ELSEVIER

Chemometrics and Intelligent Laboratory Systems 58 (2001) 3–13

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

On the difference between low-rank and subspace approximation: improved model for multi-linear PLS regression

Rasmus Bro [a,*], Age K. Smilde [b], Sijmen de Jong [c]

[a] *Chemometrics Group, Food Technology, Department of Dairy and Food Science, The Royal Veterinary and Agricultural University, DK-1958 Frederiksberg, Denmark*

[b] *Process Analysis and Chemometrics, Department of Chemical Engineering, University of Amsterdam, Amsterdam, Netherlands*

[c] *Unilever Research Vlaardingen, P.O. Box 114, 3130 AC Vlaardingen, Netherlands*

Received 13 December 2000; accepted 16 April 2001

Abstract

While both Tucker3 and PARAFAC models can be viewed as latent variable models extending principal component analysis (PCA) to multi-way data, most fundamental properties of PCA do not extend to both models. This has practical importance, which will be explained in this paper. The fundamental difference between the PARAFAC and the Tucker3 model can be viewed as the difference between so-called low-rank and subspace approximation of the data. This insight is used to pose a modification of the multi-linear partial least squares regression (N-PLS) model. The modification is found by exploiting the basic properties of PLS and of multi-way models. Compared to the current prevalent implementation of N-PLS, the new model provides a more reasonable fit to the independent data and exactly the same predictions of the dependent variables. Thus, the reason for introducing this improved model is not to obtain better predictions, but rather the aim is to improve the secondary aspect of PLS: the modeling of the independent variables. The original version of N-PLS has some built-in problems that are easily circumvented with the modification suggested here. This is of importance, for example, in process monitoring, outlier detection and also, implicitly, for jackknifing of model parameters. Some examples are provided to illustrate some of these points.

Keywords: Partial least squares; Multi-way calibration; PARAFAC; Tucker3

1. Introduction

Some properties of two-way models do not extend straightforwardly to multi-way models. In this paper, the difference between fitting low-rank multi-way models and fitting subspace-based multi-way models is discussed. A low-rank model is defined here as a model $\hat{\mathbf{M}}$ $(I \times J)$ that minimizes the sum of squares of the residuals when fitted to a data matrix $\mathbf{X}$ $(I \times J)$ but *subject to* the model being of a predefined rank. In case of a bilinear model, this can be stated as

$$\min\|\mathbf{X} - \hat{\mathbf{M}}\|^2 = \min\|\mathbf{X} - \left(\mathbf{t}_1\mathbf{p}_1^{\mathrm{T}} + \mathbf{t}_2\mathbf{p}_2^{\mathrm{T}} + \ldots + \mathbf{t}_F\mathbf{p}_F^{\mathrm{T}}\right)\|^2, \quad (1)$$

where F is the rank of the model. As long as the **t**'s and **p**'s are required to be independent, the model will be of rank F per definition, because rank can be defined as the minimum number of bilinear rank-one

* Corresponding author.
E-mail addresses: rb@kvl.dk (R. Bro), asmilde@its.chem.uva.nl (A.K. Smilde), Sijmen-de.Jong@unilever.com (S. de Jong).

0169-7439/01/$ - see front matter
PII: S0169-7439(01)00134-4

terms necessary to describe the matrix. A subspace model, on the other hand, is defined here as a model that maximizes the variation retained in the joint subspaces (of dimension F) used to define the model. These subspaces are given by $\mathbf{T}$ ($I \times F$) and $\mathbf{P}$ ($J \times F$). Subsequently, the model is defined as the joint projection of $\mathbf{X}$ onto the subspace spanned in the row-mode by $\mathbf{T}$ and in the column-mode by $\mathbf{P}$. This leads to a definition of $\hat{\mathbf{M}}$ as satisfying

$$\min\|\mathbf{X} - \hat{\mathbf{M}}\|^2 = \min\|\mathbf{X} - \mathbf{T}\mathbf{T}^{+}\mathbf{X}(\mathbf{P}\mathbf{P}^{+})^{\mathrm{T}}\|^2, \qquad (2)$$

where the superscript '+' means Moore–Penrose pseudo-inverse. For two-way analysis, the low-rank and the subspace approximation leads to the exact same model which is also equivalent to a principal component analysis (PCA) model. This is not so for multi-way analysis. The difference for multi-way models is related to the difference between the decomposition models PARAFAC [5,11] and Tucker3 [17]. The background and some important consequences of this difference will be explained.

The insight above is used to develop a modified model for multi-linear partial least squares regression (N-PLS) [4]. In recent years, this method has gained more impact [2,8,13–15,22–24,33]. It was originally developed by Bro [4], and it was later shown that the predictions obtained in the case of three-way data are identical to the method earlier suggested by Ståhle (see Refs. [9,30,32]). The exploratory aspects differ though, because the latter model introduces an orthogonalization of the score vectors. This particular orthogonalization explicitly leads to a two-way unfolded/matricized model. This makes interpretation of the model more difficult. Alternative multi-linear regression model such as Multi-way Covariates Regression have also appeared in the literature more recently [3,31].

Although N-PLS is a generalization of the two-way PLS regression method, there are certain issues which do not generalize and may appear problematic in certain settings. These problems relate to the model of the three-way $I \times J \times K$ array $\underline{\mathbf{X}}$ rather than the prediction part.

(1) *Perfect fit problems:* In the current N-PLS, a trilinear model of the independent data array $\underline{\mathbf{X}}$ is used. However, due to the sequential nature of the algorithm, such a trilinear model cannot fit, e.g. a rank-F trilinear data array with F components even in the noise-free case. This is opposed to the two-way case, where such data are perfectly fit. Another way of stating this is that the subtraction of rank-one terms during deflation of $\underline{\mathbf{X}}$ in the original N-PLS does not necessarily lower the rank of $\underline{\mathbf{X}}$ which it does indeed for two-way PLS. For noise-free data, the score and weight matrices of the model *will* span exactly the spaces spanned by the data, but the trilinear model will not fit the data exactly. This leads to problems of how to interpret and use the residuals of the $\underline{\mathbf{X}}$ model. It also leads to problems if the residuals are to be used, e.g. for estimating uncertainties, because high residuals do not necessarily imply inadequacy of a sample.

(2) *The regression part of N-PLS does not imply a trilinear model of* $\underline{\mathbf{X}}$: Regression is, per definition, related to subspaces. A regression model is defined as the regression of one or more dependent (possibly latent) variables onto a certain subspace. This also holds in situations where the model is calculated sequentially. The outcome of a regression model is a set of regression coefficients that are defined as the coordinates of the vector being regressed expressed in terms of the basis onto which it was regressed. The fundamental aspect of a latent variable regression model is therefore related to selecting an appropriate subspace for the regression, because, with respect to predictions, it is *only* the choice of subspace that matters. Hence, the regression aspect of N-PLS does not imply that the model of $\underline{\mathbf{X}}$ must be trilinear, but only that a suitable subspace is wanted.

