Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chemometrics and Intelligent Laboratory Systems 66 (2003) 101–115
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

Estimation of chemical rank of a three-way array using a two-mode subspace comparison approach

Hong-Ping Xie[a,b], Jian-Hui Jiang[a], Ning Long[a], Guo-Li Shen[a], Hai-Long Wu[a], Ru-Qin Yu[a,*]

[a] *State Key Laboratory for Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China*
[b] *Pharmic Department, Soochow University, Suzhou 215007, China*

Received 25 April 2002

Abstract

When two matrices are formed by unfolding a three-way array along two full-rank modes and when two subspaces with the same size are constructed by taking the first principal component vectors of the same mode space in these two unfolded matrices, the principal components corresponding to the chemical species should behave identically in both subspaces, while the components corresponding to the noise contribution should behave differently in these two subspaces. Based on the difference of the behavior of the chemical signal and noise components, the two-mode subspace comparison (TMSC) approach has been proposed to estimate the chemical rank of a three-way array with two full-rank modes. Two outstanding features of the proposed method have been demonstrated. It is robust to a very high degree of collinearity between the spectra or chromatograms involved, or to a very high level of noise contained in a three-way array. The method has been shown to be useful for the treatment of three-way analytical data sets obtained by high-performance liquid chromatography-diode array detector (HPLC)-DAD and excitation–emission fluorescence spectroscopy.

Keywords: Two-model subspace comparison; Chemical rank; Three-way array; HPLC-DAD; Excitation–emission spectroscopy

1. Introduction

The three-way data for a chemical system, which contains more information than second-order ones, become easily available with modern analytical instrumentation. Analysis of such data sets has been the subject of a series of analytical chemistry research [1]. The first step in three-way data analysis is to determine the number of chemical components involved in the analytical system, which is also called the chemical rank. To circumvent the difficulty in the determination of the correct chemical rank, some methods that are insensitive to the estimated component number of a three-way array have been proposed [2–7]. But these methods still require an estimated chemical rank, which is generally larger than the real one of a three-way data array studied. Therefore, searching for efficient methods of estimating the chemical rank of a three-way array is still of considerable interest.

* Corresponding author. Tel./fax: +86-731-8822782.
E-mail address: rqyu@mail.hunu.edu.cn (R.-Q. Yu).

0169-7439/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0169-7439(03)00005-4

Kruskal [8] has discussed the rank of a three-way array. A three-way array can be decomposed into a number of triads, each being a tensor product of three vectors. The minimum number of these triads, which can describe rightly the decomposition model, is called the rank of the three-way array [8] (i.e., the component number of the one). This rank may be greater than the maximal dimension of three modes of the array, which is a different characteristic feature from a second-order data set. Therefore, the method of estimating the component number of a three-way array should be different from that for a second-order data set.

The physical meaning of component number may be different for different decomposition models [9–14] of a three-way array. For PARAFAC [11] and its modification versions, the component number of a model is the chemical rank of the three-way array (i.e., the number of chemical species). Bro [15] divided the methods of estimating component number into three groups: methods based on split-half experiments, those examining the residual variation, and those utilizing the field knowledge concerning the data set being modeled.

Recently, Louwerse et al. [16] proposed two alternative generalizations of the two-way cross-validation method [i.e., the expectation maximization (EM) and the leave-bar-out (LBO) approaches] to estimate the component number of a Tucker3 model. In fact, these two methods belong to the category of methods based on examining the residual variation, and the component numbers estimated by these two methods are all the principal component numbers of the Tucker3 model rather than the chemical rank of the system as discussed above. More recently, we proposed a principal norm vector orthogonal projection (PNVOP) approach [17], which carried out the estimation of the chemical rank of a three-way array. An unfolded matrix should be formulated by unfolding a three-way array along one of its full-rank modes. When an orthogonal projection is carried out along the column space of the unfolded matrix using an orthogonal projection matrix formulated by a principal norm vector, which is the maximum Frobenius norm vector in the column space of the unfolded matrix, the mathematical rank would decrease by one for the column space of the unfolded matrix. After this kind of the projection is carried out in n circles for a three-way array containing n chemical species, the projective residual matrix would become a noise matrix. At this time, the decrease n of the mathematical rank would be equal to the chemical rank of the three-way array. According to the variation of the projective residual, one can estimate the chemical rank of a three-way array. This method is robust in resisting heteroscedastic noise. It has been successfully applied for some simulated and real data arrays. When the degree of correlation of spectra or chromatograms or the noise level in a three-way data array is too high, the PNVOP method might have some difficulties.

In this paper, an alternative approach for estimating the chemical rank of a three-way array, called two-mode subspace comparison (TMSC) method, is proposed. Two matrices are formulated by unfolding a three-way array along its two full-rank modes, say, spectral and chromatographic modes. One constructs two principal component subspaces of the same size corresponding to the same mode in these two unfolded matrices. In the case of HPLC-DAD three-way data, for example, one takes the spectral or chromatographic principal component subspace. The chemical rank is estimated by comparing these two principal component subspaces of the same mode. When the component number taken for the two subspaces is less than or equal to the chemical rank of this system, all the vectors of the two subspaces are the principal component vectors and are mainly comprised of the projections of chemical signal vectors, that is, they are the linear combinations of the same signal base vectors. Approximately, these two subspaces may be represented by each other, and their difference should be very small. Otherwise, these two subspaces should contain many noise vectors. Because of the randomness of noise, in the latter case, these two subspaces should not share the same set of the base vectors and cannot be represented by each other. At this time, the difference of these two subspaces should increase significantly. Therefore, one can estimate the chemical rank according to the residual variation of these two subspaces with the change of the estimated component number of the three-way array.

The proposed method seems to belong to the second group according to the classification of Bro [15] (i.e., methods based on examining the residual variation, although here the residual refers to that of

two subspaces, rather than the residual in the ordinary sense of model fitness measure). This method is also different from ordinary rank estimations for second-order data [18,19]. It has two outstanding advantages of resisting a very high degree of correlation of spectra or chromatograms and a very high level of noise contained in a three-way array.

2. Theory

2.1. Trilinear model

The TMSC approach is based on the trilinear model. When the three-way data set is an $(I \times J \times K)$ array $\underline{\mathbf{R}}$, a trilinear model can be expressed as [15]:

$$\underline{\mathbf{R}}_{I\times J\times K} = \sum_{n=1}^{N} \boldsymbol{x}_n \otimes \boldsymbol{y}_n \otimes \boldsymbol{z}_n + \underline{\mathbf{E}}_{I\times J\times K} \quad (1)$$

where $\boldsymbol{x}_n$, $\boldsymbol{y}_n$, and $\boldsymbol{z}_n$ are the response profiles of the nth response-active component along x, y, and z axes, respectively; N is the component number; $\otimes$ is the tensor product; and $\underline{\mathbf{E}}$ is the measurement error array.

Eq. (1) might also be expressed as three systems of matrix equations along the coordinate axes:

$$\mathbf{R}_{\cdot\cdot k} = \mathbf{X}\mathrm{diag}(\boldsymbol{z}_{(k)})\mathbf{Y}^{\mathrm{T}} + \mathbf{E}_{\cdot\cdot k} \quad (k = 1,2,\cdots,K) \quad (2)$$

$$\mathbf{R}_{\cdot j\cdot} = \mathbf{Z}\mathrm{diag}(\boldsymbol{y}_{(j)})\mathbf{X}^{\mathrm{T}} + \mathbf{E}_{\cdot j\cdot} \quad (j = 1,2,\cdots,J) \quad (3)$$

$$\mathbf{R}_{i\cdot\cdot} = \mathbf{Y}\mathrm{diag}(\boldsymbol{x}_{(i)})\mathbf{Z}^{\mathrm{T}} + \mathbf{E}_{i\cdot\cdot} \quad (i = 1,2,\cdots,I) \quad (4)$$

where the superscript T denotes the matrix transposition; $\mathbf{R}_{\cdot\cdot k}$ is the kth matrix slice of $\underline{\mathbf{R}}_{I\times J\times K}$ along the z-axis; $\mathbf{R}_{\cdot j\cdot}$ is the jth matrix slice along the y-axis; $\mathbf{R}_{i\cdot\cdot}$ is the ith matrix slice along the x-axis; $\mathrm{diag}(\boldsymbol{z}_{(k)})$ is the diagonal matrix whose diagonal elements are the corresponding ones of the kth row vector $\boldsymbol{z}_{(k)}$ of the response profile matrix $\mathbf{Z}_{K\times N}$; and $\mathbf{Z}_{K\times N}$ consists of the response profile vectors $\boldsymbol{z}_n$ $(n = 1,2,\ldots,N)$ of the N response-active components as expressed by the following equation:

$$\mathbf{Z}_{K\times N} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_N) \quad (5)$$

