ELSEVIER

Chemometrics and Intelligent Laboratory Systems 52 (2000) 33–43

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

Alternating coupled matrices resolution method for three-way arrays analysis

Yang Li, Jian-Hui Jiang, Hai-Long Wu, Zeng-Ping Chen, Ru-Qin Yu *

College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, People's Republic of China

Received 15 April 1999; accepted 6 April 2000

Abstract

An alternating coupled matrices resolution (ACOMAR) method is developed for decomposition of three-way data arrays. By utilizing alternating least squares algorithm to minimize the proposed coupled matrices resolution error, the intrinsic profiles are found. Moreover, it yields simultaneously a numerically exact solution for all analytes present in the samples. This method retains the second-order advantage of quantization for analyte(s) of interest in the presence of potentially unknown interferents. The performance of a simulated experiment and a real analytical example shows that the proposed method works well when the number of components is chosen to be equal to or greater than the actual model dimensionality. The insensitivity of the ACOMAR method to the estimated component number escapes the difficulty of determining a proper component number for the model, which is hard to handle for the PARAFAC algorithm. Furthermore, this method circumvents the two-factor degeneracy, which is intrinsic in the PARAFAC algorithm.

Keywords: Alternating coupled matrices resolution (ACOMAR); Three-way data analysis; PARAFAC algorithm

1. Introduction

With the development of second-order instruments that generate a two-way data for each sample it has become demanding and important to develop new methods to handle these data [1–4]. The attractive merit that can be derived from three-way data arrays lies in that the analysis of several components of interest can be performed in the presence of unknown interferents, commonly called the ''second-order advantage'' [4,5]. Two types of algorithms have been reported in chemical literature for decomposition of three-way data arrays. The first type approach is to resolve the arrays by directly performing an eigenanalysis, with the well-known examples of generalized rank annihilation method (GRAM) [6–8] and the direct trilinear decomposition (DTLD) method [9–11]. These methods yield direct solution to the component profiles in each order. Unfortunately, GRAM is constrained to use only one standard and one mixture sample at a time. Although DTLD method allows for a direct solution through multiple

* Corresponding author.
E-mail address: rqyu@mail.hunu.edu.cn (R.-Q. Yu).

0169-7439/00/$ - see front matter
PII: S0169-7439(00)00073-3

samples, it requires construction of two pseudo-samples to formulate an eigenproblem, which unavoidably incurs a loss of information in multiple samples. An additional problem is that these approaches may sometimes yield imaginary solutions and exhibit inflated variance. In general, these methods work well when the signal-to-noise ratio is high. The second type of method aims at fitting a trilinear model to the data by utilizing an iterative procedure [12–15]. This type of method is exemplified by the PARAFAC algorithm. These methods provide a judicious way to extract useful information of the multiple samples. The method, however, does not always converge to chemically meaningful solutions unless proper constraints are placed on the solutions [10]. Additionally, the success of the algorithm is dependent on the choice of the number of components for the model. The resolved profiles will be satisfactory only if the chosen component number equals the actual one, which leads to a predicament hard to handle in practical applications.

This paper reports a newly developed method, alternating coupled matrices resolution (ACOMAR), with some improved properties. The method aims at seeking a couple of matrices called coupled matrices, which minimize a new least squares based criterion. An advantage of the method over GRAM is that it provides a statistically plausible manner to employ multi-sample information. Another salient virtue of ACOMAR is that the resolved profiles are quite stable with respect to the estimated component number when it is chosen to be equal to or greater than the actual component number, escaping the dilemma in selection of a proper component number for the model. In addition, the proposed coupled matrices resolution (COMAR) criterion provides a natural way to avoid the problem of two-factor degeneracy (2FD) [16–18]. The performance of the ACOMAR algorithm is compared with that of the traditional PARAFAC algorithm on a simulated example and a real chemical problem in three-way data analysis.

The remainder of this paper is organized as follows. First, the theory of the proposed method is described. Second, the ACOMAR method is evaluated by treating one simulated and one real chemical data set and comparing the performance with that of PARAFAC. Finally, some results and discussion are given.

2. Theory

2.1. Trilinear model for second-order resolution

Suppose that second-order data are collected in an $I \times J \times K$ three-way data array $\underline{\mathbf{R}}$. A trilinear model for such array has the following form:

$$\underline{\mathbf{R}} = \sum_{n=1}^{N} \mathbf{x}_n \otimes \mathbf{y}_n \otimes \mathbf{z}_n + \underline{\mathbf{E}} \tag{1}$$

where $\mathbf{x}_n$, $\mathbf{y}_n$, and $\mathbf{z}_n$ are the profiles in three order, respectively, of the nth detectable species among the component(s) of interest and the interferent(s). Here, the symbol $\otimes$ denotes the tensor product. $\underline{\mathbf{E}}$ is the residual array. Let $\mathbf{R}_{\cdot k}$ and $\mathbf{E}_{\cdot k}$ be the kth slices of $\underline{\mathbf{R}}$ and $\underline{\mathbf{E}}$, respectively, along the third order. In matrix notation, the trilinear model can be written as

$$\mathbf{R}_{\cdot\cdot k} = \mathbf{X}\mathrm{diag}(\mathbf{z}_{(k)})\mathbf{Y}^{\mathrm{T}} + \mathbf{E}_{\cdot\cdot k}$$

$$k = 1,2,\ldots,K \tag{2}$$

where $\mathrm{diag}(\mathbf{z}_{(k)})$ denotes the diagonal matrix in which the corresponding diagonal elements are elements of $\mathbf{z}_{(k)}$, and $\mathbf{z}_{(k)}$ is the kth row of profile matrix $\mathbf{Z}(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N)$. The superscript T denotes the transpose of a matrix. It can be seen from Eq. (1), that the trilinear model treats the profile matrices of three orders in a symmetric way. Therefore, when $\underline{\mathbf{R}}$ is sliced along the other two orders, the trilinear model can also be written as

$$\mathbf{R}_{i\cdot\cdot} = \mathbf{Y}\mathrm{diag}(\mathbf{x}_{(i)})\mathbf{Z}^{\mathrm{T}} + \mathbf{E}_{i\cdot\cdot} \qquad i = 1,2,\ldots,I \tag{3}$$

$$\mathbf{R}_{\cdot j\cdot} = \mathbf{Z}\mathrm{diag}(\mathbf{y}_{(j)})\mathbf{X}^{\mathrm{T}} + \mathbf{E}_{\cdot j\cdot} \qquad j = 1,2,\ldots,J \tag{4}$$

The trilinear resolution aims at finding the profile matrices $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ from the measuring data $\mathbf{R}$.