(3) *Uniqueness problems:* The current N-PLS model is unique in a mathematical sense meaning that the parameters are identified for a given set of data and a given number of components. For other models, however, uniqueness usually carries a broader and more important meaning as well. Two-way PCA and PLS and multi-way Tucker models are known to provide models of the independent X-data that are unique up to scaling, permutation and rotations. This implies that these models uniquely identify the subspaces of the data. The multi-linear PARAFAC model even identifies the axes in the subspaces (*unique axes property*). An important additional implication of these different types of uniqueness properties of PCA,

Tucker3 and PARAFAC is that the uniqueness has repercussions not only for the data at hand, but also for the population from which the samples (data) are drawn. If the subspace in a PCA model is uniquely identified and reflects underlying latent phenomena specific to the population, then it follows that another PCA model fitted to new data from the same population will span the same subspace (up to the noise). Likewise, for PARAFAC, if the axes/components are uniquely identified and pertain to phenomena specific to the population, then the same axes are found when the model is fitted to different samples from the same population. However, for N-PLS, no such implications can be made. The N-PLS model of $\underline{\mathbf{X}}$ is a tri- (or multi-) linear model, which means that, like PARAFAC, it has no rotational freedom. However, unlike PARAFAC, the components found from any particular set of data will differ from those found from another set of samples from the same population *even* in the noise-free case. This is because the criterion used for finding each component (maximum covariance) is not only pertaining to the space covered by the significant variation but also the actual specific amount of different phenomena in any chosen set of sample (for finite sample sizes). For example, if the samples in a given data set mainly contain high amounts of analyte and low amounts of interferents, the first component is likely to reflect primarily the analyte variation. If the model is fitted to another set of samples from the same population, which mainly contains interferents at high levels and analytes at low levels; then, the first component will differ from the aforementioned, mainly describing the variation of the interferents. Note that this is not a problem in two-way PLS where only the subspaces of $\mathbf{X}$ are claimed unique. For two-way PLS, the actual components would also differ in the above-mentioned example, but the subspaces spanned by the complete model would be the same up to noise. For the original N-PLS, the axes will differ in both cases.

(4) *Jackknifing assessment of parameters is impossible:* Jackknifing as a means for assessing the stability of the estimated parameters was recently introduced for two-way PCA and PLS [21]. Jackknife analysis of the model parameters of $\underline{\mathbf{X}}$ in the original N-PLS model, however, is not possible. Even though the model is unique, the scores and loadings will change when estimated from different subsets as a consequence of (3). This even holds in the noise-free case. As the model of $\underline{\mathbf{X}}$ in N-PLS has no rotational freedom, there is no way of, e.g. transforming solutions towards a common solution as it is done in the two-way models. Moreover, the jackknife assessment of the model parameters should be done in terms of subspaces and not unique axis, because the subspaces are the essential entities in the regression model, as argued under (2).

These are the problems dealt with in this paper. In the following, it will be assumed that the independent variables held in the $I \times J \times K$ three-way array $\underline{\mathbf{X}}$ have a three-way structure. The three-way array will also sometimes be shown unfolded/matricized [16] to a two-way matrix. This matricized version of the data is called $\mathbf{X}$ and will always be an $I \times JK$ matrix. Thus, the first mode is always maintained as the rows, and the second and third mode combined into the column-mode of the matrix. In the following, it is assumed that the dependent variable is univariate, $\mathbf{y}(I \times 1)$. The situation with multivariate dependent variables follows immediately from this.

2. Theory

2.1. Subspace approximation versus low-rank approximation

Principal component analysis is an optimal representation of a matrix with respect to several different criteria. One interpretation of PCA is to consider it as the best low-rank approximation of a matrix in a least squares sense. Another entirely equivalent view is to see it as the best approximation of the data within a joint low-dimensional subspace. Intuitively, these two properties seem equivalent, and indeed this is true for the two-way case. For three- and higher-way data, however, the two criteria lead to two completely different models: the PARAFAC and the Tucker3 model.

The best low-rank approximation of a three-way array is a PARAFAC model. This is so because the definition of rank coincides with the definition of the PARAFAC model [18]. For a given array $\underline{\mathbf{X}}$ of size $I \times J \times K$ and rank F, a PARAFAC model describ-

ing the array is given by three component matrices: **A** ($I \times F$), **B** ($J \times F$), **C** ($K \times F$) with typical elements a_{if}, b_{jf}, and c_{kf}, respectively. These elements satisfy

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf}, \quad i = 1, \ldots, I; \quad j = 1, \ldots, J; \quad k = 1, \ldots, K; \tag{3}$$

which can also be written in terms of matrices as

$$\mathbf{X} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathrm{T}}. \tag{4}$$

The term $\mathbf{C} \odot \mathbf{B}$ is the Khatri–Rao [27] product defined as the column-wise Kronecker product of the two matrices [6]. Thus, $\mathbf{C} \odot \mathbf{B}$ will be of size $JK \times F$ and the fth column will equal the Kronecker product of the fth column of **B** and of **C**.

It is interesting and counter-intuitive that even though the PARAFAC model is the least squares low-rank trilinear solution, the model does not incorporate all the variation of $\underline{\mathbf{X}}$ within the joint subspaces of **A**, **B**, and **C**. To explain this, consider, as an example, an ordinary two-way PCA model. The PCA model is a low-rank approximation of **X** because it can be written in terms of independent rank-one contributions, the number of which defines the rank of the model

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} = \mathbf{t}_1\mathbf{p}_1^{\mathrm{T}} + \mathbf{t}_2\mathbf{p}_2^{\mathrm{T}} + \ldots + \mathbf{t}_F\mathbf{p}_F^{\mathrm{T}} + \mathbf{E}. \tag{5}$$