Similarly, $\mathrm{diag}(\boldsymbol{y}_{(j)})$ is formed by the jth row vector $\boldsymbol{y}_{(j)}$ of $\mathbf{Y}_{J\times N}$ expressed as:

$$\mathbf{Y}_{J\times N} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N) \quad (6)$$

and $\mathrm{diag}(\boldsymbol{x}_{(i)})$ is comprised of the ith row vector $\boldsymbol{x}_{(i)}$ of $\mathbf{X}_{I\times N}$, which denotes a response matrix consisting of the response profiles $\boldsymbol{x}_n$ $(n = 1,2,\ldots,N)$:

$$\mathbf{X}_{I\times N} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) \quad (7)$$

Trilinear resolutions are carried out for all components at the same time or one by one for each component according to Eqs. (2)–(4). The goal of the trilinear resolution is to resolve the response profiles $\mathbf{X}_{I\times N}$, $\mathbf{Y}_{J\times N}$, and $\mathbf{Z}_{K\times N}$ for obtaining the chemical information concerning the measured processes, and the first step of the trilinear resolution is to estimate the chemical rank N of the three-way array $\underline{\mathbf{R}}_{I\times J\times K}$.

2.2. Two-mode subspace comparison approach

By unfolding a three-way array $\underline{\mathbf{R}}(I \times J \times K)$ according to Eqs. (2)–(4), one can formulate six different unfolded matrices [17,20]. They are pairwise equivalent, and the three unfolded matrices with different properties could be expressed as:

$$\mathbf{RA}_{I\times JK} = [\mathbf{R}_{\cdot\cdot 1}, \mathbf{R}_{\cdot\cdot 2}, \cdots, \mathbf{R}_{\cdot\cdot K}] \quad (8)$$

$$\mathbf{RB}_{J\times IK} = [\mathbf{R}^{\mathrm{T}}_{\cdot\cdot 1}, \mathbf{R}^{\mathrm{T}}_{\cdot\cdot 2}, \cdots, \mathbf{R}^{\mathrm{T}}_{\cdot\cdot K}] \quad (9)$$

$$\mathbf{RC}_{K\times IJ} = [\mathbf{R}_{\cdot 1\cdot}, \mathbf{R}_{\cdot 2\cdot}, \cdots, \mathbf{R}_{\cdot J\cdot}] \quad (10)$$

When unfolding a three-way array along a full-rank mode, say I-mode as Eq. (8), the column space of the unfolded matrix **RA** involves all JK vectors contained in all K matrix slices of the three-way array $\underline{\mathbf{R}}$, and the chemical rank of column space of **RA** is equal to that of $\underline{\mathbf{R}}$ [17]. Because the I-mode is a full rank (i.e., the row space of **RA** is a full rank), the row space of the unfolded matrix **RA** should contain all information concerning the chemical rank of the three-way array $\underline{\mathbf{R}}$. Therefore, one can estimate the correct chemical rank of a three-way array through the chemical rank estimation of the row and column

spaces of its unfolded matrices along its full-rank mode.

Suppose that mode K is the sample or concentration mode, and modes I and J are the spectral and chromatographic ones of full rank, respectively. To explain the TMSC approach, take a pair of Eqs. (8) and (9) formed by unfolding $\underline{\mathbf{R}}$ along modes I and J of full rank, respectively, as an example. The column space of $\mathbf{RA}_{I\times JK}$ is constructed by all JK spectral vectors of dimension $(I\times 1)$ of K samples. These spectral vectors of dimension $(I\times 1)$ of each sample are juxtaposed side by side with each other. The row space of **RA** is formed by all chromatographic vectors of K samples, and each row vector of dimension $(1\times JK)$ of **RA** is formed by joining with the head to end all K chromatographic vectors of dimension $(1\times J)$ at the same spectral wavelength of K samples. Similar arguments can be used to describe $\mathbf{RB}_{J\times IK}$. Its column space is constructed by juxtaposing the chromatographic profiles of K samples and its row vectors are formed by joining spectral profiles of K samples.

The singular value decomposition of the matrices **RA** and **RB** gives:

$$\mathbf{RA}_{I\times JK} = \mathbf{U}_{\boldsymbol{A}}\boldsymbol{\Lambda}_{\boldsymbol{A}}\mathbf{V}_{\boldsymbol{A}}^{\mathrm{T}} \tag{11}$$

$$\mathbf{RB}_{J\times IK} = \mathbf{U}_{\boldsymbol{B}}\boldsymbol{\Lambda}_{\boldsymbol{B}}\mathbf{V}_{\boldsymbol{B}}^{\mathrm{T}} \tag{12}$$

here $\mathbf{U}_{\boldsymbol{A}}$ is an $I\times I$ orthogonal spectral matrix with each column vector denoting a spectrum for one principal component; $\mathbf{V}_{\boldsymbol{A}}$ denotes a $JK\times JK$ orthogonal chromatographic matrix, each column vector being formed by joining with the head to end K chromatogram vectors of the same principal component in K samples; $\mathbf{U}_{\boldsymbol{B}}$ is a $J\times J$ chromatographic matrix with each column vector being the chromatographic profile of one principal component; $\mathbf{V}_{\boldsymbol{B}}$ is an $IK\times IK$ spectral matrix with each column vector being formed by joining with the head to end K spectral vectors of the same principal component in K samples; $\boldsymbol{\Lambda}_{\boldsymbol{A}}$ and $\boldsymbol{\Lambda}_{\boldsymbol{B}}$ are the diagonal singular value matrices of **RA** and **RB**, respectively.

Take the spectral space as an example (i.e., the two column spaces of $\mathbf{U}_{\boldsymbol{A}}$ and $\mathbf{V}_{\boldsymbol{B}}$) to explain the TMSC method. Let the spectra of the first i principal components in $\mathbf{U}_{\boldsymbol{A}}$ and $\mathbf{V}_{\boldsymbol{B}}$ (i.e., the first i column vectors of $\mathbf{U}_{\boldsymbol{A}}$ and $\mathbf{V}_{\boldsymbol{B}}$) make two matrices $\mathbf{U}_1$ and $\mathbf{V}_1$, respectively. The column spaces of matrices $\mathbf{U}_1$ and $\mathbf{V}_1$ correspond to the two subspaces of column spaces of $\mathbf{U}_{\boldsymbol{A}}$ and $\mathbf{V}_{\boldsymbol{B}}$, respectively:

$$\mathbf{U}_1 = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_i]_{I\times i} \qquad (i = 1, 2, \ldots, I) \tag{13}$$

$$\mathbf{V}_1 = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_i]_{IK\times i} \qquad (i = 1, 2, \ldots, I) \tag{14}$$

here $\boldsymbol{u}_m$ and $\boldsymbol{v}_m$ $(m=1,2,\ldots,i)$ are the mth column vectors of $\mathbf{U}_{\boldsymbol{A}}$ and $\mathbf{V}_{\boldsymbol{B}}$, respectively. Rearrange matrix $\mathbf{V}_1$ to formulate an $(I\times Ki)$ spectral matrix $\mathbf{V}_2$, that is,

$$\mathbf{V}_2 = [\boldsymbol{v}_{1,1}, \boldsymbol{v}_{1,2}, \ldots, \boldsymbol{v}_{1,K}, \boldsymbol{v}_{2,1}, \boldsymbol{v}_{2,2}, \ldots, \boldsymbol{v}_{2,K}, \ldots, \boldsymbol{v}_{i,1}, \boldsymbol{v}_{i,2}, \ldots, \boldsymbol{v}_{i,K}]_{I\times Ki} \tag{15}$$

Here vectors $\boldsymbol{v}_{m,1},\boldsymbol{v}_{m,2},\ldots,\boldsymbol{v}_{m,K}$ are formed by rearranging vector $\boldsymbol{v}_m$ $(m=1,2,\ldots,i)$ as the following equation:

$$[\boldsymbol{v}_{m,1}, \boldsymbol{v}_{m,2}, \ldots, \boldsymbol{v}_{m,k}]_{I\times K} = \begin{bmatrix} v_m(1) & v_m(I+1) & \ldots & v_m((K-1)I+1) \\ v_m(2) & v_m(I+2) & \ldots & v_m((K-1)I+2) \\ \vdots & \vdots & \vdots & \vdots \\ v_m(I) & v_m(21) & \ldots & v_m(KI) \end{bmatrix}_{I\times K} \tag{16}$$

where $v_m(1),v_m(2),\ldots,v_m(KI)$ are all the elements of the column vector $\boldsymbol{v}_m$ of the matrix $\mathbf{V}_1$ in Eq. (14). In Eq. (16), the first to Kth column vectors express, respectively, the spectra of the mth principal components of the first to Kth sample matrices in **RB**. Eq. (15) is the spectral matrix of the first i principal components in the K sample matrices. Therefore, Eqs. (13) and (15) all are the spectral matrices of the first i principal components in these two unfolded matrices **RA** and **RB**, respectively, and their column spaces are two same spectral spaces of the first i principal components.

