ELSEVIER

Chemometrics and Intelligent Laboratory Systems 47 (1999) 189–204

Chemometrics and intelligent laboratory systems

# A new criterion for simple-structure transformations of core arrays in *N*-way principal components analysis

René Henrion [a,*], Claus A. Andersson [b]

[a] *Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, D-10117 Berlin, Germany*

[b] *Royal Veterinary and Agricultural University, Department of Food Science, Chemometrics Group, DK-1871 Copenhagen, Denmark*

**Abstract**

Among the possible (orthogonal) transformations of core arrays in *N*-way principal components analysis (PCA), the conventional approach of body diagonalization turns out not to provide the simplest structure (in the sense of minimizing the number of significant entries). As an alternative, the maximization of the variance-of-squared core entries is proposed. Both criteria are equivalent in a two-way constellation but may differ markedly for $N \geq 3$. Actually, using the variance criterion may provide more insight into the rank structure of the given data, and it is also easily applied to general rectangular core arrays. In order to clarify the relation between body diagonality and variance-of-squares, we prove the following main result of the paper: If some cubic *N*-way core array can be transformed to exact body diagonality, then the same transformation yields maximum variance-of-squared entries. This result implies the equivalence in the two-way case mentioned above. A solution algorithm is formulated and illustrated with a small numerical example. The application to data examples from environmental chemistry and chromatographic analysis is briefly discussed. 

*Keywords:* *N*-way principal components analysis; Tucker3 model; Core array; Simple-structure transformation; Body diagonality; Variance-of-squares

## 1. Introduction

*N*-way data analysis has become an efficient tool for solving chemometric problems which are based on complex (*N*-dimensional or *N*-way) data arrays as they arise, for instance, from hyphenated instrumentation. For early papers in this direction, we refer to Refs. [1,2]. Since then, a lot of contributions mainly to three-way data analysis have appeared. Chemometrically-oriented introductions to three-way analysis may be found in Refs. [3,4]. Meanwhile, at least the case $N = 4$ must be considered practically relevant (e.g., emission/excitation data from fluorescence measurements of different samples under changing conditions like pH [5]). Maybe the most important methods involved are Parallel Factor Analysis (PARAFAC) [6], Canonical Decomposition (CANDECOMP) [7] and the Tucker3 model of (three-way) Principal Components Analysis (PCA) [8], but also some variants of three-way Partial Least Squares (PLS) [9–11].

* Corresponding author. Fax: +49-30-2044975; E-mail: henrion@wias-berlin.de

0169-7439/99/$ - see front matter 
PII: S0169-7439(98)00209-3

The present paper addresses a specific problem of $N$-way PCA (for an introduction see Ref. [12]). More precisely, a new approach of transforming core arrays to simple structure is proposed and compared to the conventional diagonalization procedure.

## 2. Transformation of core arrays to simple structure

The general model of $N$-way PCA is (compare Ref. [13])

$$\operatorname{vec}\mathbf{X} \approx (\mathbf{A}_1 \otimes \cdots \otimes \mathbf{A}_N)\operatorname{vec}\mathbf{C} \tag{1}$$

Here, $\mathbf{X}$ denotes an $N$-way data array of order $(n_1, \ldots, n_N)$, the $\mathbf{A}_i$ are component matrices of orders $(n_i, s_i)$, where, usually, the $s_i$'s are small numbers for the purpose of data reduction, and $\mathbf{C}$ is the so called $N$-way core array of order $(s_1, \ldots, s_N)$. Furthermore, vec and $\otimes$ denote the vectorization operator and the Kronecker product, respectively. It is emphasized, that in the following vec will be understood as an operator unfolding the given array in a way that the first index runs fast and the last index slowly. For matrices, this corresponds to the usual stacking of columns (note that there is some inconsistency in the definition in Ref. [13], pages 30 and 363). Accordingly, we understand the Kronecker product in the sense

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} b_{11}\mathbf{A} & \cdots & b_{1m}\mathbf{A} \\ \vdots & \ddots & \vdots \\ b_{n1}\mathbf{A} & \cdots & b_{nm}\mathbf{A} \end{pmatrix}.$$

