Talanta 83 (2010) 541–548

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Direct decomposition of three-way arrays using a non-negative approximation

Jiangming Sun, Tonghua Li*, Peisheng Cong, Wenwei Xiong, Shengnan Tang, Li Zhu

Department of Chemistry, Tongji University, 1239 Siping Road, Shanghai 200092, China

ARTICLE INFO

Article history:
Received 10 May 2010
Received in revised form
20 September 2010
Accepted 27 September 2010
Available online 1 October 2010

Keywords:
Non-negative matrix approximation
Second-order calibration
PARAFAC
Kinetics

ABSTRACT

Non-negative matrix approximation (NNMA) has been used in diverse scientific fields, but it still has some major limitations. In the present study a novel trilinear decomposition method, termed three-way NNMA (TWNNMA), was developed. The method decomposes three-way arrays directly without unfolding and overcomes the restriction of locking zero elements in the deduced multiplicative update rules by adding a positive symmetric matrix. Direct trilinear decomposition was used as the TWNNMA initialization method and experimental results confirm that this greatly accelerated the convergence. An obvious advantage of TWNNMA is the uniqueness of the non-negative solution, which facilitates a better understanding of the underlying physical realities of complex data. TWNNMA was applied in complex systems such as chemical kinetics, second-order calibration and analysis of GC–MS data. The results demonstrate that TWNNMA, differing from previous trilinear decomposition methods, is comparable to existing second-order calibration methods and represents a promising resolution method for complex systems.

1. Introduction

Owing to the variety of multi-way data produced by modern analytical instrumentation and to the development of chemometrics methods, multi-way analysis has been extensively applied in chemistry in recent years. Multi-way refers to data with two or more dimensions, which are also called multi-order. In particular, a unique solution can be obtained when decomposing three-way arrays under conditions for essential uniqueness such as Kruskal's permutation lemma, which is attracting increasing interest in chemistry. There is also a property associated with second- or higher-order sample data (three- or higher-way arrays) called the second-order advantage [1]. The applicability and immense potential of the second-order advantage in analytical chemistry have become an active area of theoretical interest and of intense experimental research [2–7]. However, regardless of the algorithm used, multi-way data do not always lead to the second-order advantage [8]. Suitable algorithms for analysis of second-order data are parallel factor analysis (PARAFAC) [9], the generalized rank annihilation method (GRAM) [10], direct trilinear decomposition (DTLD) [11], mltivariate curve resolution coupled to alternating least squares (MCR–ALS) [12], alternating trilinear decomposition (ATLD) [13] and its variants self-weighted alternating trilinear decomposition (SWATLD) [14] and alternating penalty trilinear decomposition (APTLD) [15], and bilinear least squares (BLLS) [16], unfolded partial least squares(UPLS) [17] and multi-way partial least squares (N-PLS) [18] coupled to residual bilinearization (RBL) procedure [19].

Non-negative matrix approximation (NNMA), also known as non-negative matrix factorization [20], which approximates a given matrix **V** of non-negative values using an additive linear combination of two non-negative matrices **W** and **H**, is a new method for analysis of multi-way data. Decomposition is usually achieved by optimizing cost functions such as the Frobenius norm and Bregman divergence to measure the divergence between **V** and the product of **W** and **H**. There are two other widely used NNMA approaches that use alternating non-negative least squares termed positive matrix factorization (PMF) [21] and non-negative least squares (NNLS) [22]. NNMA was popularized by Lee and Seung in a simple and useful algorithm procedure based on multiplicative update rules. As often the data to be analyzed is nonnegative, and the low-rank data are further required to be nonnegative values in order to avoid contradiction of physical realities. When classical tools are used, retention of this non-negativity cannot be guaranteed. The NNMA approach to approximate a given non-negative data matrix is thus a natural choice. NNMA has been successfully used in diverse scientific fields, including signal and image processing [20,23–25], natural language processing applications such as text mining and document clustering [26–28], information retrieval [29], computational biology applications [30] such as molecular pattern discovery [31] and functional characterization of genes and biomedical informatics [32], and chemistry [33–37]. Recently, NNMA has been extended to higher-order tensors to yield a model known as non-negative PARAFAC [38], which has also been investi-

* Corresponding author. Tel.: +86 21 65983987; fax: +86 21 65983987.
E-mail addresses: jiangming.sun@hotmail.com (J. Sun), lith@tongji.edu.cn (T. Li).

0039-9140/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2010.09.035

gated as positive tensor factorization (PTF) [39], nonnegative tensor factorization (NTF) [40,41] or nonnegative tucker decomposition (NTD) [42].

NNMA has attracted much attention during the past decade but it still has three main limitations. First, similar to the rotational ambiguity of factor analysis, decomposition for NNMA is not unique in general [43]. In other words, a matrix and its inverse can be used to transform two factorization matrices, for example **WH** to $\mathbf{WDD}^{-1}\mathbf{H}$. If the two new matrices **WD** and $\mathbf{D}^{-1}\mathbf{H}$ are nonnegative, they form another parameterization of the factorization. Second, most methods used for NNMA do not guarantee algorithm convergence at a reasonable speed [44]. Third, a serious drawback of multiplicative update rules is that once an element in **W** or **H** becomes zero, it must remain zero during iteration. This locking of zero elements is restrictive, meaning that once the algorithm starts to head towards a fixed point, even if it is a poor fixed point, the process must continue in the same vein.

To attempt to overcome the non-uniqueness of NNMA, several modifications have been proposed. Hoyer showed that explicit incorporation of a sparseness constraint improved the final decomposition [45]. Alberto imposed a non-smooth constraint to control the sparseness of both factors, which resulted in better interpretability of the factors [24]. Gao et al. improved the NNMA algorithm by imposing smoothness, unimodality and sparseness properties for chemical spectra to obtain acceptable and reliable solutions [33]. To speed up NNMA convergence, Gonzalez created a modification by adding step-length parameters to accelerate the Lee and Seung multiplicative update algorithm [46]. Shepherd [47] developed a proposed gradient descent method that can accelerate convergence using suitable choices for the step size. Others have used an alternative least squares (ALS) algorithm and variants thereof instead [48]. In these algorithms, a least squares step is followed by another least squares step in an alternating process. There is usually a projection step after each least squares step to ensure non-negativity. Depending on the implementation, ALS algorithms for NNMA can be very fast. Moreover, unlike multiplicative algorithms, ALS-based NNMA does not lock elements. As observed from experimental results, by imposing sparseness, orthogonality or specific spectral properties, ambiguities in NNMA results can be reduced. NNMA convergence speed increases when a gradient descent or ALS is used. However, decomposition for NNMA is still far from unique. Moreover, a convergence theory to support the gradient descent approach and ALS is somewhat lacking [49] and in some cases ALS does not identify a non-negative solution that fits the data well. Finally, no multiplicative update rules-based algorithm has resolved the restriction of locking zero elements.