2.2. Least squares-based COMAR error

For decomposition of three-way data arrays, several procedures have been suggested. In the GRAM [6], instrument response matrices are collected from one standard and one mixture sample, as presented as follows:

$$\mathbf{R}_{\cdot\cdot 1} = \mathbf{X}\mathrm{diag}(\mathbf{z}_{(1)})\mathbf{Y}^{\mathrm{T}} \tag{5}$$

$$\mathbf{R}_{\cdot\cdot 2} = \mathbf{X}\mathrm{diag}(\mathbf{z}_{(2)})\mathbf{Y}^{\mathrm{T}} \tag{6}$$

Suppose that $\mathbf{U} = (\mathbf{Y}^{\mathrm{T}})^{+}$ and $\mathrm{diag}(\mathbf{z}_{(1)}) = \mathrm{diag}(\mathbf{z}_{(2)})\mathbf{C}$, then Eqs. (5) and (6) can be rewritten as:

$$\mathbf{R}_{..1}\mathbf{U} = \mathbf{X}\mathrm{diag}(\mathbf{z}_{(2)})\mathbf{C} \tag{7}$$

$$\mathbf{R}_{..2}\mathbf{U} = \mathbf{X}\mathrm{diag}(\mathbf{z}_{(2)}) \tag{8}$$

where $\mathbf{C}$ is a diagonal matrix whose diagonal elements are quotients of the corresponding elements of $\mathbf{z}_{(1)}$ to $\mathbf{z}_{(2)}$. Therefore, $\mathbf{U}$ and $\mathbf{C}$ can be resolved by the generalized eigenproblem:

$$\mathbf{R}_{..1}\mathbf{U} = \mathbf{R}_{..2}\mathbf{U}\mathbf{C} \tag{9}$$

This forms the mathematical formulation of GRAM. Generally, the QZ algorithm is implemented to solve the above generalized eigenproblem. Notice in the formulation, one needs the assumption that all the elements of residual matrix $\mathbf{E}$ equal zeros. Unfortunately, this is not the case in practice. It leads to the method to produce imaginary solutions and inflated variance. Moreover, GRAM is constrained to use only one standard and one mixture samples at one time. In this study, some improvements for GRAM have been done. It takes the residuals into account such that the error propagation factor can be reduced and the imaginary solutions can be avoided. Furthermore, it generalizes the method to multiple samples case.

In general analytical practice, one can assume that the profile matrix $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)$ has full column rank. Since $\mathbf{Y}$ has full column rank, it is known that there exists a matrix $\mathbf{P}$ satisfying

$$\mathbf{Y}^{\mathrm{T}}\mathbf{P} = \mathbf{I} \tag{10}$$

where $\mathbf{I}$ is an $N \times N$ unity matrix. Post-multiplying $\mathbf{R}_{..k}$ by $\mathbf{P}$ in Eq. (2) and using Eq. (10), one can obtain:

$$\mathbf{R}_{..k}\mathbf{P} = \mathbf{X}\mathrm{diag}(\mathbf{z}_{(k)}) + \tilde{\mathbf{E}}_{..k}$$

$$k = 1,2,\ldots,K \tag{11}$$

where $\tilde{\mathrm{E}}_{..k} = \mathbf{E}_{..k}\mathbf{P}$ $(k = 1, 2, \ldots, K)$, which can also be regarded as residual matrices. Therefore, it is natural to utilize least squares principle to estimate the parameters. That is, $\mathbf{P}$, $\mathbf{X}$ and $\mathbf{z}_{(k)}$ $(k = 1, 2, \ldots, K)$ can be estimated by the minimization of the following loss function,

$$\mathbf{L}_1 = \sum_{k=1}^{K} \|\mathbf{R}_{..k}\mathbf{P} - \mathbf{X}\mathrm{diag}(\mathbf{z}_{(k)})\|^2 \tag{12}$$

where $\|.\|$ denotes Frobeninus norm of a matrix, that is, $\|\mathbf{A}\|^2$ is the trace of the matrix $(A^{\mathrm{T}}A)$. There are three parameters **P, X,** and **Z** to be estimated in the above least squares based criterion and the estimated $\mathbf{P}$, $\mathbf{X}$, and $\mathbf{Z}$ minimizing the loss function $\mathbf{L}_1$ are the least squares solution for the profile matrices to be resolved. Considering that in the loss function Eq. (12), the transformation matrix $\mathbf{P}$ aims at making each column of $\mathbf{R}_{..k}$ transformed to be parallel to each column of $\mathbf{X}$, we name $\mathbf{P}$ the coupled matrix of $\mathbf{X}$ and the aforementioned loss function is called COMAR error. Currently, there is no direct solution to the Eq. (11). In the presented study based on the alternating least squares algorithm, an iterative procedure is proposed for obtaining the estimates of $\mathbf{P}$, $\mathbf{X}$ and $\mathbf{Z}$. It is an important part for the resolution procedure to exploit $\mathbf{P}$ and $\mathbf{X}$, thus the proposed method is called ACOMAR method.

As the trilinear model treats the $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ in a symmetric way, an analogous criterion to Eq. (11) can also be obtained as

$$\mathbf{L}_2 = \sum_{k=1}^{K} \|\mathbf{R}_{..k}^{\mathrm{T}}\mathbf{Q} - \mathbf{Y}\mathrm{diag}(\mathbf{z}_{(k)})\|^2 \tag{13}$$

where $\mathbf{Q}$ is a matrix satisfying

$$\mathbf{X}^{\mathrm{T}}\mathbf{Q} = \mathbf{I} \tag{14}$$

and $\mathbf{Q}$ is called the coupled matrix of $\mathbf{Y}$, too. By minimizing the criterion $\mathbf{L}_2$ one can obtain the estimates of $\mathbf{Q}$, $\mathbf{Y}$ and $\mathbf{Z}$.