The aim of $N$-way PCA is, given $\mathbf{X}$, to find the component matrices (sometimes additionally required to be column-wise orthonormal) and the core array such that the above approximation is optimal in the sense of least squares deviations. The component matrices $\mathbf{A}_i$ allow to plot the basic factors in each of the $N$ modes influencing the total variation in the array $\mathbf{X}$. The core array $\mathbf{C}$, on the other hand, indicates how factor combinations from different modes interact. For instance, in a three-way constellation ($N = 3$) with orthonormal component matrices, the squared core element $c_{121}^2$ measures the amount of data variance covered by combining the first factor of the first mode with the second factor of the second mode and the first factor of the third mode. Such consideration of interactions is not necessary in conventional two-mode PCA since the core matrix can always be diagonalized there. Hence, the information in data tables is exhausted efficiently by independent extraction of successive factors for objects and variables. The explanatory effect of interactions (say by combining the first factor of objects with the second factor of variables) can always be made zero. Things become different for data arrays of dimensions larger than two. Of course, one might still suppress interactions by restricting the model of decomposition, which is the case in the PARAFAC approach. However, such decomposition is no longer the most efficient one. Indeed, using the Tucker3 model with possible interactions between different factors of different modes, the same amount of data variation as in a PARAFAC decomposition might be explained by a smaller number of factors. On the other hand, interactions are more difficult to interpret. In particular, a generalization of the well-known bi-plots from two-way PCA to 'tri-plots' or '$N$-plots' is not straightforward. Therefore, a common strategy is to simplify the interaction structure among factors after a Tucker3 analysis as far as possible. This is the aim of simple-structure transformations of core arrays. In the ideal case, one could remove all the interactions and would arrive at the same result as with a direct PARAFAC approach. Unfortunately, this is not possible in general, so one has to be satisfied with structures simplified according to suitable criteria which will be discussed in the sequel. For an illustration of the PCA decomposition according to the Tucker3 model (1) and for an interpretation of the core elements, we refer to the data example in Section 5.

While the optimal component matrices in (1) may be determined by an alternating least squares algorithm (see Ref. [14]), the corresponding optimal core array results from them according to (compare Ref. [13])

$$\operatorname{vec}\mathbf{C} = \left(\mathbf{A}_1^{\mathrm{T}} \otimes \cdots \otimes \mathbf{A}_N^{\mathrm{T}}\right)\operatorname{vec}\mathbf{X}$$

On the other hand, neither the component matrices $\mathbf{A}_i$, nor the core array $\mathbf{C}$ are uniquely determined in the decomposition (1). Indeed, using nonsingular matrices $\mathbf{P}_i$ of orders $(s_i, s_i)$, this same decomposition transforms to (by $\mathbf{P}_i\mathbf{P}_i^{-1} = \mathbf{I}_{s_i}$):

$$\begin{aligned}\operatorname{vec}\mathbf{X} &\approx (\mathbf{A}_1 \otimes \cdots \otimes \mathbf{A}_N)\operatorname{vec}\mathbf{C}\\ &= (\mathbf{A}_1 \otimes \cdots \otimes \mathbf{A}_N)(\mathbf{I}_{s_1} \otimes \cdots \otimes \mathbf{I}_{s_N})\operatorname{vec}\mathbf{C}\\ &= (\mathbf{A}_1 \otimes \cdots \otimes \mathbf{A}_N)(\mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_N)(\mathbf{P}_1^{-1} \otimes \cdots \otimes \mathbf{P}_N^{-1})\operatorname{vec}\mathbf{C}\\ &= (\mathbf{A}'_1 \otimes \cdots \otimes \mathbf{A}'_N)\operatorname{vec}\mathbf{C}',\end{aligned}$$

where $\mathbf{A}'_i = \mathbf{A}_i\mathbf{P}_i$ are the transformed versions of the original component matrices $\mathbf{A}_i$, and $\mathbf{C}'$ is the new core array which relates to the old one through

$$\operatorname{vec}\mathbf{C}' = (\mathbf{P}_1^{-1} \otimes \cdots \otimes \mathbf{P}_N^{-1})\operatorname{vec}\mathbf{C} \tag{2}$$

Along with the transformed core $\mathbf{C}'$, the $\mathbf{A}'_i$ provide the same approximation of $\mathbf{X}$ as the original $\mathbf{A}_i$ and $\mathbf{C}$. Actually, corresponding to the manifold of possible nonsingular matrices $\mathbf{P}_i$, there is an infinite number of equally good approximations of the given data array. For simplicity and comparison to existing methods, we restrict the further presentation mainly to orthogonal transformations. Then, the inverses $\mathbf{P}_i^{-1}$ in (2) simply become the transposed matrices $\mathbf{P}_i^{\mathrm{T}}$. Furthermore, in this case, the transformed component matrices $\mathbf{A}'_i$ remain orthonormal if so were the original ones $\mathbf{A}_i$ and, hence, the entries of the transformed core $\mathbf{C}'$ may be interpreted as variance contributions of factor combinations from different (transformed) components $\mathbf{A}'_i$ as it held true for the original core $\mathbf{C}$ and the original components $\mathbf{A}_i$ (see above).