In the present study, a novel method for trilinear decomposition, termed three-way NNMA (TWNNMA), was developed. TWNNMA, based on the PARAFAC model, involves fitting the model to the least squares with a non-negative constraint for each dimension. It can decompose three-way arrays directly without unfolding or rearranging the arrays into a matrix by using the inner product of two such arrays. This approach can be used to resolve extreme spectral overlapping because of the second-order advantage of the method. To overcome the restriction of locking zero elements in deduced multiplicative update rules, a positive symmetric matrix is multiplied by each factor to form new multiplicative update rules. By altering the parameter in the symmetric matrix, different degree of smoothness can be achieved. When the new multiplicative update rules are combined with a stopping condition, TWNNMA converges to a stationary point. Good initialization can improve the speed and accuracy of the solutions for many NNMA algorithms, so different initialization methods were also investigated. The experimental results demonstrate that the residual sum of squares decreased much faster for DTLD initialization than for other initializations methods, thus accelerating TWNNMA convergence. An obvious advantage of the method is the uniqueness of the non-negative solution, which facilitates a better understanding of the underlying physical realities of complex data. TWNNMA was applied in complex systems such as chemical reaction kinetics and second-order calibration. Satisfactorily resolved kinetic concentration profiles and spectra were obtained. Compared with existing second-order calibration methods, the TWNNMA resolution results are reasonably precise. Thus, TWNNMA is a promising method for analysis of chemical reaction kinetics. The quantitative determination results demonstrate that TWNNMA is a novel method that provides the second-order advantage.

2. Notation

Throughout this paper, scalars are represented by lowercase italics and vectors by bold lowercase characters. Bold uppercase letters denote two-way matrices and underlined bold uppercase letters denote three-way arrays. The letters *I*, *J*, *K* indicate the dimensions of different modes in three-way arrays. *F* is the number of factors used in the decomposition. $\underline{\mathbf{X}}$ represents a three-way array, and the *ijk*th element of $\underline{\mathbf{X}}$ is x_{ijk} or $\underline{\mathbf{X}}_{ijk}$. **A**, **B** and **C** with dimensions $I \times N$, $J \times N$ and $K \times N$, respectively, are the three factorization matrices of $\underline{\mathbf{X}}$.

Outer product of two matrices: a special tensor product is defined whereby a three-order tensor can be obtained as the outer product of two matrices with the same column.

Inner product of two three-order tensors: a two-order tensor can be obtained as the dot product of two three-order tensors. The symbol "••" is used here to represent the inner product, which differs from the commonly used dot product in mathematics.

Terms and abbreviations are listed in the glossary (see Glossary).

3. Method and algorithm

The CANDECOMP/PARAFAC model, sometimes simply called the PARAFAC model, essentially searches for a solution to the following problem: given a tensor $\underline{\mathbf{X}}$ belonging to $\mathbf{R}^{d_1 \times \cdots \times d_k}$, find an optimal rank-*F* approximation. It can be expressed as:

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk}, \tag{1}$$

where three-way array $\underline{\mathbf{X}}$ with elements x_{ijk} are given by three loading matrices, **A** ($I \times F$), **B** ($J \times F$) and **C** ($K \times F$) with elements $\mathbf{A}_{if}$, $\mathbf{B}_{jf}$, and $\mathbf{C}_{kf}$. *F* is the number of factors. The trilinear model minimizes the sum of squares of the residual error e_{ijk}.

Fitting the PARAFAC model to least squares, the objective function can be expressed as:

$$D = \|\underline{\mathbf{X}} - \underline{\mathbf{ABC}}\|_F^2. \tag{2}$$

Being analogous to the Euclidean distance, the residual sum of squares is used to evaluate the approximation. Its lower bound is 0 and it equals 0 if and only if $\underline{\mathbf{X}} = \underline{\mathbf{ABC}}$.

Taking the derivative of objective *D* with respect to **A** gives:

$$\frac{\partial}{\partial \mathbf{A}_{ia}} D = -2\underline{\mathbf{X}} \bullet\bullet \underline{\mathbf{BC}} + 2\underline{\mathbf{ABC}} \bullet\bullet \underline{\mathbf{BC}}. \tag{3}$$

According to the gradient algorithm:

$$\mathbf{A}_{i\alpha} \leftarrow \mathbf{A}_{i\alpha} - \delta_{i\alpha} \frac{\partial}{\partial \mathbf{A}_{ia}} D, \tag{4}$$

$$\mathbf{A}_{i\alpha} \leftarrow \mathbf{A}_{i\alpha} + \delta_{i\alpha} \left(\underline{\mathbf{X}} \bullet\bullet \underline{\mathbf{BC}} - \underline{\mathbf{ABC}} \bullet\bullet \underline{\mathbf{BC}}\right). \tag{5}$$

Note that the coefficient, δ_{ia}, which can be called the learning rate.

Forcing:

$$\delta_{ia} = \frac{\mathbf{A}_{ia}}{(\underline{\mathbf{ABC}} \bullet \bullet \underline{\mathbf{BC}})_{ia}}, \tag{6}$$

gives the multiplicative rule:

$$\mathbf{A}_{ia} \leftarrow \mathbf{A}_{ia} \frac{(\underline{\mathbf{X}} \bullet \bullet \underline{\mathbf{BC}})_{ia}}{(\underline{\mathbf{ABC}} \bullet \bullet \underline{\mathbf{BC}})_{ia}}. \tag{7a}$$

Similarly,

$$\mathbf{B}_{ja} \leftarrow \mathbf{B}_{ja} \frac{(\underline{\mathbf{X}} \bullet \bullet \underline{\mathbf{AC}})_{ja}}{(\underline{\mathbf{ABC}} \bullet \bullet \underline{\mathbf{AC}})_{ja}}, \tag{7b}$$

$$\mathbf{C}_{ka} \leftarrow \mathbf{C}_{ka} \frac{(\underline{\mathbf{X}} \bullet \bullet \underline{\mathbf{AB}})_{ka}}{(\underline{\mathbf{ABC}} \bullet \bullet \underline{\mathbf{AB}})_{ka}}, \tag{7c}$$

Note that after each update of **A** and **B**, there is a normalization step.

