An Improved Trilinear Decomposition Algorithm Based on a Lagrange Operator

Jian-Zhong Lu, Hai-Long Wu,† Jian-Hui Jiang, Ning Long, Cui-Yun Mo, and Ru-Qin Yu

State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, P. R. China

An improved trilinear decomposition algorithm based on a Lagrange operator (LO) is developed in this paper, which introduces a Lagrange operator and penalty terms in the loss function to improve the performance of the algorithm. Compared to the traditional parallel factor (PARAFAC) algorithm, the algorithm not only may converge much faster, but also overcome the sensibility to estimate the number of components. A set of simulated and measured excitation/emission fluorescence data were treated by both the proposed and traditional PARAFAC algorithm to compare their efficiencies. The analytical results obtained with real chemical system containing aspirin and its metabolic products show that the trilinear decomposition methodology is a promising tool to obtain spectral and composition information from mixtures without chemical separation.

Introduction

With the development of modern analytical instruments, such as HPLC-DAD, LC-MS and EEMs, which generate the responses of a two-dimensional matrix per sample, or even multidimensional data, it is urged to develop new and effective algorithms for chemometrics as applied to the decomposition of three-way data obtained from complex systems.[1-10] Recently, a series of three-way algorithms have been developed. Among them, the rank annihilation factor analysis method (RAFA)[11,12] was applied to the decomposition of three-way data, though it could determine only one component each time.   The generalized rank annihilation method (GRAM) and direct trilinear decomposition method (DTLD),[13-16] on the other hand, were developed to permit the determination of several components simultaneously.   The parallel factor analysis (PARAFAC)[17-21] algorithm is one of the most popular methods in the decomposition of three-way data. There were also some other versions to improve PARAFAC,[10,22,23] which were attempted to provide improved results. Booksh *et al*. have used the w-PARAFAC algorithm[24] to overcome the effect of Raman scattering to PARAFAC performance. Recently, Smilde's group has developed several models to improve the decomposition of multi-way data.[25-27]

There have been many applications of three-way data analysis. Wentzell *et al*. have applied three-way data analysis to fluorescence spectroscopy in determining PAHs,[28] and Booksh *et al*. have determined hydrocarbons in ocean water using the PARAFAC approach.[29] The PARAFAC methodology has been used by Tan *et al*. to treat the kinetic system for the degradation of chlorophyll-*a* and chlorophyll-*b*.[30]

† To whom correspondence should be addressed.
E-mail: hlwu@hnu.net.cn

The main goal of a trilinear data analysis is to solve two practical problems in analytical chemistry.[5] First, it can perform a second-order calibration, that is to say, a simultaneous determination of the components of interest, even in the presence of other unknown interferents.   Secondly, it can resolve complex chemical systems, such as kinetic models, to obtain physical resolution by using "mathematical separation". The practical experimental data are used to be complex, especially when the error is not subjected to the normal distribution from sample to sample, so that there are many different loss functions defined to decompose the three-way data.   Although the principle of least squares (LS) may generally be used to define a loss function, it is feasible only when the error is subjected to the normal distribution. This is not the case, however, for many practical situations. Thus, the loss functions defined using the least-squares concept are not the only criteria for evaluating the decomposition of the three-way data. In addition, the intrinsic trilinear structure should be considered in decomposing the three-way data.

Wu *et al*. had developed an alternating trilinear decomposition (ATLD) algorithm[2,10] to decompose three-way data. Kowalski's group has made an elaborate and pertinent evaluation:[1] It (ATLD) still converges faster than the original PARAFAC algorithm (ATLD is reported as converging in less than ten iterations), and its iterative procedure uses the Moore–Penrose generalized inverse with a singular value decomposition (SVD) in the trilinear sense. As mentioned previously, some of the main drawbacks to the used of PARAFAC are (a) the occurrence of two-factor degeneracies and (b) the presence of swamps. ATLD was developed with the aim to regularize the procedure, thus avoiding these traps.   Kowalski *et al*.[1] also pointed out: These equations involve the use of the Moore–Penrose generalized inverse, denoted by the + superscript, whereas the original PARAFAC algorithm did not. Because this computation is based on the SVD, singular values less than a certain tolerance are treated as zero. This makes it

possible to perform the calculations even when the estimated component number N is greater than the number of chemical species present in the samples, reducing the rank deficiency problem that exists with PARAFAC and that can cause two-factor degeneracies. The importance of this result stems from the fact that rank determination is not always straightforward, especially in the presence of noise, and an overestimation of the number of chemical species in the sample may sometimes occur.

In this paper, the Lagrange operator method (LO) is used to construct the loss function in the optimization procedure to improve the trilinear decomposition algorithm. Penalty and Lagrange operator terms are added purposely in order to overcome the shortcomings and deficiencies of the traditional PARAFAC algorithm. Compared to the traditional PARAFAC algorithm, the LO algorithm converges much faster and is almost not sensitive toward overfactoring the model. It has been demonstrated that this algorithm is efficient and practical both in simulated data and measured excitation-emission fluorescence data.

