

An attempt to separate $A \rightarrow B$ process spectra. Separation by Minimum overLapping Method SEMILAM

Izabela K. Zawadzka, Jacek A. Soroka,* Marta J. Sawicka,
and Monika Gąsiorowska

*Section of Instrumental Analysis, Institute of Chemistry and Environmental Protection,
Szczecin University of Technology, Piastow Ave. 42, Szczecin, PL 71-065, Poland
E-mail: sorja@ps.pl*

Received 12 January 2006; Revised 15 January 2007

It is of particular importance to characterise spectrally each component in the chemical systems in which the reaction proceeds or when the two forms of the same compound exist in equilibrium (tautomerism, *cis-trans* isomerism, dissociation, etc.). Moreover, physical or chemical separation of the component that is in a smaller amount is always difficult, as well as recording its spectrum. A new simple numerical method of extraction of overlapped spectra of two components is presented. It was termed SEMILAM, which stands for 'separation by minimum overlapping method'. This method can be applied when two spectra of the same sample consisting of two species can be recorded. These spectra have to differ in composition but total concentration has to maintain constant. In other words, the data required to apply the method comes from one sample for which the two spectra were recorded at two different points of time (the composition of the sample has to change). Knowledge of real fractions found in the crude plots is sufficient to extract quantitatively the spectra of pure components. This new simple method was successfully tested on simulated spectra and then applied to a real problem.

KEY WORDS: spectra deconvolution, equilibrium processes, binary mixtures

1. Introduction

The commonly used analytical techniques without a previous separation (direct ones) often fail if the sample is a mixture. The existence of two forms of the same substance or two different compounds (for instance substrate and product) in a solution can produce false results of an analysis. Furthermore, traditional methods of mixture separation, for instance chromatography, chemical extraction or distillation, are often laborious and expensive, or may even appear

* Corresponding author.

unreliable in the case of equilibrated systems or reaction proceeding. An appropriate mathematical method of data processing (such as off-line spectra separation after measurement), which leads to the spectra extraction from the crude plot, can be a noteworthy alternative.

In order to define the stoichiometry of equilibrated mixtures one can use models based on the approximation of the experimental data by means of least squares curve fitting methods [1, 2]. Thus, many algorithms have been proposed based on spectrometric data, providing information about the equilibrium in a solution. In the recent years the following methods have been developed: (Evolving Factor Analysis) (EFA) [3], Heuristic Evolving Latent Projection (HELP) [4], SIMPLISMA [5], Least Squares Algorithm (ALS) [6], PARALLEL FACTOR analysis (PARAFAC) [7], Generalized Rank Annihilation Method (GRAM) [8] and Tri-Linear Decomposition (TLD) [9]. The EFA method evaluates the pure spectra (chromatograms) or the concentration profiles on the basis of the mixtures' plots. An extensive set of data has to be collected in a matrix and then complicated algorithms have to be applied to transform this matrix and eventually evaluate deconvoluted data. This method is commonly applied in chromatography to deconvolute overlapped peaks. The major advantage of this method is that it is completely independent on selective regions but to be applied successfully one has to have zero-concentration window, i.e. the other analytes have to contribute significantly to the signal in the zero-concentration window when compared to the contribution to the noise [10]. HELP is similar to the previous one but uses also the selective region information existing in the data [10]. SIMPLISMA is a common and useful tool in analytical chemistry although it needs huge set of the data (collected in matrix). Moreover it is necessary to select very pure rows or columns before the resolution can be performed because the method needs least square approach for the resolution of data set [11]. PERAFAC also needs huge set of data collected in a 3D-matrix, not in two-dimensional one [12]. Basically all methods mentioned above rely on the very big sets of data in order to deconvolute overlapped peaks. The developments in the field of computation of equilibrium constants from experimental data were reviewed a few years ago [13–16]. The description of the frequently used computer programs developed in order to evaluate instability constants from spectrophotometric data based on knowledge of chemical model are presented in the paper [17].

Owing to rapid development in the techniques and computer systems mixture spectra deconvolution can be carried out by means of Principal Component Analysis (PCA) method using the data from a high-class diode array spectrometer [18]. This method was established on mathematical data processing – (Singular Value Decomposition) (SVD) [19, 20]. The high cost of diode array spectrometer using high-quality optics can be regarded as a disadvantage of this method.

The H-point Curve Isolation Method (HPCIM) [21, 22] was proposed for a binary mixture analysis where A component is known and the other (B) component is unknown. With this method, it is possible to plot the spectrum of B species in a simple way, without forcing the data to fit the specific model. It is necessary to have the analyte and the sample spectra only. The method cancels the contribution of A in the sample signal, thereby giving a set of possible spectra for B species. The real B spectrum can be calculated by finding pairs of wavelengths according to the calibration model of the H-Point Standard Addition Method (HPSAM) [23, 24]. However, *a priori* knowledge of at least one component is needed to apply this method.

Still, the majority of the methods were elaborated primary for one peculiar problem encountered by the authors during their studies. This paper presents a universal method for the separation of binary mixture spectra without the necessity to fit the spectral data to a specific model. The analysed sample can consist of two forms of the same chemical substance or be a mixture of two different compounds (a substrate and a product of a chemical reaction). This method requires one only to record two spectra of one sample containing the different fractions of each component during each measurement (performed at different points of proceeding processes). The only condition that limits the use of this method is to maintain the total concentration of a sample. Similarly to HPCIM method it is possible to plot the spectrum of A and B species in a simple way. The method cancels the contribution of B or A in the sample signal, thereby giving a set of possible spectra for A or B species, respectively. The real spectra are chosen by means of statistical methods.