The proposed COMAR errors, $\mathbf{L}_1$ and $\mathbf{L}_2$ are both derived from the least squares method. One can chose one of them as the criterion for three-array analysis. In this study, $\mathbf{L}_1$ is taken as an example.

2.3. ACOMAR algorithm

According to the above-mentioned loss function, an alternating least squares algorithm is used to exploit the solution. That is, it minimizes in an alternating manner the COMAR error Eq. (12) over $\mathbf{X}$ for fixed $\mathbf{P}$ and $\mathbf{Z}$, over $\mathbf{P}$ for fixed $\mathbf{X}$ and $\mathbf{Z}$ and over $\mathbf{z}_{(k)}$ $(k = 1, 2, \ldots, K)$ for fixed $\mathbf{X}$ and $\mathbf{P}$. Details of the procedure are described below.

If $\mathbf{X}$ minimizes $\mathbf{L}_1$ (Eq. (12)) for fixed $\mathbf{P}$ and $\mathrm{diag}(\mathbf{z}_{(k)})$, it is necessary for $\mathbf{X}$ to satisfy the following condition:

$$\frac{\partial \mathbf{L}_1}{\partial \mathbf{X}} = -2\sum_{k=1}^{K}\left(\mathbf{R}_{..k}\mathbf{P} - \mathbf{X}\mathrm{diag}(\mathbf{z}_{(k)})\right)\mathrm{diag}(\mathbf{z}_{(k)}) = 0 \quad (15)$$

For fixed $\mathbf{X}$ and $\mathrm{diag}(\mathbf{z}_{(k)})$ the reasonable $\mathbf{P}$ should satisfy,

$$\frac{\partial \mathbf{L}_1}{\partial \mathbf{P}} = 2\sum_{k=1}^{K}\mathbf{R}_{..k}^{\mathrm{T}}\left(\mathbf{R}_{..k}\mathbf{P} - \mathbf{X}\mathrm{diag}(\mathbf{z}_{(k)})\right) = 0 \quad (16)$$

For fixed $\mathbf{X}$ and $\mathbf{P}$, the reasonable $\mathbf{Z}$ should satisfy,

$$\frac{\partial \mathbf{L}_1}{\partial z_{kn}} = -2\mathbf{x}_n^{\mathrm{T}}(\mathbf{R}_{..k}\mathbf{p}_n - \mathbf{x}_n z_{kn}) = 0$$

$$k = 1,2,\ldots,K \quad n = 1,2,\ldots,N \quad (17)$$

where $\mathbf{z}_{kn}$ denotes the (k, n) element of the profile matrix $\mathbf{Z}$ and $\mathbf{p}_n$ is the nth column of $\mathbf{P}$.

Thus, one can obtain the update of $\mathbf{X}$ for fixed $\mathbf{P}$ and $\mathrm{diag}(\mathbf{z}_{(k)})$ from Eq. (15),

$$\mathbf{X} = \left(\sum_{k=1}^{K}\mathbf{R}_{..k}\mathbf{P}\mathrm{diag}(\mathbf{z}_{(k)})\right) \times \left(\sum_{k=1}^{K}\left(\mathrm{diag}(\mathbf{z}_{(k)})\right)^2\right)^{-1} \quad (18)$$

From Eq. (16), the update of $\mathbf{P}$ for fixed $\mathbf{X}$ and $\mathrm{diag}(\mathbf{z}_{(k)})$ is,

$$\mathbf{P} = \left(\sum_{k=1}^{K}\mathbf{R}_{..k}^{\mathrm{T}}\mathbf{R}_{..k}\right)^{+}\sum_{k=1}^{K}\mathbf{R}_{..k}^{\mathrm{T}}\mathbf{X}\mathrm{diag}(\mathbf{z}_{(k)}) \quad (19)$$

Here, ''+'' denotes the Moore–Penrose inverse.

The updates of $\mathrm{diag}(\mathbf{z}_{(k)})$ for fixed $\mathbf{X}$ and $\mathbf{P}$ are,

$$\mathbf{z}_{(k)} = \left(\left(\mathrm{diag}\left(\mathbf{X}^{\mathrm{T}}\mathbf{R}_{..k}\mathbf{P}\right)\right)./\mathrm{diag}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)\right)^{\mathrm{T}}$$

$$k = 1,2,\ldots,K \quad (20)$$

where ./ denotes the elementwise division. Here, $\mathrm{diag}(\mathbf{A})$ is a column vector formed from the diagonal elements of matrix $\mathbf{A}$.

Having given the updating equations for these parameters, the general algorithm for the ACOMAR method can be described as follows:

1. Initialize $\mathbf{X}$ and $\mathbf{P}$.
2. Compute $\mathbf{z}_{(k)}$ $(k = 1, 2, \ldots, K)$ using Eq. (20) and subsequently scale $\mathbf{Z}$ to be columnwise normalized.
3. Compute $\mathbf{X}$ using Eq. (18) and subsequently scale $\mathbf{X}$ to be columnwise normalized.
4. Compute $\mathbf{P}$ using Eq. (19).
5. Repeat steps (2) to (4) until a stopping criterion is satisfied.

In this study, the stopping criterion is the change of the COMAR error Eq. (12) between consecutive iterations below a certain value (e.g., 10^{-6}). After the algorithm converges, one can obtain the profile matrices $\mathbf{X}$, $\mathbf{P}$, and $\mathbf{Z}$ from the given three-way arrays. However, it is notable that when two corresponding columns of two of the estimated $\mathbf{X}$, $\mathbf{P}$, and $\mathbf{Z}$ are multiplied with 1 or -1 simultaneously, the trilinear decomposition of model (2) does not alter. In other words, with the identified profiles their signs are still unsettled. A procedure to eliminate such a source of uncertainty is demanding. As known that the profiles in x order in practical applications are non-negative, one can realize the sign transformation according to the sign of the element $x_{n,\max}$ with the largest absolute value in nth column of $\mathbf{X}$. The process is carried out as follows: the nth column of $\mathbf{X}$ is multiplied by $\mathrm{sign}(x_{n,\max})$ $(n = 1, 2, \ldots, N)$, where $\mathrm{sign}(x)$ represents the sign of x. If there are several elements in the same column having the largest magnitude, the sign of the first such element is adopted. With such sign transformation for the computed $\mathbf{X}$, the true profiles matrix $\mathbf{X}$ can be recovered using the post-processing described below.