3.1. Modified multiplicative update rules

The basic and later NNMA multiplicative update rules reported do not allow any factor element to ever come up from the value of zero. In the present study, a smoothing matrix **S** was added to the above multiplicative rules. The modified multiplicative rules thus obtained are:

$$\mathbf{A}_{ia} \leftarrow \mathbf{AS}_{ia} \frac{(\underline{\mathbf{X}} \bullet \bullet \underline{\mathbf{BC}})_{ia}}{(\underline{\mathbf{ABC}} \bullet \bullet \underline{\mathbf{BC}})_{ia} + \varepsilon}, \tag{8a}$$

$$\mathbf{B}_{ja} \leftarrow \mathbf{BS}_{ja} \frac{(\underline{\mathbf{X}} \bullet \bullet \underline{\mathbf{AC}})_{ja}}{(\underline{\mathbf{ABC}} \bullet \bullet \underline{\mathbf{AC}})_{ja} + \varepsilon}, \tag{8b}$$

$$\mathbf{C}_{ka} \leftarrow \mathbf{CS}_{ka} \frac{(\underline{\mathbf{X}} \bullet \bullet \underline{\mathbf{AB}})_{ka}}{(\underline{\mathbf{ABC}} \bullet \bullet \underline{\mathbf{AB}})_{ka} + \varepsilon}, \tag{8c}$$

where ε is a pre-defined small positive number in case the denominator is zero. The positive symmetric matrix $\mathbf{S} \in \mathbf{R}^{\mathbf{N}\times\mathbf{N}}$ is defined as:

$$\mathbf{S} = \left(1-\theta\right)\mathbf{I} + \frac{\theta}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}, \tag{9}$$

where **I** is the identity matrix, **1** is a vector of ones, N is the number of factors and the parameter θ satisfies $0 \le \theta < 1$.

The effect of smoothing matrix **S** can be explained as follows. Let **A** be a non-negative matrix. Consider the transformed matrix $\mathbf{Y} = \mathbf{AS}$. If $\theta = 0$, then $\mathbf{Y} = \mathbf{A}$ and no smoothing of **A** has occurred. When θ is approximately 1, each row of matrix **Y** tends to the constant vector with all elements almost equal to the average of the elements in the corresponding row of **A**. If any element in a row of **A** is non-zero, then all elements in the corresponding row of **Y** would be equal to the same non-zero value, instead of having some values zero and others clearly non-zero. Provided that some elements of the factors become zero after a single iteration, the elements update to non-zero in the subsequent iteration. Thus, the restriction of locking zero elements is avoided in the NNMA-modified multiplicative update rules.

Note that the parameter θ controls the degree of smoothness. When $\theta = 0$, the method corresponds to the basic NNMA multiplicative update rules. The parameter θ was defined as a small positive number. After running the algorithm, it was noted that adjustment of θ during iteration could speed up convergence.

3.2. Algorithm

We developed TWNNMA under the PARAFAC model and used the residual sum of squares as the objective function. The algorithm comprises the following steps:

1. Estimate the number of factor for three-way array $\underline{\mathbf{X}}$.
2. Initialize the factor matrices **A**, **B** and **C** under the non-negativity constraint (see Section 6).
3. Calculate the reconstructed three-way array $\underline{\mathbf{ABC}}$.
4. Calculate the new factor matrix **A** using Eq. (8a) and scale **A** to be columnwise normalized.
5. Calculate the new factor matrix **B** using Eq. (8b) and scale **B** to be columnwise normalized.
6. Calculate the new factor matrix **C** using Eq. (8c).
7. Calculate the residual sum of squares and the KKT residual norm (see Section 6).
8. Repeat steps 3–7 until the KKT residual norm converges to the stopping condition.

4. Experimental

4.1. Simulated fluorescence excitation–emission kinetic data sets

Consecutive reactions with a reversible step occur widely in chemical kinetics and examples include the Lindemann–Hinshelwood mechanism for unimolecular decay and the Michaelis–Menten mechanism of enzyme catalysis. We considered the reaction scheme for the reaction type:

$$v_1 A \underset{k_{-1}}{\overset{k_1}{\longleftrightarrow}} v_2 B \xrightarrow{k_2} C.$$

In this case the concentration matrix was generated using rate constants $k_1 = 10\ \mathrm{min}^{-1}$, $k_{-1} = 0\ \mathrm{s}^{-1}$ and $k_2 = 0.7\ \mathrm{min}^{-1}$. The reaction orders were $v_1 = 1$ and $v_2 = 1$. Pure-component simulated spectra of *A*, *B* and *C* were chosen to provide a fairly high degree of correlation and low selectivity, giving matrices $\mathbf{Ex}\,(30 \times 3)$ and $\mathbf{Em}\,(20 \times 3)$. Correlation coefficients between each pair of excitation spectra were [ex1, ex2] = 0.9612, [ex1, ex3] = 0.9189 and [ex2, ex3] = 0.9374. Correlation coefficients between each pair of emission spectra were [em1, em2] = 0.9312, [em1, em3] = 0.9780 and [em2, em3] = 0.9716. The pure-component simulated spectral data sets were normalized before coupling with concentration data to generate synthetic data sets.

TWNNMA was first tested on a synthetic noise-free data set (D1). Its performance was then investigated for synthetic data sets D2, D3 and D4 to which random Gaussian noise was added with a standard deviation of 0.02, 0.05 and 0.10 units of the mean signal, respectively.

4.2. Simulated data set for second order calibration

The data set published by Olivieri [50] was used as the simulated data set for second-order calibration. This comprises simulated fluorescence excitation–emission matrix data for a set of calibration and test samples containing two calibrated analytes and a single interferent. Ten calibration samples with a random design for the concentrations of both analytes were provided, along with ten test samples containing the three analytes in random concentrations. Random Gaussian noise with 0.03 units of standard deviation was added to all signals. The data set and the programs used in Ref. [50] are available via www.chemometry.com.

4.3. GC–MS metabolomic experiment data set

Duran et al. [51] placed *Medicago truncatula* seeds in a controlled greenhouse environment for 82 days at an average temperature of 28 °C, 40% relative humidity, and a day length of 16 h to study the impact of photosynthesis to the different *M. truncatula* tissue by GC–MS analysis. Analyses were performed on a 60 m DB-5MS column under Agilent Chemstation (Hewlett Packard 6890 gas

Table 1
Loss values (the sum of squared errors) relative to the value for a one-factor model, fit values in percentages and core consistencies (CORCO) versus the number of factors in an unconstrained PARAFAC model of the data sets D1, D2, D3 and D4.

N	D1		D2		D3		D4	
	CORCO	Loss	CORCO[a]	Loss	CORCO[a]	Loss	CORCO[a]	Loss
1	100	1.000	100	1.000	100	1.000	100	1.000
2	100	0.319	100	0.327	100	0.348	100	0.431
3	100	4.49×10^{-6}	100	0.008	99.9	0.045	97.2	0.156
4	2.21×10^{-17}	1.62×10^{-5}	85.4	0.008	45.5	0.045	32.7	0.153
5	–	5.80×10^{-6}	64.5	0.008	32.8	0.045	20.2	0.150
6	–	2.53×10^{-5}	48.8	0.008	0.2	0.045	9.0	0.150
7	–	–	23.8	0.008	–	–	–	–

[a] Core consistency is not stable when the number of factors is above three. Here the CORCO is the mean value of 10 runs.

chromatograph, 5973 mass selective detector, and 6890 series injector). Samples were analyzed by injecting 1 μl with a split injection ratio of 25:1. Separations were achieved using a temperature program with a mass scanning range of 50–650 *m/z*. *M. truncatula* roots, stems and leaves from ten replicate plants generated 30 GC–MS datasets in total. We rearranged the datasets to a three-way array with sample, chromatogram and mass modes and used the data to validate the TWNNMA method. The data set was downloaded from www.noble.org/plantbio/MS/downloads.html.