The part of the data that is within the joint subspaces defined by **T** and **P** can be found by projecting **X** onto these subspaces. This part of the data can be written as

$$\mathbf{X}^{\mathrm{proj}} = \mathbf{T}\mathbf{T}^{+}\mathbf{X}(\mathbf{P}\mathbf{P}^{+})^{\mathrm{T}} \tag{6}$$

and it holds that

$$\mathbf{T}\mathbf{T}^{+}\mathbf{X}(\mathbf{P}\mathbf{P}^{+})^{\mathrm{T}} = \mathbf{T}\mathbf{P}^{\mathrm{T}}. \tag{7}$$

which can easily be verified by inserting the result of Eq. (5) into the left-hand side of Eq. (7) and noticing that **E** is completely annihilated by the projectors $\mathbf{T}\mathbf{T}^{+}$ and $\mathbf{P}\mathbf{P}^{+}$ in contrast to $\mathbf{T}\mathbf{P}^{\mathrm{T}}$ which remains unaffected. For the three-way case, the PARAFAC model is the best low-rank approximation of a three-way array. A PARAFAC model can be written in terms of the matricized three-way array as

$$\mathbf{X} = \mathbf{X}^{\mathrm{PAR}} + \mathbf{E} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathrm{T}} + \mathbf{E}. \tag{8}$$

The part of $\underline{\mathbf{X}}$ within the joint subspaces defined by **A**, **B** and **C** is calculated analogously to Eq. (6) by projecting $\underline{\mathbf{X}}$ onto $\mathbf{A}\mathbf{A}^{+}$ in the first mode, onto $\mathbf{B}\mathbf{B}^{+}$ in the second mode and onto $\mathbf{C}\mathbf{C}^{+}$ in the third mode. Using matricized notation and the Kronecker product, $\otimes$, this can be written

$$\mathbf{X}^{\mathrm{proj}} = \mathbf{A}\mathbf{A}^{+}\mathbf{X}(\mathbf{C}\mathbf{C}^{+} \otimes \mathbf{B}\mathbf{B}^{+})^{\mathrm{T}}. \tag{9}$$

Although less transparent than Eq. (6), the above simply states that the three-way array is projected onto $\mathbf{A}\mathbf{A}^{+}$ ($I \times I$) hence transformed into $\mathbf{A}\mathbf{A}^{+}\mathbf{X}$. This transformed $I \times J \times K$ array is then projected onto $\mathbf{B}\mathbf{B}^{+}$ ($J \times J$) in the second mode. That is, the $I \times J \times K$ array $\mathbf{A}\mathbf{A}^{+}\mathbf{X}$ is simply rearranged to a $J \times I \times K$ array **M** and projected as $\mathbf{B}\mathbf{B}^{+}\mathbf{M}$. Finally, this array is rearranged into a $K \times I \times J$ array **N** and projected onto $\mathbf{C}\mathbf{C}^{+}$ ($K \times K$) yielding $\mathbf{C}\mathbf{C}^{+}\mathbf{N}$. This is the projected array $\mathbf{X}^{\mathrm{proj}}$ of Eq. (9) rearranged appropriately.

However, the projection of $\underline{\mathbf{X}}$ onto these subspaces does *not* equal the PARAFAC model. The expression in Eq. (9) is equal to the so-called Tucker3 model with loading matrices **A**, **B** and **C** [17]. The Tucker3 model provides the best model of the data in terms of the truncated bases, i.e. the best sub-space approximation. That is, within the subspaces generated by **A**, **B** and **C**, the Tucker3 model retains the maximal amount of variation in the model of $\underline{\mathbf{X}}$. The PARAFAC model is a restriction of the Tucker3 model. It also stays within the subspaces defined by **A**, **B** and **C** but it does not retain all the variation defined in Eq. (9).

The Tucker3 model can also be written as

$$\mathbf{X} = \mathbf{A}\mathbf{G}(\mathbf{C} \otimes \mathbf{B})^{\mathrm{T}} + \mathbf{E} \tag{10}$$

where

$$\mathbf{G} = \mathbf{A}^{+}\mathbf{X}(\mathbf{C}^{+} \otimes \mathbf{B}^{+})^{\mathrm{T}} \tag{11}$$

and is called the core array of size $F \times FF$. It is calculated in its matricized version here, but the core is a three-way array of size $F \times F \times F$. Inserting Eq. (11) into Eq. (10) directly provides Eq. (9). This representation of the Tucker3 model sheds some light on

the similarity between the Tucker3 model and the truncated two-way singular value decomposition (SVD). The so-called core array, **G**, corresponds to the matrix of singular values. In SVD, the matrix of singular values is diagonal. This means that the projection or regression onto the subspace coincides with a low-rank approximation, because only singular vectors of identical subscripts are multiplied together. This is different for the three-way case. In the three-way case, the core array will not be super-diagonal in general (even though the model possesses rotational freedom, none of these equivalent rotated models will have a super-diagonal core).

If the core array *is* super-diagonal, then this implies that only loadings vectors with the same component number interact. This is then the PARAFAC model. However, as the core array arising as the regression of the data onto the best-fitting subspace is not super-diagonal in general, then there is no simple relation between low-rank (PARAFAC) and subspace (Tucker3) fitting.

For any set of loading matrices **A**, **B** and **C**—be they actual solutions to a model or not—the model in Eq. (10) will fit better than Eq. (8), except in the case of perfect PARAFAC fit. This is because of the fact that PARAFAC is a restricted version of Tucker3 and can be seen from the inequality

$$\|\mathrm{vec}\mathbf{X} - (\mathbf{C}\otimes\mathbf{B}\otimes\mathbf{A})\mathrm{vec}\mathbf{G}\|^2 \leq \|\mathrm{vec}\mathbf{X} - (\mathbf{C}\otimes\mathbf{B}\otimes\mathbf{A})\mathrm{vec}\mathbf{I}\|^2 \quad (12)$$

which holds for any set of appropriately sized **A**, **B** and **C** when **G** is the solution to $\min\|\mathrm{vec}\mathbf{X} - (\mathbf{C}\otimes\mathbf{B}\otimes\mathbf{A})\mathrm{vec}\mathbf{G}\|^2$. This part corresponds to the fit of Eq. (10) and $\|\mathrm{vec}\mathbf{X} - (\mathbf{C}\otimes\mathbf{B}\otimes\mathbf{A})\mathrm{vec}\mathbf{I}\|^2$ corresponds to the fit of a PARAFAC model. The inequality follows because the left-hand side of the inequality can be seen as a standard multiple linear regression problem for fixed **A**, **B** and **C**, where the regression coefficients, vec**G**, can be found freely, whereas on the right-hand side, the corresponding parameters are fixed as the vectorized superdiagonal array vec**I** with ones on the superdiagonal.