Theory

The trilinear model can be expressed as

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

where $\mathbf{x}_n$, $\mathbf{y}_n$, $\mathbf{z}_n$ are the profiles in three orders. It can also be expressed as

$$\mathbf{R}_{..k}, \mathbf{X}\mathrm{diag}\,(\mathbf{z}_{(k)})\mathbf{Y}^{\mathrm{T}} + \mathbf{E}_{..k}, \qquad k = 1,2,...,K, \tag{2}$$

where $\mathbf{R}_{..k}$ and $\mathbf{E}_{..k}$ are the kth slices of $\underline{\mathbf{R}}$ and $\underline{\mathbf{E}}$, respectively, along the third order, $\mathbf{X}$ and $\mathbf{Y}$ are the $I \times N$ matrix of the spectra of the relative intensity and the $J \times N$ matrix of the weighted chromatographic coefficients, respectively. $\mathbf{z}_{(k)}$ denotes the kth row of the $K \times N$ relative concentration matrix. Diag $(\cdot)$ denotes the diagonal matrix of order $N \times N$ in which the corresponding diagonal elements are elements of a vector.

Suppose that $\mathbf{P}(I \times N)$ and $\mathbf{Q}(J \times N)$ satisfy

$$\mathbf{P}^{\mathrm{T}}\mathbf{X} = \mathbf{I}_N \tag{3}$$

and

$$\mathbf{Y}^{\mathrm{T}}\mathbf{Q} = \mathbf{I}_N, \tag{4}$$

where $\mathbf{P}$ and $\mathbf{Q}$ belong to subspaces of $\mathbf{X}$ and $\mathbf{Y}$, respectively.
$\mathbf{U}_X(I \times N)$ and $\mathbf{U}_Y(J \times N)$ are the sets of orthonormal bases spanning the common column and row subspaces, respectively, of $\mathbf{R}_{..k}$ and satisfy

$$\mathbf{P} = \mathbf{U}_X\mathbf{G} \tag{5}$$

and

$$\mathbf{Q} = \mathbf{U}_Y\mathbf{B}. \tag{6}$$

Also let $\mathbf{G}(N \times N)$ and $\mathbf{H}(N \times N)$ satisfy

$$\mathbf{A}^{\mathrm{T}}\mathbf{G} = \mathbf{I}_N \tag{7}$$

and

$$\mathbf{B}^{\mathrm{T}}\mathbf{H} = \mathbf{I}_N. \tag{8}$$

Suppose $\overline{\mathbf{R}}_{..k} = \mathbf{U}_X^{\mathrm{T}}\mathbf{R}_{..k}\mathbf{U}_Y$ and $\overline{\mathbf{E}}_{..k} = \mathbf{U}_X^{\mathrm{T}}\mathbf{E}_{..k}\mathbf{U}_Y$, where $\overline{\mathbf{R}}_{..k}$ is an $N \times N$ matrix. From Eq. (2) one obtains

$$\mathbf{G}^{\mathrm{T}}\overline{\mathbf{R}}_{..k}\mathbf{H} = \mathrm{diag}(\mathbf{z}_{(k)}) + \mathbf{G}^{\mathrm{T}}\overline{\mathbf{E}}_{..k}\mathbf{H}. \tag{9}$$

Then, the loss function is obtained as

$$\mathbf{LF} = \sum_{k=1}^{K} \|\mathbf{G}^{\mathrm{T}}\overline{\mathbf{R}}_{..k}\mathbf{H} - \mathrm{diag}(\mathbf{z}_{(k)})\|_F^2, \tag{10}$$

$$\text{s.t:}\quad \|\mathbf{G}^{\mathrm{T}}\mathbf{A} - \mathbf{I}_N\|_F = 0, \tag{11}$$

$$\|\mathbf{B}^{\mathrm{T}}\mathbf{H} - \mathbf{I}_N\|_F = 0. \tag{12}$$

In order to transform a constrained problem to an unconstrained one, a Lagrange operator is applied to construct a new Lagrange augmented function as follows:

$$\mathbf{LF} = \sum_{k=1}^{K} \|\mathbf{G}^{\mathrm{T}}\overline{\mathbf{R}}_{..k}\mathbf{H} - \mathrm{diag}(\mathbf{z}_{(k)})\|_F^2 - v_1\|\mathbf{G}^{\mathrm{T}}\mathbf{A} - v_2\|\mathbf{B}^{\mathrm{T}}\mathbf{H} - \mathbf{I}_N\|_F$$
$$+ \lambda\,\|\mathbf{G}^{\mathrm{T}}\mathbf{A} - \mathbf{I}_N\|_F^2 + \lambda\|\mathbf{B}^{\mathrm{T}}\mathbf{H} - \mathbf{I}_N\|_F^2, \tag{13}$$

where both the second and third negative terms are the Lagrange operator ones and the subsequent positive terms are the penalty ones. v_1 (or v_2) and λ are constants.