4.4. Software

All calculations were performed in the Octave environment (http://www.octave.org) using a 2.80-GHz Intel (R) Core 2 Quad Q9650 computer with Windows Server 2008 R2. The codes for the DTLD and core consistency diagnostic were from the Matlab N-way Toolbox [52]. The code for HOSVD was from TP Tool [53]. Other calculation codes were written by the authors.

5. Results

5.1. Kinetics analysis

Kinetics research is widely used to identify chemical reaction mechanisms in complex reaction system such as biochemical reaction networks. Multivariate curve resolution, which is traditionally used in kinetics, can be used to describe processes without explicit use of the underlying chemical model. In the present study, TWNNMA was used to decompose simulated kinetics directly without any information except for the native non-negative property.

TWNNMA requires a certain number of responsive factors in advance, so in this case the core consistency diagnostic and the residuals of the least squares fit of the three-way arrays were used to estimate the number of factors. These two parameters stabilize when the correct number of factors has been reached (see Section 6). Data sets D1, D2, D3 and D4 were analyzed and the results obtained for the core consistency diagnostic and loss function values are listed in Table 1. As shown, the correct number of factors should be three for data decomposition using TWNNMA.

Fig. 1 shows profiles of the three factors from data decomposition by TWNNMA for different noise levels. Fig. 1A shows resolved kinetic concentration profiles. Fig. 1B and C shows factors corresponding to simulated excitation and emission spectra, respectively. All the kinetic concentration profiles are fairly well resolved and the corresponding excitation and emission spectra are almost entirely overlapping. Because severe noise was added to dataset D4, the concentration profile obtained is slightly worse than for the other data sets.

Table 2 shows information related to the kinetics application for different noise levels. TWNNMA shows acceptable fits for data set D1, D2 and D3. For data set D4, the lack of fit obtained is slightly worse. Nevertheless, the square of the correlation coefficient between actual and resolved concentrations and excitation and emission spectra indicates acceptable goodness of fit.

5.2. Quantitative determination

As indicated in Ref. [50], the number of factors was set to three. A separate component profiles in each dimension was then produced after TWNNMA decomposition of the three-way data set comprising calibration and test sample data. Because TWNNMA decomposition only provides relative values, the analyte concentrations in the prediction samples were obtained after calibration was performed in a procedure similar to a calibration curve for one component against standard concentrations. Predicted concentrations for both analytes, the root mean square error of prediction, and figures of merit such as sensitivity, selectivity and limit of detection [54–56] are listed in Table 3. PARAFAC, ATLD, APTLD, SWATLD, BLLS/RBL and TWNNMA show similar sensitivity and selectivity in this particular example. TWNNMA is quite precise according to the root mean square error of prediction. Moreover, TWNNMA have a low limit of detection.

5.3. Analysis of GC–MS data

Non-processed MS files from GC–MS analysis were converted in NetCDF format to Octave. After data preprocessing such as chromatogram alignment and baseline correction, retention time from 26.03 min to 26.17 min (Fig. 2A) was selected for study. In contrast with only phosphoric acid (91.2% match) recognized by searching in the NIST 2005 mass spectral library directly, two components eluted successively and were identified as phosphoric acid (NIST 05, 92.7% match) and glycerol (NIST 05, 91.8% match) respectively when TWNNMA was used (Fig. 2B and C). Fig. 2D demonstrates the perceptible change of relative concentration of two components in *M. truncatula* roots, stems and leaves. The change trends also reflect the influence of photosynthesis to different *M. truncatula* tissue.

6. Discussion

6.1. Determining the number of factors

There are three main ways to determine the correct number of factors in a PARAFAC model: (1) split-half analysis, (2) judging of residuals, and (3) comparison with external knowledge of the data being modeled. Among these techniques, the core consistency diagnostic [57], which reflects the resemblance between Tucker3 core fits and the super-diagonal PARAFAC core and suggests whether a PARAFAC model is a valid model for a three-way array, has been commonly applied in estimating the number of factors [4–7].

The core consistency diagnostic is an effective tool for determining the appropriate number of factors in a PARAFAC model. However, it has been shown using purely synthetic data that the core consistency diagnostic is not yet complete [50]. The residu-

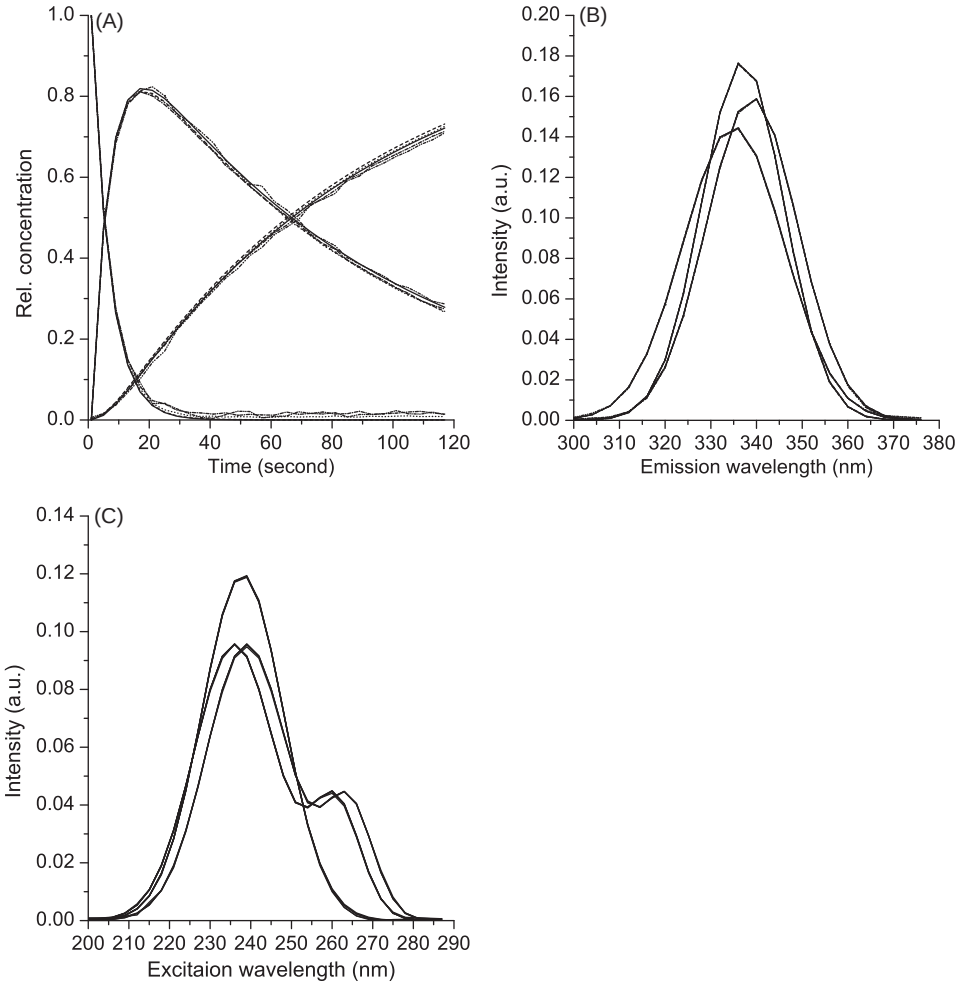

Fig. 1. Kinetic profiles (A) and fluorescence excitation (B) and emission (C) spectra recovered for simulated kinetics data set D1 (dash line), D2 (dot line), D3 (dash dot line) and D4 (dash dot dot line). Solid lines are the true profiles.

