ELSEVIER

Chemometrics and Intelligent Laboratory Systems 60 (2002) 239–251

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

Three-way alternating least squares using three-dimensional tensors in MATLAB®

Ernst Bezemer, Sarah C. Rutan *

Department of Chemistry, Virginia Commonwealth University, PO Box 842006, Richmond, VA 23284-2006, USA

Abstract

This paper describes an improved three-way alternating least-squares multivariate curve resolution algorithm that makes use of the recently introduced multi-dimensional arrays of MATLAB®. Multi-dimensional arrays allow for a convenient way to apply chemically sound constraints, such as closure, in the third dimension. The program is designed for kinetic studies on liquid chromatography with diode array detection but can be used for other three-way data analysis. The program is tested with a large number of synthetic data sets and its flexibility is demonstrated, especially when non-trilinear data sets are fit. In this case, the algorithm finds a solution with a better fit than direct trilinear decomposition (DTD). When trilinear data are used, the optimal fit is not as good as when a direct decomposition method is used. Most real data sets, however, have some degree of non-trilinearity. This makes this method a better choice to analyze non-trilinear, three-way data than direct trilinear decomposition. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Multivariate curve resolution; Alternating least squares; Two-way data analysis; Three-way data analysis; Kinetics; Trilinearity; Closure

1. Introduction

Recent analytical hardware developments have resulted in commonly available hyphenated instruments, such as GC-MS and HPLC-DAD, which give rise to very large data sets. The information of interest is the retention profiles and the spectra corresponding to each of the pure components, allowing for subsequent qualitative or quantitative analysis depending on the goals of the investigator. Additional data are obtained when investigating reaction kinetics, where samples are studied at regular time intervals to observe concentration differences in the various reactive species. In a kinetic study, the retention profiles and pure component spectra remain the same in theory, while only the concentrations of the species change. Thus, besides the retention profiles and the pure component spectra, a third form of information can be abstracted, the kinetic profiles.

These multi-dimensional data sets are often useful when peak overlap occurs. One or more peaks of interest might coelute with other compounds or baseline effects and noise might prevent accurate integration of a signal. Chemometric methods allow for isolation of a peak out of a complex background mathematically, increasing the accuracy of the determination of the concentration [1]. Second-order information given by a diode array detector, in the case of

* Corresponding author. Tel.: +1-804-828-7517; fax: +1-804-828-8599.

E-mail address: srutan@saturn.vcu.edu (S.C. Rutan).

0169-7439/02/$ - see front matter
PII: S0169-7439(01)00199-X

liquid chromatography, or by mass spectral detection with gas chromatography, can reduce the ambiguity in the resolution results [2,3]. When observing kinetic behavior, there is additional information in the third order, which can further improve the accuracy of the results.

The current chemometric algorithms that use three-way data sets stack the two-way data sets together resulting in a large two-dimensional data matrix, even when a three-way model is applied [4–7]. This results in unnatural formatting of the data set and can be avoided. The algorithm described here makes use of three-dimensional data tensors, resulting in a true three-way alternating least-squares algorithm. This has only become possible with MATLAB® version 5 or higher, which for the first time includes the capability of faithfully representing multi-dimensional arrays [8].

With the use of a true three-way data representation, flexible implementation of various constraints (i.e. non-negativity, closure, unimodality, trilinearity) is facilitated, in any of the three data directions.

2. Theory

The literature on multi-way analysis is not consistent in the notation of variables of different orders. The notation used in this paper is as follows: Bold outlined letters ($\mathbb{D}$) are used for three-dimensional tensors. Bold capital letters (**C**) are used for matrices. Small bold letters (**p**) are used for vectors. Capital italic letters (K) are used for the size of a dimension and the total number of measurements. Small italic letters (k) are used for scalars. Small italic letters with subscripts (d_{rsk}) are used to indicate matrix or tensor elements.

A two-way bilinear data matrix, obtained from a hyphenated instrument, can be described as the inner product of the concentration profiles and the pure component spectra according to Eq. (1) [9],

$$\mathbf{D} = \mathbf{C} \cdot \mathbf{S}^{\mathrm{T}} + \mathbf{E} \tag{1}$$

where **D** ($T \times S$) is the data matrix, **C** ($T \times N$) consists of the concentration profiles for each of the pure components, **S** ($S \times N$) are the pure component spectra and **E** ($T \times S$) is the error matrix. The variables T, S and N represent the number of chromatographic time points, the number of measured wavelengths and the number of components, respectively.

To decompose the data matrix via an alternating least-squares method, one of the profile matrices has to be estimated to provide an initial guess for the algorithm. This is where most of the various chemometric algorithms differ [10]. Some algorithms use a form of principal component analyses [11–16], while others have used a variety of other mathematical tools to generate the initial guesses, ranging from orthogonal projections to Kalman filtering [17–22]. With one set of pure component profiles estimated (the concentration profiles in this example), the second set of profiles can be calculated using Eq. (2).

$$\mathbf{S} = (\mathbf{C}^{\mathrm{T}} \cdot \mathbf{C})^{-1} \cdot \mathbf{C}^{\mathrm{T}} \cdot \mathbf{D} \tag{2}$$

This process can be constrained and iterated upon until a set minimal improvement of the fit error is reached [23].

A three-way kinetic trilinear data set $\mathbb{D}$ ($T \times S \times K$) can be described as the inner product of the retention profiles **R** ($T \times N$), the pure component spectra **S** ($S \times N$) and the kinetic profiles **K** ($K \times N$) according to Eq. (3) [24],

$$\mathbb{D} = \mathbf{R} \otimes \mathbf{S} \otimes \mathbf{K} + \mathbb{E} \tag{3}$$

where $\mathbb{E}$ is the error tensor, K is the number of kinetic time points and N is the number of components. Fig. 1 shows a graphical representation of the tensor multi-

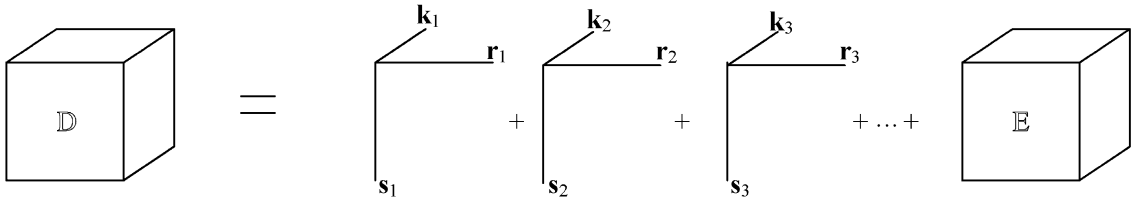

Fig. 1. Tensor multiplication in the trilinear model.

plication and is known as the trilinear model. In the ideal case, the data tensor can be unambiguously decomposed into the corresponding matrices using direct trilinear decomposition (DTD) or a generalized rank annihilation method (GRAM) [25,26]. However, for real data, where there are retention time shifts, intermolecular interactions and other effects that cause the data set to be non-trilinear, this direct approach cannot be used and an iterative method has to be employed.