2. Theory

The method described in the paper can be used when two different spectra, i.e. S_1 and S_2 are known. The spectra have to be recorded for the same sample at different points of the processes under investigation while the total component concentration of the sample has to maintain constant.



$$[A] + [B] = \text{const.}$$

The principal feature of the spectrum in the range of validity of Lambert–Beer’s law is its digital resolution; the spectrum is registered as a vector with a constant-repeated steps i.e. an independent variable. In this case, the spectra can be described as:

$$\begin{aligned} S_1 &= x_{A1}S_A + (1 - x_{A1})S_B \\ S_2 &= x_{A2}S_A + (1 - x_{A2})S_B, \end{aligned} \tag{1}$$

where: S_1, S_2 – spectra of samples measured at some points of the above mentioned processes, frequently overlapped;

S_A, S_B – spectra of pure components (component A and B, respectively);

x_{A1}, x_{A2} – the fractions of spectrum S_A in the crude spectra (in S_1 and S_2 , respectively);

and $x_{A1} \neq x_{A2}$;

The classic method of separation relies on a mathematical transformation depending on:

- the spectrum of binary system consisting of the bands of A and B components is described as:

$$S = xS_A + (1 - x)S_B, \quad x - S_A \text{ fraction in } S \quad (2)$$

- and if the spectrum S_A is known, then:

$$S_B = (S - x \cdot S_A)/(1 - x), \quad (3)$$

which is a key-step of the ‘extraction’ procedure available in MEDSON M48 software [25]. By cancelling the contribution of A in a binary plot one can obtain a pure B spectrum. It is necessary to choose the proper x only to achieve this. If both spectra of pure components are unknown (S_A and S_B) the two unknowns have to be found. The fractions of S_A spectrum in both spectra, i.e. S_1, S_2 , specifically x_{A1} and x_{A2} have to be found see equation (1). If S_1, S_2, x_{A1} and x_{A2} are known, equation (1) becomes a system of two equations with two unknowns. The univocal solution gives S_A and S_B . Finding fractions of the spectra of pure components in binary mixtures plots usually poses a serious problem because one can create an unlimited set of possible solutions (figure 3). However, only one of them is true. Therefore, a test has to be discovered to indicate the true answer. In order to use this method reliably the spectrum recorded during the process should not change more than by 1–2%, otherwise the obtained error will be too large. In other words, the sample composition should not change more than by 1–2% during the spectrum recording. It was assumed that all the possible processes in the sample were inhibited during the spectra measurement. The method was elaborated under the assumption that the spectra of binary mixtures comprise a linear combination of pure component bands. First, to elaborate this method, synthetic spectra were used. The columnar vectors with 226 rows of absorbance-energy dependence were used. These artificial spectra can be

described as a sum of Gaussian curves:

$$S(i) = \sum_{k=1}^n A_k \exp \left[-\frac{(w_i - w_{0,k})^2}{2 \cdot \sigma_k^2} \right], \quad (4)$$

where: A_k – maximum amplitude of the k -band;

w_i – wavelength;

$w_{0,k}$ – position of the k -band;

σ_k – parameter of a band-width;

i – the number of a spectrum point (indicating wavelength, wave-number, etc.);

k – number of bands,

and they were generated by means of our own software. The applied, synthetic spectra of binary mixtures comprise a linear combination of the generated spectra of pure components. Particular plots were generated due to procedure: $a\%$ of spectrum S_A to was added to $(100-a)\%$ of spectrum S_B . Afterwards, on the basis of the S_1 and S_2 spectra and after finding the accurate x_{A1} and x_{A2} by means of numerical processing, the original spectra i.e. S_A and S_B were calculated. The system of equations with two unknowns (1) was solved, i.e. the pure S_A and S_B spectra were extracted.

During the investigations it was assumed that in the set of potential solutions the true, pure component spectra should be as dissimilar as possible. Certainly, these predicted spectra should also fulfil the basic criteria of ‘physical reality’, for instance the signals must not be below zero (negative absorption), etc. The statistical methods were used to achieve the goal. One tried to find a number, which would univocally characterise the extracted spectra. Therefore, to find appropriate fractions (x_{A1} , x_{A2}) it was decided to employ the correlation feature based on the rule that the most different spectra should therefore have the lowest mutual correlation and overlapping.

To prove that the method is correct and to estimate the error of the method experiments on synthetic data were conducted. In order to demonstrate a practical application of this method it was in fact applied to a real problem. At the same time the influence of the positions of the overlapping spectra on results was checked.

3. Results and discussion

3.1. Synthetic Spectra

Two spectra were generated referring to S_A and S_B in equation (1); each one as a singular Gaussian function based on equation (4) for $k = 1$. In addition, the bands were differently positioned in relation to each other, i.e. $w_{0,k}$ see equation (4) was changed (figure 1). Afterwards, these spectra were numerically

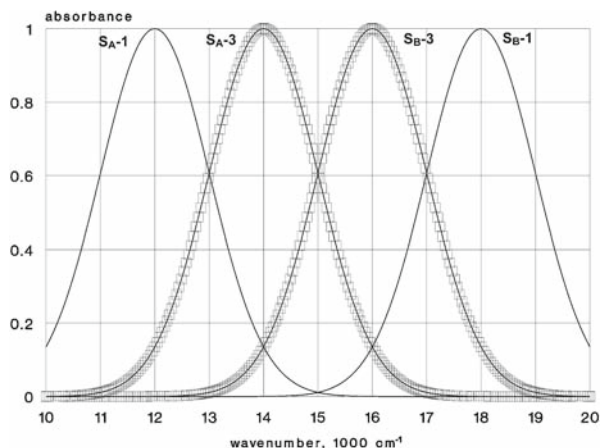


Figure 1. Two pairs of pure spectra: pair S_A-1 and S_B-1 : minor overlapping degree, pair S_A-3 and S_B-3 : medium overlapping degree.