A reasonable choice of a particular solution in (1) would require the core array to have as few significant entries as possible in order to arrive at a model with a minimum number of describing factors. Doing so, the interpretational effort of the results obtained may be considerably reduced. For the purpose of illustration, consider the following three-mode core arrays of order (2,2,2), in unfolded form (i.e., the third index refers to the slice left or right to the separation line while the first two indices are read in the slices as for usual matrices):

$$\left(\begin{array}{cc|cc} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{array}\right);\quad \left(\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right);\quad \left(\begin{array}{cc|cc} \sqrt{2} & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \end{array}\right) \tag{3}$$

All these cores can be transformed into each other by using appropriate orthogonal matrices in (2). It is clear that the structure simplifies from the left to the right: in the situation of the very left core one would have to interpret eight equally important factor combinations of the $N$-way PCA model. This number reduces to four in the second and to two in the third core. Sometimes, additional knowledge about the model allows to fix specific core elements as zero and to consider restricted core arrays from the very beginning of analysis. This approach is discussed in Ref. [15] and it has been applied to a selected calibration problem of analytical chemistry in Ref. [16]. In general, however, the insight into the problem structure is rather limited, so premature restrictions of the core might not be advisable. Instead, one can admit a completely loaded core as the output of any $N$-way PCA algorithm and afterwards use the degree of freedom in the decomposition (1) discussed above, in order to find transformation matrices $\mathbf{P}_i$, such that the new core resulting from Eq. (2) has a simple structure. In the following, we restrict considerations to cubic core arrays of order $(s, \ldots, s)$. This restriction is not necessary for the approach to be described here, but it allows comparison with existing methods. In the sense of the discussion above, one may imagine several criteria for measuring 'simple structure'. In Ref. [17], the simple structure was formulated as a slice-wise diagonality of the (three-way) core array. The theoretical argument behind this is, that

in case of a possible exact slice-wise diagonalization, the Tucker3 model reduces to a PARAFAC model as soon as one renounces the orthogonality of the components. For instance, both, the second and third core in (3) are slice-wise diagonal.

On the other hand, with orthogonality constraints imposed on the components—and this may have certain advantages—Tucker3 reduces to PARAFAC (also with orthogonal components) only in case that the core has so-called body diagonal shape. By this, it is meant that the entries of $\mathbf{C}$ satisfy $\mathbf{C}_{i_1 \dots i_N} = 0$ if not $i_1 = \cdots = i_N$. None of the cores in (3) are body diagonal, since, in all cases there are nonzero entries outside the left upper and right lower corners of the unfolded arrays. Indeed, in this data example, there does not exist any orthogonal transformation of the given cores to exact body diagonality. Body diagonality is a desirable property of the core in that it avoids interaction between unequal components from different modes. From a more practical standpoint, body diagonality allows superposition and joint interpretation of component plots. As in the example, exact body diagonalization of core arrays fails in most cases. At least, one can try to fit body diagonality as close as possible, which amounts to maximize the sum of squared body diagonal entries $\text{diag} = \sum_{i=1}^{s} C^2_{i,\dots,i}$. The total sum of squared core entries being invariant under the transformation (2) with orthogonal $\mathbf{P}_i$, this means to minimize the squared off-diagonal entries, hence, a body diagonal shape of the core is approached.

In (3), one computes the values $\text{diag}_1 = 1$, $\text{diag}_2 = 2$, $\text{diag}_3 = 2$ for the succeeding cores. Actually, the value 2 represents the maximum of diag among all possible transformations (2) with orthogonal $\mathbf{P}_i$, so the second and third core are not only slice-wise diagonal, but they have maximum body diagonal shape at the same time. If exact body diagonality was possible here, then one should obtain $\text{diag} = 4$, a value which is equal to the total sum of squares in the cores. An algorithm for (maximum) body diagonalization of three-way core arrays was suggested in Ref. [18]. In Ref. [19], theoretical bounds for the success of body diagonalization of three-way core arrays were derived. For the special case of cores of order (2,2,2)—which is important in exploratory diagram analysis of components—a degree of 80–90% of body diagonality (= diag divided by the total sum of squared entries) may be expected on the average. This makes diagonalization a useful approach for obtaining simple structure of cores.

Simple structure of the core can be understood, however, in a sense different from diagonality. It seems natural to look for transformations providing the smallest number of significant core entries, or equivalently, the largest number of negligible (if not zero) entries. This is a direct formulation of minimizing the effort of interpretation of components. It is intuitively clear, that this aim is not automatically realized by body diagonalization since the latter restricts not only the number of significant elements but simultaneously the shape of the core. Renouncing the diagonality shape, one has hope to find cores with fewer significant entries although not necessarily located on the diagonal. Although, due to its simplicity, the example (3) is not capable of completely highlighting this aspect, it suffices to demonstrate that maximum body diagonality is not directly related with simple structure. As already stated above, both the second and third core in (3) have the same degree of body diagonality while the structure of the third core is much simpler with only two significant entries as compared to the second core. Much more evident examples will be provided in the following sections.

## 3. Variance-of-squares

### *3.1. Definition of the criterion*

As a quantitative criterion directly oriented towards maximizing the number of negligible entries in the core, we propose to use the variance of the squared entries of the core. More precisely, we define

$$\text{var} = \sum_{i_1=1}^{s} \cdots \sum_{i_N=1}^{s} \left( C^2_{i_1 \dots i_N} - \overline{C} \right)^2 \tag{4}$$

where

$$\overline{C} = \sum_{i_1=1}^{s} \cdots \sum_{i_N=1}^{s} C_{i_1 \ldots i_N}^2 / s^N = [\mathrm{vec}\,\mathbf{C}]^{\mathrm{T}} [\mathrm{vec}\,\mathbf{C}] / s^N \tag{5}$$

is the mean of squared entries.