There are several multi-way data analysis techniques such as PARAFAC, Tucker3 and N-PLS [4–6]. All these decomposition methods are based on principal component analysis of the data set. PARAFAC is a constrained form of Tucker3 and Tucker3 is a constrained form of N-PLS [27]. These three algorithms operate by regression of the dependent variables onto the scores and have the core tensor $\mathbb{H}$ ($N_1 \times N_2 \times N_3$) in common as shown in Eq. (4).

$$\mathbb{D} = \mathbf{R} \otimes \mathbb{H} \otimes (\mathbf{S}^{\mathrm{T}} \cdot \mathbf{K}^{\mathrm{T}}) + \mathbb{E} \tag{4}$$

The core matrix may have a different number of components in each direction but can be constrained to a super identity tensor resulting in the PARAFAC model and a trilinear solution.

The algorithm described in this paper is an alternating least-squares method, which is a multivariate curve resolution technique. It does not use a core matrix as found in PARAFAC and Tucker3, but the data are dissected as shown in Eq. (5),

$$\mathbb{D} = \mathbb{R} \otimes \mathbb{S} \otimes \mathbf{K} + \mathbb{E} \tag{5}$$

where $\mathbb{D}$ ($T \times S \times K$) is the data tensor, $\mathbb{E}$ ($T \times S \times K$) is the error, $\mathbb{R}$ ($T \times N \times K$) contains the retention profiles, $\mathbb{S}$ ($S \times N \times K$) contains the spectral profiles and $\mathbf{K}$ ($K \times N$) contains the kinetic profiles. This model still assumes bilinearity regarding the spectral and retention profiles so that the spectrum of each component is the same at each chromatographic time point. In the trilinear case, the retention and spectral profiles for each component are identical in all kinetic experiments and the model becomes identical to Eq. (3).

Unlike the earlier mentioned multi-way techniques, this algorithm is not based on principal component analysis even though the initial guesses can be generated that way. It uses the data structure and chemically relevant constraints to find a chemically reasonable solution. The choice of the number of components is not as crucial for this method as it is in PARAFAC, Tucker3, N-PLS or DTD models. This is an advantage in kinetic studies where not all components are present at the same time, since intermediates and products might only appear long after the starting materials are gone. In this work, we have used principal component analysis techniques such as singular value decomposition (SVD) and evolving factor analysis (EFA) to generate the initial guesses. These methods have been described earlier [28,29]. The MCR-ALS algorithm by Gargallo et al. [7] models only the predetermined number of components by working with a data matrix reproduced from the scores and loading matrices. This ignores information that might be hidden in a higher number of principal components and can be a potential limitation when using non-trilinear data since in that case, the rank of the data tensor can be higher than the number of real components. In the present work, instead of PCA compression, an extra unconstrained component can be added to account for the baseline and noise of the measurements.

To best explain the functionality of the algorithm, an example of kinetic measurements obtained using a LC-DAD is used. The data set is organized in a three-dimensional cube where each slice of the cube is a chromatogram with a spectral profile at each time point and each slice is a consecutive injection of an aliquot of the evolving chemical system.

A block diagram of the algorithm is shown in Fig. 2. As in two-way ALS, in step 1, the data tensor ($\mathbb{D}$) is divided (slice by slice) by each slice in the initial guess tensor ($\mathbb{C}_{\mathrm{initial}}$) (or multiplied by the pseudo-inverse of the initial guess tensor) to obtain the spectral profile tensor ($\mathbb{S}_{\mathrm{calc}}$). Each of the K slices of the spectral profile tensor is a ($S \times N$) matrix with estimates for the pure component spectra on each row. Thus, each component can have a unique spectral profile for each kinetic measurement. This would apply, for example, when molecular interactions distort the fluorescence spectrum [30].

After applying the appropriate constraints in steps 2 and 3, the data tensor is divided in step 4 by the newly calculated spectral tensor resulting in the concentration profile tensor ($\mathbb{C}_{\mathrm{calc}}$). As in the case for the spectral