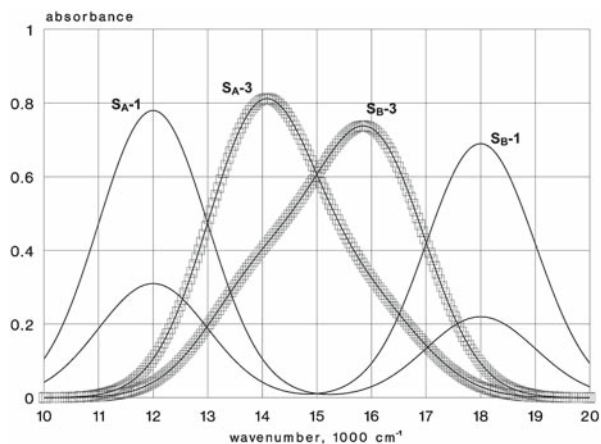


Figure 2. Two pairs of mixtures spectra.

mixed with different proportions, i.e. 78% of spectrum S_A was added to 22% of spectrum S_B and then 31% of spectrum S_A was added to 69% of spectrum S_B :

$$S_1 = 78\%S_A + 22\%S_B$$

$$S_2 = 31\%S_A + 69\%S_B.$$

Thus, disturbed shapes were obtained referring to S_1 and S_2 in equation (1). The way in which the bands were positioned in relation to each other strongly affects the shape of the spectra under investigation (figure 2).

Based on the assumption that the differences between the separated spectra are maximal, the correlation coefficient (R) can be a test providing a

correct deconvolution. Subsequently, x_{A1} and x_{A2} , which affect the lowest correlation between the extracted spectra, should be found. Theoretically R could be the test providing the correct deconvolution. However, R can equal minus one (-1) while the lowest correlation is when R equals zero. Due to this fact finding the R -minimum would not give the correct deconvolution. Moreover, even if $|R|$ is used, then the statistics itself is discontinuous in its first derivative so it is difficult to fit a function to it. Therefore, R^2 seemed to be the best test. Assuming that the extracted spectra of pure components are as different as possible, the spectra (S_A and S_B) characterised by possibly the lowest R^2 have to be found. The extracted spectra can be described as follows:

$$\begin{aligned} S'_A &= [S_1 - (1 - x_{A1})S_B]/x_{A1} \\ S'_B &= (S_2 - x_{A2}S_A)/(1 - x_{A2}). \end{aligned} \quad (5)$$

If the boundary conditions are:

$$\text{for } x_{A1} = 0 \text{ then } S'_A = S_2$$

$$\text{for } x_{A2} = 1 \text{ then } S'_B = S_1$$

the final solution is:

$$\begin{aligned} S'_A &= [S_1 - (1 - x_{A1})S_2]/x_{A1} \\ S'_B &= (S_2 - x_{A2}S_1)/(1 - x_{A2}), \end{aligned} \quad (6)$$

where: S'_A – extracted spectrum of A compound;

S'_B – extracted spectrum of B compound.

Owing to initial knowledge about non-overlapped and overlapped spectra and on the basis of equation (6) the mean fractions x_{A1} and x_{A2} can be calculated as follows:

$$x_{A1} = (S_1 - S_2)/(S'_A - S_2)$$

$$x_{A2} = (S'_B - S_2)/(S'_B - S_1).$$

Using the above relationships and for S_1 , S_2 and S'_A , S'_B selected as:

$$S_1 = 78\% S_A + 22\% S_B$$

$$S_2 = 31\% S_A + 69\% S_B$$

$$S'_A = 100\% S_A$$

$$S'_B = 100\% S_B$$

the mean theoretical values are $x_{A1} = 0,681159$ and $x_{A2} = 0,397436$.

The aim was to find x_{A1} and x_{A2} , which were fractions of A component in the spectra. On the basis of the disturbed plots one can predict the original shapes, i.e. to extract pure compound spectra from the plots of binary mixtures. To

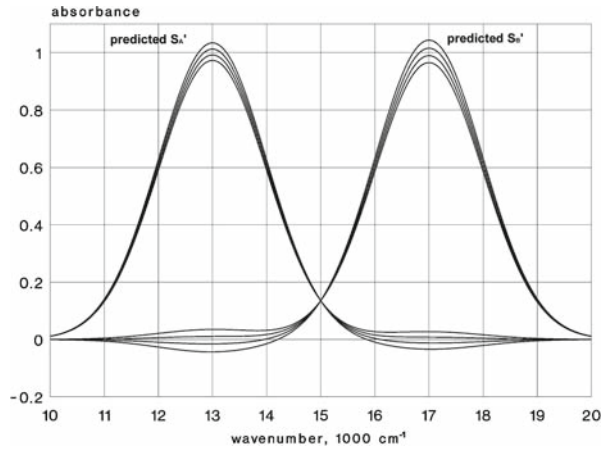


Figure 3. Few possible solutions generated by software written by the authors of the presented paper.

achieve this, on the basis of equation (6) both x_{A1} and x_{A2} were step-changed from 0 to 1 by means of the software written by the authors of the present paper. By proceeding like this (i.e. changing x_{A1} and x_{A2} from 0 to 1) all the possible solutions were tested (figure 3), whereas the pair, which gave the lowest value of the test, was chosen as the solution. In another words the pair x_{A1} and x_{A2} that caused the minimal value of R^2 was selected.