Eq. (4), as a numerical entity, is identical with the quartimax criterion defined for simple structure transformations of loading matrices in factor analysis [20], but it must not be confused with these. Optimizing the two-way quartimax measure, which relates to loading matrices rather than core matrices, does not necessarily produce simple core arrays which is the aim of the current discussion. Therefore, we keep the name 'variance-of-squares' criterion in order to avoid any confusion with concepts from factor analysis.

The justification of the variance-of-squares criterion relies on the following simple observation: If a set of vectors $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is restricted to have non-negative components $\boldsymbol{x}_i$ which sum up to a constant value, then the variance of the components attains its maximum at those vectors having exactly one component different from zero. In order to translate this result to the context of core arrays, consider now the set of vectors $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ which are vectorizations of squared entries of core arrays related by transformations of type (2). Obviously, the components $\boldsymbol{x}_i$ are non-negative, and they sum up to a constant value, since the sum of squares in a core array does not change under the considered transformation (2) with orthogonal $\mathbf{P}_i$ (compare also the first statement in the proof of the Theorem in Section 3.3). Therefore, maximum variance of the $\boldsymbol{x}_i$, which is the maximum variance of the squared core entries, aims at reducing the number of nonzero core elements to one. In Eq. (3), the third core has maximum variance-of-squares among all possible transformations. The concrete values $\mathrm{var}_1 = 0$, $\mathrm{var}_2 = 2$, $\mathrm{var}_3 = 6$ for the three cores reflect quite well the increasing simplicity of their structure.

Note that, although the justification given above relates to orthogonal transformation matrices $\mathbf{P}_i$ in (2), the variance-of-squares criterion itself may be applied to general nonsingular transformation matrices $\mathbf{P}_i$. In the special case of orthogonal transformations which we focus on in this paper, the objective of maximizing the variance-of-squares measure becomes similar to a special case of the three-mode Orthomax criterion proposed by Kiers [21]. The three-mode Orthomax measure is optimized successively for each of the three modes by maximizing the ORMAX matrix operator

$$\mathrm{ORMAX}(\boldsymbol{\Lambda}, \gamma) = \sum_{l=1}^{r} \left( \sum_{i=1}^{m} \lambda_{il}^4 - \frac{\gamma}{m} \left( \sum_{i=1}^{m} \lambda_{il}^2 \right)^2 \right) \tag{6}$$

with $\lambda_{il}$ denoting the element in the $i$th row and the $j$th column of the matrix $\boldsymbol{\Lambda}$. A scalar $\gamma$ weighs the squared mean of the squared column entries of $\boldsymbol{\Lambda}$. In the three-mode Orthomax approach, ORMAX is applied alternatingly to the three unfoldings of the core to yield an overall optimization. Setting $\gamma = 0$ for all three modes, the criterion simplifies to the three-mode Quartimax measure. This situation entails that the squared mean values of the squared entries are neglected, causing the sum of the fourth powers of the core elements to be maximized. Similarly, for orthogonal transformation matrices the variance-of-squares measure (4) will have an invariant mean value of the squares, implicitly resulting in maximization of the fourth powers of the core elements (compare (7)). Whereas, the three-mode Quartimax procedure operates on the unfoldings of the core, the variance-of-squares procedure addresses the problem by optimizing the core directly. Also, both approaches differ when general nonsingular (not just orthogonal) transformations are allowed, since the mean of squares is no longer invariant and, hence, the maximization of the variance-of-squares is no longer equivalent to the maximization of fourth powers then. In terms of understanding the effects of core transformations, we prefer the variance-of-squares measure since variance has an intuitive meaning for analysts while the Quartimax measure is somewhat abstract. Recently, an approach for simultaneous optimization of the orthogonality of the core and the component matrices has been proposed, see Ref. [22].

### 3.2. On the relations between body diagonality and variance-of-squares

Let us now check the relations between the body diagonality and variance-of-squares criteria. The second and the third core in (3) show that maximum body diagonality does not automatically provide maximum variance-of-squares. The example was for three-way cores, but what about the simpler two-way case ($N = 2$), where body diagonality reduces to conventional diagonalization of square matrices? The answer is given by the Corollary to the Theorem in Section 3.3: In the two-way case, the maximization of 'body diagonality' implies the maximization of the variance-of-squares of a quadratic core matrix. In other terms: For $N = 2$, there is no gain by introducing the variance criterion, and the core simplification is completely achieved by singular value decomposition, which is an admissible transformation in the sense of (2). This equivalence in the two-way case might explain why the consideration of the quite natural variance-of-squares criterion has been ignored so far in favour of different diagonality criteria.