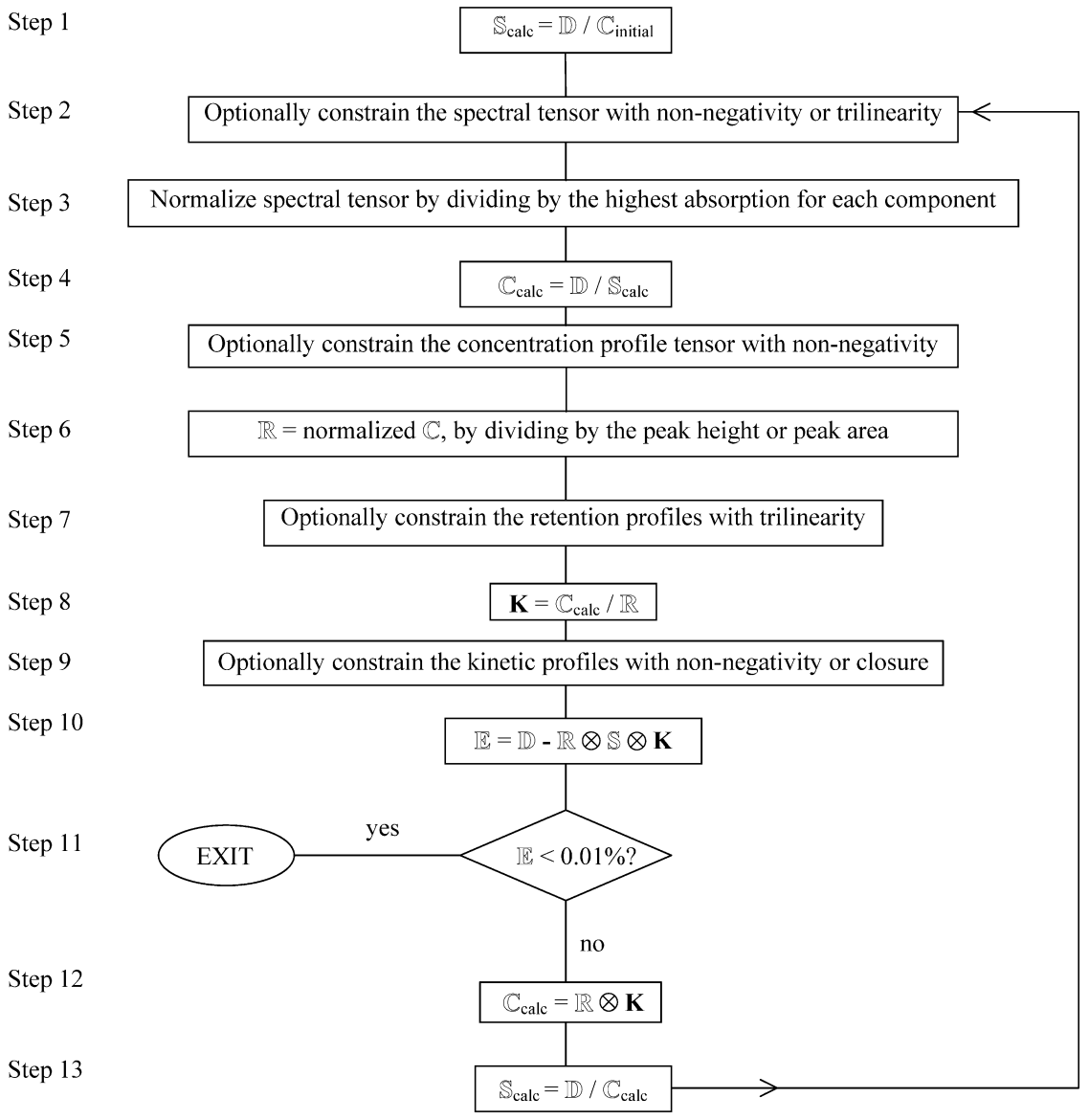

Fig. 2. Block diagram of the three-way MCR-ALS program.

tensor, each component is allowed to have a unique chromatographic profile at each kinetic point. This occurs fairly commonly in chromatography where retention time or peak shape may change due to column aging or temperature variations. This is a related approach to that found in PARAFAC2 [4]. However, to take advantage of the third-order information, closure and trilinearity constraints are used here. Without these constraints, the data slices are independent and the algorithm performs two-way ALS on each of the slices. The ability of this algorithm to provide the option of being able to relax the constraints on three-way data from complete trilinearity to bilinearity assists in dealing with non-trilinear data.

Gargallo et al. [7] implemented a different approach in the MCR-ALS algorithm to specifically

handle peak shifts. The peak maxima are sought out and aligned. Alternatively, the trilinearity constraint is relaxed to accommodate non-trilinear data.

An important assumption made is that it is possible to isolate the retention profiles $\mathbb{R}$ ($T \times N \times K$) by rearranging Eq. (5) to Eq. (6),

$$\mathbb{D} = \mathbb{C} \otimes \mathbb{S} + \mathbb{E} \qquad \text{with } \mathbb{C} = \mathbb{R} \otimes \mathbf{K} \tag{6}$$

where $\mathbb{C}$ ($T \times N \times K$) contains the concentration profiles and $\mathbf{K}$ ($K \times N$) is a matrix containing the concentrations of each species at each of the sample times. The bilinearity assumption is only valid when the retention profile (or peak shape) is not concentration dependent and thus a concentration profile is merely the retention profile multiplied by a scalar. These scalars are calculated in step 8 and form the kinetic profiles after applying the concentration-based constraints in steps 5 and 7 and normalizing the retention profiles in step 6.

After applying constraints to the kinetic profile in step 9, the quality of the fit is checked for convergence in steps 10 and 11. If convergence has not yet been achieved, the concentration profile ($\mathbb{C}$) is reconstructed from the retention profile tensor ($\mathbb{R}$) and the kinetic profile matrix ($\mathbf{K}$) in step 12 and is consecutively used to calculate a new spectral tensor in step 13. This process is iterated until a minimum error is reached or a set number of iterations have occurred.

3. Validation of the method

3.1. Generation of the validation data set

The algorithm was tested using a large number of different synthetic three-way data sets with two overlapping components. A typical data set would be composed of the three matrices shown in Fig. 3. Two components of equal height with Gaussian peak shapes were used for the retention profiles (Fig. 3a). Simple first-order kinetics (exponential decay) was used to simulate the kinetic profiles (Fig. 3b). The spectral profiles were generated by adding a Gaussian function to an exponential decay function (Fig. 3c).

The characteristic parameters for the simulated data are as follows. The chromatographic axis is 100 units with 0.5 resolution resulting in 201 data points. The wavelength window is also 100 with intervals of 0.5. The total kinetic time window was also 100 units with a first-order rate constant of 0.07. However, only eight

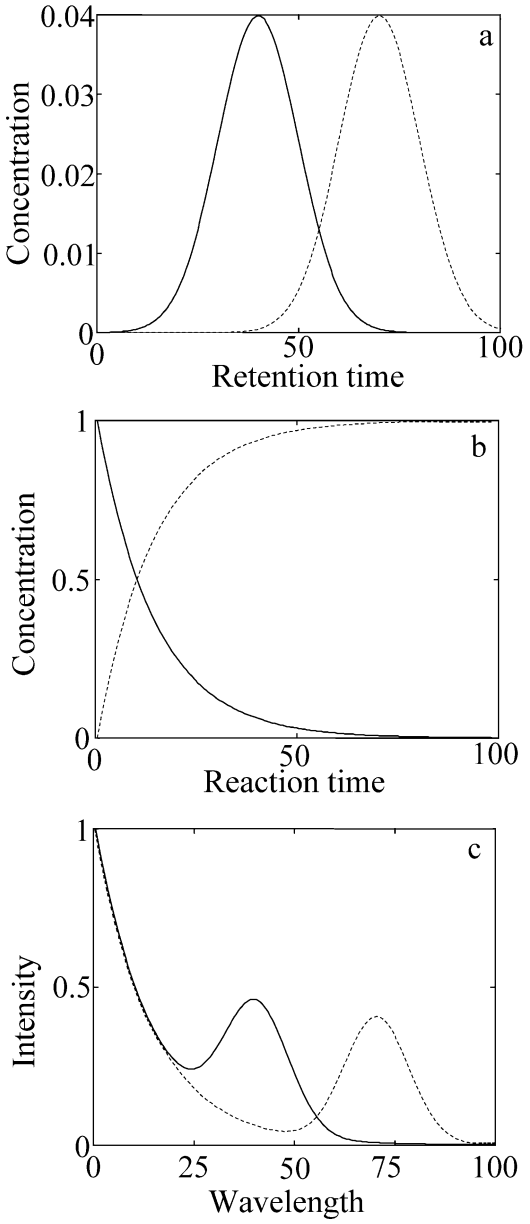

Fig. 3. Typical (a) retention profiles, (b) kinetic profiles and (c) spectral profiles for the generation of the three-way data sets. The profile of the reactant is indicated by a solid line (——), while the profile of the product is indicated by a dashed line (– –).

kinetic measurements were taken with an exponentially increasing time interval. The retention times were 40 and 70, for the reactant and product components, respectively. The standard deviation of the Gaussian retention peaks was 10. The pure component spectra were composed of an exponential decay function ($k = 0.07$) plus a Gaussian function. For the reactant, this Gaussian had a mean of 40 and a standard deviation of 10 and the product species had a mean of 70 and a standard deviation of 10. The Gaussian functions were multiplied by 10 before being added to the exponential decay function to form the spectrum. Lastly, normally distributed random numbers with a standard deviation of 10% of the maximum value of the data set were added to simulate noise. Three identical validation data sets were generated with different noise components. The fit errors of these three data sets were averaged to minimize the influence of the difference in the random fluctuations of the noise component.