After these investigations it emerged that the test used for the calculations (R^2) did not have a minimum at a point, which would make it possible to extract the spectra properly (see table 1 column IV and V). Attempts were made to extract spectra on the basis of other tests:

- the so-called R^2 cov coefficient:

$$R^2_{\text{cov}} = \frac{R^2}{\text{cov}(\mathbf{x}', \mathbf{y}')} = \frac{\text{cov}(\mathbf{x}', \mathbf{y}')}{\text{var}(x') \cdot \text{var}(y')} \quad (7^*)$$

- covariance: $\text{cov}(\mathbf{x}', \mathbf{y}')$ (8^{*})

^{*}) see further explanation.

The purpose of these calculations was the same—to find a minimum of applied function. Taking into account the studies (results in tables 1, 3 and 4) the covariance turned out to be the best determinant.

To conduct deconvolution the use of covariance as the test gave the best results. In all the cases the pairs x_{A1} and x_{A2} were found for the covariance below zero. This was because the sum $\sum \mathbf{x}'\mathbf{y}'$ was always smaller than $\frac{1}{n} \sum \mathbf{x}'\mathbf{y}'$. Subsequently, it was checked if the minimum of $\sum \mathbf{x}'\mathbf{y}'$ can become a test. Finding the minimum of the above mentioned sum gave the same results as the covariance (see table 1 columns VIII–XI). Consequently, for spectra extraction

Table 1
Obtained results in first experiment – efficiency of applied tests and influence of the position of the bands

I**)	II	III	IV	V*)	VI	VII*)	VIII	IX*)	X	XI*)
No	Fraction	mean values	R^2	Error [%]	R2cov	Error [%]	cov(\mathbf{x}' , \mathbf{y}')	Error [%]	$\sum \mathbf{x}'\mathbf{y}'$	Error [%]
1	x_{A1}	0.6812	0.8338	22.40	0.6889	1.13	0.6802	0.15	0.6802	0.15
(minor)	x_{A2}	0.3974	0.0526	86.76	0.3913	1.53	0.3982	0.20	0.3982	0.20
2	x_{A1}	0.6812	0.8843	29.82	0.6912	1.47	0.6801	0.16	0.6801	0.16
(small)	x_{A2}	0.3974	0.0862	78.31	0.3896	1.96	0.3983	1.11	0.3983	1.11
3	x_{A1}	0.6812	0.6727	1.25	0.6727	1.25	0.6727	1.25	0.6727	1.25
(medium)	x_{A2}	0.3974	0.4041	1.69	0.4041	1.69	0.4041	1.69	0.4041	1.69
4	x_{A1}	0.6812	0.6133	9.97	0.6133	9.97	0.6133	9.97	0.6133	9.97
(strong)	x_{A2}	0.3974	0.4511	13.51	0.4511	13.51	0.4511	13.51	0.4511	13.51

*) Error calculated as: $\frac{|\text{mean value} - \text{obtained value}|}{\text{mean value}} \cdot 100\%$,

**) No indicates overlapping degree of \mathbf{S}_A and \mathbf{S}_B and as a consequence of \mathbf{S}_1 and \mathbf{S}_2 (see figures 1, 2 where examples for No 1 and 3 are shown).

obtaining a minimum of $\sum \mathbf{x}'\mathbf{y}'$ is satisfactory. This gives very easy opportunity to separate the plots of the mixtures. It is enough to find the pair x_{A1} and x_{A2} giving the sum $\sum \mathbf{x}'\mathbf{y}'$ minimal. If the vectors \mathbf{x}' and/or \mathbf{y}' have negative and positive values (derivative spectroscopy) the best test of the same power can be $\sum (\mathbf{x}'\mathbf{y}')^2$. The procedure aimed at complete separation of binary mixture spectra into particular compound bands was worked out. Even though it did not give mean values it still provided an opportunity to extract overlapped spectra in a simply and cheap way. As it can be seen in the results the positions of the spectra of pure components, and as a consequence, the position of overlapped spectra also affect the results. Figures 1 and 2 present two cases: the first concerning row number 1 in table 1 (No 1) labelled \mathbf{S}_A-1 , \mathbf{S}_B-1 in figure 1 illustrating pure component spectra and \mathbf{S}_1-1 , \mathbf{S}_2-1 illustrating overlapped spectra (figure 2). The second case concerns row number 3 (No 3) labelled in figure 1 \mathbf{S}_A-3 , \mathbf{S}_B-3 (pure) and in figure 2 \mathbf{S}_1-3 , \mathbf{S}_2-3 (overlapped). The data collected in table 1 row by row illustrate different positions of spectra \mathbf{S}_A and \mathbf{S}_B to each other. The row numbered as four indicates spectra overlapped in considerable part in contrast to the row number 1 which denotes minor overlapping. The best results were obtained, what was expected, if pure spectra did not overlap notably. If you consider column XI, i.e. the error for $\sum \mathbf{x}'\mathbf{y}'$ as the test, it increases with a decrease in distance between the maxima of \mathbf{S}_A and \mathbf{S}_B . For data in 4th row of table 1 the error is too large to consider the result a reliable one. Therefore, the attempt to separate spectra with minor distance between positions of their maxima, on the basis of their derivatives with respect to the wavelength, was undertaken. This could give slim “bands” so the problem with too large self-overlapping of $\mathbf{S}_A/\mathbf{S}_B$

Table 2
An attempt to employ derivatives – results.