A misleading feature of the example in (3) is that the core with maximum variance-of-squares (third core) is contained in—although not identical with—the set of cores having maximum body diagonality. This is not true in general. In order to obtain a more general impression, consider Fig. 1 where variance vs. diagonality plots for three different cores each subject to 5000 random orthogonal transformations are given. Here, the plots (a), (b), and (c) refer to transformations of the cores

$$\mathbf{C} = \left(\begin{array}{cc|cc} 1 & \alpha & 1 & \beta \\ \alpha & 1 & \beta & 1 \end{array}\right),$$

where $\alpha = \beta = 0$ in (a), $\alpha = \beta = -0.1$ in (b), and $\alpha = -0.1$, $\beta = 0.1$ in (c). Obviously, Fig. 1(a) relates to transformations of the cores in (3) since, for $\alpha = \beta = 0$, $\boldsymbol{C}$ is equal to the second core there. As a consequence, the three cores of (3) are contained in the plot of Fig. 1(a) as points with the coordinates (diag, var) = (1,0), (2,2), and (2,6), respectively. Note that the vertical line, joining the last two of these points, represents an infinite number of transformed cores with maximum body diagonality but with varying values for the variance-of-squares. Such a phenomenon is not stable since an arbitrarily small perturbation of the core entries (e.g., the parameters $\alpha$, $\beta$) will destroy this vertical line, and a constellation as in Fig. 1(b) and (c) is likely to occur. Here, the qualitative relationship between the diagonality and variance criteria is quite different: In Fig. 1(b), maximum body diagonality implies maximum variance-of-squares (which was not true in Fig. 1(a)), while in Fig. 1(c), the maxima of diag and var are completely unrelated: indeed, the maximum of diag leads only to a value of var, which is less than half the maximum of var. In contrast to Fig. 1(a), both situations are stable with respect to small perturbations of the core entries (due to the fact that var and diag are continuous functions of the core), hence, both of them are typically observed.

Now, the question arises, under which conditions does the one or the other situation occur. As the main result in this direction the following statement, which even relates to general (cubic) $N$-way cores, is proved in the Theorem of Section 3.3: If the given core array may be transformed according to (2) to exact (!) body diagonality, then the resulting diagonal core array has maximum variance-of-squares at the same time. This result is mainly of theoretical interest in that it connects the relation between both criteria with the structure of the core array. By contraposition, one concludes that neither the cores in Fig. 1(a) nor those in Fig. 1(c) can be transformed to exact body diagonality (since there are transformations providing maximum diagonality but not maximum variance-of-squares). From the practical point of view, one has to take into account of course that a transformation of cubic $N$-way cores to exact body diagonality is possible for $N \geq 3$ in exceptional cases only.

By the way, Fig. 1 also shows, that even minimum diagonality can lead to maximum variance-of-squares. At least for two-way matrices of order (2,2), this is not surprising since the diagonal elements may be placed as

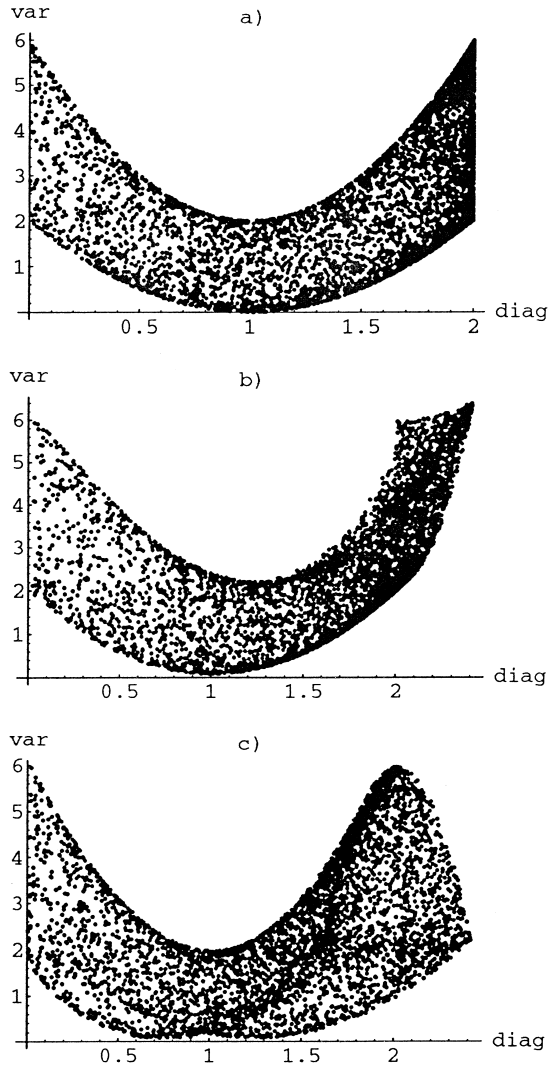


Fig. 1. Plot of variance-of-squared entries vs. body diagonality for 5000 random orthogonal transformations of three different core arrays.

well on the anti-diagonal without changing the variance. For higher order, this is no longer true as can be seen from the simple two-way example:

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

This matrix has evidently minimum diagonality with variance-of-squares equal to 2, whereas maximum diagonality is attained (after a similarity transformation) when the eigenvalues 2, $-1$, $-1$ are placed on the diagonal. This gives a variance-of-squares equal to 14, which must be the maximum according to the Corollary proved in Section 3.3. Consequently, in this example, minimum diagonality yields a variance-of-squares value which is far from the maximum.