als for least squares fits of three-way arrays to the trilinear model is another important parameter that stabilizes when the correct number of factors has been reached. In the present study, the core consistency diagnostic did not always allow selection of the correct number of factors, but analysis of the residual fit led to an appropriate number of factors in the two applications. Therefore, we suggest that analysis of the residual fit should be used as an alternative approach to the core consistency diagnostic.

6.2. Initialization

NNMA algorithms are sensitive to the initialization procedure. Most algorithms use simple random initialization. Wild et al. showed that a centroid initialization built from centroid decomposition was a much better alternative to random initialization [58]. Langville recommended the use of results from a fast ALS-type NNMA algorithm for initialization [48]. In all cases, good initialization can improve the convergence speed and solution accuracy of many NNMA algorithms. Moreover, good initialization can avoid some convergence problems.

In the present study, several initialization methods were proposed. Besides random initialization, HOSVD initialization and DTLD initialization were investigated. To ensure non-negativity, absolute values of the loading vectors obtained from HOSVD and DTLD were used as starting values in TWNNMA. As observed from Fig. 3, HOSVD initialization converged faster than random initialization and the residual sum of squares and KKT residual norm (see Section 6.3) for DTLD initializa-

Table 2
Lack of fit and the square of the correlation coefficient between actual and resolved profiles for the analytes (indicated as 1, 2 and 3) using TWNNMA.

Data set	Noise level[a]	%Lack of fit[b]	Concentration			Emission spectra			Excitation spectra		
			1	2	3	1	2	3	1	2	3
D1	0	0.33	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
D2	0.02	1.28	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
D3	0.05	3.15	0.999	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000
D4	0.10	6.18	0.998	0.999	0.999	1.000	1.000	1.000	1.000	0.998	1.000

[a] Random Gaussian noise was added to data sets D2, D3, D4 with a standard deviation of 0.02, 0.05 and 0.10 units of the mean signal, respectively.

[b] Lack of fit $(\%) = 100 \times \sqrt{\sum_{i,j,k} r_{ijk}^2 / \sum_{i,j,k} d_{ijk}^2}$ where r_{ijk} are residuals and d_{ijk} are the elements of the raw three-way data set.

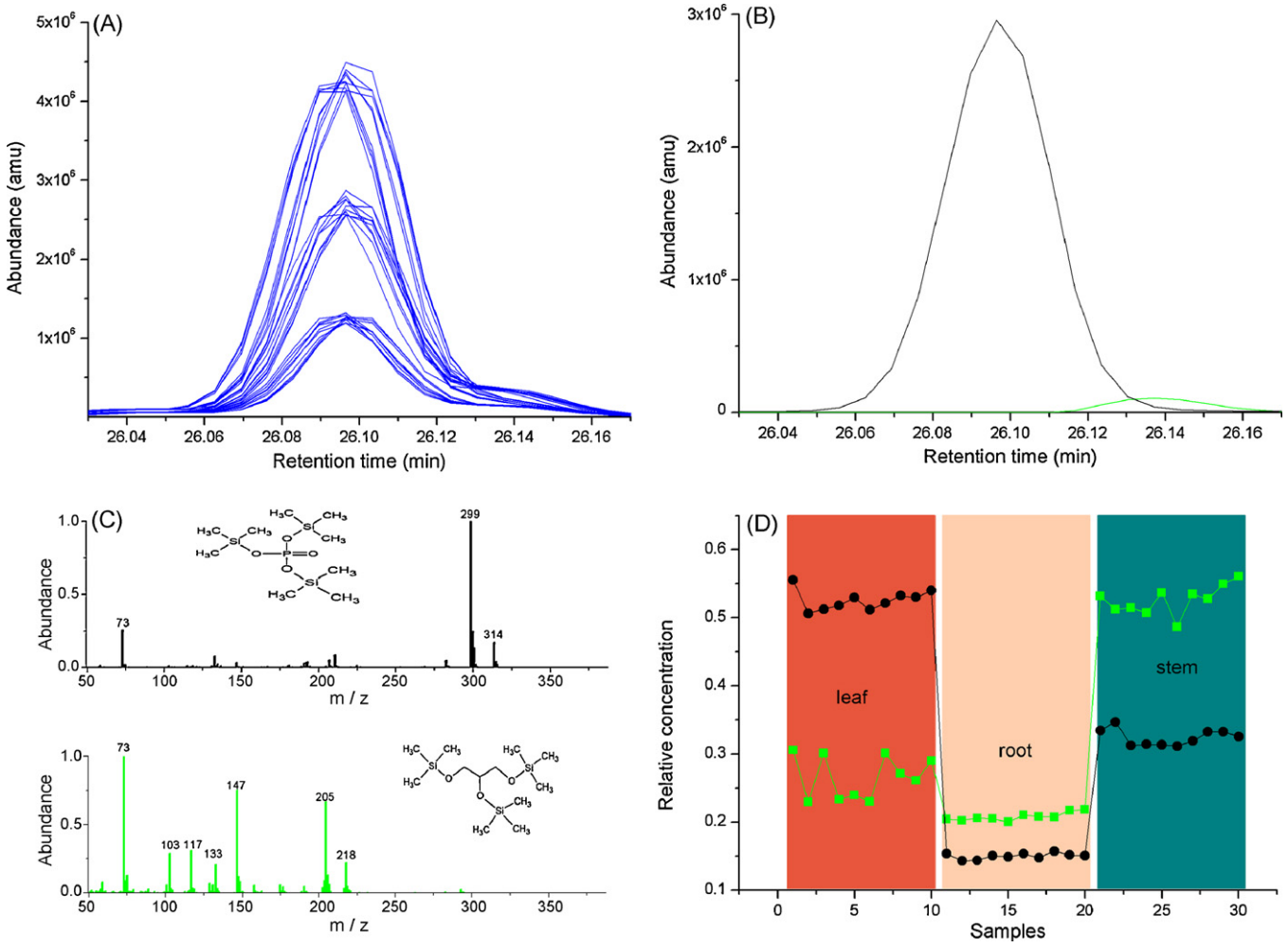

Fig. 2. (A) The total ion chromatogram of *Medicago truncatula* roots, stems and leaves (retention time from 26.02 min to 26.17 min) and TWNNMA results: (B) chromatographic loadings; (C) mass-spectrum loading (derivatives); and, (D) concentration loading for leaves (sample nos. 1–10), roots (sample nos. 11–20) and stems (sample nos. 21–30). Note that the color black and green refer to component 1 and component 2 respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