The difference between the Tucker3 and the PARAFAC model is very significant and it points to why it is sometimes difficult to use the PARAFAC model in practice. The Tucker3 model is, in a sense, the natural extension of PCA with respect to many of the intuitively experienced properties of PCA. The Tucker3 model simply re-expresses the array in terms of truncated bases. As such, the Tucker3 model will always 'work', in the same sense as *any* two-way matrix can be arbitrarily well modeled with a PCA model. If the (pseudo-) rank of the variation in each mode is lower than the actual dimension,[1] then a Tucker3 model *is* valid per definition in the same sense that a PCA model is for two-way data. The PARAFAC model, on the other hand, is a very special model. It models the data as a restricted version of the subspace spanned by the component matrices. It may often not be particularly suited for approximating an array if the systematic variation does not approximately follow a trilinear model. For example, chromatographic data with significant retention time shifts are often not well modeled with PARAFAC directly. On the other hand, though, that does not imply that PARAFAC is only applicable for very tri- or multi-linear data. PARAFAC was developed for and has been extensively used for data that are very far from being trilinear (psychometric, sensometric and similar data). Even though curve-resolution is not possible in those cases, the use of underlying tri- or multi-linear latent variables is a very reasonable approximation. And indeed, very stable and reproducible structures can often be extracted from such data [10,12,19,25,26].

2.2. Multi-linear partial least squares regression

The original N-PLS model provides a *low-rank* approximation of the independent data and it is tempting to assume that a low-rank structure must be present in order for the model to work well. However, it will be shown that N-PLS is more correctly seen as a subspace model. This leads to a more efficient model of the independent data. That the N-PLS model is in fact a subspace model also explains why

[1] The (pseudo-) rank can, e.g. be assessed for each mode by assessing the rank of the two-way array obtained by unfolding/matricizing the array such that the mode of interest constitutes the rows and the two remaining modes together constitute the columns. The (pseudo-) rank of this matrix will equal the dimensionality of the systematic variation in that mode. In practice, the rank can sometimes be difficult to assess, but in principle, the useful rank for each mode can be determined this way.

N-PLS works well even for data that are not well modeled by a trilinear (PARAFAC) model.

One specific version of the original N-PLS algorithm is shown in Table 1. This algorithm only treats the situation with univariate dependent variable, but is easily extended to the multivariate case [4,6]. The algorithm shown here is the one suggested by de Jong [9] in which no explicit deflation of $\underline{\mathbf{X}}$ is performed in the algorithm. It will be shown later, that this algorithm, which provides predictions identical to the deflated version, is a useful starting point for a modified algorithm.

2.3. New multi-linear PLS model

In N-PLS regression, the subspace (set of score vectors) is found sequentially by adding one component at a time. Thus, rank-one trilinear components are found sequentially. Merely adding these components leads to a low-rank trilinear model of $\underline{\mathbf{X}}$ with the same structure as a PARAFAC model. This is similar to ordinary two-way PLS where adding the successively calculated bilinear components leads to a bilinear model with the same structure as a PCA model. It is therefore not surprising that N-PLS, just as two-way PLS, has been viewed as a low-rank model. However, as discussed above, there is no relation between a low-rank model and a subspace-based model for multi-way data. The central aim in PLS is to find a set of score vectors spanning a reasonable subspace. Additionally, requiring the columns to be related only with columns of other component matrices with the same number (trilinearity) is a further restriction, not of the regression part of the model, but *of the model of the independent data*.

In N-PLS, the model of $\underline{\mathbf{X}}$ is given by the score matrix $\mathbf{T}$ and the loading matrices $\mathbf{W}^{\mathrm{J}}$ (second mode weights) and $\mathbf{W}^{\mathrm{K}}$ (third mode weights). Assuming a trilinear model of $\underline{\mathbf{X}}$ leads to the problems mentioned in the Introduction. As has been argued here, there is no reason why the model of $\underline{\mathbf{X}}$ should be trilinear. Even if the data *are* low-rank trilinear, N-PLS cannot model the data with a similar low-rank trilinear model.

In order to devise an improved model, it is helpful to shortly consider current models and possible alternatives. Currently, there are several different algorithms in the literature that all provide identical predictions but through widely different models of $\underline{\mathbf{X}}$.

(1) At one extreme, there is the deflation approach of Ståhle [32] using a matricized model of $\underline{\mathbf{X}}$ by introducing loading vectors of length JK. This approach could also be extended to higher-order arrays and thus work in the same situations as N-PLS. With respect to predictions, nothing would be lost, and thus, it is a viable approach. However, it is a premise of this work, that a non-matricized model is sought. This especially makes sense because a non-matricized model is assumed in the calculations of the scores even in the Ståhle approach. More importantly, the use of loadings in each variable-mode enables a simpler interpretation of the model.

(2) Other variants of additional loading vectors could also be introduced. For example, it would be natural to introduce trilinear loading vectors, i.e. loading vectors that are estimated given the score vector. This would provide a better-fitting model of $\underline{\mathbf{X}}$ compared to using the weights (as for two-way PLS). However, the introduction of loadings would not provide a solution to any of the problems mentioned in the Introduction. Further, it would complicate the algorithm numerically and algebraically and would not lead to orthogonal score vectors, which has tradi-

Table 1
The original multi-linear PLS1 algorithm. Takes (usually) centered $\mathbf{X}$ ($I \times JK$) and $\boldsymbol{y}$ ($I \times 1$) as input. The algorithm shown is the original version but without deflation of $\mathbf{X}$ as suggested by de Jong [9]

1. $f = 1$; $\mathbf{X}_0 = \mathbf{X}$; $\boldsymbol{y}_0 = \boldsymbol{y}$ (f: component counter)

Determine rank-one model of $\mathbf{X}$*-residual*
2. Calculate $\mathbf{Z}$ ($J \times K$) as $\boldsymbol{y}_{\mathrm{f}-1}{}^{\mathrm{T}}\mathbf{X}_{\mathrm{f}-1}$ ($1 \times JK$) rearranged to proper size
3. Determine $\boldsymbol{w}^{\mathrm{J}}$ and $\boldsymbol{w}^{\mathrm{K}}$ as the first left and right singular vectors of $\mathbf{Z}$
4. Calculate $\boldsymbol{t}$ as $\mathbf{X}_{\mathrm{f}-1}(\boldsymbol{w}^{\mathrm{K}} \otimes \boldsymbol{w}^{\mathrm{J}})$
5. Store $\mathbf{T}_{\mathrm{f}} := \boldsymbol{t}$; $\mathbf{W}_{\mathrm{f}}^{\mathrm{J}} := \boldsymbol{w}^{\mathrm{J}}$; $\mathbf{W}_{\mathrm{f}}^{\mathrm{K}} := \boldsymbol{w}^{\mathrm{K}}$

Regression part
6. $\mathbf{b}_{\mathrm{f}} = (\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{T}}\boldsymbol{y}$

Deflation
8. $\boldsymbol{y}_{\mathrm{f}} = \boldsymbol{y}_0 - \mathbf{T}\mathbf{b}_{\mathrm{f}}$
9. $f = f + 1$. Continue from 2 until proper description of $\boldsymbol{y}_0$

tionally been the reason for introducing the loading vectors.