I	II	III	IV	V*)	VI	VII*)
Derivative order	Fraction	Theoretical values	Obtained values $\sum x'y'$	Error [%]	Obtained values $\sum (x'y')^2$	Error [%]
I	x_{A1}	0.681159	0.3850	43.48	0.8189	20.22
	x_{B2}	0.397436	0.7439	87.17	0.2921	26.50
II	x_{A1}	0.681159	0.0006	99.91	0.8715	27.94
	x_{B2}	0.397436	0.9994	151.46	0.2532	36.29
III	x_{A1}	0.681159	0.0001	99.99	1.0000	46.81
	x_{B2}	0.397436	0.9999	151.59	0.0990	75.09
IV	x_{A1}	0.681159	0.0001	99.99	1.0000	46.81
	x_{B2}	0.397436	0.9999	151.59	0.0000	100.00

*) Error calculated as: $\frac{|\text{mean value} - \text{obtained value}|}{\text{mean value}} \cdot 100\%$.

spectra would be solved. To achieve this the most self-covered spectra (see table 1 last row) were numerically processed to their derivatives by means of Medson M48 software [25]. Four new pairs of curves were obtained i.e. the first, second, third and fourth derivatives were done. Unfortunately, the attempt completely failed as it is shown in table 2. The non-derivative error for this pair (see row 4 in table 1) was about 10% whereas for derivatives (table 2) exceeded 150%. The lowest error was obtained if the derivative order equalled 1 (20%).

Nevertheless, if self-overlapping is not too significant the method can be used successfully, even if the fractions x_{A1} and x_{A2} obtained by means of the proposed here procedure are not exactly mean values. This can be due to the properties of the tests used in the calculations. The explanation is as follows. If a test is defined as:

$$R^2 = \frac{\text{cov}(\mathbf{x}'\mathbf{y}')^2}{\text{var}(\mathbf{x}')\text{var}(\mathbf{y}')}$$

and expanding numerator and dominator as follows:

$$\begin{aligned} \mathbf{x}' &= \mathbf{x} + a\mathbf{y} \\ \mathbf{y}' &= \mathbf{y} + b\mathbf{x} \quad \text{and } a, b \geq 0 \\ \text{cov}(\mathbf{x}', \mathbf{y}') &= \sum \mathbf{x}'\mathbf{y}' - \frac{1}{n} \sum \mathbf{x}' \sum \mathbf{y}' \\ \text{cov}(\mathbf{x}', \mathbf{y}') &= \sum (\mathbf{x} + a\mathbf{y})(\mathbf{y} + b\mathbf{x}) - \frac{1}{n} \sum (\mathbf{x} + a\mathbf{y}) \sum (\mathbf{y} + b\mathbf{x}) \\ &= (1 + ab)\text{cov}(\mathbf{x}, \mathbf{y}) + b\text{var}(\mathbf{x}) + a\text{var}(\mathbf{y}) \\ \text{var}(\mathbf{x}') &= \sum (\mathbf{x}')^2 - \frac{1}{n} (\sum \mathbf{x}')^2 \\ \text{var}(\mathbf{x}') &= \sum (\mathbf{x} + a\mathbf{y})^2 - \frac{1}{n} [\sum (\mathbf{x} + a\mathbf{y})]^2 = \text{var}(\mathbf{x}) + a^2\text{var}(\mathbf{y}) + 2a\text{cov}(\mathbf{x}, \mathbf{y}) \end{aligned}$$

$$\text{var}(\mathbf{y}') = \text{var}(\mathbf{y}) + b^2\text{var}(\mathbf{x}) + 2b\text{cov}(\mathbf{x}, \mathbf{y})$$

$$R^2 = \frac{[(1+ab)\text{COV}(\mathbf{x},\mathbf{y})+b\text{var}(\mathbf{x})+a\text{var}(\mathbf{y})]^2}{[\text{var}(\mathbf{x})+a^2\text{var}(\mathbf{y})+2a\text{COV}(\mathbf{x},\mathbf{y})][\text{var}(\mathbf{y})+b^2\text{var}(\mathbf{x})+2b\text{COV}(\mathbf{x},\mathbf{y})]}$$

where:

\mathbf{x} – vector representing pure A component spectrum (non-distributed),

\mathbf{y} – vector representing pure B component spectrum (non-distributed),

\mathbf{x}' – vector representing disturbed A component spectrum (for instance measured during proceeding process),

\mathbf{y}' – vector representing disturbed B component spectrum (for instance measured during proceeding process),

a, b – coefficients,

n – number of vectors values.

It leads to the conclusion that:

$\text{var}(\mathbf{x}) > 0$ for each $\{\mathbf{x}\}$

$\text{var}(\mathbf{y}) > 0$ for each $\{\mathbf{y}\}$

$\text{cov}(\mathbf{x},\mathbf{y})$ – negative or positive.

Therefore, even if covariance (\mathbf{x},\mathbf{y}) reaches its minimum, i.e. $\text{cov}(\mathbf{x},\mathbf{y}) = \min$ and as a test would give the mean values the covariance $(\mathbf{x}',\mathbf{y}')$, i.e. $\text{cov}(\mathbf{x}', \mathbf{y}')$ can have different minima if the coefficients a and b change. When the spectrum is measured while the process is proceeding it is impossible to obtain mean values. Under such conditions only approximated values can be obtained.