When the number i of principal components is less than or equal to the chemical rank N, each principal component is constructed by a projection of N pure spectral vectors $\boldsymbol{s}_1,\boldsymbol{s}_2,\ldots,\boldsymbol{s}_N$, that is, each principal component is a linear combination of $\boldsymbol{s}_1,\boldsymbol{s}_2,\ldots,\boldsymbol{s}_N$.

Since there exists noise in the system, the noise vector also makes a contribution to the principal component vectors. Now, any principal component vectors $\boldsymbol{u}_m$ and $\boldsymbol{v}_{m,k}$ of Eqs. (13) and (15), respectively, can be expressed as:

$$\boldsymbol{u}_m = c_{m,1}\boldsymbol{s}_1 + c_{m,2}\boldsymbol{s}_2 + \ldots + c_{m,N}\boldsymbol{s}_N + c_{m,0}\boldsymbol{s}_{u,0} \quad (m = 1,2,\ldots,i) \tag{17}$$

$$\boldsymbol{v}_{m,k} = c_{mk,1}\boldsymbol{s}_1 + c_{mk,2}\boldsymbol{s}_2 + \ldots + c_{mk,N}\boldsymbol{s}_N + c_{mk,0}\boldsymbol{s}_{v,0} \quad (m = 1,2,\ldots,i;\quad k = 1,2,\ldots,K) \tag{18}$$

where $c_{m,n}$ and $c_{mk,n}$ ($n = 1,2,\ldots,N$) are the combination coefficients of the pure vector $\boldsymbol{s}_n$ ($n = 1,2,\ldots,N$) for the principal component spectra $\boldsymbol{u}_m$ and $\boldsymbol{v}_{m,k}$, respectively; and $\boldsymbol{s}_{u,0}$ and $\boldsymbol{s}_{v,0}$ are noise vectors contributing to $\mathbf{U}_A$ and $\mathbf{V}_B$, respectively. One notices that here the difference of the noise contributions in the formation of $\mathbf{U}_A$ and $\mathbf{V}_B$ is specified. The combination coefficients express the contribution of the pure spectral vectors and the noise to the principal component vectors of the spectra. The following relationships hold:

$$\| c_{m,0}\boldsymbol{s}_{u,0} \|_2 \ll \| c_{m,n}\boldsymbol{s}_n \|_2 \qquad (n = 1,2,\ldots,N) \tag{19}$$

$$\| c_{mk,0}\boldsymbol{s}_{v,0} \|_2 \ll \| c_{mk,n}\boldsymbol{s}_n \|_2 \qquad (n = 1,2,\ldots,N) \tag{20}$$

where $\| \bullet \|_2$ is the Frobenius norm of a vector. When only taking into consideration the signal base vectors for vectors $\boldsymbol{u}_m$ and $\boldsymbol{v}_{m,k}$ in Eqs. (17) and (18), respectively, these two principal component vectors of the spectra have identical base vectors and all $\boldsymbol{v}_{m,k}$ can be represented by vectors $\boldsymbol{u}_m$ ($m = 1,2,\ldots,i$), that is,

$$\boldsymbol{v}_{m,k} = c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_i\boldsymbol{u}_i \quad (m = 1,2,\ldots,i;\quad k = 1,2,\ldots,K) \tag{21}$$

Here again, $c_1,c_2,\ldots,c_i$ are combination coefficients. In reality, there must be a noise contribution to the principal component vectors $\boldsymbol{v}_{m,k}$ and $\boldsymbol{u}_m$ of the signals. The randomness of the noise distribution makes the accumulation of the noise in these principal components be different, that is, $\boldsymbol{s}_{u,0}$ and $\boldsymbol{s}_{v,0}$ in Eqs. (17) and (18), respectively, are not identical. In Eq. (21), the noise involved in all $\boldsymbol{u}_m$ ($m = 1,2,\ldots,i$) cannot represent completely the noise contribution involved in $\boldsymbol{v}_{m,k}$. So Eq. (21) should also contain a noise term $\boldsymbol{s}_0$, which is not identical to $\boldsymbol{s}_{v,0}$ in Eq. (18), that is,

$$\boldsymbol{v}_{m,k} = c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_i\boldsymbol{u}_i + c_0\boldsymbol{s}_0 \quad (m = 1,2,\ldots,i;\quad k = 1,2,\ldots,K) \tag{22}$$

For different terms in Eq. (22), the following relationship holds:

$$\| c_j\boldsymbol{u}_j \|_2 \gg \| c_0\boldsymbol{s}_0 \|_2 \qquad (j = 1,2,\ldots,i) \tag{23}$$

As all the vectors in Eq. (22) are normalized, Eq. (23) becomes:

$$| c_j | \gg | c_0 | \qquad (j = 1,2,\ldots,i) \tag{24}$$

One constructs the orthogonal projective matrix $\mathbf{P}$ using all the vectors $\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_i$ of the matrix $\mathbf{U}_1$ expressed in Eq. (13):

$$\mathbf{P} = \mathbf{I} - \mathbf{U}_1\mathbf{U}_1^{\mathrm{T}} \tag{25}$$

where $\mathbf{I}$ is an identity matrix. When the vector $\boldsymbol{v}_{m,k}$ projects along $\mathbf{U}_1$ with $\mathbf{P}$, the residual vector $\boldsymbol{rs}_{m,k}$ for $\boldsymbol{v}_{m,k}$ will be expressed as:

$$\begin{aligned}\boldsymbol{rs}_{m,k} &= \boldsymbol{P}\boldsymbol{v}_{m,k} \\ &= [\mathbf{I} - (\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_i)(\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_i)^{\mathrm{T}}] \\ &\quad \times (c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_i\boldsymbol{u}_i + c_0\boldsymbol{s}_0)\end{aligned} \tag{26}$$

Because $\{\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_i\}$ is an orthogonal set of vectors, and the normalized noise vector $\boldsymbol{s}_0$ is not orthogonal to this set, Eq. (26) can be written as:

$$\begin{aligned}\boldsymbol{rs}_{m,k} &= (-c_{0,1})\boldsymbol{u}_1 + (-c_{0,2})\boldsymbol{u}_2 + \ldots \\ &\quad + (-c_{0,i})\boldsymbol{u}_i + c_0\boldsymbol{s}_0\end{aligned} \tag{27}$$

where $c_{0,j} = c_0\boldsymbol{u}_j^{\mathrm{T}}\boldsymbol{s}_0$ ($j = 1,2,\ldots,i$). In Eq. (27), one has:

$$\| \boldsymbol{rs}_{m,k} \|_2 \leq (| c_{01} | + | c_{02} | + \ldots + | c_{0i} | + | c_0 |) \tag{28}$$

As the correlation between the signal and noise is very small, and considering Eq. (24), the value of the right side of the inequality (Eq. (28)) should be very small. In other words, the residual of the vector $\boldsymbol{v}_{m,k}$ must be very small.