First,

$$\frac{\partial\|\mathbf{G}^{\mathrm{T}}\mathbf{A} - \mathbf{I}_N\|_F}{\partial\mathbf{G}} = \frac{\partial\sqrt{\mathbf{trace}((\mathbf{G}^{\mathrm{T}}\mathbf{A} - \mathbf{I}_N)^{\mathrm{T}}(\mathbf{G}^{\mathrm{T}}\mathbf{A} - \mathbf{I}_N))}}{\partial\mathbf{G}}$$
$$= \frac{1}{2\sqrt{\mathbf{trace}((\mathbf{G}^{\mathrm{T}}\mathbf{A} - \mathbf{I}_N)^{\mathrm{T}}(\mathbf{G}^{\mathrm{T}}\mathbf{A} - \mathbf{I}_N))}} \times \frac{\partial\mathbf{trace}((\mathbf{G}^{\mathrm{T}}\mathbf{A} - \mathbf{I}_N)^{\mathrm{T}}(\mathbf{G}^{\mathrm{T}}\mathbf{A} - \mathbf{I}_N))}{\partial\mathbf{G}}$$
$$= \frac{1}{2\|\mathbf{G}^{\mathrm{T}}\mathbf{A} - \mathbf{I}_N\|_F} \times 2\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{G} - \mathbf{I}_N) = \frac{1}{\|\mathbf{G}^{\mathrm{T}}\mathbf{A} - \mathbf{I}_N\|_F}\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{G} - \mathbf{I}_N) \tag{14}$$

is deduced before the iterative formula of $\mathbf{G}$ is obtained.

Suppose $\mathrm{MT} = 2\|\mathbf{G}^{\mathrm{T}}\mathbf{A} - \mathbf{I}_N\|_F$, $\mathrm{NT} = 2\|\mathbf{B}^{\mathrm{T}}\mathbf{H} - \mathbf{I}_N\|_F$. Then, the needed equations are obtained from the LF by iteration.

$$\frac{\partial\mathbf{LF}}{\partial\mathbf{G}} = \sum_{k=1}^{K}\overline{\mathbf{R}}_{..k}\mathbf{H}[\mathbf{H}^{\mathrm{T}}\,\overline{\mathbf{R}}_{..k}\mathbf{G} - \mathrm{diag}(\mathbf{z}_{(k)})] - \frac{2v_1}{\mathrm{MT}} \times$$
$$\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{G} - \mathbf{I}_N) + 2\lambda\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{G} - \mathbf{I}_N) = 0. \tag{15}$$

The following expression for $\mathbf{G}$ is obtained:

$$\mathbf{G} = (\sum_{k=1}^{K}\overline{\mathbf{R}}_{..k}\mathbf{H}\mathbf{H}^{\mathrm{T}}\overline{\mathbf{R}}_{..k} + (\lambda - \frac{v_1}{\mathrm{MT}})\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1} \times$$
$$(\sum_{k=1}^{K}\overline{\mathbf{R}}_{..k}\mathbf{H}\mathrm{diag}(\mathbf{z}_{(k)}) + (\lambda - \frac{v_1}{\mathrm{MT}})\mathbf{A}). \tag{16}$$

Similarly, an expression for $\mathbf{H}$ can also be obtained:

$$\mathbf{H} = (\sum_{k=1}^{K}\overline{\mathbf{R}}_{..k}\mathbf{G}\mathbf{G}^{\mathrm{T}}\overline{\mathbf{R}}_{..k} + (\lambda - \frac{v_2}{\mathrm{NT}})\mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1} \times$$
$$(\sum_{k=1}^{K}\overline{\mathbf{R}}_{..k}\mathbf{G}\mathrm{diag}(\mathbf{z}_{(k)}) + (\lambda - \frac{v_2}{\mathrm{NT}})\mathbf{B}). \tag{17}$$

Table 1   Comparison of resolved concentrations obtained by LO and by PARAFAC method ($\times 10^{-6}$ g/ml)

Sample	Concentration taken			Concentration found by LO			Concentration found by PARAFAC		
	SA	2,5-DBA	PABA	SA	2,5-DBA	PABA	SA	2,5-DBA	PABA
1	0.0912	0.0000	0.0000	0.0884	−0.0001	−0.0003	0.0877	−0.0002	−0.0003
2	0.0000	0.0768	0.0000	0.0001	0.0772	−0.0004	0.0001	0.0774	−0.0003
3	0.0000	0.0000	0.0571	0.0010	0.0006	0.0596	0.0009	0.0007	0.0597
4	0.0912	0.0768	0.0000	0.0895	0.0766	−0.0002	0.0894	0.0766	−0.0002
5	0.0912	0.0000	0.0571	0.0914	0.0001	0.0581	0.0911	−0.0000	0.0581
6	0.0000	0.0768	0.0571	−0.0006	0.0746	0.0561	−0.0003	0.0748	0.0562
7	0.0912	0.0768	0.0571	0.0926	0.0781	0.0586	0.0927	0.0780	0.0586
8	0.0912	0.1056	0.0897	0.0917	0.1045	0.0889	0.0922	0.1045	0.0888
9	0.1064	0.0960	0.0734	0.1057	0.0948	0.0722	0.1059	0.0947	0.0721
10	0.1216	0.0768	0.0734	0.1209	0.0765	0.0723	0.1211	0.0763	0.0723
11	0.1064	0.1056	0.0000	0.1097	0.1084	−0.0001	0.1097	0.1083	−0.0001