To confirm the results the same Gaussian curves were mixed in different proportions, i.e. 15%/85% (\mathbf{S}_1) and 77%/23% (\mathbf{S}_2). The mean values are:

$$x_{A1} = 0,805195$$

$$x_{A2} = 0,270588$$

The data collected in table 3 confirmed the results from the previous example. To verify the results obtained for simple artificial experiments, and to check if the proposed method can be applied in more complicated cases, another experiment was conducted using synthetic spectra approximated by dual Gaussian curves. The applied procedure is in every case identical:

Table 3
Obtained results in second experiment– efficiency of applied tests and influence of the position of the bands.

I**)	II	III	IV	V*)	VI	VII*)	VIII	IX*)
No	Fraction	mean values	R^2	Error [%]	R^2_{cov}	Error [%]	$\frac{\sum x'y'}{\sum (x'y')^2}$	Error [%]
1	x_{A1}	0.805195	1	24.19	0.8134	1.02	0.8042	0.12
Minor	x_{A2}	0.270588	0	100.00	0.2638	2.51	0.2714	0.30
2	x_{A1}	0.805195	1	24.19	0.8158	1.32	0.804	0.15
Small	x_{A2}	0.270588	0	100.00	0.2619	3.22	0.2715	0.33
3	x_{A1}	0.805195	0.7962	1.12	0.7962	1.12	0.7962	1.12
Medium	x_{A2}	0.270588	0.278	2.73	0.278	2.73	0.278	2.73
4	x_{A1}	0.805195	0.7326	9.02	0.7326	9.02	0.7326	9.02
Strong	x_{A2}	0.270588	0.3307	22.21	0.3307	22.21	0.3307	22.21

*) Error calculated as: $\frac{|\text{mean value} - \text{obtained value}|}{\text{mean value}} \cdot 100\%$,

***) No indicates overlapping degree of S_A and S_B and as a consequence of S_1 and S_2 .

- synthetic spectra were generated (by means of equation 4 for $k = 2$);
- to obtain disturbed shapes, the spectra were added in different proportions: $S_1 - 32\%/68\%$ and $S_2 - 79\%/21\%$ (figure 4);
- the set of possible results was formed by means of self-written software.
- on the basis of this data the pair of values, i.e. x_{A1} and x_{A2} which caused the lowest test (tests) was chosen.

The influence of different positions of maxima of S_A and S_B bands to each other was tested too by changing $w_{0,k}$ during generating S_A and S_B (see equation 4). It influenced the shape of S_1 and S_2 (figure 4).

The application of the developed method for more complex spectra gave the following results:

The data collected in table 4 confirmed the results from previous examples and showed a possibility to apply the described method in more complicated cases.

3.2. Real data

The solvatochromic investigation can be one of such cases when physical or chemical separation of the component that is in a smaller amount is difficult as well as recording its spectrum but it is of particular importance to cha-

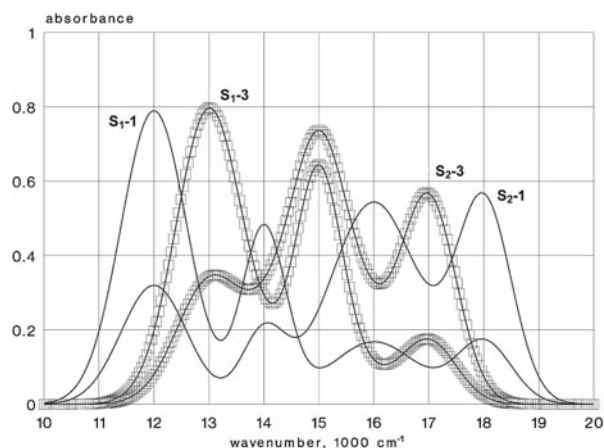


Figure 4. Two pairs of the crude plots where pure spectra were approximated by dual Gaussian curves.

Table 4

Obtained results in third experiment – efficiency of applied tests and influence of the position of the bands.

I**)	II	III	IV	V*)	VI	VII*)	VIII	IX*)
No	Fraction	mean values	R^2	Error [%]	R^2_{cov}	Error [%]	$\sum x'y'$ or $\sum (x'y')^2$	Error [%]
1 minor	x_{A1}	0.594938	0.9945	67.17	0.6303	5.95	0.5942	0.12
	x_{A2}	0.308823	0.1663	46.15	0.2778	10.04	0.31	0.39
2 small	x_{A1}	0.594938	0.5901	0.81	0.5901	0.81	0.5901	0.81
	x_{A2}	0.308823	0.31	0.39	0.31	0.39	0.31	0.39
3 medium	x_{A1}	0.594938	0.0979	83.54	0.0979	83.54	0.0979	83.54
	x_{A2}	0.308823	0.4193	35.78	0.4193	35.78	0.4193	35.78

*) Error calculated as: $\frac{|\text{mean value} - \text{obtained value}|}{\text{mean value}} \cdot 100\%$,

**) No indicates overlapping degree of S_A and S_B and as a consequence of S_1 and S_2 .

racterise spectrally this component. The equilibrium of the ionized and unionized forms in solution has its impress on the spectrum because coming from both forms bands are overlapped. Investigating merocyanines, ionised, solvatochromic form appears only *in situ* in alkaline solution and it is impossible to completely shift equilibrium constant into the direction of a desirable form. For scholars it is essential to know the exact position of the maximum of the long-wave band descending from the ionized (solvatochromic) form, so that the desirable band has to be extracted from a binary mixture spectrum.