When the rearranged matrix $\mathbf{V}_2$ (Eq. (15)) of the matrix $\mathbf{V}_1$ projects along $\mathbf{U}_1$ (Eq. (13)) with $\mathbf{P}$, from Eq. (26), one has:

$$\begin{aligned}\mathbf{RS} &= \mathbf{PV}_2 \\ &= [\boldsymbol{rs}_{1,1}, \boldsymbol{rs}_{1,2}, \ldots, \boldsymbol{rs}_{1,K}, \boldsymbol{rs}_{2,1}, \boldsymbol{rs}_{2,2}, \ldots, \boldsymbol{rs}_{2,K}, \ldots, \\ &\quad \boldsymbol{rs}_{i,1}, \boldsymbol{rs}_{i,2}, \ldots, \boldsymbol{rs}_{i,K}]_{I\times Ki} \end{aligned} \tag{29}$$

For the Frobenius norm of the residual matrix $\mathbf{RS}$ of the matrix $\mathbf{V}_2$, one has:

$$(\|\mathbf{RS}\|_2)^2 \le \sum_{m=1}^{i}\sum_{k=1}^{k} \|\boldsymbol{rs}_{m,k}\|_2 \tag{30}$$

From Eqs. (28) and (30), one might make the following conclusion. When the number i of principal components is less than or equal to the chemical rank N of a three-way array, which implies that the size of two spectral subspaces of the unfolded matrices **RA** and **RB** is less than or equal to N and the vectors of these two subspaces correspond to principal components of the spectra, the Frobenius norm of the projective residual matrix **RS** should be very small. That is to say, the difference between these two subspaces is very small.

When $i>N$, the vectors $\boldsymbol{u}_{N+1}, \boldsymbol{u}_{N+2}, \ldots, \boldsymbol{u}_i$ in Eq. (13) do not have the same set of base vectors as the rearranged vectors of $\{\boldsymbol{v}_{N+1,1}, \boldsymbol{v}_{N+1,2}, \ldots, \boldsymbol{v}_{N+1,K}, \boldsymbol{v}_{N+2,1}, \boldsymbol{v}_{N+2,2}, \ldots, \boldsymbol{v}_{N+2,K}, \ldots, \boldsymbol{v}_{i,1}, \boldsymbol{v}_{i,2}, \ldots, \boldsymbol{v}_{i,K}\}$ in Eq. (15), which are formed by rearranging vectors $\{\boldsymbol{v}_{N+1}, \boldsymbol{v}_{N+2}, \ldots, \boldsymbol{v}_i\}$ in Eq. (14) according to Eq. (16), as these vectors all are the noise ones. These vectors cannot be represented by each other, or an expression similar to Eq. (22) does not hold for these vectors. The Frobenius norm of the residual vectors of these vectors projecting along $\mathbf{U}_1$ with $\mathbf{P}$ should increase significantly. Therefore, one can estimate the chemical rank N according to the variation of Frobenius norm of the projective residual matrix of two spectral principal component subspaces of the unfolded matrices **RA** and **RB** with the subspace size (i.e., the component number).

In the J-mode spaces of full rank of the unfolded matrices **RA** and **RB** (i.e., the chromatographic spaces denoted by the column space of the matrices $\mathbf{V}_A$ and $\mathbf{U}_B$ in Eqs. (11) and (12), respectively), one can also use the TMSC method to estimate the chemical rank of the three-way array, which is similar to the spectral spaces expressed by the column space of the matrices $\mathbf{U}_A$ and $\mathbf{V}_B$.

3. The TMSC algorithm

Suppose that I, J, and K modes are the spectral, chromatographic, and sample (or concentration) ones, respectively. The TMSC algorithm for the spectral space consists of the following steps:

1. Unfold the three-way array $\underline{\mathbf{R}}_{I\times J\times K}$ along the two modes of spectral and chromatographic profiles to form the unfolded matrices $\mathbf{RA}_{I\times JK}$ and $\mathbf{RB}_{J\times IK}$ according to Eqs. (8) and (9), respectively.
2. Make the singular value decompositions for **RA** and **RB** to obtain two spectral matrices $\mathbf{U}_A$ and $\mathbf{V}_B$ of principal components according to Eqs. (11) and (12), respectively.
3. Let the component number $i=1$, and take the two spectral submatrices $(I\times i)$ $\mathbf{U}_1$ and $(IK\times i)$ $\mathbf{V}_1$ of the matrices $\mathbf{U}_A$ and $\mathbf{V}_B$ according to Eqs. (13) and (14), respectively.
4. Rearrange the $(IK\times i)$ matrix $\mathbf{V}_1$ to form the $(I\times Ki)$ matrix $\mathbf{V}_2$ according to Eqs. (15) and (16).
5. Construct a projection matrix $\mathbf{P}$ with the matrix $\mathbf{U}_1$ according to Eq. (25), and project $\mathbf{V}_2$ along $\mathbf{U}_1$ with $\mathbf{P}$ to obtain the residual matrix **RS** of $\mathbf{V}_2$ according to Eq. (29).
6. Calculate the Frobenius norm $FN(i)$ of **RS**, and record this value and the corresponding component number i. Let the component number $i=i+1$ and then the algorithm is returned back to step 3.
7. Plot $FN(i)$ versus i. Examine the trend of the curve to see whether a sudden or a sharp turning point of increase of $FN(i)$ with an increase of i occurs; if so, terminate the calculation. Estimate the chemical rank of the three-way array as the value of i in the $FN(i)$ versus i curve corresponding to the turning point.

The algorithm can easily be extended to the case of the two chromatographic spaces corresponding to the two matrices $\mathbf{V}_A$ and $\mathbf{U}_B$ in Eqs. (11) and (12), respectively. All programs were written in MATLAB.

4. Experimental

4.1. The simulated HPLC-DAD data

Three groups of data have been simulated. Each group of data contains three chemical species. The pure spectral profile for each of the three components is simulated as follows:

$$s = k_1 gs(4i-3, m_1, n_1) + k_2 gs(4i-3, m_2, n_2) \quad (i = 1, 2, \cdots, 50) \tag{31}$$

where $gs(x,m,n) = \exp[-(x-m)^2/(2n^2)]$, and the pure chromatographic profile for each of the three components is expressed as:

$$c = kgs(4i-3, m, n) \qquad (i = 1, 2, \cdots, 20) \tag{32}$$

All the parameters for the spectral and chromatographic profiles expressed in Eqs. (31) and (32), respectively (i.e., k_1, m_1, n_1, k_2, m_2, n_2, k, m, and n) are given in Table 1. The correlation coefficient for the aforementioned spectral or chromatographic profiles between components i and j is listed as R_{ij} in Table 1.

The relative concentrations of the three components in the 10 samples contained in a three-way data array are simulated by random numbers of uniform distribution in the region $[0,C_i]$, where $i = 1,2,3$ for components 1, 2, and 3, respectively, and C_i denotes the maximum relative concentration of component i. The parameters C_1, C_2, and C_3 for three groups of the data arrays are shown in Tables 2–4. The matrices **X** of spectral mode, **Y** of the chromatographic mode, and **Z** of the concentration mode are formed according to Eqs. (7), (6) and (5), respectively.

The homoscedastic noises added in the first, second, and third groups of the three-way arrays are simulated by random numbers of normal distribution with zero mean and standard deviations shown in Tables 2–4, respectively. The heteroscedastic noise with relative intensities of different ratios with respect to the signals (these ratios shown in Tables 2–4) is involved in the three groups of data arrays. The response matrices $\mathbf{R}_{..k}$ ($k = 1,2,\ldots,10$) of the 10 samples in each group of the data are formulated according to Eq. (2). All the three-way data simulated are ($50 \times 20 \times 10$) arrays.

Table 1
Spectral and chromatographic parameters and correlation coefficients for three groups of the simulated HPLC-DAD data