To test the influence of changes in the various parameters, each of the parameters of the synthetic data was varied at one time. The noise level was varied from 1% to 25% in 1% increments. The concentration ratio of the two components was varied from 1:1 to 20:1 with increments of 1. The retention time of the product species which is the second "eluting" peak, was varied from 41 to 80 with increments of 1. The standard deviation of the retention profiles was varied from 2 to 30 with increments of 1. The mean of the Gaussian function used for the spectrum of the product ranged from 41 to 70 with increments of 1, and the number of kinetic time points was varied from 2 to 20 in steps of 1.

3.2. Generation of the initial estimates

An initial guess tensor was generated for each data set. To increase the speed of the analysis, the primary initial guess was not obtained by EFA for every kinetic slice. The EFA result from analysis of the data set with the characteristic parameters was utilized instead. These two profiles had approximate triangular shapes and formed the initial guess matrix together with a small constant (0.0001) profile functioning as the extra background component. Two-way MCR-ALS was performed on the middle kinetic slice (e.g. the fourth kinetic measurement out of eight) with a non-negativity constraint applied to the spectral and concentration profiles of the two components. The background component was left unconstrained. The two-way ALS result was used as initial guesses for each adjacent slice (in this case, the third and fifth kinetic measurement) and two-way ALS was performed on those slices. This approach was used for each of the K slices and is schematically depicted in Fig. 4 for a data set with five kinetic measurements.

3.3. Validation of the program

The data analysis was done without assuming any prior knowledge of the retention, spectral and kinetic profiles. The only input used for the algorithm was the data set, the number of components and various chemically relevant constraints. The three-way ALS algorithm used non-negativity constraints on the spectral and retention profiles of the two components and left the third component (the background) unconstrained. The application of the closure and trilinearity constraints was of special interest; therefore, all data sets were analyzed using (1) no additional constraints (no), (2) just closure (cl), (3) only trilinearity (tr) and (4) both trilinearity and closure (t + c). Although all the validation data sets had a noise component, the ability of the algorithm to reproduce the synthetic data set as it was before the noise was added was evaluated. This was accomplished by determining the percentage lack of fit between the tensor created by multiplying the optimal retention, kinetic and spectral profiles of the first two components, leaving out the background component, with the synthetic data set before the noise component was added, according to Eq. (7).

$$\text{percentage lack of fit} = 100\sqrt{\frac{\sum_{t,s,k}\left(d_{t,s,k}^{\text{optimal}} - d_{t,s,k}^{\text{synthetic}}\right)^2}{\sum_{t,s,k}\left(d_{t,s,k}^{\text{input}}\right)^2}} \quad (7)$$

This results in a fit error that represents the overall fit of the data tensor, but not the fit for each individual profile.

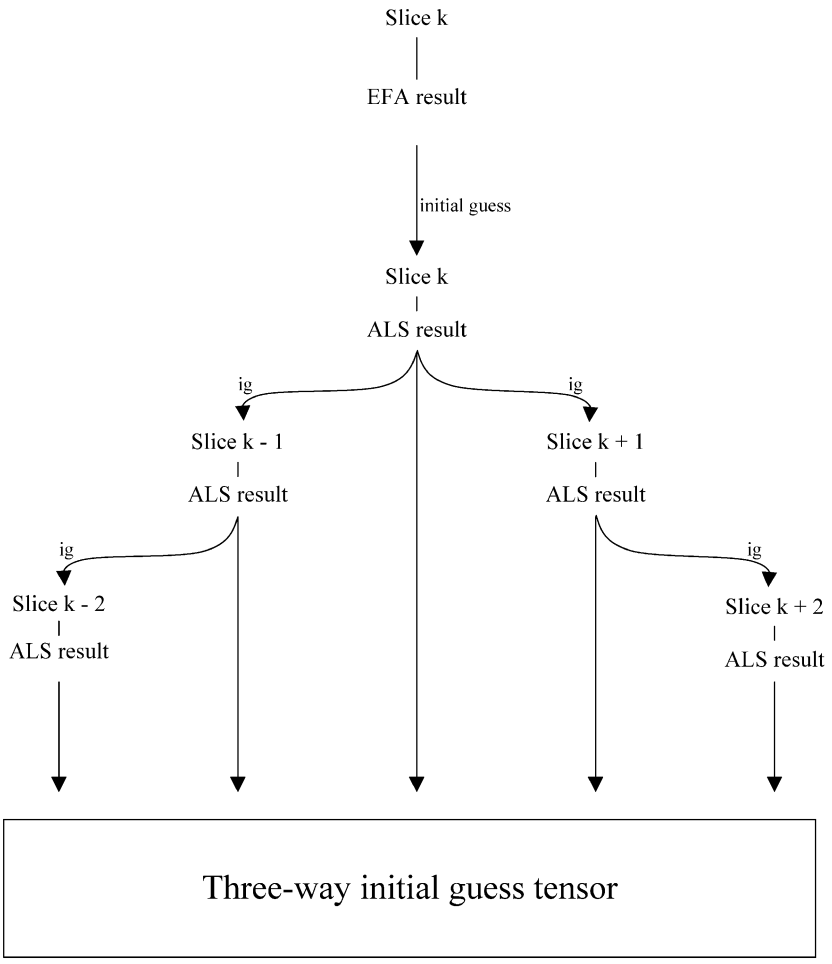

Fig. 4. Procedure for generating the three-way initial guess tensor. The index k refers to the midpoint measurement in the reaction profile. The results from the ALS fit for this matrix, $\mathbf{D}_k$, are used as initial guesses for the preceding and subsequent matrices in sequence.

