



# Chemical rank estimation for second-order calibration by discrete Fourier transform coupled with robust statistical analysis



Yong-Jie Yu <sup>a,c,1</sup>, Hai-Yan Fu <sup>b,1</sup>, Hui-Wen Gu <sup>a</sup>, Yong Li <sup>a</sup>, Chao Kang <sup>a</sup>, Yi-Peng Wang <sup>c</sup>, Hai-Long Wu <sup>a,\*</sup>

<sup>a</sup> State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China

<sup>b</sup> College of Pharmacy, South-Central University for Nationalities, Wuhan 430074, China

<sup>c</sup> Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou 450001, China

## ARTICLE INFO

### Article history:

Received 27 September 2014

Received in revised form 1 December 2014

Accepted 13 December 2014

Available online 23 December 2014

### Keywords:

Chemical rank estimation

Discrete Fourier transformation

Second-order calibration

Overlapped fluorescence spectra resolution

## ABSTRACT

The accurate estimation of the underlying number of components in complex samples is critical in data analysis. A new chemometric strategy was developed in this study to determine accurately the number of underlying components in complex samples. First, discrete Fourier transformation was used to project the eigenvectors from the singular value decomposition to the frequency space. A robust statistical analysis based on iterative *t*-test was then employed to eliminate the outliers in the Fourier coefficients of each eigenvector. Finally, ANOVA was used to differentiate the meaningful components from noise. Simulated and published fluorescence datasets were used to demonstrate the strategy. Results indicate that the proposed strategy accurately and efficiently estimated the number of underlying components in the analyzed dataset. Moreover, the performance of the proposed method was comparable with the well-known core consistency diagnostic and Monte Carlo simulation coupled with frequency location methods. The new technique coupled with second-order calibration was successfully used to resolve the problem of seriously overlapped fluorescence spectra in the accurate quantification of fluoroquinolone antibiotics in tap water samples. Second-order advantage was achieved.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

Interpreting an overlapped chemical signal provided by a second-order instrument, such as gas chromatography–mass spectrometry, high-performance liquid chromatography–mass spectrometry, and excitation-emission matrix fluorescence, is a challenging task in complex sample analysis. In this kind of analysis, the accurate number of components in a dataset necessary to provide valuable resolutions is difficult to determine [1–5]. To date, methods for overlapped signal resolution have been extensively studied, and chemometric methodologies that include multivariate curve resolution [6–10] and second-order calibration [11–13] have been widely accepted in various reports. Significant advantages of these methods include separation of overlapped chemical signals using a mathematical separation strategy and accurate quantification of analytes. However, successful applications of these methods largely depend on chemical rank estimation, which continues to be a challenging task in practical applications.

The accurate estimation of the number of underlying components in samples is a primary step in chemometric methods. Either

underestimation or overestimation of chemical rank in analyzed samples results in invalid solutions for methods that are sensitive to the number of components, such as the well-known parallel factor analysis [14]. Meanwhile, methods that are not sensitive to the number of components also encounter difficulty in obtaining valuable solutions when the employed number of components is less than the underlying components. These methods include alternating trilinear [15] and self-weighted alternating trilinear decompositions (SWATLD) [16] and algorithm combination method (ACM) [17]. Notably, accurate determination the number of components is also a critical step for these methods. Our experience indicated that both quantitative and qualitative results can be improved when excessive factors that corresponding to small components were introduced [17]. Nevertheless, chemical rank estimation for second-order calibration has not been paid much attention. Therefore, high-quality chemical rank estimation methods should be developed for practical applications.

Analysts usually use some techniques that are built based on the singular value decomposition of the bilinear structure of the recorded signal to roughly estimate the number of underlying components. Typical examples of these methods are represented by Malinowski's RE and empirical indicator functions [18] and those based on singular values [19]. These methods work well when the spectra of compounds are not seriously overlapped, and thus valuable information of component

\* Corresponding author. Tel./fax: +86 731 88821818.

E-mail address: [hlwu@hnu.edu.cn](mailto:hlwu@hnu.edu.cn) (H.-L. Wu).

<sup>1</sup> These authors contributed equally.

number is provided during data analysis. In practical applications, however, methods based on singular values exhibit limitation in accurately separating singular values of meaningful components from those corresponding to noise.

The methods that have been proposed to address the chemical rank estimation problem in second-order calibration can be categorized in two groups according to the employed strategies [20–34]. Methods of the first group estimate chemical rank by searching for components that maintain a trilinear structure. Methods that belong to the first group include the well-known core consistency diagnostic (CORCONDIA) [21] and SWATLD coupled with Monte Carlo simulation (MCS) [30]. The strategies of these methods are based on the fact that core consistency value or relative concentration value rapidly decreases to zero when excessive factors are employed. By contrast, methods of the second group are mainly built upon a vector space projection strategy, such as ADD-ONE-UP [23], principal norm vector orthogonal projection [24,25], pseudo-sample extraction projection technique [26], and region-based moving window subspace projection technique coupled with MCS [27,28]. The idea behind these methods is that eigenvectors obtained by singular value decomposition (SVD) contain more useful information than eigenvalues. A subspace spanned by primary eigenvectors obtained from various types of augmented matrices of three-way data array is highly coincident. Inconsistency in a spanned subspace significantly increases when excessive eigenvectors that correspond to noise factors are introduced.