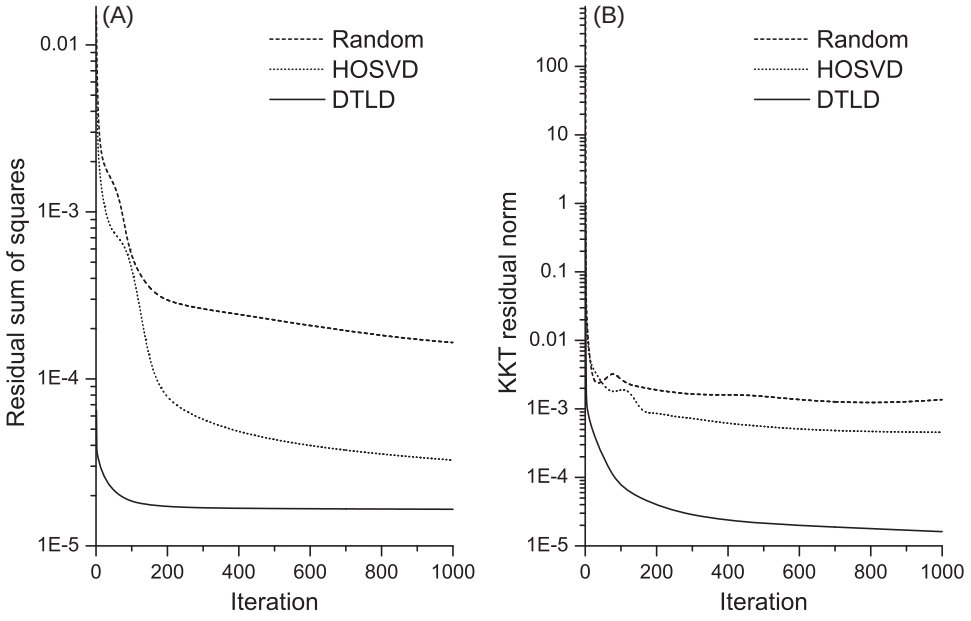

Fig. 3. Convergence graph comparing different initialization methods for data set D2. The solid line, dot line and dash line denote the change of SSQ (A) and KKT norm (B) with the initialization of DTLD, HOSVD and random, respectively.

Table 3
Results for TWNNMA and existing second-order calibration methods.

Test sample	Nominal[a]	PARAFAC[a]	ATLD[a]	APTLD[a]	SWATLD[a]	BLLS/RBL[a]	U-PLS/RBL[a]	N-PLS/RBL[a]	TWNNMA
Analyte 1									
1	2.95	2.97	2.91	2.95	2.95	2.94	2.94	2.96	2.96
2	3.05	3.07	3.08	3.09	3.09	3.06	3.06	3.09	3.05
3	1.04	1.05	1.12	1.05	1.05	1.06	1.06	1.07	1.04
4	2.39	2.41	2.52	2.42	2.42	2.37	2.36	2.40	2.40
5	2.39	2.46	2.40	2.46	2.46	2.44	2.45	2.46	2.44
6	2.34	2.36	2.47	2.37	2.37	2.39	2.39	2.41	2.35
7	2.77	2.81	2.72	2.79	2.79	2.75	2.75	2.75	2.79
8	1.15	1.20	1.14	1.22	1.22	1.19	1.19	1.21	1.18
9	2.34	2.35	2.52	2.39	2.39	2.38	2.38	2.40	2.34
10	3.46	3.47	3.40	3.44	3.44	3.41	3.41	3.43	3.46
RMSEP[b]		0.03	0.09	0.04	0.04	0.03	0.04	0.05	0.02
SEN[c]		0.68	0.68	0.68	0.68	0.68	0.68	–	0.68
SEL[d]		0.20	0.20	0.20	0.20	0.19	–	–	0.20
LOD[e]		0.20	0.42	0.21	0.21	0.15	0.15	–	0.09
Analyte 2									
1	2.99	3.03	3.24	2.97	2.97	3.06	3.00	3.04	2.98
2	3.66	3.75	3.90	3.69	3.69	3.87	3.70	3.86	3.69
3	1.43	1.51	1.36	1.45	1.45	1.52	1.45	1.51	1.48
4	3.60	3.72	3.33	3.64	3.64	3.62	3.46	3.62	3.68
5	1.76	1.86	1.91	1.78	1.78	1.77	1.78	1.77	1.80
6	3.64	3.68	3.57	3.67	3.67	3.70	3.58	3.68	3.66
7	1.35	1.45	1.28	1.38	1.38	1.32	1.35	1.29	1.39
8	2.19	2.27	2.48	2.29	2.29	2.35	2.28	2.35	2.22
9	2.57	2.63	2.28	2.56	2.56	2.60	2.53	2.59	2.59
10	3.30	3.37	3.22	3.32	3.32	3.17	3.14	3.15	3.33
RMSEP[b]		0.08	0.20	0.04	0.04	0.10	0.08	0.10	0.04
SEN[c]		0.27	0.26	0.27	0.27	0.28	0.28	–	0.27
SEL[d]		0.17	0.18	0.17	0.17	0.18	–	–	0.21
LOD[e]		0.51	1.10	0.54	0.54	0.40	0.40	–	0.20

[a] The results for PARAFAC, ATLD, APTLD, SWATLD, BLLS/RBL, U-PLS/RBL, N-PLS/RBL were obtained by Olivieri et al. [50].

[b] The root mean square errors of prediction is determined as: RMSEP $= \left[\sum (y_i - \widehat{y}_i)^2/N\right]^{1/2}$, where N is the sample size, y_i and $\widehat{y}_i$ are the actual and predicted concentrations respectively.

[c] The sensitivity (SEN) is determined as: $\mathrm{SEN}_n = k_n \left\{ \left[(B_{sus}^T P_{b,uns} B_{sus}) \times (C_{sus}^T P_{c,uns} C_{sus}) \right]^{-1} \right\}_{nn}^{-1/2}$.

[d] The selectivity (SEL) is obtained by dividing SEN by k_n.

[e] Limit of detection (LOD) was calculated as: $\mathrm{LOD} = \delta(\alpha, \beta) \times \left[\sqrt{\sum (C_i - \widehat{C}_i)^2/n - 2}/k \right] \times \sqrt{1 + 1/n + \bar{x}^2/\sum (x_i - \bar{x})^2}$.

tion decreased much faster than for the other two initializations methods.