	First group	Second group	Third group
Parameters for pure spectral profiles expressed in Eq. (31)			
Component 1	$k_1 = 0.5$, $m_1 = 60$, $n_1 = 10$	$k_1 = 0.5$, $m_1 = 68$, $n_1 = 10$	$k_1 = 0.4$, $m_1 = 50$, $n_1 = 8$
	$k_2 = 0.4$, $m_2 = 110$, $n_2 = 12$	$k_2 = 0.3$, $m_2 = 122$, $n_2 = 10$	$k_2 = 0.3$, $m_2 = 80$, $n_2 = 14$
Component 2	$k_1 = 0.5$, $m_1 = 70$, $n_1 = 22$	$k_1 = 0.5$, $m_1 = 70$, $n_1 = 8$	$k_1 = 0.5$, $m_1 = 90$, $n_1 = 10$
	$k_2 = 0.3$, $m_2 = 130$, $n_2 = 20$	$k_2 = 0.3$, $m_2 = 120$, $n_2 = 7$	$k_2 = 0.2$, $m_2 = 120$, $n_2 = 10$
Component 3	$k_1 = 0.5$, $m_1 = 55$, $n_1 = 20$	$k_1 = 0.5$, $m_1 = 72$, $n_1 = 10$	$k_1 = 0.5$, $m_1 = 120$, $n_1 = 11$
	$k_2 = 0.3$, $m_2 = 100$, $n_2 = 15$	$k_2 = 0.3$, $m_2 = 118$, $n_2 = 10$	$k_2 = 0.2$, $m_2 = 160$, $n_2 = 8$
Parameters for pure chromatographic profiles expressed in Eq. (32)			
Component 1	$k = 0.5$, $m = 35$, $n = 10$	$k = 0.5$, $m = 28$, $n = 10$	$k = 0.4$, $m = 20$, $n = 5$
Component 2	$k = 0.5$, $m = 40$, $n = 5$	$k = 0.5$, $m = 30$, $n = 7$	$k = 0.5$, $m = 40$, $n = 8$
Component 3	$k = 0.5$, $m = 45$, $n = 10$	$k = 0.5$, $m = 32$, $n = 10$	$k = 0.5$, $m = 60$, $n = 6$
Correlation coefficients			
Spectra[a]	$R_{12} = 0.7929$,	$R_{12} = 0.9639$,	$R_{12} = 0.2840$,
	$R_{13} = 0.7588$,	$R_{13} = 0.9402$,	$R_{13} = 0.3750$,
	$R_{23} = 0.7370$	$R_{23} = 0.9617$	$R_{23} = 0.2294$
Chromatograms[a]	$R_{12} = 0.7177$,	$R_{12} = 0.9447$,	$R_{12} = 0.2540$,
	$R_{13} = 0.7242$,	$R_{13} = 0.9297$,	$R_{13} = 0.3208$,
	$R_{23} = 0.7480$	$R_{23} = 0.9449$	$R_{23} = 0.2530$

[a] The correlation coefficient for spectral or chromatographic profiles between components i and j is expressed as R_{ij}.

Table 2
Rank estimation for the first group of the simulated HPLC-DAD data of three components

C_1 = 1.0, C_2 = 1.0,[a] homoscedastic noise 0.20%, and heteroscedastic noise 0.20%						
C_3[a]	0.80	0.10	0.04	0.03	0.02	0.01
Rank of SS[b]	3	3	3	3	3	2
Rank of CS[c]	3	3	3	3	3	2
C_1 = 1.0, C_2 = 1.0, C_3 = 1.0,[a] and homoscedastic noise 0.20%						
Heteroscedastic noise (%)	0.60	1.0	5.0	10.0	15.0	20.0
Rank of SS[b]	3	3	3	3	3	[d]
Rank of CS[c]	3	3	3	3	3	3
C_1 = 1.0, C_2 = 1.0, C_3 = 1.0,[a] and heteroscedastic noise 0.20%						
Homoscedastic noise (%)	0.50	1.00	2.00	3.00	4.00	6.00
Rank of SS[b]	3	3	3	3	3	[d]
Rank of CS[c]	3	3	3	3	3	[d]

[a] The maximum relative concentration.
[b] The rank of the spectral space.
[c] The rank of the chromatographic space.
[d] The rank estimation failed.

4.2. The fluorescence spectral data for anthracene-type compounds

Eight mixture samples of anthracene, 9,10-dimethylanthracene, 1,2:5,6-dibenzanthracene, and 2-aminoanthracene were prepared in cyclohexane solvent,

Table 3
Rank estimation for the second group of the simulated HPLC-DAD data of three components

C_1 = 1.0, C_2 = 1.0,[a] homoscedastic noise 0.20% and heteroscedastic noise 0.20%						
C_3[a]	0.80	0.20	0.15	0.12	0.10	0.05
Rank of SS[b]	3	3	3	3	3	2
Rank of CS[c]	3	3	3	3	3	2
C_1 = 1.0, C_2 = 1.0, C_3 = 1.0,[a] and homoscedastic noise 0.20%						
Heteroscedastic noise (%)	0.60	1.0	2.0	3.0	4.0	5.0
Rank of SS[b]	3	3	3	3	3	3
Rank of CS[c]	3	3	3	3	3	2
C_1 = 1.0, C_2 = 1.0, C_3 = 1.0,[a] and heteroscedastic noise 0.20%						
Homoscedastic noise (%)	0.10	0.30	0.60	0.80	1.00	1.30
Rank of SS[b]	3	3	3	3	3	[d]
Rank of CS[c]	3	3	3	3	3	[d]

[a] The maximum relative concentration.
[b] The rank of the spectral space.
[c] The rank of the chromatographic space.
[d] The rank estimation failed.

Table 4
Rank estimation for the third group of the simulated HPLC-DAD data of three components

C_1 = 1.0, C_2 = 1.0,[a] homoscedastic noise 0.20%, and heteroscedastic noise 0.20%						
C_3[a]	0.80	0.10	0.05	0.02	0.01	0.005
Rank of SS[b]	3	3	3	3	3	2
Rank of CS[c]	3	3	3	3	3	2
C_1 = 1.0, C_2 = 1.0, C_3 = 1.0,[a] and homoscedastic noise 0.20%						
Heteroscedastic noise (%)	0.50	1.0	5.0	10.0	15.0	20.0
Rank of SS[b]	3	3	3	3	3	3
Rank of CS[c]	3	3	3	3	3	3
C_1 = 1.0, C_2 = 1.0, C_3 = 1.0,[a] and heteroscedastic noise 0.20%						
Homoscedastic noise (%)	0.50	1.00	5.00	10.0	15.0	20.0
Rank of SS[b]	3	3	3	3	2 or 3	[d]
Rank of CS[c]	3	3	3	3	3	[d]

[a] The maximum relative concentration.
[b] The rank of the spectral space.
[c] The rank of the chromatographic space.
[d] The rank estimation failed.

and the component concentrations of each sample are shown in Table 5. The fluorescence spectra were recorded using a HITACHI F4500 fluorescence spectrophotometer with a wavelength scan speed of 1200 nm/min. A range of 250–418 nm of excitation wavelength with the interval of 4 nm and a range of 352–500 nm of emission wavelength with the interval of 4 nm were used. The effect of Rayleigh scattering was corrected by background subtraction using a solvent blank. A $(38 \times 43 \times 8)$ data array was obtained for this experiment.

4.3. The excitation–emission fluorescence spectra of the dye mixtures

The fluorescence dyes of acridine, fluorescein, and rhodamine B, coexisting in a liquid laser, were used to

Table 5
The component concentrations of the samples of anthracene-type compounds ($\times 10^{-2}$ ppm)

Sample	1	2	3	4	5	6	7	8
Anthracene	0.00	0.00	1.76	1.76	1.76	5.28	5.28	5.28
9,10-Dimethyl-anthracene	1.52	3.04	0.00	1.52	3.04	0.00	1.52	3.04
1,2:5,6-Di-benzanthracene	2.00	4.00	2.00	4.00	0.00	4.00	0.00	2.00
2-Amino-anthracene	1.60	4.80	4.80	0.00	1.60	1.60	4.80	0.00

Table 6
The component concentrations of the fluorescence dye samples (10^{-3} g/l)

Sample	1	2	3	4	5	6
Acridine	0.00	0.00	0.00	0.00	0.24	0.12
Fluorescein	0.12	0.00	0.12	0.24	0.12	0.24
Rhodamine B	0.00	0.11	0.22	0.11	0.22	0.22

prepare six mixture samples with different component concentrations shown in Table 6. The spectra were recorded with a HITACHI 850 fluorescence spectrophotometer with the following parameters: a wavelength scan speed of 240 nm/min, a range of 450–600 nm of excitation wavelength, a range of 480–620 nm of emission wavelength, and wavelength intervals all at 5 nm. The effect of Rayleigh scattering was corrected by background subtraction using a solvent blank. A ($31 \times 29 \times 6$) three-way array was obtained.

5. Results and discussions

5.1. The simulated HPLC-DAD data arrays

The three groups of the three-way HPLC-DAD arrays with the low, middle, and high degrees of correlation among the coefficients have been simulated to examine the influences of correlation, relative concentration, and homoscedastic and heteroscedastic noise on the TMSC method.

For the first group of data arrays with the middle degree of correlation coefficients of the spectra and chromatograms, the results of the chemical rank estimation with TMSC are shown in Table 2. When both the homoscedastic and heteroscedastic noise are 0.20%, the chemical ranks of the three-way arrays with both the concentration ratios C_3/C_1 and C_3/C_2 down to 0.02 can be estimated accurately by the TMSC method. For the influence of noise, when the relative concentrations of three components all are the same, the chemical rank estimation using TMSC is reasonably accurate for data arrays with a heteroscedastic noise level of up to 15.0% and a homoscedastic noise level of up to 4.0%. However, a too high level of noise would result in the failure of estimating chemical rank, that is, residual variation does not show a turning point of sharp increase with the increase of component number in a residual plot.