(3) It is possible to simply skip deflation of $\underline{\mathbf{X}}$ as suggested in de Jong's [9] version of N-PLS. This can be beneficial and very important, e.g. for numerical reasons. The non-deflating algorithm provides the same predictions as the original deflating one. Because it is the trilinear model (and deflation) of $\underline{\mathbf{X}}$ that is problematic with respect to modeling of $\underline{\mathbf{X}}$, it is suggested to use the non-deflating algorithm as a starting point for the improved N-PLS.

It is clear that it is the added constraint of trilinearity of the model of $\underline{\mathbf{X}}$ that is problematic. It is therefore natural to exchange the trilinear model with a subspace model instead. In the original N-PLS, the model of $\underline{\mathbf{X}}$ is given as

$$\mathbf{X} = \mathbf{T}(\mathbf{W}^{\mathrm{K}} \odot \mathbf{W}^{\mathrm{J}})^{\mathrm{T}} + \mathbf{E}. \quad (13)$$

This model can simply be exchanged with a model where no trilinearity is imposed, but where the $\underline{\mathbf{X}}$ data are merely projected onto the three subspaces defined by $\mathbf{T}$, $\mathbf{W}^{\mathrm{J}}$, and $\mathbf{W}^{\mathrm{K}}$. This model is defined as

$$\mathbf{X} = \mathbf{TG}(\mathbf{W}^{\mathrm{K}} \otimes \mathbf{W}^{\mathrm{J}})^{\mathrm{T}} + \mathbf{E}. \quad (14)$$

where the core $\mathbf{G}$ is defined as in Eq. (11)

$$\mathbf{G} = \mathbf{T}^{+}\mathbf{X}\left((\mathbf{W}^{\mathrm{K}})^{+} \otimes (\mathbf{W}^{\mathrm{J}})^{+}\right)^{\mathrm{T}}. \quad (15)$$

Using a subspace model of $\underline{\mathbf{X}}$ will solve all the problems mentioned and still provide a non-matricized model, i.e. a model where each set of weights and scores pertain to one mode only. In practice, the basic algorithm is maintained without the deflation of $\underline{\mathbf{X}}$. When the weights, scores and regression coefficients have been determined using the original algorithm, the residuals in the $\underline{\mathbf{X}}$ space are then found from these parameters using Eqs. (14) and (15). As this step is performed after the model has been fitted, it does not affect the way the weights, scores and regression coefficients are calculated; only the way the residuals are calculated. Therefore, these parameters are identical to the ones found in the original non-deflating N-PLS.

3. Practical aspects of the new N-PLS model

The new model and algorithm for N-PLS has a number of interesting features. Some of the important ones will be described here.

3.1. Same regression coefficients and predictions as the original N-PLS

A practical important aspect of the new algorithm is how the predictions relate to the predictions of the old algorithm. It holds that the N-PLS model produces the exact same regression coefficients and same predictions as the original N-PLS as nothing is changed with respect to the prediction part. The only change is the actual model $\hat{\underline{\mathbf{X}}}$ of $\underline{\mathbf{X}}$, but as this model is calculated separately from the main algorithm as explained in Table 1, the same predictions will be obtained. As for the original N-PLS, regression coefficients working on the original data $\mathbf{X}$ can be calculated straightforwardly [9,30] and are identical to these.

3.2. Better fit of independent data

When the number of components is higher than one, the fit to $\underline{\mathbf{X}}$ will be better for the new algorithm. It may be possible to construct synthetic cases where the fits will be identical, but for all practical purposes, the fit will always be better. This does not necessarily imply that there will be a large difference. That depends on the data. However, even for data that *are* trilinear and perfectly described by PARAFAC, the original N-PLS model does not manage to describe the systematic variation in the data, which is also shown next.

3.3. Perfect trilinear data are perfectly modeled with new N-PLS model

Consider a data set that is perfectly trilinear for which a regression model is built for predicting some appropriate property. Assume for simplicity that all (PARAFAC) components are necessary for predicting the dependent variable. If the data are rank F, an F-component PARAFAC model can model the data

perfectly; say $\underline{\mathbf{X}} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathrm{T}}$. It also holds that an N-PLS model can predict the relevant property using an F-component model. This can be shown as follows.

If **A**, **B** and **C** have full column rank, then rank $r(\underline{\mathbf{X}}) = r(\mathbf{X}) = r(\mathbf{A}) = F$.

Proof. The matrices **B** and **C** have full column rank; hence, their k-rank equals F and $JK \geq F$. The product $\mathbf{C} \odot \mathbf{B}$ ($JK \times F$) has full column rank if $k_{\mathrm{B}} + k_{\mathrm{C}} \geq F + 1$ [29], where k_{B} is the k-rank of **B**. Therefore, $(\mathbf{C} \odot \mathbf{B})^{\mathrm{T}}$ has full row rank. Then, the rank $r(\mathbf{X}) = r(\mathbf{A}) = F$ [20]. Hence, $r(\underline{\mathbf{X}}) = r(\mathbf{X}) = r(\mathbf{A}) = F$. □

In N-PLS, each score vector $\boldsymbol{t}$ is a linear combination of **X** variables: $\boldsymbol{t} = \mathbf{X}(\boldsymbol{w}^{K} \otimes \boldsymbol{w}^{\mathrm{J}})$. When **X** is not deflated, then the weights for a new component are found as the solution of

$$\max \mathbf{y}^{\mathrm{T}}(\mathbf{I} - \mathbf{T}\mathbf{T}^{+})\mathbf{X}(\mathbf{w}^{\mathrm{K}} \otimes \mathbf{w}^{\mathrm{J}}) = \max \mathbf{y}^{\mathrm{T}}(\mathbf{I} - \mathbf{T}\mathbf{T}^{+})\mathbf{t}, \quad (16)$$

where **T** holds the score vectors found so far. Clearly for the maximizing $\boldsymbol{t}$, it must hold true that $(\mathbf{I} - \mathbf{T}\mathbf{T}^{+})\boldsymbol{t} \neq \mathbf{0}$. In other words, $\boldsymbol{t}$ must have a component outside the column space of **T**. As a consequence, the score matrix has full column rank. For an F-factor N-PLS model, it therefore holds that range(**T**) = range(**X**) = range(**A**) and the property $\boldsymbol{y}$ can equally well be predicted from the PARAFAC scores **A**, as from the N-PLS scores **T**.