Note: The concentrations of LO or PARAFAC are regressed against the concentrations taken as $C_{taken} = a \times Z + b$, where C_{taken} is the concentration vector taken, Z is the relative concentration vector decomposed by LO or PARAFAC, a and b are the parameters of regression. Then, the concentration vector found by LO or PARAFAC is obtained: $C_{found} = a \times Z + b$.

The iterative formula of $\mathbf{Z}$ is deduced as follows:

$$\frac{\partial \mathbf{LF}}{\partial z_{kn}} = -2(\mathbf{g}_n^T \overline{\mathbf{R}}_{.k} \mathbf{h}_n - z_{kn}) = 0, \tag{18}$$

$$z_{kn} = \mathbf{g}_n^T \overline{\mathbf{R}}_{.k} \mathbf{h}_n, \ k = 1,2,..., K, \ n = 1,2,..., N. \tag{19}$$

From

$$\frac{\partial \mathbf{LF}}{\partial \mathbf{A}} = 2\lambda \mathbf{G}(\mathbf{G}^T\mathbf{A} - \mathbf{I}_N) - \frac{2v_1}{MT}\mathbf{G}(\mathbf{G}^T\mathbf{A} - \mathbf{I}_N) = 0, \tag{20}$$

one obtains

$$\mathbf{A} = (\mathbf{G}^T)^{-1}. \tag{21}$$

In fact, $\mathbf{A}$ can also be deduced from Eq. (7), and similarly one can write

$$\mathbf{B} = (\mathbf{H}^T)^{-1}. \tag{22}$$

Although MT and NT include $\mathbf{G}$ and $\mathbf{H}$, MT and NT are constants. Thus, during the iteration process, one obtains the value of MT from $\mathbf{G}$ and $\mathbf{A}$ obtained in the previous iteration cycle; in a similar way, one obtains the value of NT from the corresponding $\mathbf{H}$ and $\mathbf{B}$.

The steps of the iteration include the following: (1) Initialize $\mathbf{G}$, $\mathbf{H}$, MT, NT, and fix the values of v_1 (or v_2) and λ, which will be discussed later. (2) Obtain $\mathbf{Z}$ from $\mathbf{G}$ and $\mathbf{H}$. (3) Obtain $\mathbf{A}$ and $\mathbf{B}$ from $\mathbf{G}$ and $\mathbf{H}$, respectively. (4) Obtain the normalized column wisely $\mathbf{G}$ and $\mathbf{H}$ from MT and NT, respectively. (5) Obtain MT from $\mathbf{A}$ and $\mathbf{G}$, and obtain NT from $\mathbf{B}$ and $\mathbf{H}$. (6) Repeat steps 2 – 5 until a stopping criterion is satisfied,

$$\left| \frac{\mathbf{LF}^m - \mathbf{LF}^{(m-1)}}{\mathbf{LF}^{(m-1)}} \right| < \varepsilon, \tag{23}$$

where m is the iteration number.

In fact, this method is an ALS-type algorithm. Because the $\mathbf{X}$ and $\mathbf{Y}$ matrices rather than $\mathbf{G}$ and $\mathbf{H}$ matrices are needed finally, $\mathbf{X}$ and $\mathbf{Y}$ are obtained from $\mathbf{G}$ and $\mathbf{H}$, respectively, and normalized as

$$\mathbf{x}_n = \mathbf{U}_x \mathbf{a}_n / \|\mathbf{a}_n\|, \tag{24}$$

$$\mathbf{y}_n = \mathbf{U}_y \mathbf{b}_n / \|\mathbf{b}_n\|, \tag{25}$$

where $\mathbf{a}_n$ and $\mathbf{b}_n$ are the nth columns of $\mathbf{A}$ and $\mathbf{B}$, respectively, as for $\mathbf{Z}$, the corresponding equation is

$$\mathbf{z}_k = \mathrm{diag}(\mathbf{X}^+ \mathbf{R}_{.k}(\mathbf{Y}^T)^+), \ k=1,2...,K. \tag{26}$$

$\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ should be postprocessed according to the symmetry property of the trilinear model.[10]