The residual plots of chemical rank estimation for the first group of the three-way arrays with the different relative concentrations $C_3 - s$ are shown in Figs. 1 and 2. The chemical rank estimated in the spectral space is equal to that estimated in the chromatographic one. From Figs. 1 and 2, one notices that

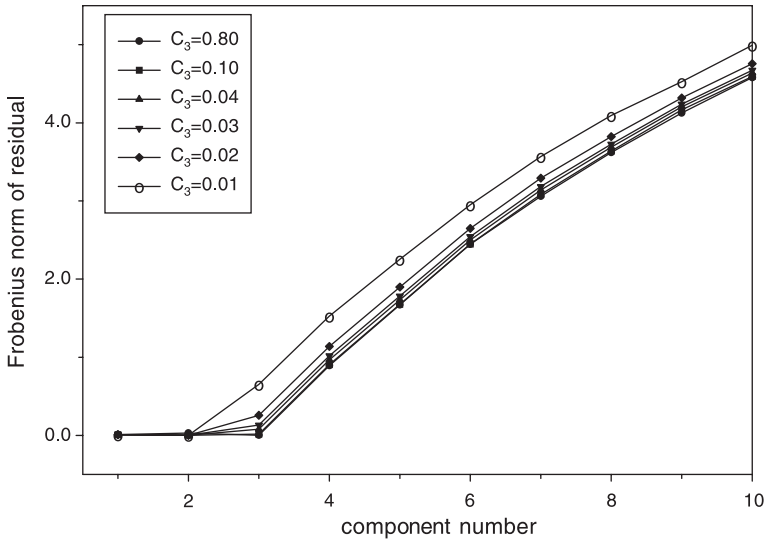

Fig. 1. Rank estimation in the spectral space for the first group of the simulated HPLC-DAD arrays with different maximum relative concentrations $C_3 - s$ and the fixed C_1 and C_2 both as 1.0 and homo- and heteroscedastic noise levels both as 0.20%.

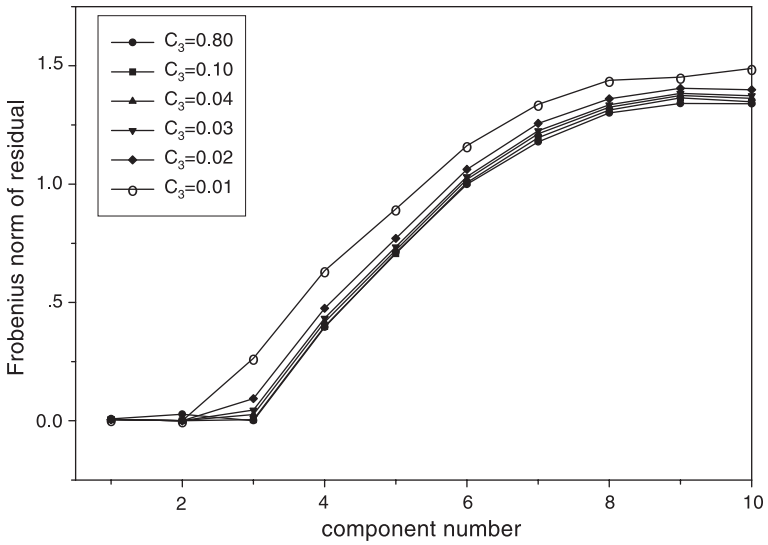

Fig. 2. Rank estimation in the chromatographic space for the first group of the simulated HPLC-DAD arrays with different maximum relative concentrations $C_3 - s$ and the fixed C_1 and C_2 both as 1.0 and homo- and heteroscedastic noise levels both as 0.20%.

when the relative concentration of the third component is less than or equal to 0.02, if the component number chosen is less than or equal to the chemical rank 3 of the three-way array, the Frobenius norm of the projective residual vector or matrix is very small; otherwise, the Frobenius norm increases rapidly with the component number chosen, which denotes that the noise vectors are added to the two subspaces compared using the TMSC method. When C_3 is too small (i.e., there is a minor component in the system), such as 0.01, one cannot obtain a correct result of estimating chemical rank. For 0.01 of C_3, the chemical rank estimated (2) is not equal to the real chemical rank (3) of the three-way array.

Table 3 shows the results of the second group of the HPLC-DAD data sets with the high degree of

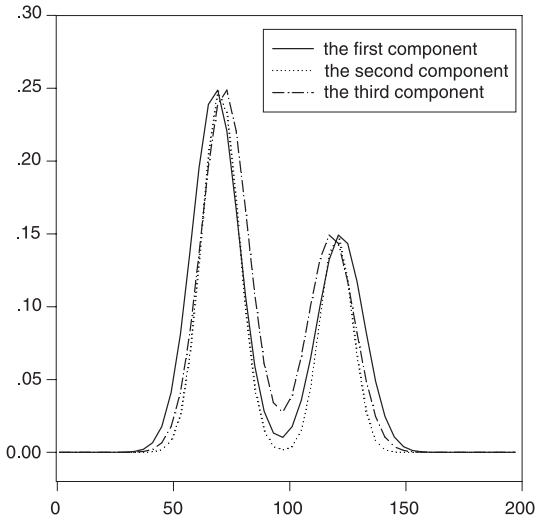

Fig. 3. The spectral profiles for the second group of simulated HPLC-DAD data arrays.

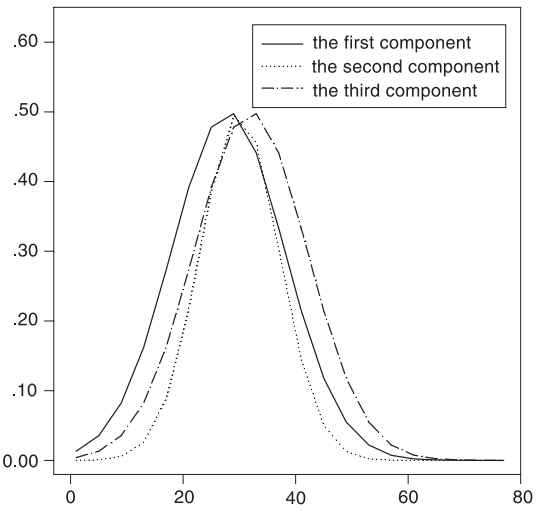

Fig. 4. The chromatographic profiles for the second group of simulated HPLC-DAD data arrays.

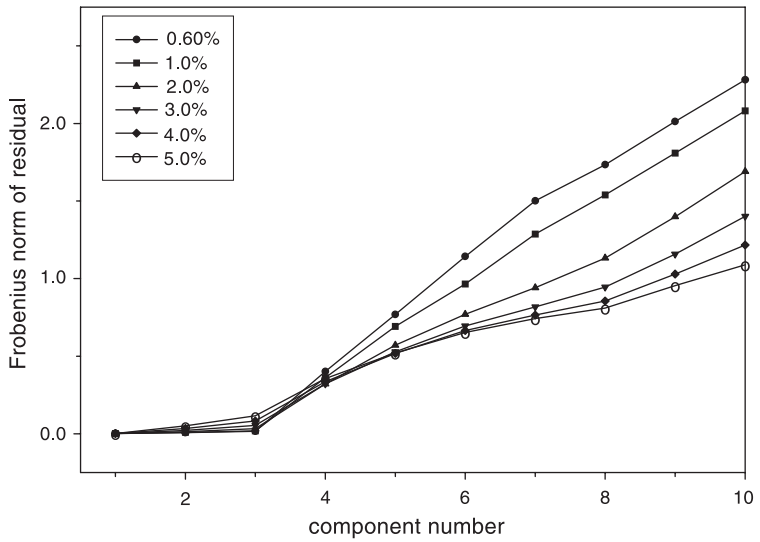

Fig. 5. Rank estimation in the spectral space for the second group of the simulated HPLC-DAD arrays with different heteroscedastic noise levels and a fixed homoscedastic noise of 0.2%, and the maximum relative concentrations C_1, C_2, and C_3 all fixed as 1.0.

