498	0.497	0.499	0.497	0.501	0.501	0.502	0.500
		$\bar{x}_3 - t$	0.493	0.492	0.496	0.493	0.493	0.495	0.495	0.497	0.498	0.497	0.496
		$\bar{x}_3 + t$	0.506	0.504	0.504	0.503	0.501	0.503	0.499	0.505	0.504	0.507	0.504
		s_{c_3}	0.006	0.009	0.005	0.007	0.007	0.005	0.004	0.006	0.004	0.008	0.006
		$s_{\bar{x}_3}$	0.003	0.003	0.005	0.007	0.006	0.005	0.003	0.006	0.004	0.007	0.006

Discussion

It can be seen from Table 2 that the means of the determined concentrations of the 1st component are in most cases higher than the actual value. This is probably in connection with the lower (half as much) absorptivity of this component when compared to the others. From the equal absorptivities of the 2nd and 3rd components it could be expected that the deviations of the determination of these components should be approximately equal. In fact, the results of the determination of the 2nd component are in most cases lower than the actual value, whereas in the case of the 3rd component the means of the calculated concentrations lie around the actual values. This fact may be explained by back influence of the first component, the absorption band of which is nearer to the absorption band of the 2nd component.

The magnitude of the absorptivities and the positions of the absorption bands influence also the values of the estimates of standard deviations; thus the relations

$$s_{c1} \gg s_{c2} > s_{c3}$$

and

$$s_{x1} \gg s_{x2} > s_{x3}$$

are valid. For the 2nd set of data the second inequality does not hold.

Higher values of the estimates of the standard deviations s_{x1} for the 1st component cause a widening of the confidence interval of the determination of this component. As the introduced random errors have been given by normal distribution, the actual concentrations are mostly within the confidence intervals calculated for the particular sets of data. The number of exceptions corresponds roughly to the chosen level of significance ($\alpha = 0.05$). From the evaluation of the whole set ($m = 100$) it can be seen that the actual concentrations of the particular components are in all cases within the calculated confidence intervals.

The advantage of the method of least squares in solving an overdetermined system of linear absorption equations lies in the fact that it allows to calculate the matrices of the absorbance residues \mathbf{E}_A and \mathbf{E}_B respectively, from which the blunders in the measurement of the absorbances can be found directly and the experimental errors can be estimated.

Symbols

$g = 1, 2, \dots, k$	sequence of the analytical position
$i = 1, 2, \dots, m$	sequence of the mixture (sample)
$j = 1, 2, \dots, n$	sequence of the component
k	number of analytical positions
m	number of mixtures (samples)
n	number of components
\mathbf{A}	absorbance matrix of mixtures ($k \times m$ matrix)
\mathbf{B}	absorbance matrix of samples ($k \times m$ matrix)
\mathbf{C}	matrix of the component concentrations in mixtures ($n \times m$ matrix)
\mathbf{C}^T	transpose of the matrix \mathbf{C} ($m \times n$ matrix)
\mathbf{E}_A	matrix of the residues of mixture absorbances ($k \times m$ matrix)
\mathbf{E}_B	matrix of the residues of sample absorbances ($k \times m$ matrix)
\mathbf{I}	unit matrix ($k \times k$ matrix)

K	matrix of absorptivities ($k \times n$ matrix)
K^T	transpose of the matrix K ($n \times k$ matrix)
Q	matrix for calculation of the concentrations ($n \times k$ matrix)
R	matrix for calculation of the residual absorbances ($k \times k$ matrix)
X	matrix of the component concentrations in samples ($n \times m$ matrix)
a_{gt}	element of the matrix A (absorbance of the i th mixture at the g th analytical position)
c_{ji}	element of the matrix C (concentration of the j th component in the i th mixture)
k_{gj}	element of the matrix K (absorptivity of the j th component at the g th analytical position)
x_{jt}	element of the matrix X (concentration of the j th component in the i th sample)
\bar{x}_j	mean of the determined concentrations of the component j
s_{cj}	standard deviation of the concentration
s_{zj}	estimate of the standard deviation of the concentration
$s_{\bar{x}_j}$	estimate of the standard deviation of the mean \bar{x}_j
$A_j(\nu)$	absorbance of the component j at wavenumber ν
A_{0j}	absorbance of the component j at wavenumber ν_{0j}
ν_{0j}	wavenumber corresponding to the maximum of the band of the component j
$\Delta\nu_{1/2}(j)$	the half-band width of the component j
g_{0j}	sequence number of the analytical position corresponding to the maximum of the band of the component j
k_{0j}	absorptivity of the component j at analytical position g_{0j}

References

1. Vasilev, A. F., *Zavod. Lab.* **31**, 677 (1965).
2. Przybylski, Z., *Chem. Anal.* (Warsaw) **13**, 453 (1968).
3. Šustek, J., Livař, M., and Schiessl, O., *Chem. Listy* **66**, 168 (1972).
4. Bauman, R. P., *Absorption Spectroscopy*, pp. 364–433. Wiley, New York, 1967.
5. Kössler, I., *Kvantitativní infračervená spektrometrická analýza*. (Quantitative Infrared Spectrometric Analysis.) Pp. 13–108. Státní nakladatelství technické literatury. (State Publishing House of Technical Literature.) Prague, 1970.
6. Meehan, E. J., *Treatise on Analytical Chemistry*, Part I, Vol. 5, pp. 2753–2803. (Kolthoff, I. M. and Elving, P. J., Editors.) Wiley, New York, 1966.
7. Schilt, A. A. and Jaselskis, B., *Treatise on Analytical Chemistry*, Part I, Vol. 5, pp. 2943–3055. (Kolthoff, I. M. and Elving, P. J., Editors.) Wiley, New York, 1966.
8. Smith, A. L., *Treatise on Analytical Chemistry*, Part I, Vol. 6, pp. 3535–3743. (Kolthoff, I. M. and Elving, P. J., Editors.) Wiley, New York, 1965.
9. Papoušek, D. and Plíva, J., *Collect. Czech. Chem. Commun.* **30**, 3007 (1965).

Translated by J. Kovačičová