In the original version of N-PLS, the trilinear N-PLS model will not fit the $\underline{\mathbf{X}}$ data perfectly. This is so because PARAFAC is unique and therefore also the only rank F trilinear model that can fit the data perfectly. The original N-PLS model of $\underline{\mathbf{X}}$ is also rank F, but as the N-PLS model is not identical to the PARAFAC model, the fit will be different and, hence, worse. However, the *new* N-PLS model will also fit the data perfectly because the score and weight matrices span the same spaces as the component matrices of PARAFAC. To see this, observe the following. For trilinear $\underline{\mathbf{X}}$, the data follow the model

$$\mathbf{X} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathrm{T}}. \quad (17)$$

The N-PLS weights $\boldsymbol{w}^{\mathrm{J}}$ and $\boldsymbol{w}^{\mathrm{K}}$ follow from the SVD of $\mathbf{Z} = \mathbf{B}\mathbf{D}\mathbf{C}^{\mathrm{T}}$, where **D** is a diagonal matrix with the vector $\boldsymbol{d} = \boldsymbol{y}^{\mathrm{T}}(\mathbf{I} - \mathbf{T}\mathbf{T}^{+})\mathbf{A}$ along the diagonal. This follows from Eqs. (16) and (17). Notice that **D**, i.e. the diagonal $\boldsymbol{d}$, changes with every dimension, unlike **B** and **C**. As a result, for any dimension f, $\boldsymbol{w}^{\mathrm{J}}$ and $\boldsymbol{w}^{\mathrm{K}}$ will lie in the space spanned by the columns of **B** and **C**, respectively. As **Z** changes for each successive dimension, we conjecture that the dominant left and right singular vectors $\boldsymbol{w}^{\mathrm{J}}$ and $\boldsymbol{w}^{\mathrm{K}}$ do not lie completely in the spaces spanned by their predecessors, i.e. for the full F-factor N-PLS model we have rank $r(\mathbf{W}^{\mathrm{J}}) = r(\mathbf{W}^{\mathrm{K}}) = F$. While we have no proof for this conjecture, we can ascertain the weaker condition that for successive components, the composite weight vectors $\boldsymbol{w}^{\mathrm{K}} \otimes \boldsymbol{w}^{\mathrm{J}}$ form an independent set of vectors. This follows because the composite weight vectors give rise to the independent set of score vectors **T**. So we have $r(\mathbf{W}^{\mathrm{K}} \odot \mathbf{W}^{\mathrm{J}}) = r(\mathbf{T}) = F$ which is a necessary requirement for the above conjecture. Proceeding from the assumption that the columns of the N-PLS weight matrices $\mathbf{W}^{\mathrm{J}}$ and $\mathbf{W}^{\mathrm{K}}$ are a basis for the column-space of **B** and **C**, respectively, we have $\mathbf{W}^{\mathrm{J}}(\mathbf{W}^{\mathrm{J}})^{+}\mathbf{B} = \mathbf{B}$ and $\mathbf{W}^{\mathrm{K}}(\mathbf{W}^{\mathrm{K}})^{+}\mathbf{C} = \mathbf{C}$, as we have $\mathbf{T}\mathbf{T}^{+}\mathbf{A} = \mathbf{A}$. As consequence, the Tucker3-type modeling of trilinear $\mathbf{X} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathrm{T}}$, projecting $\underline{\mathbf{X}}$ onto $\mathbf{T}\mathbf{T}^{+}$ in the first mode, onto $\mathbf{W}^{\mathrm{J}}(\mathbf{W}^{\mathrm{J}})^{+}$ in the second mode and onto $\mathbf{W}^{\mathrm{K}}(\mathbf{W}^{\mathrm{K}})^{+}$ in the third mode, leaves $\underline{\mathbf{X}}$ completely intact.

A simulated example will be given to illustrate how the change to subspaces leads to a more reasonable model of $\underline{\mathbf{X}}$.

Let

$$\mathbf{A} = \begin{bmatrix} 3 & 9 \\ 4 & 8 \\ 2 & 7 \\ 3 & 6 \\ 1 & 5 \\ 2 & 4 \\ 9 & 1 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 1 \\ 4 & 2 \\ 5 & 1 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 3 \\ 3 & 2 \\ 2 & 3 \\ 8 & 4 \end{bmatrix},$$

$$\mathbf{r} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Let the $7 \times 5 \times 6$ array $\underline{\mathbf{X}}$ be given as

$$x_{ijk} = \sum_{f=1}^{2} a_{if} b_{jf} c_{kf}, \quad i = 1, \ldots, 7;$$

$$j = 1, \ldots, 5; \quad k = 1, \ldots, 6;$$

and let the 7×1 vector $\boldsymbol{y}$ be given as

$$y_i = r_1 \mathbf{t}_{i1} + r_2 \mathbf{t}_{i2}. \quad i = 1, \ldots, 7;$$

Thus, the independent data are perfectly trilinear. A two-component PARAFAC model can fit $\underline{\mathbf{X}}$ perfectly but using the original version of N-PLS, only 88.95% of the variation in $\underline{\mathbf{X}}$ is explained using centered data. The fit to $\boldsymbol{y}$ is perfect as expected. Using the modified version of N-PLS, 100% of the variation in $\underline{\mathbf{X}}$ is explained. The model of the independent data are now of a Tucker3-like structure with a core array with non-zero off-diagonal elements. Centering these data is not mandatory, as there are no offsets, but the conclusion remains the same when uncentered data are used.

3.4. Rotational freedom enabling resampling-assessment of parameters

As discussed by Martens and Martens [21], it is feasible to investigate the stability of a model by assessing the variability in the parameters when estimated from different subsets. Before assessments can be made, any mathematical indeterminacies must be eliminated or fixed. For the two-way PLS model, there is a rotational freedom in the bilinear model. This means that even though (possibly noise-free) data are used, scores, loadings, etc., obtained from different subsets will differ. By the use of suitable rotations, this problem was alleviated in two-way PLS. However, the parameters in the original trilinear N-PLS model cannot be rotated. Rotating a trilinear model will result in a lower fit, and, hence, change the model. For the new N-PLS model, on the other hand, rotational freedom exists, because the structure is equivalent to a Tucker3 model. The N-PLS model of $\underline{\mathbf{X}}$ can be written as

$$\mathbf{X} = \mathbf{T}\mathbf{G}(\mathbf{W}^{\mathrm{K}} \otimes \mathbf{W}^{\mathrm{J}})^{\mathrm{T}} + \mathbf{E}. \tag{18}$$