Random initialization is simple to implement but DTLD initialization can greatly speed up TWNNMA convergence. Thus, we recommend DTLD for TWNNMA initialization.

6.3. Stopping condition

Several implementations of the NNMA multiplicative update rules use a maximum number of iterations as the stopping criterion. However, setting of a fixed number is not a mathematical method for controlling the number of iterations executed because the most appropriate value for maximum iteration is problem-dependent. Some researchers checked the difference between recent iterations. If the difference is small enough, then the procedure stops. However, such stopping conditions do not reveal whether a solution is close to a stationary point or not. Langville et al. [48] proposed an angular convergence measure. Once an angle is less than a certain threshold, the algorithm stops because the factors have converged satisfactorily. Lin [49] proposed a stopping condition that simultaneously checks for stationarity at each iteration, and fits nicely into his projected gradient algorithm. Moreover, Lin proved that NNMA multiplicative update rules converge to a stationary point [59].

Here we extend the Lin stopping condition to TWNNMA under the Karush–Kuhn–Tucker (KKT) optimality condition [60]. Thus, the matrices $\mathbf{A}^{k+1}$, $\mathbf{B}^{k+1}$ and $\mathbf{C}^{k+1}$ resulting from the iterative procedures should satisfy:

$$\|\nabla_A f(\mathbf{A}^{k+1}, \mathbf{B}^k, \mathbf{C}^k)\|_F \le \overline{\varepsilon_A}, \tag{10a}$$

$$\|\nabla_B f(\mathbf{A}^{k+1}, \mathbf{B}^{k+1}, \mathbf{C}^k)\|_F \le \overline{\varepsilon_B}, \tag{10b}$$

$$\|\nabla_C f(\mathbf{A}^{k+1}, \mathbf{B}^{k+1}, \mathbf{C}^{k+1})\|_F \le \overline{\varepsilon_C}, \tag{10c}$$

where $\nabla f(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is defined as:

$$\nabla_A f(\mathbf{A}, \mathbf{B}, \mathbf{C}) = (\underline{\mathbf{X}}_{(1)} \bullet \bullet \underline{\mathbf{BC}} - \underline{\mathbf{ABC}}_{(1)} \bullet \bullet \underline{\mathbf{BC}}), \tag{11a}$$

$$\nabla_B f(\mathbf{A}, \mathbf{B}, \mathbf{C}) = (\underline{\mathbf{X}}_{(2)} \bullet \bullet \underline{\mathbf{AC}} - \underline{\mathbf{ABC}}_{(2)} \bullet \bullet \underline{\mathbf{AC}}), \tag{11b}$$

$$\nabla_C f(\mathbf{A}, \mathbf{B}, \mathbf{C}) = (\underline{\mathbf{X}}_{(3)} \bullet \bullet \underline{\mathbf{AB}} - \underline{\mathbf{ABC}}_{(3)} \bullet \bullet \underline{\mathbf{AB}}). \tag{11c}$$

We set

$$\overline{\varepsilon_A} = \overline{\varepsilon_B} = \overline{\varepsilon_C} = \max(10^{-3}, \varepsilon)\|\nabla f\left(\mathbf{A}^1, \mathbf{B}^1, \mathbf{C}^1\right)\|_F, \tag{12}$$

at the start, where ε is the tolerance. For ease of expression, we define the KKT residual norm as the sum of the left-hand-side of Eqs. (10a)–(10c).

If the KKT residual norm is not greater than the right-hand-side of Eq. (12), then the iterative procedure for TWNNMA stops. We set tolerance 10^{-3} in the three cases of present study. Implement of TWNNMA showed that KKT residual norm achieved stop condition after a few iterations (<100).

7. Conclusion

A novel trilinear decomposition method, TWNNMA, was proposed. The method resolves the main drawbacks of most algorithms used for NNMA. Three applications in chemical kinetics, second-order calibration and analysis of GC–MS data demonstrated that TWNNMA would be a promising resolution method for complex systems. Determination of the correct number of factors for three-way arrays was also investigated. The results suggest that analysis of the residual fit can be used as an alternative approach to the core consistency diagnostic. DTLD initialization can speed up the convergence of TWNNMA greatly. We also propose a stopping condition that reveals whether the TWNNMA solution is close to a stationary point or not. TWNNMA, which decomposes three-way arrays directly without unfolding, provides a new decomposition method for three-way arrays. This decomposition technique will contribute to the development of resolution methods for three-way arrays.

Acknowledgements

The authors would like to acknowledge financial supports by the National Natural Science Foundation of China (20675057 and 20705024).