Although several studies used the aforementioned methods in chemical rank estimation of three-way data, determining the number of components in a complex sample analysis is still a challenging task. For instance, CORCONDIA is not efficient in chemical rank estimation in practical applications, which might be inconvenient for large-scale sample analysis. Methods based on vector space projection are more applicable for simulations rather than real samples, because eigenvectors corresponding to meaningful components are obviously different from those representing random noise in simulations. In practical applications, eigenvectors of components are not readily separated from noise factors because constructing pseudo-samples in these methods creates additional problems, such as enhancing the effects of non-linear factors.

However, the idea of vector space projection, in which eigenvectors contain more useful information than eigenvalues, is valuable in diagnosing chemical rank for practical applications [33]. This finding is due to the high computational efficiency of eigenvectors with the use of a computer. Yu et al. developed an MCS coupled with frequency location (MCS-FL) [34] to obtain accurately the number of components in three-way data. This method was used successfully to detect chemical components in mixtures. However, MCS-FL focused on providing a number of meaningful components instead of the necessary chemical rank for second-order calibration. Moreover, the method is not suitable for current trilinear algorithms. Additionally, the adopted frequency localization strategy did not fully take advantage of the distribution characteristics in frequency analysis. This strategy was performed by obtaining the sum of squared Fourier coefficients after normalization to distinguish components and noise.

In the present study, a novel chemical rank estimation method based on discrete Fourier transform coupled with robust statistical analysis (DFT-RSA) was developed. First, the eigenvectors from SVD were projected to the frequency space by DFT to obtain the corresponding Fourier coefficient vectors. A robust statistical method was then employed to differentiate the useful vectors from noise. Simulation and a published fluorescence dataset that quantified local anesthetic drugs, namely procaine and tetracaine, in human plasma were employed to verify the performance of the proposed method. The results indicate that chemical rank was estimated accurately using the DFT-RSA. Finally, the proposed method was successfully used in combination with ACM to resolve the problem of seriously overlapped fluorescence spectra. The methods were applied to quantify accurately three widely used fluoroquinolone

antibiotic agents, namely, danofloxacin (DAN), sarafloxacin (SAR), and ofloxacin (OFL).

## 2. Methodology

### 2.1. DFT-RSA

Fig. 1 shows the workflow of DFT-RSA, which mainly composed three stages. First, the DFT was employed to project eigenvectors from the SVD to the frequency space, and the corresponding Fourier coefficients were obtained. A robust statistical analysis based on iterative *t*-test was then used to eliminate the outliers of the Fourier coefficients for each eigenvector. Finally, ANOVA was used to determine the significant eigenvectors and the Fourier coefficient distributions of which in the frequency space were significantly different from noise.

#### 2.1.1. Three-way data pre-processing

In second-order calibration, a second-order tensor (matrix) is recorded for each sample. This tensor is calculated using Eq. (1):

$$\mathbf{X}_{\cdot,k} = \mathbf{A} \text{diag}(\mathbf{c}_{(k)}) \mathbf{B}^T \quad (1)$$

where  $\mathbf{X}_{\cdot,k}$  represents the response matrix of the *k*th sample with a size of  $I \times J$ , and *I* and *J* are the numbers of excitation and emission channels, respectively.  $\mathbf{A}$  is the excitation matrix with a size of  $I \times N$ , where *N* is the number of underlying components,  $\mathbf{B}$  denotes the emission matrix with a size of  $J \times N$ , and  $\mathbf{c}_{(k)}$  contains the corresponding concentrations in the

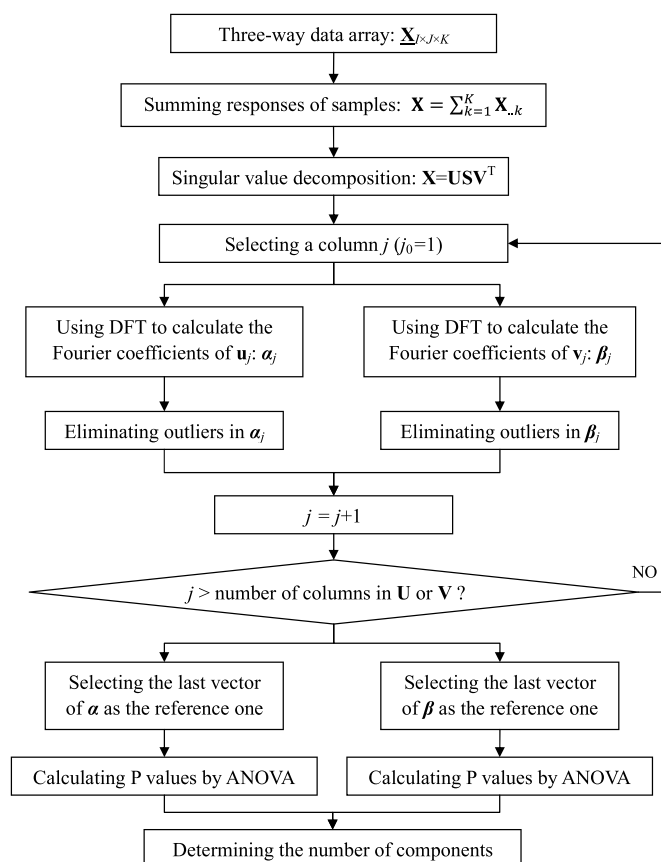


Fig. 1. Framework of chemical rank estimation based on DFT-RSA.

Download English Version:

<https://daneshyari.com/en/article/1179381>

Download Persian Version:

<https://daneshyari.com/article/1179381>

[Daneshyari.com](https://daneshyari.com)