### 3.3. Theoretical results

Now, we prove the statements referred to above. To this aim, let $\mathbf{C}$ denote a cubic $N$-way core array of order $(s,\dots,s)$. Given any $N$-tuple of orthogonal matrices $\mathbf{P}_1,\dots,\mathbf{P}_N$ of common order $(s,s)$, the following functions are introduced:

$$T(\mathbf{P}_1,\cdots,\mathbf{P}_N) = \mathrm{vec}^{-1}\left[\left(\mathbf{P}_1^{\mathrm{T}} \otimes \cdots \otimes \mathbf{P}_N^{\mathrm{T}}\right)\mathrm{vec}\,\mathbf{C}\right]$$

$$\mathrm{var}(\mathbf{P}_1,\cdots,\mathbf{P}_N) = \sum_{i_1=1}^{s} \cdots \sum_{i_N=1}^{s} \left(T^2_{i_1\cdots i_N}(\mathbf{P}_1,\cdots,\mathbf{P}_N) - \bar{T}(\mathbf{P}_1,\cdots,\mathbf{P}_N)\right)^2$$

Here, $\mathrm{vec}^{-1}$ refers to the operator which assigns to each vector with $s^N$ components the uniquely defined $N$-way array of order $(s,\dots,s)$ the vectorization of which gives this vector. Obviously, $T(\mathbf{P}_1,\dots,\mathbf{P}_N)$ is exactly the transformed core array $\mathbf{C}'$ from (2). In the second definition, $\bar{T}(\mathbf{P}_1,\dots,\mathbf{P}_N)$ denotes the mean of squares of the transformed core array (compare (5)), so var $(\mathbf{P}_1,\dots,\mathbf{P}_N)$ is the variance of squared entries in the transformed core array.

**Theorem 3.1**. *If there exist orthogonal matrices* $\mathbf{P}_i^*$ $(i=1,\dots,N)$ *of common order* $(s,s)$ *such that* $T(\mathbf{P}_1^*,\dots,\mathbf{P}_N^*)$ *is body diagonal, then* $var(\mathbf{P}_1^*,\dots,\mathbf{P}_N^*)$ *maximizes the expression* $var$ $(\mathbf{P}_1,\dots,\mathbf{P}_N)$ *among all N-tuples of orthogonal matrices* $\mathbf{P}_1,\dots,\mathbf{P}_N$ *of common order* $(s,s)$.

**Proof**. We start with the obvious observation that the mean of squares of a core array is invariant under the transformation $T$. In fact, due to the orthogonality of the $\mathbf{P}_i$, one has

$$\begin{aligned}
\bar{T}(\mathbf{P}_1,\cdots,\mathbf{P}_N) &= \left[\mathrm{vec}\,T(\mathbf{P}_1,\cdots,\mathbf{P}_N)\right]^{\mathrm{T}}\left[\mathrm{vec}\,T(\mathbf{P}_1,\cdots,\mathbf{P}_N)\right]/s^N \\
&= \left[\left(\mathbf{P}_1^{\mathrm{T}} \otimes \cdots \otimes \mathbf{P}_N^{\mathrm{T}}\right)\mathrm{vec}\,\mathbf{C}\right]^{\mathrm{T}}\left[\left(\mathbf{P}_1^{\mathrm{T}} \otimes \cdots \otimes \mathbf{P}_N^{\mathrm{T}}\right)\mathrm{vec}\,\mathbf{C}\right]/s^N \\
&= \left[\mathrm{vec}\,\mathbf{C}\right]^{\mathrm{T}}\left(\left(\mathbf{P}_1\mathbf{P}_1^{\mathrm{T}}\right) \otimes \cdots \otimes \left(\mathbf{P}_N\mathbf{P}_N^{\mathrm{T}}\right)\right)\left[\mathrm{vec}\,\mathbf{C}\right]/s^N \\
&= \left[\mathrm{vec}\,\mathbf{C}\right]^{\mathrm{T}}\left[\mathrm{vec}\,\mathbf{C}\right]/s^N \\
&= \bar{C},
\end{aligned}$$

where $\bar{C}$ refers to the mean of squares of $\mathbf{C}$ (see (5)). Therefore, the variance criterion, as a function of the chosen transformation, written as

$$\begin{aligned}
\mathrm{var}(\mathbf{P}_1,\cdots,\mathbf{P}_N) &= \sum_{i_1=1}^{s} \cdots \sum_{i_N=1}^{s} \left(T^2_{i_1\cdots i_N}(\mathbf{P}_1,\cdots,\mathbf{P}_N) - \bar{C}\right)^2 \\
&= \sum_{i_1=1}^{s} \cdots \sum_{i_N=1}^{s} T^4_{i_1\cdots i_N}(\mathbf{P}_1,\cdots,\mathbf{P}_N) + s^N\bar{C}^2 \\
&\quad - 2\bar{C}\sum_{i_1=1}^{s} \cdots \sum_{i_N=1}^{s} T^2_{i_1\cdots i_N}(\mathbf{P}_1,\cdots,\mathbf{P}_N) \\
&= \sum_{i_1=1}^{s} \cdots \sum_{i_N=1}^{s} T^4_{i_1\cdots i_N}(\mathbf{P}_1,\cdots,\mathbf{P}_N) - s^N\bar{C}^2 \qquad (7)
\end{aligned}$$