References

[1] K.S. Booksh, B.R. Kowalski, Anal. Chem. 66 (1994) 782–791.
[2] A.C. Olivieri, Anal. Chem. 80 (2008) 5713–5720.
[3] M.M. Sena, M.G. Trevisan, R.J. Poppi, Talanta 68 (2006) 1707–1712.
[4] R.S. Valverde, M.D.G. Garcia, M.M. Galera, H.C. Goicoechea, Talanta 70 (2006) 774–783.
[5] D.B. Gil, A.M. de la Peña, J.A. Arancibia, G.M. Escandar, A.C. Olivieri, Anal. Chem. 78 (2006) 8051–8058.
[6] S.H. Zhu, H.L. Wu, A.L. Xia, Q.J. Han, Y. Zhang, R.Q. Yu, Talanta 74 (2008) 1579–1585.
[7] B. Hemmateenejad, Z. Rezaei, S. Zaeri, Talanta 79 (2009) 648–656.
[8] G.M. Escandar, N.M. Faber, H.C. Goicoechea, A. Muñoz de la Peña, A.C. Olivieri, R.J. Poppi, Trends Anal. Chem. 26 (2007) 752–765.
[9] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149–171.
[10] E. Sanchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496–499.
[11] E. Sanchez, B.R. Kowalski, J. Chemom. 4 (1990) 29–45.
[12] R. Tauler, Chemom. Intell. Lab. Syst. 30 (1995) 133–146.
[13] H.L. Wu, M. Shibukawa, K. Oguma, J. Chemom. 12 (1998) 1–26.
[14] Z.P. Chen, H.L. Wu, J.H. Jiang, Y. Li, R.Q. Yu, Chemom. Intell. Lab. Syst. 52 (2000) 75–86.
[15] A.L. Xia, H.L. Wu, D.M. Fang, Y.J. Ding, L.Q. Hu, R.Q. Yu, J. Chemom. 19 (2005) 65–76.
[16] M. Linder, R. Sundberg, Chemom. Intell. Lab. Syst. 42 (1998) 159–178.
[17] S. Wold, P. Geladi, K. Esbensen, J. Øhman, J. Chemom. 1 (1987) 41–56.
[18] R. Bro, J. Chemom. 10 (1996) 47–62.
[19] J. Ohman, P. Geladi, S. Wold, J. Chemom. 4 (1990) 79–90.
[20] D.D. Lee, H.S. Seung, Nature 401 (1999) 788–791.
[21] P. Paatero, U. Tapper, Environmetrics 5 (1994) 111–126.
[22] R. Bro, S.D. Jong, J. Chemom. 11 (1997) 393–401.
[23] P. Sajda, S. Du, T.R. Brown, IEEE Trans. Med. Imag. 23 (2004) 1453–1465.
[24] P.M. Alberto, J.M. Carazo, K. Kochi, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 403–415.
[25] B. Gokberk, H. Dutagaci, A. Ulas, IEEE Trans. Syst. Man Cybern. B: Cybern. 38 (2008) 155–173.
[26] W. Xu, X. Liu, Y. Gong, Document clustering based on non-negative matrix factorization, in: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, 2003, pp. 267–273.
[27] F. Shahnaz, M.W. Berry, V.P. Pauca, R.J. Plemmons, Inf. Process. Manage. 42 (2006) 373–386.
[28] M. Chagoyen, P. Carmona-Sqez, H. Shatkay, J.M. Carazo, A. Pascual-Montano, BMC Bioinformatics 7 (2006) 41–60.
[29] B. Xu, J. Lu, G. Huang, A constrained non-negative matrix factorization in information retrieval, in: Proceedings of the IEEE International Conference on Information Reuse and Integration, Las Vegas, 2003, pp. 273–277.
[30] K. Devarajan, PLoS Comput. Biol. 4 (2008) e1000029.
[31] J.P. Brunet, P. Tamayo, T.R. Golub, J.P. Mesirov, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 4164–4169.
[32] H. Kim, H. Park, Bioinformatics 23 (2007) 1495–1502.
[33] H.T. Gao, T.H. Li, K. Chen, W.G. Li, X. Bi, Talanta 66 (2005) 65–73.
[34] S.A. Astakhov, H. Stogbauer, A. Kraskov, P. Grassberger, Anal. Chem. 78 (2006) 1620–1627.
[35] F. Guimet, R. Boque, J. Ferre, Chemom. Intell. Lab. Syst. 81 (2006) 94–106.
[36] I. Kopriva, I. Jeric, A. Cichocki, Chemom. Intell. Lab. Syst. 97 (2009) 170–178.
[37] J.C. Hoggard, J.H. Wahl, R.E. Synovec, G.M. Mong, C.G. Fraga, Anal. Chem. 82 (2010) 689–698.
[38] L.H. Lim, P. Comon, J. Chemom. 23 (2009) 432–441.
[39] M. Welling, M. Weber, Pattern Recogn. Lett. 22 (2001) 1255–1261.
[40] D. FitzGerald, M. Cranitch, E. Coyle, Non-negative tensor factorization for sound source separation, in: Proceedings of Irish Signals and Systems Conference, Dublin, 2005, pp. 8–12.
[41] A. Cichocki, R. Zdunek, S. Choi, Non-negative tensor factorization using alpha and beta divergences, in: International Conference on Acoustics, Speech, and Signal Processing, Honolulu, Hawaii, 2007, pp. 1393–1396.
[42] M Mørup, L.K. Hansen, S.M. Arnfred, Neural Comput. 20 (2008) 2112–2131.
[43] D. Donoho, V. Stodden, in: S. Thrun, L. Saul, B. Schölkopf (Eds.), Advances in Neural Information Processing Systems, vol. 16, MIT Press, Cambridge, Massachusetts, 2004, pp. 1141–1149.
[44] M.W. Berry, M. Browne, A.N. Langville, Comput. Stat. Data Anal. 52 (2007) 155–173.
[45] P.O. Hoyer, J. Mach. Learn. Res. 5 (2004) 1457–1469.
[46] E.F. Gonzalez, Y. Zhang, Accelerating the Lee–Seung algorithm for non-negative matrix factorization, Technical report TR-05-02, Department of Computational and Applied Mathematics, Rice University, 2005.
[47] N. Rao, S.J. Shepherd, Extracting characteristic patterns from genome – wide expression data by non-negative matrix factorization, in: Proceedings of the IEEE Computational Systems Bioinformatics Conference, Stanford, 2004, pp. 570–571.
[48] A.N. Langville, C.D. Meyer, R. Albright, J. Cox, D. Duling, Algorithms, initializations, and convergence for the nonnegative matrix factorization, Technical Report Math 81706, North Carolina State University, 2006.
[49] C.J. Lin, Neural Comput. 19 (2007) 2756–2779.
[50] A.C. Olivieri, H.L. Wu, R.Q. Yu, Chemom. Intell. Lab. Syst. 96 (2009) 246–251.
[51] A.L. Duran, J. Yang, L. Wang, L.W. Sumner, Bioinformatics 19 (2003) 2283–2293.
[52] C.A. Andersson, R. Bro, Chemom. Intell. Lab. Syst. 52 (2000) 1–4.
[53] S. Nagy, Z. petres, P. Baranyi, TP Tool – a MATLAB toolbox for TP model transformation, in: 8th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics, 2007, pp. 483–495.
[54] A. Garcia-Reiriz, P.C. Damiani, A.C. Olivieri, Talanta 71 (2007) 806–815.
[55] L.A. Currie, Pure Appl. Chem. 67 (1995) 1699–1723.
[56] A.C. Olivieri, Anal. Chem. 77 (2005) 4936–4946.
[57] R. Bro, H.A.L. Kiers, J. Chemom. 17 (2003) 274–286.
[58] S. Wild, J. Curry, A. Dougherty, Pattern Recognit. 37 (2004) 2217–2232.
[59] C.J. Lin, IEEE Trans. Neural Netw. 18 (2007) 1589–1596.
[60] D.P. Bertsekas, Nonlinear Programming, second ed., Athena Scientific, Belmont, 1999.

Glossary

ALS: alternating least squares.
APTLD: alternating penalty trilinear decomposition.
ATLD: alternating trilinear decomposition.
BLLS: bilinear least squares.
CANDECOMP: canonical decomposition.
CORCO: core consistencies.
DTLD: direct trilinear decomposition.
GRAM: generalized rank annihilation.
HOSVD: higher order singular value decomposition.
KKT: Karush–Kuhn–Tucker.
LOD: limit of detection.
MCR–ALS: multivariate curve resolution coupled to alternating least squares.
N-PLS: multi-way partial least squares.
NNLS: non-negative least squares.
NNMA: non-negative matrix approximation.
NTF: non-negative tensor factorization.
NTD: non-negative tucker decomposition.
PARAFAC: parallel factor analysis.
PMF: positive matrix factorization.
PTF: positive tensor factorization.
RBL: residual bilinearization.
RMSEP: root mean square errors of prediction.
SWATLD: self-weighted alternating trilinear decomposition.
SEL: selectivity.
SEN: sensitivity.
SSQ: the residual sum of squares.
TWNNMA: three-way non-negative matrix approximation.
UPLS: unfolded-partial least squares.