In case of real spectra the below described procedure was employed:

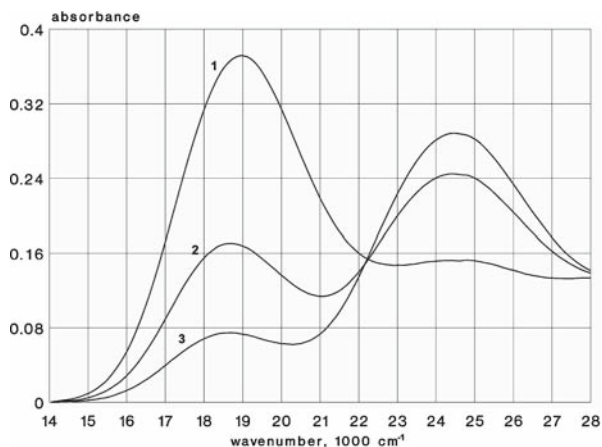


Figure 5. Spectra of solvatochromic studies, base-acid equilibrium.

Table 5
Obtained results in fourth experiment – application of new method to the real case.

I	II	III	IV	V
Pair of spectra	Fraction	R^2	R^2_{cov}	$\sum x'y' / \text{or } \sum (x'y')^2$
1 & 2	x_{A1}	0.6882	0.5822	0.3771
	x_{A2}	0.1891	0.4347	0.4347
1 & 3	x_{A1}	0.9975	0.682	0.4714
	x_{A2}	0.194	0.1939	0.1939

- spectra of solvatochromic merocyanine dye and its precursor were measured in methanol (ionized – hemicyanine and unionized – merocyanine form) [26]; 1 cm quartz cell, Specord M40.
- three spectra with different ionisation degree were measured (see figure 5) which gave the possibility to examine three pairs of overlapped plots i.e. pair 1 and 2 could become S_1 and S_2 , next pair 1 and 3 could become S_1 and S_2 and at last pair 2 and 3 could play a role of S_1 and S_2 ; nevertheless, spectra 2 and 3 seemed to be too similar so this pair was excluded from the experiment (it is essential to rely on two spectra differing considerably); next, the sets (for two experiments) of possible results were formed by means of our own software; the pairs of fractions giving the lowest tests were chosen.

Table 5 presents the results.

And the following Figure shows the extracted spectra due to minimum of $\sum x'y'$ as the test:

Determination of the exact position of solvatochromic band by means of Medson M48 software [25] gave the following results:

- if the method is applied for pairs of spectra 1&2 or 1&3 (see figure 5) the same numerical value of the position of solvatochromic band, i.e. S'_A (see figure 6) is obtained: 19067 cm^{-1} . The differences between this value (the real position of band $S'_A - 19067\text{ cm}^{-1}$) and the apparent positions obtained directly from spectra 1–3 (see figure 5) are 95, 390 and 388 cm^{-1} , respectively. The spectra were measured in methanol with different concentration of anhydrous potassium carbonate (partly dissolved) which changed the polarity of the medium (solvatochromic dye appears *in-situ* by adding K_2CO_3). Consequently, the measured spectra (figure 5) illustrate not only base-acid equilibrium but also solvatochromic, hypsochromic shift of the band. Nevertheless, this effect does not rule out the use of the described method.
- Similarly calculated location for hemicyanine non-solvatochromic band, i.e. S'_B (see figure 6) equals 24429 cm^{-1} and the differences between this value and that obtained from the apparent positions, read directly from spectra 1 to 3 (see figure 5) are +203, +6 and -33 cm^{-1} , respectively. The opposite direction of differences results from overlapping of the big absorption band located in short wave-length part of spectrum rather than from solvatochromic properties.

As the results show the overlapped plots can strongly affect the exact position of solvatochromic band. They also demonstrate that the best results can be obtained when the recorded spectra, i.e. S_1 and S_2 differ from each other considerably. Although it is hard to estimate the error, studies on synthetic spectra allow applying this method to real cases when overlapped bands do not have maxima in close distance.

4. Conclusion

The attempt to separate A→B process spectra was carried out with full success. When it is impossible to measure pure component A and/or B spectrum, but only the spectra of binary mixtures (A+B) can be recorded, it is possible to predict on the basis of these mixtures' spectra the form of spectrum of one individual element. The attempt to predict pure component bands, based on two binary mixtures spectra (with different fraction of each component), has emerged as possible. It was essential to rely on statistical methods or functions. From a set of the possible spectra the correct ones were selected on the basis of statistical tests or functions. The minimum value of $\sum (\mathbf{x}'\mathbf{y}')^2$ turned out as sufficient and the best one. The method turned out to be very easy to apply.