This model has rotational freedom because any set of non-singular matrices $\mathbf{Q}(F \times F)$, $\mathbf{R}(F \times F)$, $\mathbf{S}(F \times F)$ can be used to rotate the model according to

$$\begin{aligned}\mathbf{X} &= \mathbf{T}\mathbf{Q}\mathbf{Q}^{-1}\mathbf{G}(\mathbf{W}^{\mathrm{K}}\mathbf{S}\mathbf{S}^{-1} \otimes \mathbf{W}^{\mathrm{J}}\mathbf{R}\mathbf{R}^{-1})^{\mathrm{T}} + \mathbf{E} \\ &= \mathbf{T}\mathbf{Q}\mathbf{Q}^{-1}\mathbf{G}(\mathbf{S}^{-1} \otimes \mathbf{R}^{-1})^{\mathrm{T}}(\mathbf{W}^{\mathrm{K}}\mathbf{S} \otimes \mathbf{W}^{\mathrm{J}}\mathbf{R})^{\mathrm{T}} + \mathbf{E} \\ &= \tilde{\mathbf{T}}\tilde{\mathbf{G}}(\tilde{\mathbf{W}}^{\mathrm{K}} \otimes \tilde{\mathbf{W}}^{\mathrm{J}})^{\mathrm{T}} + \mathbf{E}. \end{aligned} \tag{19}$$

where $\tilde{\mathbf{T}} = \mathbf{T}\mathbf{Q}$, $\tilde{\mathbf{W}}^{\mathrm{J}} = \mathbf{W}^{\mathrm{J}}\mathbf{R}$, $\tilde{\mathbf{W}}^{\mathrm{K}} = \mathbf{W}^{\mathrm{K}}\mathbf{S}$ and $\tilde{\mathbf{G}} = \mathbf{Q}^{-1}\mathbf{G}(\mathbf{S}^{-1} \otimes \mathbf{R}^{-1})^{\mathrm{T}}$. This shows that the weights and scores can be rotated by appropriate counter-rotation of the core array. The original N-PLS model equals a restricted version of the new model with the core array fixed to a super-diagonal array with ones in the super-diagonal. With such a core array, no rotations are possible because the counter-rotated array will not equal a super-diagonal array but rather have elements in all positions. This follows directly from the uniqueness of the PARAFAC model [18,28].

Because the new N-PLS model is subspace based, the rotational freedom allows for the parameters of the model to be post-rotated. Hence, the stability across subset estimates can be assessed. The actual rotation task can be accomplished in several different ways and is not the subject of this paper. Suffice here to say that ordinary orthogonal rotations cannot be used because the weights and scores are not orthogonal in N-PLS.

It will be shown with a real data set that the introduction of rotational freedom enables a more reasonable ground for parameter comparison. For 268 sugar samples dissolved in water, a fluorescence landscape was obtained. The wet-chemical quality parameter color was determined according to standard procedures. The data have been described in detail elsewhere [7]. An N-PLS model was built using the new algorithm and three components found to be suitable, assessed by a segmented cross-validation with eight contiguous segments and no scaling. Two different models were fitted: one using the first 10 and last 100 samples (sample set 1) and one using the first 100 and the last 10 samples (sample set 2). These two models approximately span the same type of samples but in widely different proportions due to the difference in sampling time (sample number is also a time scale reflecting the three months of operation of the sugar factory). The emission mode weights of these two models are shown in Fig. 1. Note that these weights are identical for both the old and new algorithm. As is readily seen, the raw weights are not similar (left). However, using the new algorithm, it is possible to transform the weights because of the rotational freedom of the model. Upon transformation (right), the weights are indeed very similar indicating that the subspaces spanned in the two different subsets are stable. The transformation matrix was obtained by

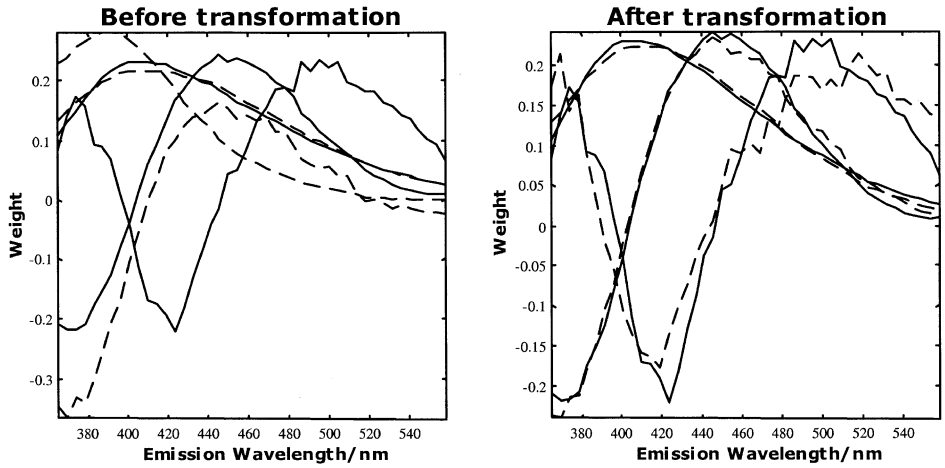

Fig. 1. Emission weights from N-PLS model on two different subset of samples. Left, the raw weights are shown with unbroken lines corresponding to sample set 1 and the dashed lines corresponding to sample set 2. To the right, the same weights are shown but upon transforming ('rotating') the sample set 2 weights to optimal agreement with sample set 1 weights.

simply regressing the weight matrix of sample set 2 onto the matrix from sample set 1. This way, the weights from sample set 2 stays in their own subspaces but are represented so that they are as close to the weights from sample set 1 as possible.

3.5. Consequences for multi-way dependent data

If the dependent data are also multi-way, then it follows immediately that the N-PLS model of the array $\underline{Y}$ must also be a subspace model. Even if the dependent data are known to be approximately trilinear, the N-PLS model will not be able to model such data through a trilinear model. This was indirectly shown in Section 3.3 where the trilinear independent data could not be modeled by a trilinear N-PLS model. Hence, instead of using a PARAFAC-like model of $\underline{Y}$, a Tucker3-like model is used in exactly the same way as the model of $\underline{\mathbf{X}}$ is changed. The result will be a better-fitting model with more reasonable least squares properties.

4. Conclusion

The theoretical difference between low-rank and subspace approximation has been described. A logical implication of this difference is the suggested modification of the calculation of the model of $\underline{\mathbf{X}}$ in N-PLS leading to a model with more well-defined and sound properties than the original N-PLS model.