To validate the performance of the program with regards to the quality of prediction of the chromatographic, spectral and kinetic profiles, the individual response profiles were compared to the original synthetic profiles by calculating the dissimilarity according to Eq. (8) [31],

$$\mathrm{dis}(\mathbf{p}_{\mathrm{calc}},\ \mathbf{p}_{\mathrm{synth}}) = \sqrt{1 - r^2(\mathbf{p}_{\mathrm{calc}},\ \mathbf{p}_{\mathrm{synth}})} \qquad (8)$$

where $\mathbf{p}_{\mathrm{calc}}$ is the calculated profile vector, $\mathbf{p}_{\mathrm{synth}}$ is synthetic profile vector and r is the correlation coefficient between the two vectors.

The algorithm was tested for stability by decreasing the resolution, increasing the spectral overlap, changing the number of time points and increasing the noise level.

4. Results and discussion

The DTD algorithm by Booksh et al. [25] was implemented in MATLAB® and applied to the same three-way data sets as the MCR-ALS algorithm. The fit error, as expected with trilinear data, is not dependent on the resolution or any of the other factors tested other than the noise and the non-trilinear peak shift [31]. The fit errors as calculated from the optimal profiles found by DTD for all replicates of all data sets

except for the noise-dependent and peak-shifted data sets were approximately 1.0%.

The first six series of simulated experiments all have a trilinear structure. This means that the spectral profiles and retention profiles have the same shape and position in each kinetic experiment. The unconstrained algorithm allows for these profiles to be different, while the trilinearity constraint forces the profiles for each component to be equivalent for each kinetic experiment.

The influence of the chromatographic resolution was evaluated by changing the retention time of the

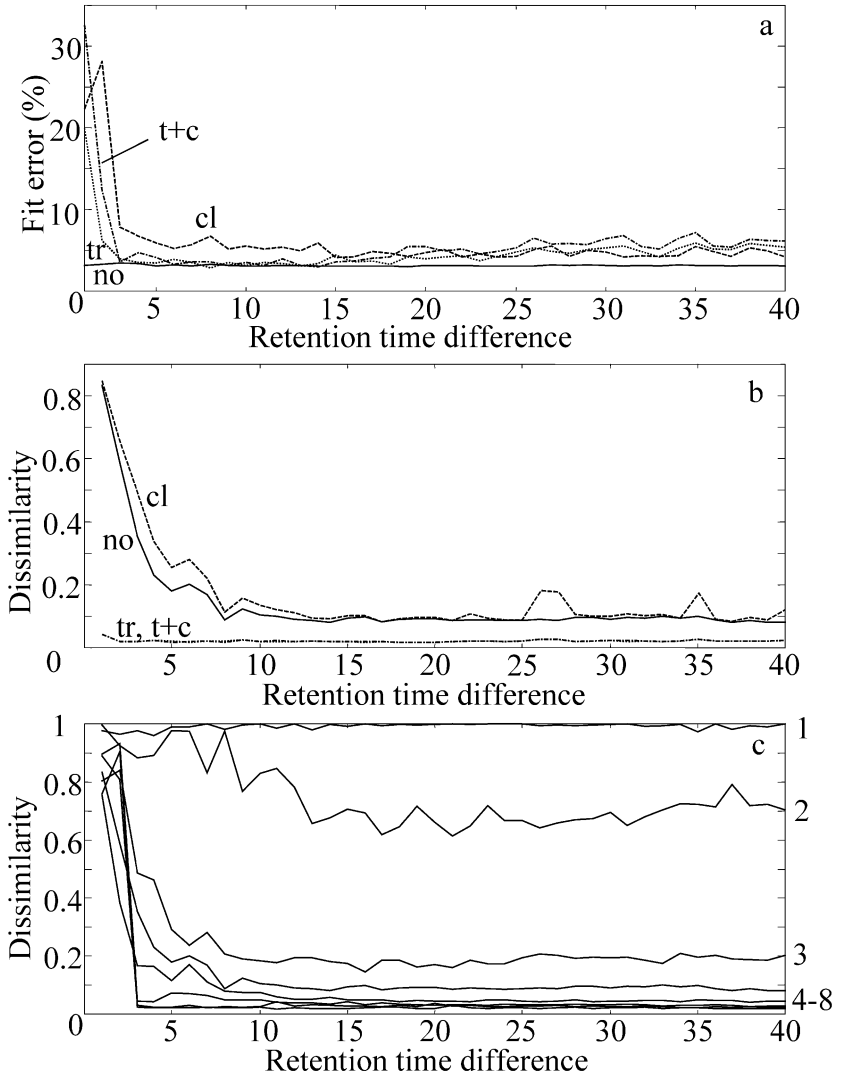

Fig. 5. The fit error (a) as defined in Eq. (7) as a function of the retention time difference between the two components when no (no) constraints are applied (——), only closure (cl) is applied (– –), just trilinearity (tr) is applied (· · · · · ·) or trilinearity and closure (t + c) are applied (–·–). The dissimilarity (b) between the reproduced retention profile of the product in the fourth kinetic experiment and the synthetic retention profile as defined by Eq. (8) when no (no) constraints are applied (——), only closure (cl) is applied (– –), just trilinearity (tr) is applied (· · · · · ·) or trilinearity and closure (t + c) are applied (–·–). The dissimilarity (c) as defined by Eq. (8) of the retention profiles of the product as a function of retention time difference and reaction time when no constraints are applied. The numbers on the right hand side indicate the number of the kinetic run.

Table 1
Comparison of the fit qualities

	Retention time difference	Peak width	Spectral difference	Molar aborptivity ratio	Noise level	Non-trilinearities
Overall fit error	no, (cl, tr, t+c)	no, (tr, t+c), cl	no, cl, (tr, t+c)	no, (cl, tr, t+c)	no, cl, tr, t+c	no, cl, tr, t+c
Spectral profile	(tr, t+c), (cl, no)	(tr, t+c), (no, cl)	(tr, t+c), (no, cl)	(tr, t+c), (no, cl)	(tr, t+c), (cl, no)	(tr, t+c), (cl, no)
Retention profile	(tr, t+c), (no, cl)	(tr, t+c), (no, cl)	(tr, t+c), (no, cl)	(tr, t+c), (no, cl)	(tr, t+c), (no, cl)	(no, cl), (tr, t+c)
Kinetic profile	(t+c, tr, cl), no	(t+c, tr), cl, no	(t+c, tr, cl), no	(t+c, tr, cl), no	(t+c, tr, cl), no	(t+c, tr, cl), no

Order (from best to worst) of the fit quality with the application of no constraints (no), closure (cl), trilinearity (tr) and trilinearity and closure (t+c). When fits are similar, the constraints are given within parentheses.