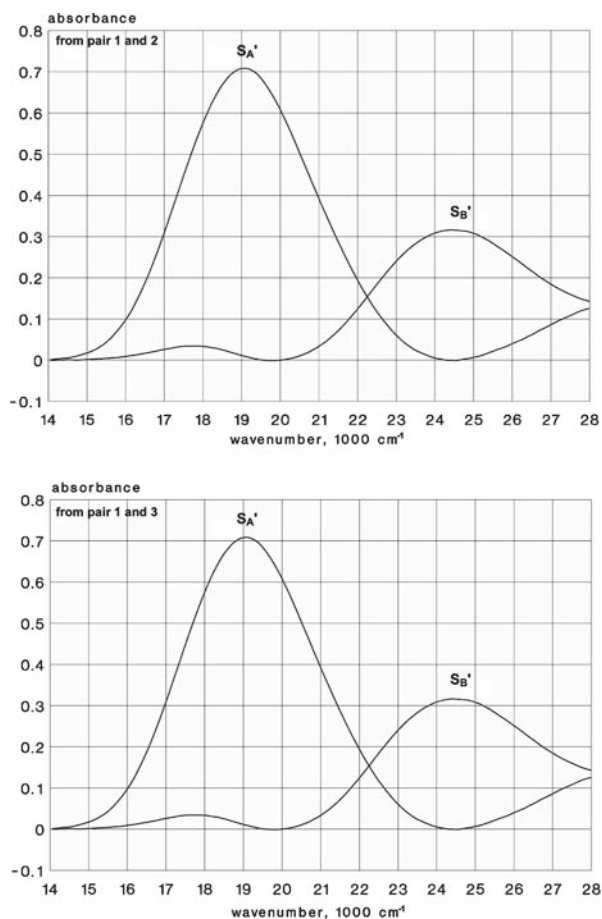


Figure 6. Bands of ionized – hemicyanine (S'_A) and unionized – merocyanine form (S'_B).

Owing to the knowledge of initial spectra (using synthetic spectra) the processed algorithm was checked and the definition of the method's error was possible to evaluate. The obtained errors for $\sum (x'y')^2$ as the test are about 1% depending on different positions of S_A and S_B bands to each other, so this allows to call this method applicable. Of course, too significant overlapping degree excludes the use of this method due to large error.

A well known method HPSAM [23, 24] is similar to the one presented here but in contrast to our method it requires much data, i.e. absorbance of a sample, absorbance of a sample with known amounts of A and B species and two wavelengths selected previously in a defined manner (for pure sample B). The here presented method was named SEMILAM. It is simple to use, it does not need as much data as HPSAM [23, 25] and there is no need to fit data to models as in other methods described in the Introduction. There are many approaches

for dealing with overlapped spectra but majorities of them are based on the earlier prepared models (PARAFAC [7]) or require large sets of data (SIMPLISMA [5]). If what we want to obtain from a spectrum is a precise position of a band coming from A and/or B element (as in the solvatochromic research), then the here developed method can be considered as satisfactory.

References

- [1] G. Scatchard, *Ann. N.Y. Acad. Sci.*, 51, (1949) 660.
- [2] D.M. Crothers, *Biopolymers*, 6 (1968) 575.
- [3] M. Maeder and A.D. Zuberbuhler, *Anal. Chim. Acta*, 181 (1986) 287.
- [4] Y.Z. Liang and O.M. Kvalheim, *J. Chemom.*, 7 (1993) 277.
- [5] W. Windig and J. Guliment, *Anal. Chem.*, 63 (1991) 1425.
- [6] R. Tauler and D. Barcelo, *Trends Anal. Chem.*, 12 (1993) 319.
- [7] B. Bro, *Chemom. Intell. Lab. Syst.*, 38 (1997) 149.
- [8] R. Tauler and B. Kowalski, *J. Chemom.*, 9 (1995) 31.
- [9] M. Gui, S.C. Ruan and A. Aghodjan, *Anal. Chem.*, 67 (1995) 3293.
- [10] B. Grung and O.M. Kvalheim, *Chemom. Intell. Lab. Syst.*, 29 (1995) 75–87.
- [11] K. De Braekeleer and D.L. Massart, *Chemom. Intell. Lab. Syst.*, 39 (1997) 127–141.
- [12] R. Bro *Chemom. Intell. Lab. Syst.*, 38 (1997) 149–171.
- [13] M. Meloun, J. Havel and E. Hogfeldt, *Computation of Solution Equilibria*, (Ellis Horwood, Chichester, 1994).
- [14] D.J. Leggett (ed.), *Computational Methods for Determination of Formation Constants*. (Plenum Press, New York, 1985).
- [15] P. Gans, A. Sabatini and A. Vacca, *Talanta*, 43 (1996) 1739.
- [16] L. Alderighi, P. Gans, A. Lenco, D. Peters, A. Sabatini and A. Vacca, *Coord. Chem. Rev.*, 184 (1999) 311.
- [17] H. Abdollahi and S. Zeinali, *Talanta*, 62 (2004) 151–163.
- [18] <http://www.chem.sc.edu/analytical/chem621/4-2001-res.pdf>.
- [19] T. Masters, *Neural, Novel And Hybrid Algorithms for Time Series Predictions*, (John Wiley, New York, 1995).
- [20] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes, The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1986).
- [21] F. Blasco-Gomez, F. Bosch Reig, P. Campins-Falco, C. Molins-Leguna and R.R. Heraez-Hernandez, *Anal. Chem.*, 72 (2000) 2559.
- [22] F. Blasco-Gomez, P. Campins-Falco, F. Bosch Reig and L. Guomin, *Analyst*, 123 (1998) 2857.
- [23] F. Bosch Reig and P. Campins-Falco, *Analyst*, 113 (1988) 1011.
- [24] F. Bosch Reig, P. Campins-Falco, J. Verdu-Andreas and C. Molins-Leguna, *Anal. Chem. Acta*, 302 (1995) 322.
- [25] Medson Co Ltd, Poznań-Paczkowo, Poland.
- [26] M.J. Sawicka, J.A. Soroka, K.B. Soroka, E.K. Wróblewska, M. Gąsiorowska and I.K. Zawadzka, *Conference Materials*, UAM Poznań, Poland, 14 th May 2004.