It was found in experiments that, according to different selections of component number, there are two kinds of cases arising for the ACOMAR method. One case is that the estimated component number N is equal to the actual one M. In this case, the estimated $\mathbf{X}$ and $\mathbf{Z}$ are the true profiles in the first and the third order, respectively. Subsequently, $\mathbf{Y}$ is obtained from

R, **X** and **Z** by least squares regression. In the second case where the estimated component number N is greater than the actual one M, the algorithm produces N profiles in **X**. However, some of these profiles are numerically identical and there are only M profiles, which is different from each other, in the **X** obtained. These M profiles give accurate estimate for the actual profile matrix **X**. Therefore, to obtain the final solution of **X** one can select M columns, which are different from each other, from the computed **X**. At the same time, the corresponding M columns in **Z** are also selected to yield the final estimate of **Z**. With these **X** and **Z** finally resolved, the profile matrix **Y** can be obtained immediately by least squares regression.

Some components of chemical meaning may not be resolved using ACOMAR when the chosen component number is not large enough. On the other hand, some meaningless columns generated by noise may be included in the solution. To resolve the remaining components and handle the general meaningless columns, one can calculate the following criterion after the computation:

$$E(\mathbf{X},\mathbf{Y},\mathbf{Z}) = \sum_{k=1}^{K} \|\mathbf{R}_{..k} - \mathbf{X}\mathrm{diag}(\mathbf{z}_{(k)})\mathbf{Y}^{\mathrm{T}}\|^2 \qquad (21)$$

If the value of the above criterion is greater than a certain value ε (e.g., 10^{-6}), one can compute $\mathbf{R} = \mathbf{R} - \sum_{n=1}^{N} \mathbf{x}_n \otimes \mathbf{y}_n \otimes \mathbf{z}_n$, and repeat the above ACOMAR algorithm with **R** being the new response matrix, until the value of the criterion (Eq. (21)) is smaller than ε. Using such procedures, all the chemical components can be completely resolved. There is another simpler scheme to handle, the case where all components are not completely resolved. That is, one can intentionally chose a sufficiently large component number. Then ACOMAR can resolve the real profiles of all chemical components. Although the solutions obtained might contain some meaningless columns generated by noise, they are easy to be discriminated from the desired ones in that they contribute little to the trilinear model. Accordingly, such columns should not be selected in post-processing.

Using the procedure described above the ACOMAR algorithm yields the resolved $\mathbf{X}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$, $\mathbf{Y}(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)$ and $\mathbf{z}_{(k)}$ $(k = 1, 2, ..., K)$. Since column vectors of **X** and **Y** represent the estimated spectra profiles for all components, it is clear that the method gives simultaneous resolution of all the components when the chosen component number N is no less than the true one M.

3. Experimental

The data analysis and simulation were carried out in the Matlab (version 5.1) environment. Two experiments are offered as examples showing the performance of the proposed method. The first experiment is computer simulation and the second example is the chemical analysis for three dyes by using excitation–emission fluorescence spectroscopy.

3.1. Simulation

A simulated excitation–emission spectra data set was constructed from a pool of five simulated sam-

Table 1
Compositions of five simulated samples and estimated concentration profiles by ACOMAR method and PARAFAC method

Sample	True concentration				ACOMAR				PARAFAC			
	Component				Component				Component			
	1	2	3	4	1	2	3	4	1	2	3	4
1	0.1612	0.3638	0.5638	0	0.1629	0.3642	0.5639	0.0009	0.1565	0.3639	0.5649	0.0047
2	0.0346	0.4927	0.7144	0	0.0343	0.4927	0.7142	0.0007	0.0388	0.4927	0.7149	0.0013
3	0.4996	0.7882	0.0082	0.4190	0.5004	0.7880	0.0082	0.4193	0.4985	0.7881	0.0079	0.4153
4	0.5000	0.0328	0.4081	0.6892	0.4974	0.0322	0.4084	0.6885	0.5063	0.0330	0.4060	0.6937
5	0.6879	0.0507	0.0712	0.5911	0.6889	0.0505	0.0712	0.5917	0.6852	0.0505	0.0701	0.5584

Table 2
Effect of the different selection of component number N on the resolution errors in three orders of ACOMAR method

N	**X** profile				**Y** profile				**Z** profile			
	1[a]	2[a]	3[a]	4[b]	1[b]	2[c]	3[c]	4[c]	1[d]	2[a]	3[a]	4[d]
4	7.4795	2.1998	1.7415	3.9905	5.1095	4.5863	3.3488	6.6005	1.6036	1.3373	4.1111	1.4988
5	7.4795	2.1998	1.7415	3.9905	5.1095	4.5863	3.3488	6.6005	1.6036	1.3373	4.1111	1.4988
6	7.4795	2.1998	1.7415	3.9905	5.1095	4.5863	3.3488	6.6005	1.6036	1.3373	4.1111	1.4988
8	7.4795	2.1998	1.7415	3.9905	5.1095	4.5863	3.3488	6.6005	1.6036	1.3373	4.1111	1.4988
10	7.4795	2.1998	1.7415	3.9905	5.1095	4.5863	3.3488	6.6005	1.6036	1.3373	4.1111	1.4988
12	7.4027	2.1989	1.7415	3.7393	4.7341	4.3662	3.3163	5.8494	1.4791	1.1919	3.9555	1.3887
14	7.4027	2.1989	1.7415	3.7393	4.7341	4.3662	3.3163	5.8494	1.4791	1.1919	3.9555	1.3887
16	7.4027	2.1989	1.7415	3.7393	4.7341	4.3662	3.3163	5.8494	1.4791	1.1919	3.9555	1.3887
18	7.1301	2.1991	1.7415	2.7769	3.2648	3.5326	3.1704	3.9304	1.0006	0.9323	3.3613	0.9571
20	7.1301	2.1991	1.7415	2.7769	3.2648	3.5326	3.1704	3.9304	1.0006	0.9323	3.3613	0.9571