correlation among the coefficients. The spectral and chromatographic profiles with the high degree of correlation are plotted in Figs. 3 and 4, respectively. In Figs. 5 and 6, one can estimate the correct chemical rank of 3 in both the spectral and chromatographic spaces for the second group of the simulated HPLC-DAD data arrays with the component relative concentrations $C_1 = 1.0$, $C_2 = 1.0$, and $C_3 = 1.0$, and the homoscedastic noise level of 0.2% and heteroscedastic noise less than or equal to 4.0%. However, when the heteroscedastic noise level is increased up to 5.0%, the chemical ranks estimated in the spectral and chromatographic spaces should be different from each other, such as 3 and 2 of the chemical ranks

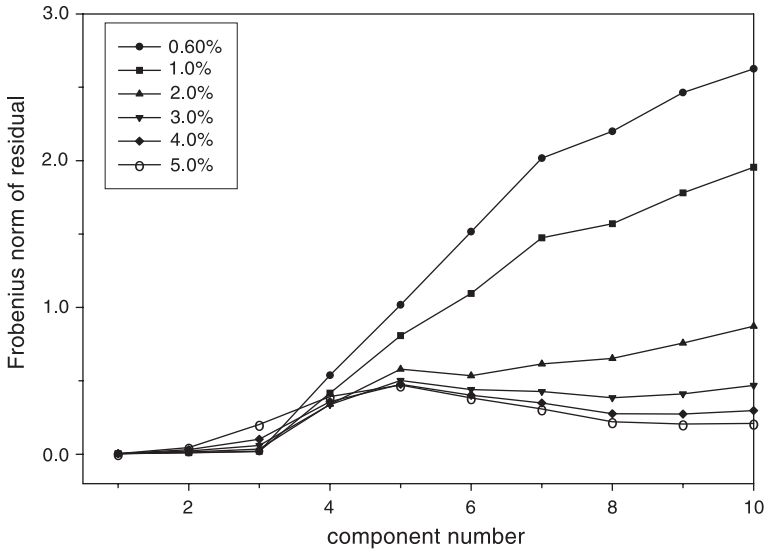

Fig. 6. Rank estimation in the chromatographic space for the second group of the simulated HPLC-DAD arrays with different heteroscedastic noise levels and a fixed homoscedastic noise of 0.2%, and the maximum relative concentrations C_1, C_2, and C_3 all fixed as 1.0.

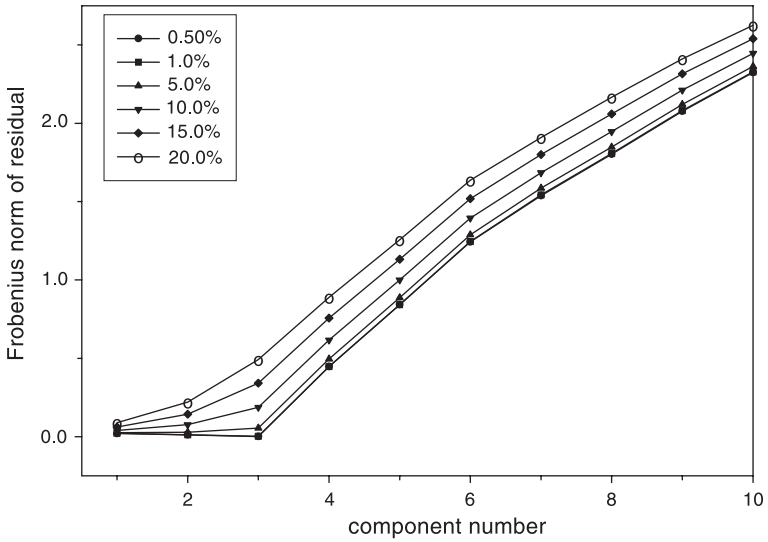

Fig. 7. Rank estimation in the spectral space for the third group of the simulated HPLC-DAD arrays with different homoscedastic noise levels and a fixed heteroscedastic noise of 0.2%, and the maximum relative concentrations C_1, C_2, and C_3 all fixed as 1.0.

estimated in the chromatographic and spectral spaces, respectively, for 5.0% of the heteroscedastic noise level, which might result from different levels of noise in the two spaces. From Table 3, one notices that an accurate value can be obtained for the chemical component number of the data array with a very high degree of correlation of the pure spectral or chromatographic profiles.

Relatively low correlation coefficients are beneficial for chemical rank estimation of a three-way array. Table 4 shows the results for the third group of data arrays with low correlation coefficients of spectra or

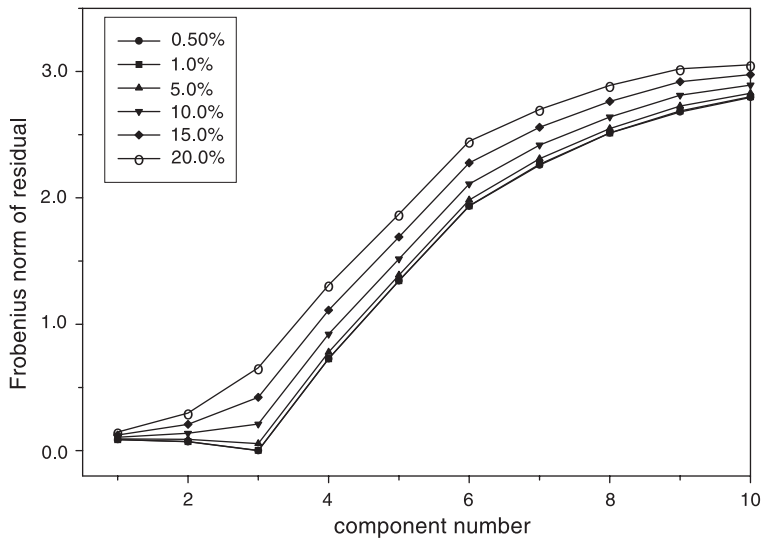

Fig. 8. Rank estimation in the chromatographic space for the third group of the simulated HPLC-DAD arrays with different homoscedastic noise levels and a fixed heteroscedastic noise of 0.2%, and the maximum relative concentrations C_1, C_2, and C_3 all fixed as 1.0.

chromatograms. Figs. 7 and 8 show Frobenius norm plots of residuals in the spectral and chromatographic spaces, respectively, for the third group of the simulated HPLC-DAD arrays with different homoscedastic noise levels and a fixed heteroscedastic noise of 0.20%, together with C_1, C_2, and C_3 all fixed as 1.0. One observes an outstanding feature of the TMSC method in resisting very high degrees of noise under such conditions, even when the noise is heteroscedastic. From the two figures, it can be seen that when the homoscedastic noise level is increased, the residuals of the two subspaces compared would increase, and when it reaches 15.0%, the turning point of residual variation is not sharp enough, that is, one may estimate a chemical rank of 3 or 2. Such a noise level seems to be the threshold, as for a homoscedastic noise level of 20.0%, one can hardly estimate correctly the chemical rank of a three-way array.

Compared to the principal norm vector orthogonal projection approach [17], the proposed TMSC method has two outstanding features: for a three-way array with a very high degree of correlation of spectra or chromatograms, its chemical rank can also be accurately estimated; a relatively high level of noise, even when the noise is heteroscedastic, can be resisted in obtaining the correct value of chemical rank of a three-way array. However, the TMSC method has a more strict restriction for a three-way array, that is, there are two modes of full rank in the data array. The principal norm vector orthogonal projection method only requires one mode of full rank in the array.

In real applications, one could estimate the chemical rank of a three-way array along any one of the two full-rank modes, such as spectral or chromatographic mode for an HPLC-DAD data array, and excitation or emis-

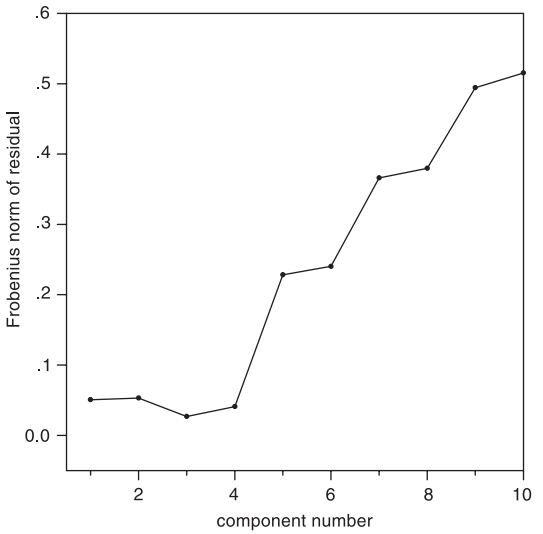

Fig. 10. Rank estimation for the fluorescence data of anthracene-type compounds in the excitation spectral space.