product species from 41 to 80, while the retention time of the reactant is held at 40. The results of the algorithm to reproduce the synthetic data are shown in Fig. 5a. At a very small difference in retention time between the two components, the algorithm minimizes on one major component with a spectral profile that is the average of the two pure component spectra. This is due to poor initial estimates. The EFA followed by two-way ALS results in only one major component when the two profiles are nearly identical. Once the resolution is increased, the fit error is in the range of 3–6%. As expected, the use of more constraints decreases the fit error. In Fig. 5b, the dissimilarity between the reproduced retention profile of the product in the fourth kinetic experiment and the synthetic retention profile is shown when various constraints are applied. Unlike the overall fit error, the application of the trilinearity constraint results in a better prediction of the individual profiles. This is also an expected result since these constraints enable the algorithm to find a chemically sound solution that is not necessarily the global minimum. Fig. 5c shows the dissimilarity between the reproduced retention profiles and the synthetic retention profiles of the product when no constraints are applied. The intensity of the signal for the product component is very small in the first two kinetic runs which results in dissimilar profiles, since the synthetic profile is a Gaussian, while the reproduced profile is mainly noise. The dissimilarities of all the profiles show the same trends as the overall fit error for the entire data tensor. Therefore, only the figures for the fit errors are given, but the overall results are summarized in Table 1.

The peak width is part of the chromatographic resolution but was tested separately by changing the standard deviation of both Gaussian functions from 2 to 30. At a standard deviation of 2, the peaks are baseline separated, while at 30, they have considerable overlap (Rs = 0.25). The results of this simulation are shown in Fig. 6. As expected, when the peak widths are increased, the resulting overlap and lower resolution increases the fit errors. However, it can also be seen that with increasing overlap, the additional

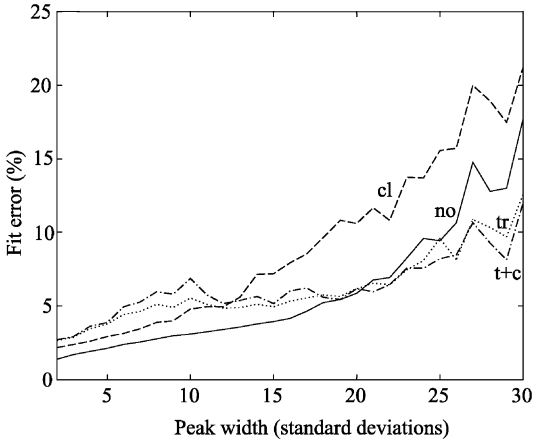

Fig. 6. The fit error as defined in Eq. (7) as a function of the peak width in standard deviations of the two components when no (no) constraints are applied (——), only closure (cl) is applied (– –), just trilinearity (tr) is applied (· · · · · ·) or trilinearity and closure (t+c) are applied (–·–).

third-order information, exploited by the trilinearity constraint, results in a better fit as compared to when the trilinearity constraint is not used. This is also true for the individual profiles where the fit quality is higher over the entire range using the trilinearity constraint.

To test the algorithm on components with similar spectra, the spectral profile of the product was altered by changing the position of the Gaussian function. The position of the Gaussian of the reactant was 40, hence the position of the Gaussian of the product was changed from 41 to 80. The results of this simulation are shown in Fig. 7. When the spectral profiles are nearly identical, the algorithm converges on two components, each with a double-peaked retention profile. This, however, is again an artifact of poor initial estimates. The application of a unimodality constraint would improve the algorithms' ability to find a better solution; however, this constraint is as of yet not implemented into the algorithm. The reproduction of the individual profiles is once again better when using the trilinearity constraint.

Another characteristic of a pure component spectrum might be a very low molar absorptivity. This effect is simulated by reducing the peak height of the product, by varying the intensity ratio (reactant to product) from 1:1 to 20:1. The results of this simulation are shown in Fig. 8. With the absorbance of the

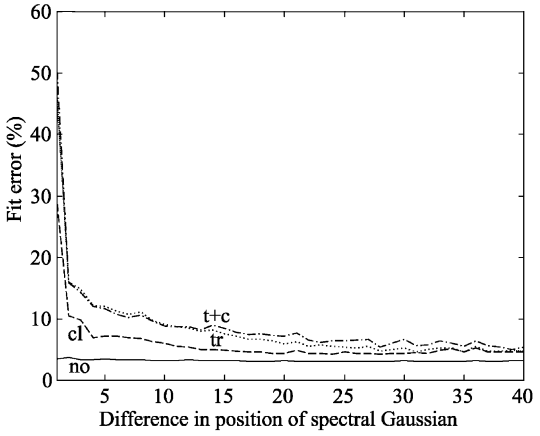

Fig. 7. The fit error as defined in Eq. (7) as a function of spectral difference between the Gaussians of the two components when no (no) constraints are applied (——), only closure (cl) is applied (– –), just trilinearity (tr) is applied (· · · · · ·) or trilinearity and closure (t+c) are applied (–·–).

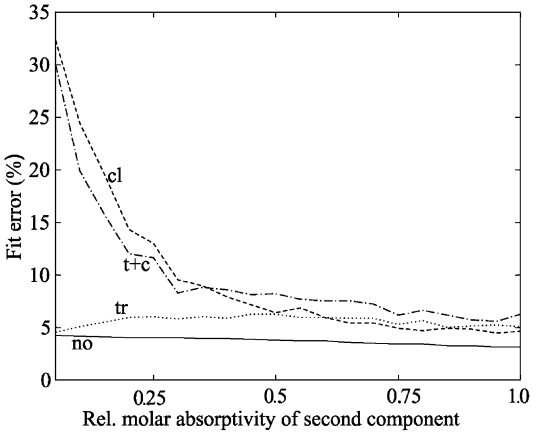

Fig. 8. The fit error as defined in Eq. (7) as a function of the ratio of the molar absorptivities of the two components when no (no) constraints are applied (——), only closure (cl) is applied (– –), just trilinearity (tr) is applied (· · · · · ·) or trilinearity and closure (t+c) are applied (–·–).

product at a 20th of that of the reactant, even at the maximum in the kinetic profile, it is well below the 10% noise level. The application of the closure constraint increases the fit error since this constraint is dependent on the uncertainty in the peak amplitude. As with the previous simulations, the reproduction of the individual profiles is once again better when using the trilinearity constraint.

Instrument noise can be a large factor in the precision of quantification of a component. The algorithm was tested with data that had a random noise component added ranging from 1% to 25% of the maximum value of the data set. The results of this simulation are shown in Fig. 9. There is quadratic dependence of the fit error with the noise level, which appears linear over this limited range. The increasing noise level decreases the quality of the fit, as is to be expected. The reproduction of the individual profiles is once again better when using the trilinearity constraint.