[a] Numbers in this column is multiplied with 10^{-6}.
[b] Numbers in this column is multiplied with 10^{-5}.
[c] Numbers in this column is multiplied with 10^{-7}.
[d] Numbers in this column is multiplied with 10^{-3}.

ples. Four excitation spectral profiles of component designated as $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{s}_3$, $\mathbf{s}_4$ were generated by:

$$\mathbf{s}_{1,i} = 0.2\mathrm{gs}(2i-1,30,30) + 0.5\mathrm{gs}(2i-1,70,10) \quad i = 1,2,\ldots,50$$

$$\mathbf{s}_{2,i} = 0.6\mathrm{gs}(2i-1,20,10) + 0.3\mathrm{gs}(2i-1,80,20) \quad i = 1,2,\ldots,50$$

$$\mathbf{s}_{3,i} = 0.7\mathrm{gs}(2i-1,40,10) + 0.2\mathrm{gs}(2i-1,90,20) \quad i = 1,2,\ldots,50$$

$$\mathbf{s}_{4,i} = 0.7\mathrm{gs}(2i-1,50,25) \quad i = 1,2,\ldots,50$$

where $\mathrm{gs}(x, a, b)$ refers to the value at x of Gaussian function with center a and standard deviation b, i.e. $\mathrm{gs}(x, a, b) = \exp\{-(x-a)^2/2b^2\}$.

Four emission spectral profiles of component c_1, c_2, c_3, c_4 were produced by:

$$c_{1,i} = 0.5\mathrm{gs}(4i-3,40,5) \qquad i = 1,2,\ldots,20$$

$$c_{2,1} = 0.5\mathrm{gs}(4i-3,30,10) \qquad i = 1,2,\ldots,20$$

$$c_{3,i} = 0.5\mathrm{gs}(4i-3,50,10) \qquad i = 1,2,\ldots,20$$

$$c_{4,i} = 0.5\mathrm{gs}(4i-3,40,9) \qquad i = 1,2,\ldots,20$$

The first two simulated samples contained only $\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_3$, and the remaining three samples contained all the four species. Their concentrations were uniformly distributed in the range of 0–1, as shown in Table 1. The three-way responses were constructed according to Eq. (1). Normally distributed

Table 3
Compositions of three samples and estimated concentration profiles by ACOMAR method and PARAFAC method (relative concentration)

Sample	True concentration			ACOMAR			PARAFAC		
	Component			Component			Component		
	Fluorescein	Rhodamine B	Acridine red	Fluorescein	Rhodamine B	Acridine red	Fluorescein	Rhodamine B	Acridine red
1	0.6667	0.3333	0	0.6719	0.3137	0.0052	0.6728	0.3152	0.0025
2	0.3333	0.6667	0.8944	0.3405	0.7094	0.9133	0.3387	0.7080	0.9112
3	0.6667	0.6667	0.4472	0.6578	0.6311	0.4074	0.6577	0.6320	0.4120

noise with a mean of zero and standard deviation of 0.2% were added. Both the proposed method and the traditional PARAFAC algorithm were performed on the simulated data for comparison. The effect of different selection of component number on the resolution error for the ACOMAR method was investigated and the results are shown in Table 2.

3.2. Real excitation–emission data

Three samples that contain different amounts of fluorescein, rhodamine B and acridine red were analyzed using fluorescence spectroscopy (excitation 450–600 nm, emission 480–620 nm, 5-nm intervals). The fluorescence spectra were recorded under computer-controlled regime of a Hitachi M850 spectralfluorometer with excitation slit width of 5 nm, an emission slit width of 5 nm and a scanspeed of 240 nm/min. All chemical substances were of analytical grade. The stock solutions were prepared with 0.01 mol/l NaOH. The concentrations of each component are shown in Table 3. It should be noticed that the Rayleigh scattering in the responses is not multilinear in its nature. The effect of Rayleigh scattering was eliminated by measuring a blank and subtracting it from the sample measurement before data treatment. The collected data arrays ($31 \times 29 \times 3$) were treated using the developed method as well as the PARAFAC algorithm.

All computer programs were written in Matlab and all calculations were carried out on a personal computer (Pentium II processor). The algorithm for PARAFAC used in the investigation is the version by Krijnen [19]. In order to gain a reasonable comparison, the starting values and the stopping criterion are selected for PARAFAC and ACOMAR algorithms as follows. The starting values of $\mathbf{X}$ and $\mathbf{P}$ are selected as the first N singular vectors of $\sum_{k=1}^{K}\mathbf{R}_{..k}\mathbf{R}_{..k}^{\mathrm{T}}$ and $\sum_{k=1}^{K}\mathbf{R}_{..k}^{\mathrm{T}}\mathbf{R}_{..k}$, respectively [7]. The stopping criterion is to terminate the iterations when the error between two iteration is below 10^{-6} or total computational epochs are greater than a predefined number, as 3000 in both simulated and real data analysis.

4. Results and discussion

Selection of the component number N to be greater than the actual one is a plausible way in practical application of the ACOMAR method. This is due to the following reasoning. It is found in extensive experiments that the performance and convergence rate of ACOMAR method are both stable with respect to overestimation of the component number. Furthermore, selection of a component number greater than the actual one ensures that all components underlying the data are resolved. In the presented study, the resolution results when component number is chosen to be equal to and to be greater than the actual one are both inspected.