Experimental

Simulated HPLC-DAD data

A set of four-component HPLC-DAD data were simulated:
    X=0:0.5:40;
    a1=peak(2,X,6,3.50)+peak(1,X,20,3.50);
    a2=peak(2,X,9,3.88)+peak(1,X,22,3.88);
    a3=peak(2,X,12,3.48)+peak(1,X,30,3.48);
    a4=peak(2,X,16,3.48)+peak(1,X,35,3.48).
Here, a1, a2, a3 and a4 are the simulated data of spectra.
Let:
    Y=0:0.5:60;
    b1=peak(1,Y,20,4);
    b2=peak(1,Y,22,6);
    b3=peak(1,Y,30,4);
    b4=peak(1,Y,40,4).
Here, b1, b2, b3 and b4 are the simulated data of the chromatographic profiles. The Gaussian function used is

$$\mathrm{peak}(a, \mathbf{x}, u, cta) = a \times \exp(-(\mathbf{x} - u)^2/(2 \times cta^2)). \tag{27}$$

A concentration matrix containing 6 different samples was constructed as follows:
c=[1 3 0 0;0 1 5 0;1 0 1 8;2 1 9 1;1 5 7 1;4 2 8 2].
Thus, a simulated cube $\underline{\mathbf{R}}$ ($81 \times 121 \times 6$) was obtained (Fig. 1).

Measured excitation-emission fluorescence data

A system containing three compounds, *i.e.* salicylic acid (SA), 2,5-dihydroxybenzoic acid (2,5-DBA), and *p*-aminobenzoic acid (PABA), dissolved in distilled water, and a series of
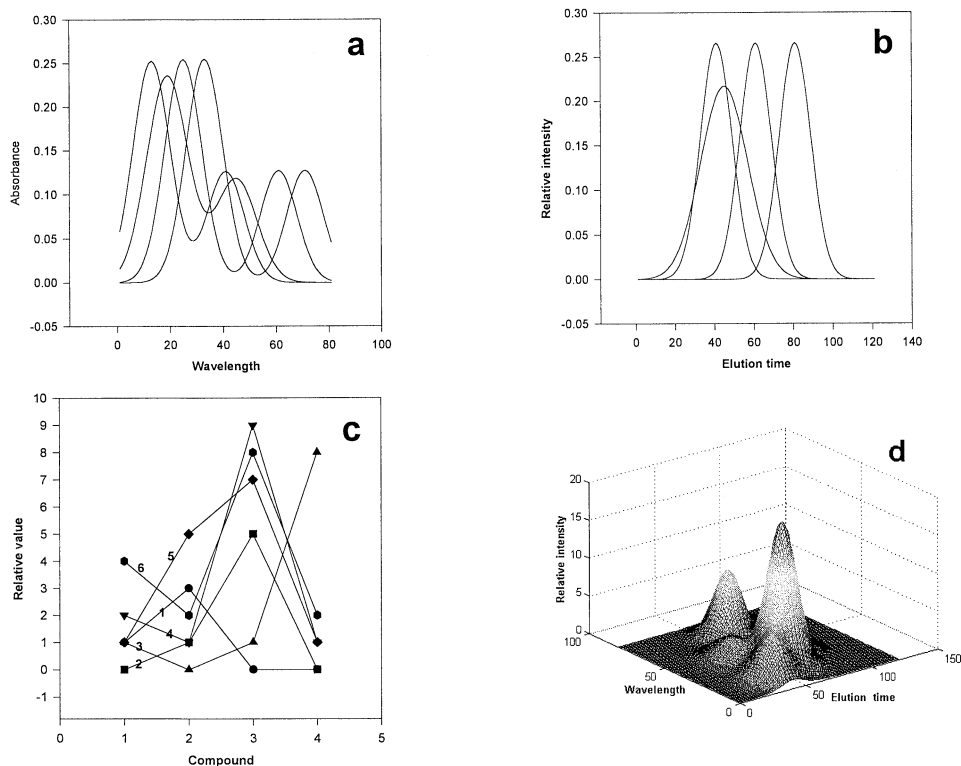
Fig. 1   Profiles of simulated HPLC-DAD data. (a) Simulated spectra; (b) simulated chromatogram profiles; (c) simulated concentration profiles and (d) HPLC-DAD profiles of the 4th mixture sample.
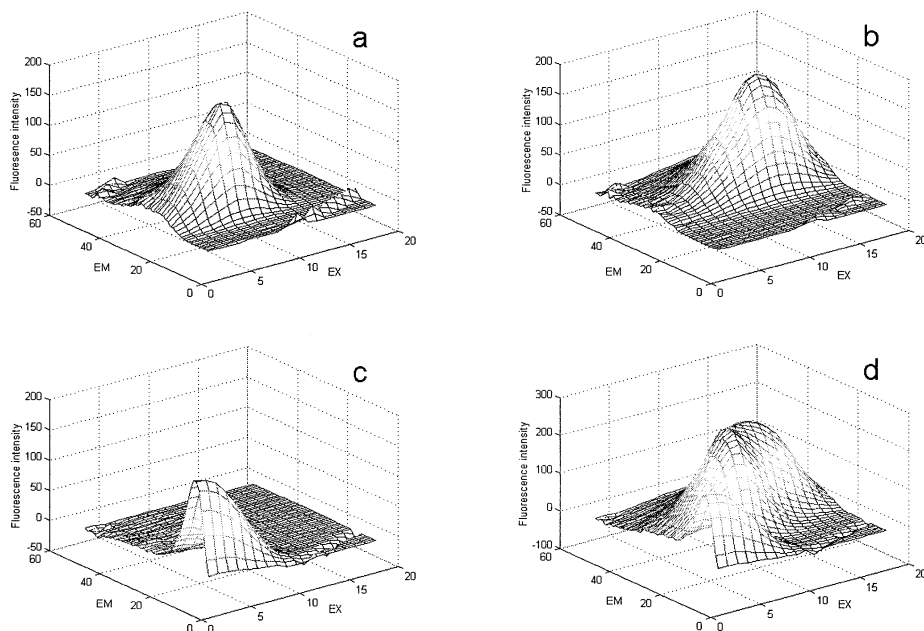