The number of injections determines the number of kinetic data points. The algorithm was tested using data sets ranging from 2 to 20 kinetic data points. The points were taken at exponentially increasing time intervals to maintain the maximum concentration change between measurements. The results of this simulation are shown in Fig. 10. The data set becomes dramatically larger with the number of measurements,

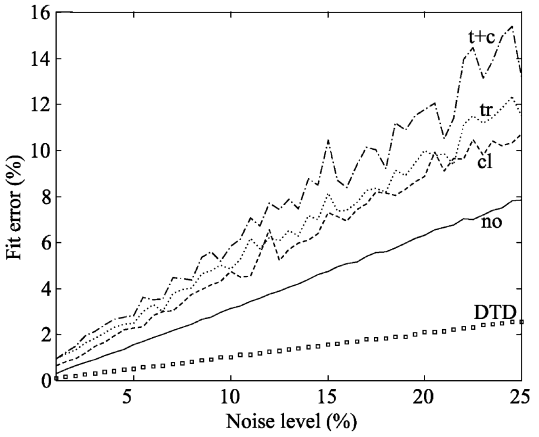

Fig. 9. The fit error as defined in Eq. (7) as a function of the noise added to the synthetic data set when no (no) constraints are applied (——), only closure (cl) is applied (– –), just trilinearity (tr) is applied (· · · · · ·) or trilinearity and closure (t + c) are applied (–·–) and the fit error by DTD (□□).

and so does the absolute sum of squared error. However, the relative error gets smaller when increasing the number of data points. The advantage of going to larger data sets is minimal, due to the redundancy of the additional scans. However, increasing the number of kinetic measurements will improve the precision of the determination of rate constants from the extracted kinetic profiles.

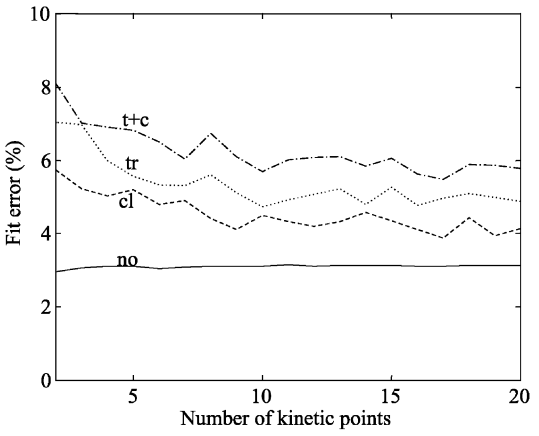

Fig. 10. The fit error as defined in Eq. (7) as a function of the number of data points in the third dimension (kinetic measurements), when no (no) constraints are applied (——), only closure (cl) is applied (– –), just trilinearity (tr) is applied (· · · · · ·) or trilinearity and closure (t + c) are applied (–·–).

So far, all data sets still have an underlying trilinearity, because of the way they are generated. To investigate how the algorithm is able to deal with non-trilinear data, the retention time of the product was shifted to earlier retention times for each succeeding kinetic measurement. This resulted in more overlap between the two components as reaction time progresses. The results of this simulation are shown in Fig. 11. Applying trilinearity constraints on a non-trilinear data set dramatically increases the lack of fit. DTD still results in a better fit than the ALS algorithm with the trilinearity constraint applied. However, a peak shift of 1 point per kinetic experiment causes a fit error that is larger than that of the ALS algorithm without trilinearity. The third-order information present in the non-trilinear data set can still be used by applying the closure constraint. A second option is to assume common spectral profiles, since even though the retention profiles change, the spectral profile still stays the same within each kinetic experiment. This is the case in this synthetic data set but also in typical LC-DAD data. Applying bilinearity (bl) in only the spectral direction (meaning the spectrum for each component is the same in each chromatographic and kinetic measurement) results in a better fit than even the unconstrained algorithm. The large increase in the fit error for large peak shifts when closure is applied is

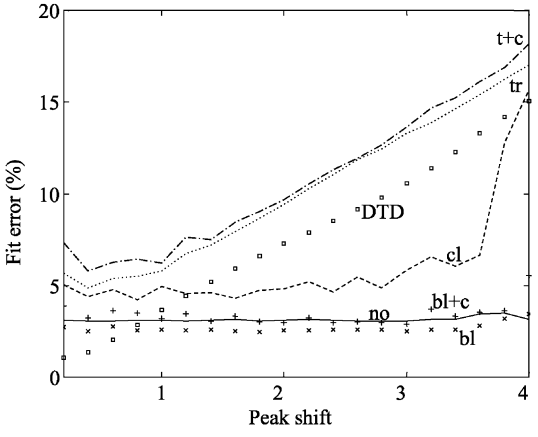

Fig. 11. The fit error as defined in Eq. (7) as a function of increasing peak shift in non-trilinear data when no (no) constraints are applied (——), only closure (cl) is applied (– –), just trilinearity (tr) is applied (· · · · · ·), trilinearity and closure (t + c) are applied (–·–), bilinearity (bl) in spectral dimension (× ×), bilinearity in spectral dimension and closure (bl + c) (+ +) and the fit error by DTD (□□).

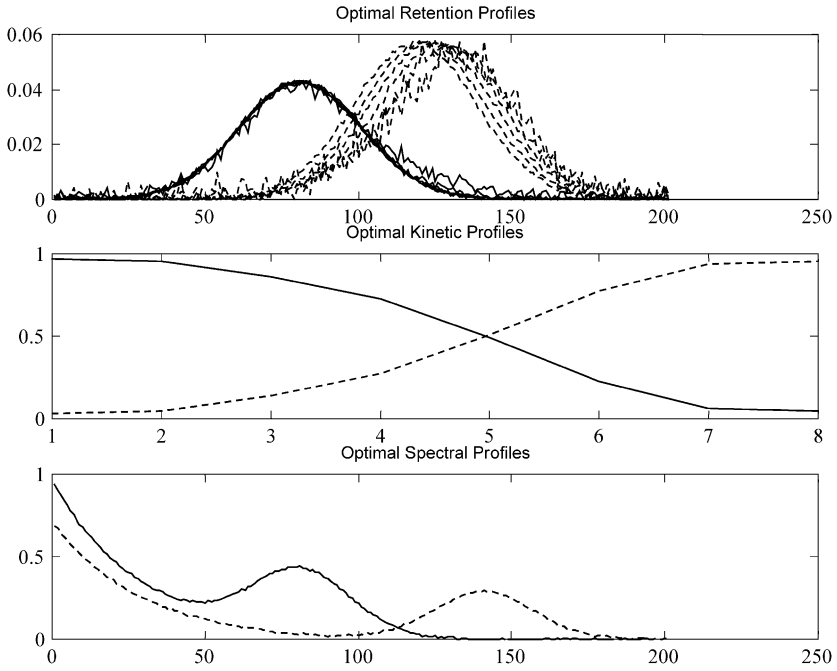

Fig. 12. Typical output of the algorithm for non-trilinear data when closure and bilinearity in the spectral profiles are used. The profile of the reactant is indicated by a solid line (——), while the profile of the product is indicated by a dashed line (– –).