4.1. Simulated example

Firstly, in the proposed ACOMAR method the number of component is chosen to be four, which is the true dimensionality of the underlying model. The resolved spectral profiles in excitation and emission orders are depicted against the actual ones in Fig. 1a and b, and the estimated concentration profiles are shown in Table 1. For comparison the PARAFAC algorithm was also carried out in the case of the chosen component number N to be four. Fig. 2a,b and Table 1 show the estimated profiles in three orders, respectively. It was found that both methods performed well. These results indicate that the ACOMAR method works as well as the PARAFAC algorithm in the case when the model dimensionality is correctly chosen. To illustrate the insensitivity to the chosen component number of the proposed method, the simulated data were analyzed with different number of components selected for the model. The relationship between the resolution errors (i.e., the Euclidean distances between the resolved profiles and the actual ones) and the chosen component number in three orders were investigated. An inspection of the results in Table 2 reveals that the ACOMAR method performed well when the number of component is increased from the actual dimensionality to significantly larger values. Therefore, with the ACOMAR method one needs not to determine the component number accurately. One only needs to choose a number greater than the possible model dimensionality. In utmost, it is feasible to take the smaller one of I and J as the estimate. The characteristic is appealing from the practical point of view since it avoids the dilemma of determining the component number before resolu-

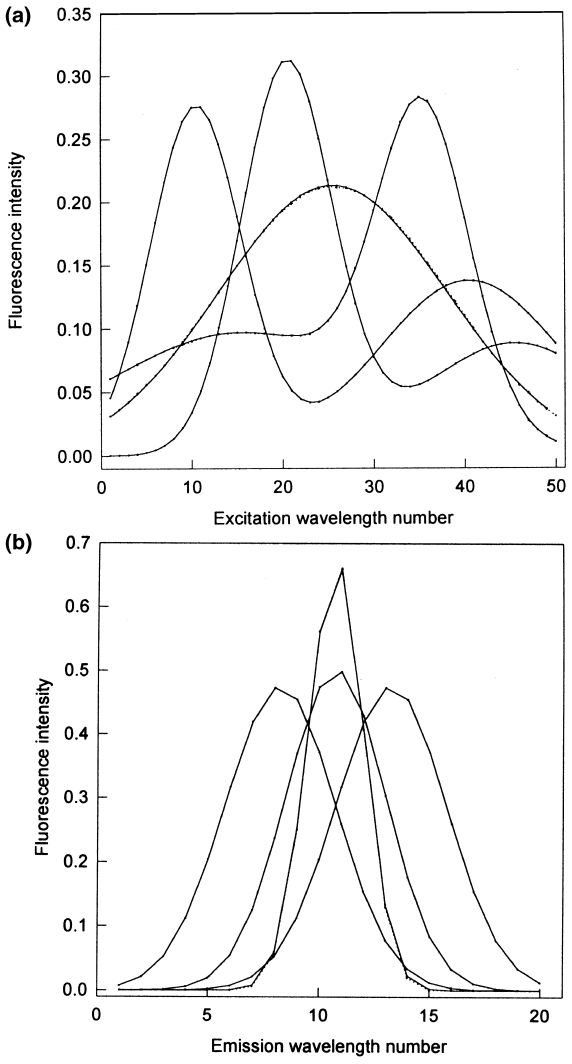

Fig. 1. (a) Resolved (dotted line) and actual (solid line) excitation spectral profiles for simulated data set using the proposed ACOMAR method. ($N = 4$). (b) Resolved (dotted line) and actual (solid line) emission spectral profiles for simulated data set using proposed ACOMAR method. ($N = 4$).

tion which is rather difficult to handle for the PARAFAC algorithm. Another satisfactory characteristic is that the convergence rate of ACOMAR method is also stable to the component number chosen. On the contrary, the PARAFAC algorithm converges much slowly in the case of overestimation of the model dimensionality.

4.2. Real excitation–emission data

Fluorescein, rhodamine B and acridine red are fluorescent dyes co-existing in liquid laser. For verifying performance of the ACOMAR and PARAFAC algorithms, the excitation and emission spectra were

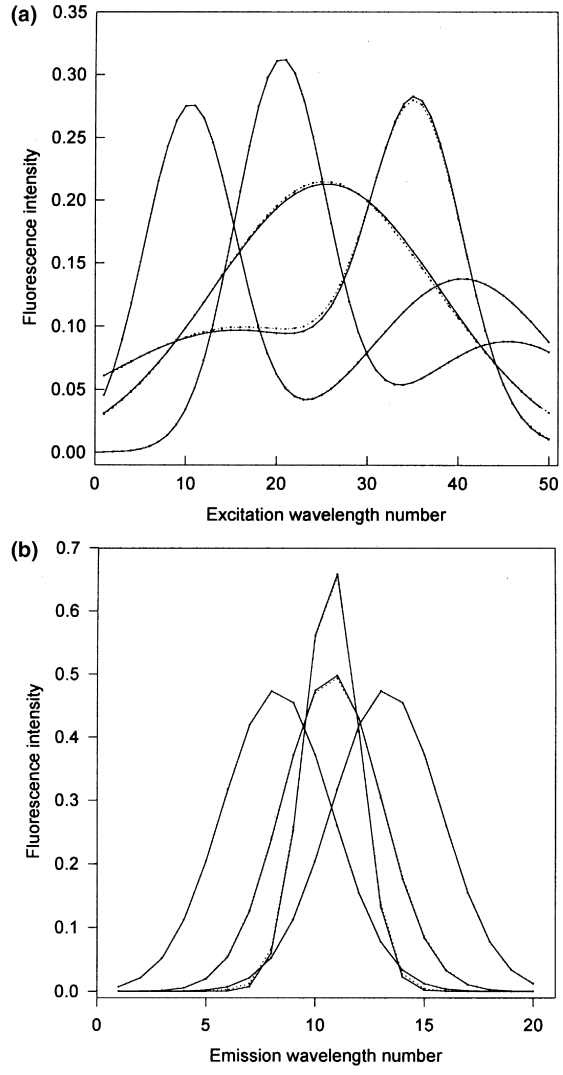

Fig. 2. (a) Resolved (dotted line) and actual (solid line) excitation spectral profiles for simulated data set using the PARAFAC algorithm. ($N = 4$). (b) Resolved (dotted line) and actual (solid line) emission spectral profiles for simulated data set using PARAFAC algorithm. ($N = 4$).