Fig. 2   Spectra of the measured excitation-emission fluorescence data. (a) Spectrum of SA; (b) spectrum of 2,5-DBA; (c) spectrum of PABA and (d) mixture spectrum of the 9th sample.

solutions of different concentrations were prepared using a pH 7 phosphate buffer solution, which are shown in Table 1. The fluorescence spectra were recorded using a HITACHI F-4500 fluorescence spectrometer.   The excitation and emission wavelengths were set from 265 to 350 nm and from 305 to 500 nm, respectively, with a fixed interval of 5 nm. The slit was 5 nm and the scan rate was 1200 nm/min. In order to eliminate the effect of Rayleigh scattering, the data of the blank solution were subtracted from the original data. All of the programs were written in MATLAB. The profiles of SA, 2,5-DBA, PABA and the 9th mixture sample are shown in Fig. 2.
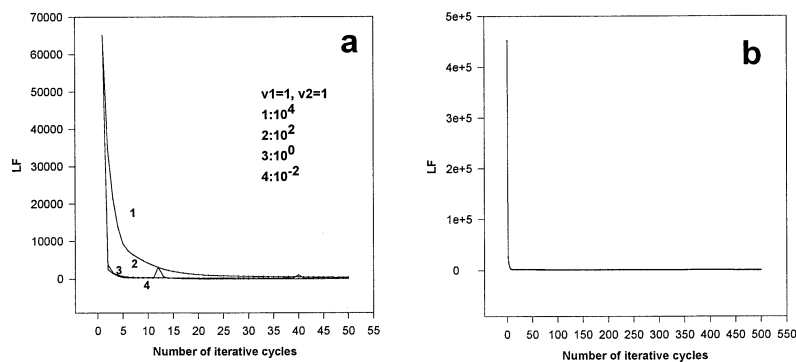
Fig. 3   Loss function (LF) *vs.* the iterative cycles. (a) Convergence curves of LO under different λ and (b) convergence curve of PARAFAC.
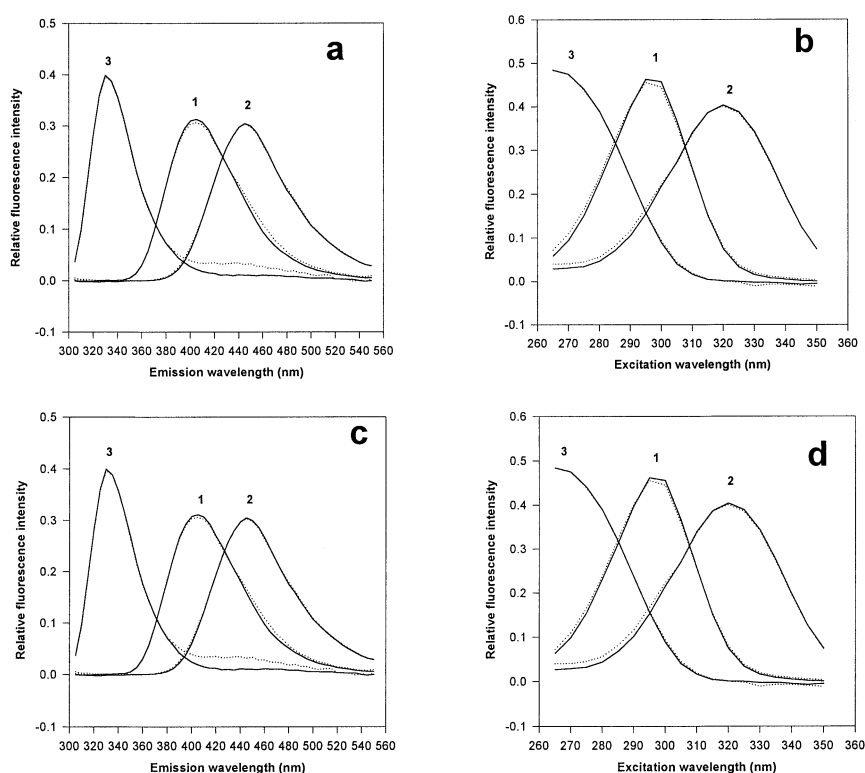

Fig. 4   Comparison of the resolved and actual spectra for the case of $N=N^*=3$. The solid lines are the resolved spectra from the measured EEMs data; the dotted lines are the actual spectra. (a) and (b) are resolved by LO algorithm; (c) and (d) are resolved by PARAFAC algorithm. 1, SA; 2, 2,5-DBA; 3, PABA.