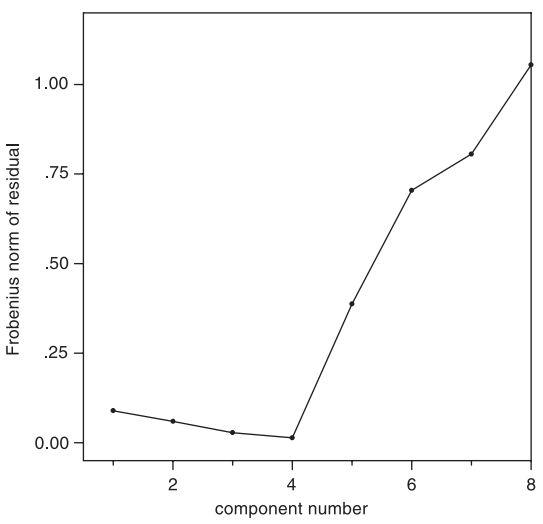

Fig. 9. Rank estimation for the fluorescence data of anthracene-type compounds in the emission spectral space.

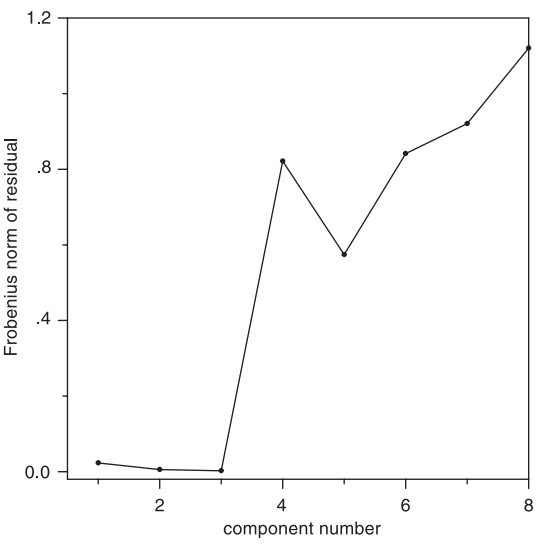

Fig. 11. Rank estimation for the fluorescence dye data in the emission spectral space.

sion mode for an excitation–emission fluorescence three-way array. The estimation of chemical rank along two full-rank modes could be used as a double check for each other. Certainly, if the concentration mode is of full rank, the TMSC method could also be used along this mode. It is necessary to notice that the two-mode subspace comparison approach is valid not only for a trilinear data set but also for a three-way data array with the two full-rank modes. The former is a more strict condition than the latter. The TMSC method only requires that a data array has two full-rank modes.

5.2. The fluorescence three-way data array of anthracene-type compounds

The correct chemical rank 4 of this data array has been obtained using the TMSC approach in the emission and excitation spectral spaces as shown in Figs. 9 and 10, respectively. When the sizes of two subspaces compared are increased from 1 to 4, the variance of the Frobenius norms of the projective residual matrices in the excitation spectral space is much different from those corresponding to the emission space, which implies that there are different noise levels and backgrounds in these two mode spaces. Accurate analytical results are easily obtainable by using the resolution methods of a three-way array.

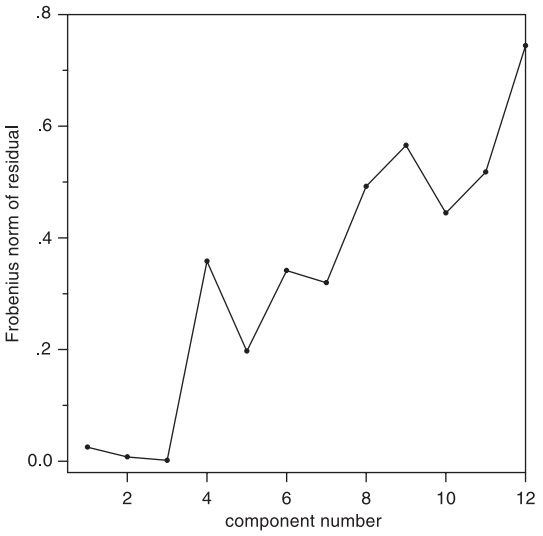

Fig. 12. Rank estimation for the fluorescence dye data in the excitation spectral space.

5.3. The fluorescence dye data

The chemical rank of the fluorescence dye three-way array of the three components was estimated using the TMSC method in the emission and excitation spectral spaces, respectively. The results are shown in Figs. 11 and 12. From these two figures, a correct chemical rank of 3 can be obtained. The knowledge of correct chemical rank makes it easy to resolve the concentration and spectral profiles using the resolution methods of a three-way array.

6. Conclusions

Two matrices are formed by unfolding a three-way array along two full-rank modes (i.e., the spectral or chromatographic modes). Two subspaces are constructed by using the first i principal component vectors of the same mode space (i.e., the spectral or chromatographic spaces) of the two unfolded matrices. When the principal component number i taken is less than or equal to the chemical rank n of this three-way array, these two subspaces are very similar to each other, and the Frobenius norm of the residual matrix formed by the one subspace projecting along another one tends to be very small. Otherwise, this norm trends to increase rapidly with the component number i taken. The chemical rank of a three-way array can be estimated according to the variation of this norm with the component number i. This approach, called two-mode subspace comparison (TMSC), can accurately estimate the chemical rank of a data array with a very high degree of correlation of spectral or chromatographic profiles or with a very high level of noise, including that of heteroscedastic noise. The chemical ranks of two real systems (i.e., anthracene type compounds and fluorescence dye samples) have been estimated successfully using the TMSC approach.

Acknowledgements

This research was supported by grants from the National Natural Science Foundation of China (grant nos. 20075006, 29735150, and 29975007).

References

[1] B.K. Lavine, Anal. Chem. 72 (2000) 91R–97R.
[2] J.H. Jiang, H.L. Wu, Z.P. Chen, R.Q. Yu, Anal. Chem. 71 (1999) 4254–4262.
[3] J.H. Jiang, H.L. Wu, Y. Li, R.Q. Yu, J. Chemom. 13 (1999) 557–578.
[4] J.H. Jiang, H.L. Wu, Y. Li, R.Q. Yu, J. Chemom. 14 (2000) 15–36.
[5] Z.P. Chen, H.L. Wu, Y. Li, R.Q. Yu, Anal. Chem. Acta 423 (2000) 187–196.
[6] Y. Li, J.H. Jiang, H.L. Wu, Z.P. Chen, R.Q. Yu, Chemom. Intell. Lab. Syst. 52 (2000) 33–43.
[7] Z.P. Chen, Y. Li, R.Q. Yu, J. Chemom. 15 (2001) 149–167.
[8] J.B. Kruskal, Linear Algebra Appl. 18 (1977) 95–138.
[9] P.M. Kroonenberg, Three-Mode Principal Component Analysis: Theory and Application, DSWO Press, Leiden, 1989.
[10] P. Geladi, Chemom. Intell. Lab. Syst. 7 (1989) 11–30.
[11] R.A. Harshman, UCLA Work. Pap. Phon. 16 (1970) 1–84.
[12] J.D. Carrol, J.J. Chang, Psychometrika 35 (1970) 283–319.
[13] L. Tucker, in: C. Harris (Ed.), Implications of Factor Analysis of Three-Way Matrices for Measurement of Charge, Problems of Measuring Charge, University of Wisconsin Press, Madison, WI, 1963, pp. 122–137.
[14] S. Wold, P. Geladi, K. Esbensen, J. Ohman, J. Chemom. 1 (1987) 41–56.
[15] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149–171.
[16] D.J. Louwerse, A.K. Smilde, H.A.L. Kiers, J. Chemom. 13 (1999) 491–510.
[17] H.P. Xie, J.H. Jiang, G.L. Shen, R.Q. Yu, Comput. Chem. 26 (2002) 183–190.
[18] E.R. Malinowski, Factor Analysis, Wiley-Interscience, New York, 1991.
[19] H.L. Shen, Y.Z. Liang, O.M. Kvalheim, R. Manne, Chemom. Intell. Lab. Syst. 51 (2000) 49–59.
[20] J.A. Westerhuis, T. Kourti, J.F. Macgregor, J. Chemom. 13 (1999) 379–413.