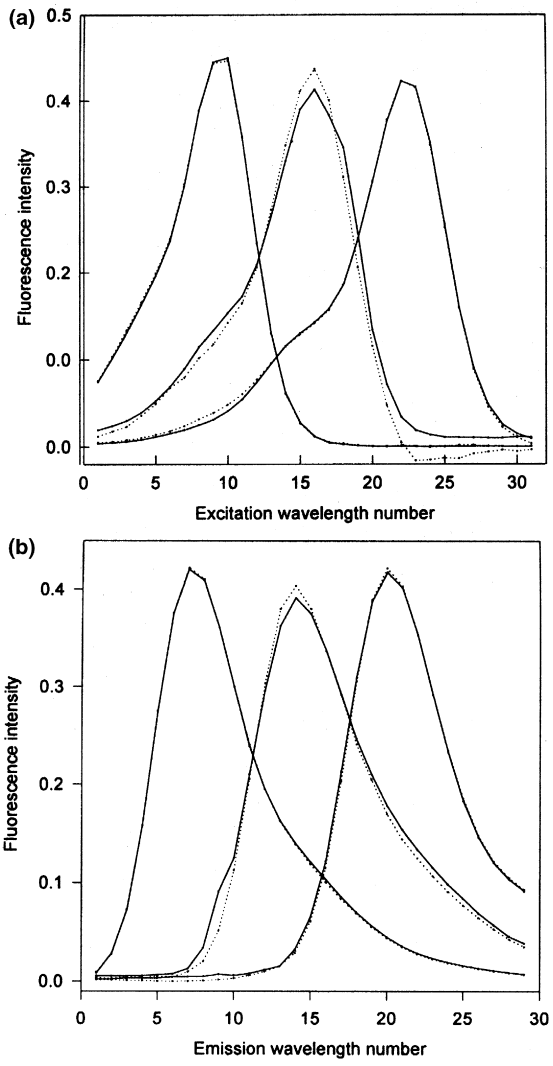

Fig. 3. (a) Resolved (dotted line) and actual (solid line) excitation spectral profiles for real data set using the ACOMAR method. ($N = 3$). (b) Resolved (dotted line) and actual (solid line) emission profiles for real data set using the proposed ACOMAR method. ($N = 3$).

measured on the above pure compounds. The effect of Rayleigh scattering should be eliminated in a multilinear decomposition if possible [17]. Therefore, a blank was measured and subtracted from the sample measurement before data treatment.

The data set of three mixtures samples was analyzed using both methods. Different component numbers were chosen. Firstly, it was set to the true component number ($N = 3$). One can observe that the estimated profiles show small discrepancy from those experimentally measured and the PARAFAC algorithm also gives good performance, as shown in Figs. 3a,b, 4a,b and Table 3. Then the component number is chosen to be four. Taking the excitation and emis-

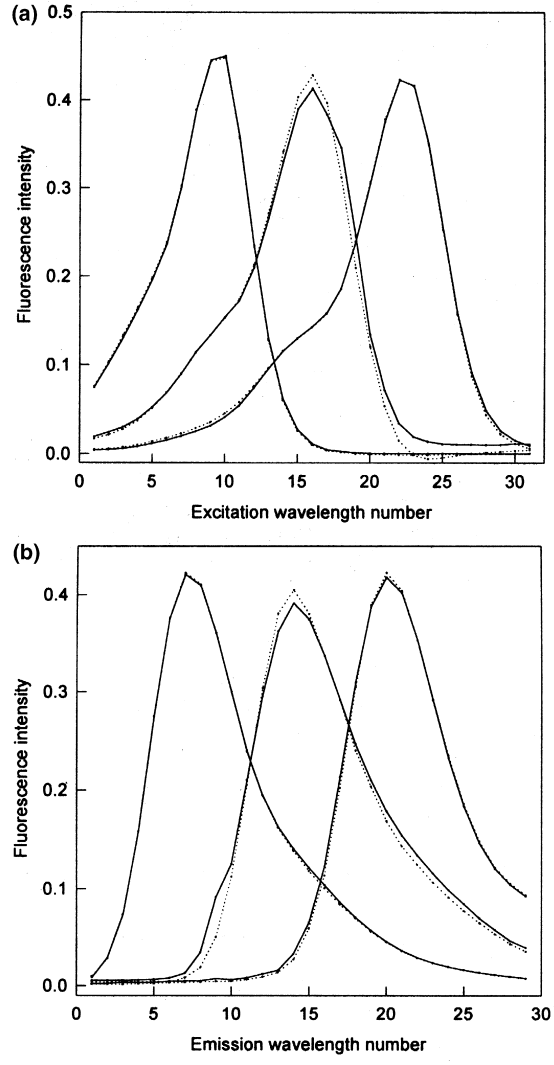

Fig. 4. (a) Resolved (dotted line) and actual (solid line) excitation spectral profiles for real data set using the PARAFAC algorithm. ($N = 3$). (b) Resolved (dotted line) and actual (solid line) emission spectral profiles for real data set using the PARAFAC algorithm. ($N = 3$).

sion spectra as examples, the profiles estimated by the ACOMAR show acceptable resolution results, as shown in Fig. 5a and b. The results using PARAFAC algorithm, however, deviated greatly from those experimentally prepared (Fig. 6a and b). The above results demonstrate again that the ACOMAR method is more stable to the overestimation of the model dimensionality.