Results and Discussion

Simulated HPLC-DAD data

In this algorithm, there were three parameters: v_1, v_2 and λ. In fact, v_1 and v_2 could have the same value because of the symmetry of the algorithm. For the convenience of the calculation, both v_1 and v_2 were taken as being fixed and equal to 1. Suppose that N^* is the actual component number of the system, and that N is a set component number; if $N \geqslant N^*$, λ could approach 0 infinitely. The penalty terms ensured the fast convergence. The Lagrange operator terms made the loss function to possess some desired characteristics, which could overcome the shortcoming that the condition number of the Hessian matrix (Eqs. (7) and (8)) would become infinity when λ is set to a very small value and the algorithm would lead to a false result. Thus, the algorithm could solve not only the convergence problem, but also the ill-conditioned problem of the loss function when the penalty terms tended to its limit. For example, when λ is set to 10^{-10}, or even a much smaller value, the Lagrange operator algorithm could also obtain the correct answer.

When λ was set as a small number ($\lambda < 10^2$), the loss function (LF) could decline after two cycles of iteration to obtain the correct result in less than 50 times. Otherwise, if λ was set as a large number, it would decline after no more than 6 cycles of iteration. On the other hand, the algorithm is not sensitive to an estimation of the component number; for example, whenever
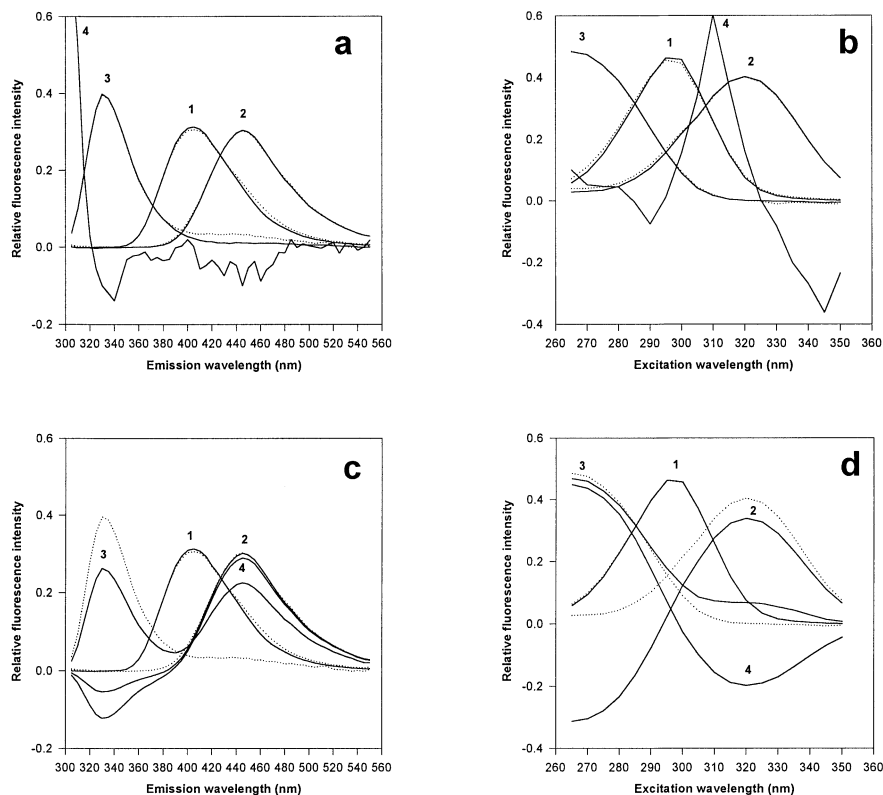
Fig. 5   Comparison of the resolved and actual spectra for the case of overfactoring ($N=4>N^*$). The solid lines are the resolved spectra from the measured EEMS data; the dotted lines are the actual spectra. (a) and (b) are resolved by the LO algorithm; (c) and (d) are resolved by the PARAFAC algorithm. 1, SA; 2, 2,5-DBA; 3, PABA; 4, background.

$N \geqslant N^*$, it could obtain the correct answer. The PARAFAC, however, is sensitive to an estimation of the component number. Only when $N = N^*$ could it obtain the correct answer and converge very slowly in 500 times, as shown in Fig. 3.