In the following, we shall make use of the known or easy to verify relations

$$(\mathbf{Q}\otimes\mathbf{R})\operatorname{vec}\mathbf{S}=\operatorname{vec}(\mathbf{QSR}^{\mathrm{T}}) \tag{8}$$

$$\sum_{ij}\mathbf{Q}_{ij}^{4}\le\operatorname{tr}[\mathbf{Q}^{\mathrm{T}}\mathbf{Q}]^{2} \tag{9}$$

between Kronecker product, matrix product, vectorization and trace of matrices **Q**, **R**, **S** with suitable orders. Now, for a cubic $N$-way array **M** of order $(s,\ldots,s)$, define its unfolding $u(\mathbf{M})$ to be the uniquely determined matrix of order $(s,s^{N-1})$ such that vec **M** = vec $u(\mathbf{M})$. Then, using (8), one gets

$$\begin{aligned}\operatorname{vec}u\big(T(\mathbf{P}_1,\cdots,\mathbf{P}_N)\big)&=\operatorname{vec}T(\mathbf{P}_1,\cdots,\mathbf{P}_N)=\big(\mathbf{P}_1^{\mathrm{T}}\otimes\cdots\otimes\mathbf{P}_N^{\mathrm{T}}\big)\operatorname{vec}u(\mathbf{C})\\&=\operatorname{vec}\big[\mathbf{P}_1^{\mathrm{T}}u(\mathbf{C})(\mathbf{P}_2\otimes\cdots\otimes\mathbf{P}_N)\big]\end{aligned}$$

Consequently, $u(T(\mathbf{P}_1,\ldots,\mathbf{P}_N))=\mathbf{P}_1^{\mathrm{T}}u(\boldsymbol{C})(\mathbf{P}_2\otimes\cdots\otimes\mathbf{P}_N)$ which implies

$$\begin{aligned}\operatorname{tr}\Big[u\big(T(\mathbf{P}_1,\cdots,\mathbf{P}_N)\big)^{\mathrm{T}}u\big(T(\mathbf{P}_1,\cdots,\mathbf{P}_N)\big)\Big]^{2}&=\operatorname{tr}\Big[\big(\mathbf{P}_2^{\mathrm{T}}\otimes\cdots\otimes\mathbf{P}_N^{\mathrm{T}}\big)u(\mathbf{C})^{\mathrm{T}}u(\mathbf{C})(\mathbf{P}_2\otimes\cdots\otimes\mathbf{P}_N)\Big]^{2}\\&=\operatorname{tr}\big(\mathbf{P}_2^{\mathrm{T}}\otimes\cdots\otimes\mathbf{P}_N^{\mathrm{T}}\big)\Big[u(\mathbf{C})^{\mathrm{T}}u(\mathbf{C})\Big]^{2}(\mathbf{P}_2\otimes\cdots\otimes\mathbf{P}_N)\\&=\operatorname{tr}\Big[u(\mathbf{C})^{\mathrm{T}}u(\mathbf{C})\Big]^{2}\end{aligned} \tag{10}$$

On the other hand, for the particular choice of transformation matrices providing body diagonality (see statement of the theorem), one has

$$\operatorname{tr}\Big[u\big(T(\mathbf{P}_1^{*},\cdots,\mathbf{P}_N^{*})\big)^{\mathrm{T}}u\big(T(\mathbf{P}_1^{*},\cdots,\mathbf{P}_N^{*})\big)\Big]^{2}=\sum_{i_1=1}^{s}\cdots\sum_{i_N=1}^{s}T_{i_1\cdots i_N}^{4}(\mathbf{P}_1^{*},\cdots,\mathbf{P}_N^{*}) \tag{11}$$

This follows from the fact that body diagonality of $T(\mathbf{P}_1^{*},\ldots,\mathbf{P}_N^{*})$ implies its unfolded copy $u(T(\mathbf{P}_1^{*},\ldots,\mathbf{P}_N^{*}))$ to have in each column and each row at most one entry different from zero. Combining (7), (9), (10) (which in particular holds for the transformation matrices $\mathbf{P}_i^{*}$) and (11), one arrives at