once again an artifact due to the quality of the initial estimates, because in these cases, the sequence of the components in the chromatogram is switched during the automatic generation of the initial estimates. This problem can be solved by inspection of the two-way ALS results and reorganizing the columns in the initial guess matrix to a consistent sequence. Unlike all previous experiments, the application of the trilinearity constraint now increases the dissimilarity between the true synthetic profiles and the reproduced profiles.

Fig. 12 shows the typical output from the program using non-trilinear data when the closure constraint as well as bilinearity in the spectral profiles are used. The profiles for each slice are overlaid in each window. This way, it is easy to interpret whether a peak shift is occurring or a spectral profile has changed. In the top window, the retention time shift of the product can be seen. The retention profiles becomes noisier as the intensity of that component decreases. The middle window shows the kinetic profile and the bottom window shows the spectral profiles of the two species. In this case, the spectral profiles are matched for each kinetic experiment, resulting in identical spectra for each component for every reaction time.

5. Conclusion

This new algorithm makes use of the representation of multidimensional matrices in MATLAB® to perform an iterative alternating least-squares technique on three-way data sets contained in three-dimensional tensors. This results in a more convenient way to implement the algorithm than the common method of stacking a three-way data set into a two-dimensional matrix. Moreover, this method also allows for a convenient application of the closure constraint in the third dimension. The closure constraint makes use of the third-order information without assuming trilinearity.

The analysis of the computer-generated data shows that this algorithm is able to reproduce the three-way synthetic data for reasonably well-separated peaks and dissimilar spectral profiles. Increasing spectral overlap, decreasing resolution and increasing noise level increases the lack of fit as is to be expected. When

analyzing poorly resolved chromatograms or components with nearly identical spectral profiles, the quality of the initial estimates is important for finding the global minimum.

The use of constraints tends to increase the overall fit error of the total data matrix; however, applying a trilinearity constraint to trilinear data results in a better fit of the individual profiles. The application of the closure constraint has only a slight influence on the prediction of the retention and spectral profiles but in general improves the prediction of the kinetic profiles.

The use of the closure constraint in the third dimension gives the method a clear advantage over previously used multivariate–multi-way methods for analysis of kinetic data and allows for higher flexibility.

Acknowledgements

The authors acknowledge financial support from the National Science Foundation (CHE-9709437) and the Jeffress Trust (J-517).

References

[1] B.K. Lavine, Anal. Chem. 70 (1998) 209R.
[2] K. de Braekeleer, A. de Juan, D.L. Massart, J. Chromatogr. 832 (1999) 67.
[3] J.S. Salau, M. Honing, R. Tauler, D. Barcelo, J. Chromatogr. 795 (1998) 3.
[4] R. Bro, C.A. Andersson, H.A.L. Kiers, J. Chemom. 13 (1999) 295.
[5] P.M. Kroonenburg, Three-Mode Principal Component Analysis, Theory and Applications, DSWO Press, Leiden, 1983.
[6] R. Bro, J. Chemom. 10 (1996) 47.
[7] R. Gargallo, R. Tauler, F. Cuesta Sanchez, D.L. Massart, Trends Anal. Chem. 15 (1996) 279.
[8] MATLAB 5.3, Mathworks, 1999.
[9] T.S. Blyth, E.F. Robertson, Matrices and Vector Spaces, Chapman & Hall, New York, 1986.
[10] F. Cuesta Sanchez, B. van den Bogaert, S.C. Rutan, D.L. Massart, Chemom. Intell. Lab. Syst. 34 (1996) 139.
[11] T.D. Jarvis, J.H. Kalivas, Anal. Chim. Acta 26 (1992) 13.
[12] G.A. Bakken, J.H. Kalivas, Anal. Chim. Acta 300 (1995) 173.
[13] H.R. Keller, D.L. Massart, Anal. Chim. Acta 246 (1991) 379.
[14] H.R. Keller, D.L. Massart, J.O. de Beer, Anal. Chem. 65 (1993) 471.
[15] J. Toft, O.M. Kvalheim, Chemom. Intell. Lab. Syst. 19 (1993) 65.
[16] F. Cuesta Sanchez, J. Toft, O.M. Kvalheim, D.L. Massart, Anal. Chim. Acta 314 (1995) 131.
[17] F. Cuesta Sanchez, M.S. Khots, D.L. Massart, J.O. de Beer, Anal. Chim. Acta 285 (1994) 181.
[18] F. Cuesta Sanchez, M.S. Khots, D.L. Massart, Anal. Chim. Acta 290 (1994) 249.
[19] F. Cuesta Sanchez, J. Toft, B. van den Bogaert, D.L. Massart, Anal. Chem. 68 (1996) 79.
[20] J.S. Vanslyke, P.D. Wentzell, Anal. Chem. 63 (1991) 2512.
[21] P.D. Wentzell, S.J. Vanslyke, Anal. Chim. Acta 257 (1992) 173.
[22] S.J. Vanslyke, P.D. Wentzell, Chemom. Intell. Lab. Syst. 20 (1993) 183.
[23] R. Tauler, A. Smilde, B. Kowalski, J. Chemom. 9 (1994) 31.
[24] E. Sanchez, B.R. Kowalski, J. Chemom. 4 (1990) 29.
[25] K.S. Booksh, Z. Lin, Z. Wang, B.R. Kowalski, Anal. Chem. 66 (1994) 2561.
[26] B.E. Wilson, E. Sanchez, B.R. Kowalski, J. Chemom. 3 (1989) 493.
[27] H.A.L. Kiers, Psychometrika 56 (1991) 449.
[28] M.A. Sharaf, D.L. Illman, B.R. Kowalski, Chemometrics, Wiley, New York, 1986.
[29] H.R. Keller, D.L. Massart, Chemom. Intell. Lab. Syst. 12 (1992) 209.
[30] T.L. Cecil, S.C. Rutan, J. Chromatogr. 556 (1991) 495.
[31] A. de Juan, S.C. Rutan, R. Tauler, Chemom. Intell. Lab. Syst. 40 (1998) 19.