Acknowledgements

With this paper, the first author would like to formally thank Elsevier Science for deciding that he should receive the 3rd Elsevier Chemometrics award 2000. This is a tremendous honor and the first author would like to dedicate the award to Prof. Lars Munck at KVL whose support and enthusiasm for his associates and for chemometrics in general is overwhelming and a source of inspiration. The algorithm used in this paper is available from the home page of Food Technology at http://www.models.kvl.dk as part of the *N*-way toolbox [1]. The first author thanks for financial support through LMC (Center for Advanced Food Studies), EC project NwayQual GRD1-1999-10377 and AQM (Advanced Quality Monitoring) supported by the ministries of research and industry.

References

[1] C.A. Anderson, R. Bro, The N-way toolbox for MATLAB, Chemom. Intell. Lab. Syst. 52 (2000) 1–4.
[2] I.E. Bechmann, Second-order data by flow injection analysis with spectrophotometric diode-array detection and incorpo-

rated gel-filtration chromatographic column. Talanta 44 (1997) 585–591.
[3] R. Boqué, A.K. Smilde, Monitoring and diagnosing batch processes with multiway covariates regression models, AIChE J. 45 (1999) 1504–1520.
[4] R. Bro, Multiway calibration, Multi-linear PLS. J. Chemom. 10 (1996) 47–61.
[5] R. Bro, PARAFAC. Tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149–171.
[6] R. Bro, Multi-way Analysis in the Food Industry. Models, Algorithms, and Applications, Ph.D. thesis, University of Amsterdam (NL), 1998.
[7] R. Bro, Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis, Chemom. Intell. Lab. Syst. 46 (1999) 133–147.
[8] R. Bro, H. Heimdal, Enzymatic browning of vegetables. Calibration and analysis of variance by multiway methods, Chemom. Intell. Lab. Syst. 34 (1996) 85–102.
[9] S. de Jong, Regression coefficients in multilinear PLS, J. Chemom. 12 (1998) 77–81.
[10] P. Geladi, H. Bergner, L. Ringqvist, From experimental design to images to particle size histograms to multiway analysis, An example of peat dewatering, J. Chemom. 14 (2000) 197–211.
[11] R.A. Harshman, Foundations of the PARAFAC procedure: models and conditions for an 'explanatory' multi-modal factor analysis, UCLA Work. Pap. Phonetics 16 (1970) 1–84.
[12] R.A. Harshman, P. Ladefoged, L. Goldstein, Factor analysis of tongue shapes, J. Acoust. Soc. Am. 62 (1977) 693–707.
[13] K. Hasegawa, M. Arakawa, K. Funatsu, Rational choice of bioactive conformations through use of conformation analysis and 3-way partial least squares modeling, Chemom. Intell. Lab. Syst. 50 (2000) 253–261.
[14] K. Hasegawa, M. Arakawa, K. Funatsu, 3D-QSAR study of insecticidal neonicotinoid compounds based on 3-way partial least squares model, Chemom. Intell. Lab. Syst. 47 (1999) 33–40.
[15] H. Heimdal, R. Bro, L.M. Larsen, L. Poll, Prediction of polyphenol oxidase activity in model solutions containing various combinations of chlorogenic acid, (−)-epicatechin, O_2, CO_2, temperature and pH by multiway analysis, J. Agric. Food. Chem. 45 (1997) 2399–2406.
[16] H.A.L. Kiers, Towards a standardized notation and terminology in multiway analysis, J. Chemom. 14 (2000) 105–122.
[17] P.M. Kroonenberg, Three-mode Principal Component Analysis. Theory and Applications, DSWO Press, Leiden, 1983.
[18] J.B. Kruskal, Rank, decomposition, and uniqueness for 3-way and N-way arrays, in: R. Coppi, S. Bolasco (Eds.), Multiway Data Analysis, Elsevier, Amsterdam, 1989, pp. 8–18.
[19] P. Ladefoged, R.A. Harshman, L. Goldstein, L. Rice, Generating vocal tract shapes from formant frequencies, J. Acoust. Soc. Am. 64 (1978) 1027–1035.
[20] G. Marsaglia, G.P.H. Styan, Equalities and inequalities for ranks of matrices, Linear Multilinear Algebra 2 (1974) 269–292.
[21] H. Martens, M. Martens, Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR), Food Qual. Preference 11 (2000) 5–16.
[22] J. Nilsson, S. de Jong, A.K. Smilde, Multiway calibration in 3D QSAR, J. Chemom. 11 (1997) 511–524.
[23] J. Nilsson, E.J. Homan, A.K. Smilde, C.J. Grol, H. Wikström, A multiway 3D QSAR analysis of a series of (*S*)-*N*-[(1-ethyl-2-pyrrolidinyl)methyl]-6-methoxybenzamides, J. Comput.-Aided Mol. Des. 12 (1998) 81–93.
[24] L. Nørgaard, R. Bro, PLS regression in the food industry. A study of N-PLS regression and variable selection for improving prediction errors and interpretation, in: M. Tenenhaus, A. Morineau (Eds.), Les Methode PLS, Symposium International PLS'99. Cisia-Ceresta, 1999, pp. 187–202.
[25] K.P. Ossenkopp, D.S. Mazmanian, The measurement and integration of behavioral variables: aggregation and complexity as important issues, Neurobehav. Toxicol. Teratol. 7 (1985) 95–100.
[26] K.P. Ossenkopp, L. Sorenson, D.S. Mazmanian, Factor analysis of open-field behavior in the rat (*rattus norvegicus*)—application of the 3-way PARAFAC model to a longitudinal data set, Behav. Processes 31 (1994) 129–144.
[27] C.R. Rao, S. Mitra, Generalized Inverse of Matrices and its Applications, Wiley, New York, 1971.
[28] N.D. Sidiropoulos, R. Bro, On the uniqueness of multilinear decomposition of N-way arrays, J. Chemom. 14 (2000) 229–239.
[29] N.D. Sidiropoulos, R. Bro, G.B. Giannakis, Parallel factor analysis in sensor array processing, IEEE Trans. Signal Process. 48 (2000) 2377–2388.
[30] A.K. Smilde, Comments on multilinear PLS, J. Chemom. 11 (1997) 367–377.
[31] A.K. Smilde, H.A.L. Kiers, Multiway covariates regression models, J. Chemom. 13 (1999) 31–48.
[32] L. Ståhle, Aspects of the analysis of three-way data, Chemom. Intell. Lab. Syst. 7 (1989) 95–100.
[33] K.D. Zissis, R. Brereton, S. Dunkerley, R.E.A. Escott, Two-way, unfolded three-way and three-mode partial least squares calibration of diode-array HPLC chromatograms for the quantification of low-level pharmaceutical impurities, Anal. Chim. Acta 384 (1999) 71–81.