Measured excitation-emission fluorescence data

The real EEMs data of three compounds, *i.e.* SA, 2,5-DBA and PABA were obtained, which include the products of the metabolism of aspirin. By using conventional fluorescence spectroscopy, it is very difficult to obtain the fluorescence spectra of these three fluorescing compounds without chemical separation. The spectra resolved by PARAFAC and the Lagrange operator algorithm were compared with the actual spectra. When the component number was equal to the actual component number ($N(=3)=N^*$), both methods could give the correct results (Fig. 4). When the component number was set to be greater than the actual component number ($N(=4)>N^*$), the LO method could still provide the correct results, while the PARAFAC method failed to provide the correct resolution (Fig. 5). It can be seen that the Lagrange operator algorithm is more stable than the PARAFAC algorithm in resisting the influence of estimating the component number. The value of λ can be set to 1 for most practical situations.

Finally, the concentration matrix was obtained by the Lagrange operator algorithm and compared with the resolved concentration by the PARAFAC algorithm (see Table 1).

Conclusion

When a trilinear decomposition algorithm based on the Lagrange operator was used to improve the performance of the traditional PARAFAC algorithm, it was demonstrated that the proposed method converged faster and did not require an exact estimation of the real number of components. Thus, the proposed algorithm overcomes the shortcoming of the traditional PARAFAC algorithm, which is sensitive to overfactoring and converges slowly. By adding the Lagrange operator and the penalty terms, the performance of the algorithm is improved. The feasibility of the proposed algorithm has been tested using simulated and measured excitation/emission data.

References

1. C. M. Fleming and B. R. Kowalski, in "*Encyclopedia of Analytical Chemistry*", ed. R. A. Myers, **2000**, Wiley and Sons, Chichester, 9736.
2. H. L. Wu, R. Q. Yu, and K. Oguma, *Anal. Sci.*, **2001**, *17*, i483.
3. R. A. Harshman, *UCLA Working Papers Phonetics*, **1970**, *16*, 1.
4. P. Geladi, *Chemom. Intell. Lab. Syst.*, **1989**, *7*, 11.
5. K. S. Booksh and B. R. Kowalski, *Anal. Chem.*, **1994**, *66*,

782A.

6. E. Sanchez, B. R. Lindberg, and B. R. Kowalski, *J. Chemom.*, **1990**, *4*, 29.

7. A. K. Smilde, *Chemom. Intell. Lab. Syst.*, **1992**, *15*, 143.

8. H. A. L. Kiers and W. P. Krijnen, *Psychometrika*, **1991**, *56*, 147.

9. E. Malinowski, "*Factor Analysis in Chemistry*", 2nd ed., **1991**, Wiley, New York.

10. H. L. Wu, M. Shibukawa, and K. Oguma, *J. Chemom.*, **1998**, *12*, 1.

11. A. Lorber, *Anal. Chim. Acta*, **1984**, *164*, 293.

12. B. Grung and O. M. Kvalheim, *Chemom. Intell. Lab. Syst.*, **1995**, *29*, 213.

13. B. E. Wilson, E. Sanchez, and B. R. Kowalski, *J. Chemom.*, **1988**, *3*, 493.

14. E. Sanchez and B. R. Kowalski, *Anal. Chem.*, **1986**, *58*, 496.

15. D. W. Millican and L. B. McGown, *Anal. Chem.*, **1990**, *62*, 2242.

16. K. S. Booksh, Z. Lin, Z. Wang, and B. R. Kowalski, *Anal. Chem.*, **1994**, *66*, 2561.

17. C. J. Appellof and E. R. Davidson, *Anal. Chem.*, **1981**, *53*, 2053.

18. B. C. Mitchell and D. S. Burdick, *J. Chemom.*, **1994**, *8*, 155.

19. H. A. L. Kiers and A. K. Smilde, *J. Chemom.*, **1995**, *9*, 179.

20. R. Bro, *Chemom. Intell. Lab. Syst.*, **1997**, *38*, 149.

21. J. L. Beltran, J. Guiteras, and R. Ferrer, *Anal. Chem.*, **1998**, *70*, 1949.

22. W. P. Krijnen, "*The Analysis of Three-way Arrays by Constrained PARAFAC Methods*", **1993**, Leiden, DSWO Press.

23. J. H. Jiang, H. L. Wu, Y. Li, and R. Q. Yu, *J. Chemom.*, **2000**, *14*, 15.

24. R. D. JiJi and K. S. Booksh, *Anal. Chem.*, **2000**, *72*, 718.

25. J. A. Westerhuis and A. K. Smilde, *J. Chemom.*, **2001**, *15*, 485.

26. R. Bro, A. K. Smilde, and S. De Jong, *Chemom. Intell. Lab. Syst.*, **2001**, *58*, 3.

27. A. K. Smilde, J. A. Westerhuis, and R. Boque, *J. Chemom.*, **2000**, *14*, 301.

28. P. D. Wentzell, S. S. Nair, and R. D. Guy, *Anal. Chem.*, **2001**, *73*, 1408.

29. K. S. Booksh, A. R. Muroski, and M. L. Myrick, *Anal. Chem.*, **1996**, *68*, 3539.

30. Y. X. Tan, J. H. Jiang, H. L. Wu, Y. Li, H. Cui, and R. Q. Yu, *Anal. Chim. Acta*, **2000**, *412*, 195.