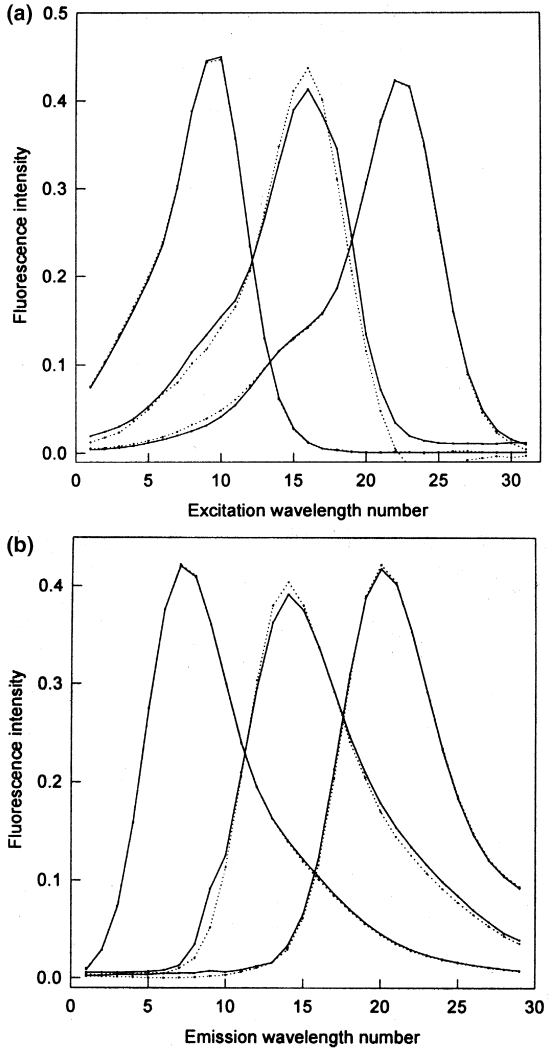

Fig. 5. (a) Resolved (dotted line) and actual (solid line) excitation spectral profiles for real data set using the proposed ACOMAR met hod. ($N = 4$). (b) Resolved (dotted line) and actual (solid line) emission spectral profiles for real data set using the proposed ACOMAR method. ($N = 4$).

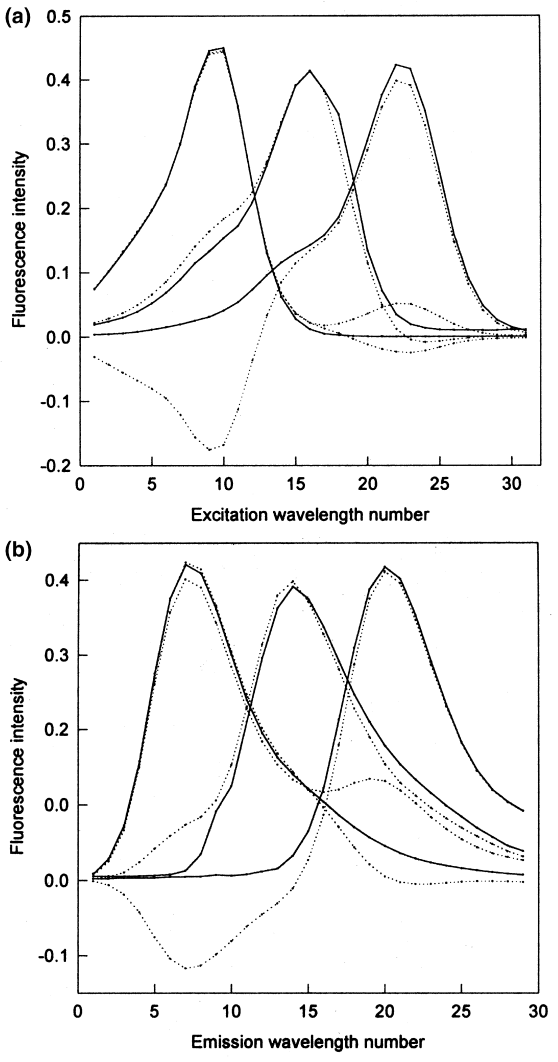

Fig. 6. (a) Resolved (dotted line) and actual (solid line) excitation spectral profiles for real data set using the PARAFAC algorithm. ($N = 4$). (b) Resolved (dotted line) and actual (solid line) emission spectral profiles for real data set using the PARAFAC algorithm. ($N = 4$).

5. Conclusions

The ACOMAR method has been proposed for decomposition of three-way data arrays. Its perfor-

mance was compared with that of the PARAFAC algorithm by simulation and by practical application to fluorescence excitation–emission data. The presented results show that the proposed method can yield simultaneously solutions with acceptable accuracy for all the analytes present in the samples. The insensitivity to the estimated component number is of practical value, which avoids the selection for the proper one.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of People's Republic of China (Grant No. 29735150). The authors would like to thank an anonymous reviewer for valuable suggestions.

References

[1] A. Lorber, Anal. Chim. Acta 164 (1984) 293.
[2] P. Geladi, Chemom. Intell. Lab. Syst. 7 (1989) 11.
[3] A.K. Smilde, Chemom. Intell. Lab. Syst. 15 (1992) 143.
[4] C.-N. Ho, G.D. Christian, E.R. Davidson, Anal. Chem. 50 (1978) 1108.
[5] E. Sanchez, B.R. Kowalski, J. Chemom. 2 (1988) 265.
[6] B. Wilson, E. Sanchez, B.R. Kowalski, J. Chemom. 8 (1989) 493.
[7] E. Sanchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496.
[8] K. Faber, A. Lorber, B.R. Kowalski, J. Chemom. 11 (1997) 95.
[9] E. Sanchez, B.R. Kowalski, J. Chemom. 4 (1990) 29.
[10] K.S. Booksh, Z. Lin, Z. Wang, B.R. Kowalski, Anal. Chem. 66 (1994) 2561.
[11] M. Gui, S.C. Rutan, A. Agbodjan, Anal. Chem. 67 (1995) 3293.
[12] P.J. Gemperlize, K.H. Miller, T.L. West, J.E. Weinstein, J.G. Hamilton, J.T. Bray, Anal. Chem. 64 (1990) 523A.
[13] A.K. Smilde, D.A. Doornbos, J. Chemom. 12 (1998) 1.
[14] H.-L. Wu, M. Shibukawa, K. Oguma, J. Chemom. 12 (1998) 1.
[15] B.C. Mitchell, D.S. Burdick, J. Chemom. 8 (1994) 155.
[16] W.S. Rayens, B.C. Mitchell, Chemom. Intell. Lab. Syst. 38 (1997) 173.
[17] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149.
[18] J.B. Kruskal, R.A. Harshman, M.E. Lundy, Multiway Data Analysis, Elsevier, Amsterdan, 1989.
[19] W.P. Krijnen, The Analysis of Three-Way Arrays by Constrained PARAFAC Methods, DSWO Press, Leiden, 1993.