$$\begin{aligned}\operatorname{var}(\mathbf{P}_1,\cdots,\mathbf{P}_N)&\le\operatorname{tr}\Big[u\big(T(\mathbf{P}_1,\cdots,\mathbf{P}_N)\big)^{\mathrm{T}}u\big(T(\mathbf{P}_1,\cdots,\mathbf{P}_N)\big)\Big]^{2}-s^{N}\overline{C}^{2}\\&=\operatorname{tr}\Big[u(\mathbf{C})^{\mathrm{T}}u(\mathbf{C})\Big]^{2}-s^{N}\overline{C}^{2}\\&=\operatorname{tr}\Big[u\big(T(\mathbf{P}_1^{*},\cdots,\mathbf{P}_N^{*})\big)^{\mathrm{T}}u\big(T(\mathbf{P}_1^{*},\cdots,\mathbf{P}_N^{*})\big)\Big]^{2}-s^{N}\overline{C}^{2}\\&=\sum_{i_1=1}^{s}\cdots\sum_{i_N=1}^{s}T_{i_1\cdots i_N}^{4}(\mathbf{P}_1^{*},\cdots\mathbf{P}_N^{*})-s^{N}\overline{C}^{2}\\&=\operatorname{var}(\mathbf{P}_1^{*},\cdots,\mathbf{P}_N^{*})\end{aligned}$$

Since the $\mathbf{P}_i$'s were chosen arbitrarily among all orthogonal matrices of common order $(s,s)$, this last inequality proves that var $(\mathbf{P}_1,\ldots,\mathbf{P}_N)$ is maximized by the transformation matrices $\mathbf{P}_i*$. □

**Corollary 3.2**. *In the two-way case* ($N=2$), *maximization of diagonality implies maximization of the variance-of-squares of a quadratic core matrix*.

**Proof**. Since any (square) matrix **C** may be transformed to diagonal shape via a singular value decomposition $\mathbf{P}^{\mathrm{T}}\mathbf{CQ}$ with orthogonal **P** and **Q**, and since this is an admissible transformation in the sense of (2) (recall that $\operatorname{vec}(\mathbf{P}^{\mathrm{T}}\mathbf{CQ})=(\mathbf{P}^{\mathrm{T}}\otimes\mathbf{Q}^{\mathrm{T}})\operatorname{vec}\mathbf{C}$), the same transformation yields maximum variance-of-squares according to the Theorem. □

## 4. A transformation algorithm

In this section, we formulate an algorithm in order to find the optimal orthogonal matrices $\mathbf{P}_i$ in (2) transforming the given core array $\mathbf{C}$, which is the output of any $N$-way PCA algorithm, into one with a maximum variance-of-squares value. We omit the theoretical derivation of the algorithm and refer to Ref. [23].

1. Set $\mathbf{C}^{\text{new}} := \mathbf{C}$ (= original core array) and $\mathbf{P}_i^{\text{new}} := \mathbf{I}_{s_i}$ $(i = 1, \ldots, N)$
2. Set $j := 0$
3. Set $j := j + 1$, $\mathbf{C}^{\text{old}} := \mathbf{C}^{\text{new}}$, $\mathbf{P}_j^{\text{old}} := \mathbf{P}_j^{\text{new}}$ and compute an orthogonal matrix $\mathbf{P}$ such that $\mathbf{P}^{\mathrm{T}}\mathbf{A}$ becomes a symmetric matrix, where the general entry of $\mathbf{A}$ is $(1 \le k \le s_j;\ 1 \le l \le s_j)$

$$A_{kl} = \sum_{i_1=1}^{s_1} \cdots \sum_{i_{j-1}=1}^{s_{j-1}} \sum_{i_{j+1}=1}^{s_{j+1}} \sum_{i_N=1}^{s_N} \left( C^{2\,\text{old}}_{i_1 \cdots i_{j-1} l i_{j+1} \cdots i_N} - \overline{C} \right) C^{\text{old}}_{i_1 \cdots i_{j-1} l i_{j+1} \cdots i_N} C^{\text{old}}_{i_1 \cdots i_{j-1} k i_{j+1} \cdots i_N}$$

   Define $\mathbf{C}^{\text{new}}$ by $\text{vec}\,\mathbf{C}^{\text{new}} := (\mathbf{I}_{s_1} \otimes \ldots \otimes \mathbf{I}_{s_{j-1}} \otimes \mathbf{P} \otimes \mathbf{I}_{s_{j+1}} \otimes \mathbf{I}_{s_N})\text{vec}\,\mathbf{C}^{\text{old}}$ and $\mathbf{P}_j^{\text{new}} := \mathbf{P}_j^{\text{old}}\mathbf{P}$. If $j < N$, then goto 3.
4. If $\text{var}(\mathbf{C}^{\text{new}})$ differs significantly from $\text{var}(\mathbf{C}^{\text{old}})$, then goto 2.
5. Stop.

The final $\mathbf{C}^{\text{new}}$ is the optimally transformed core array and the final $\mathbf{P}_j$ are the corresponding transformation matrices for the transition from $\mathbf{C}$ to $\mathbf{C}^{\text{new}}$ via (2). The decisive step in this algorithm is the symmetrification of $\mathbf{P}^{\mathrm{T}}\mathbf{A}$ in 3. This can be realized by a singular value decomposition of $\mathbf{A}$ by means of orthogonal matrices $\mathbf{U}$, $\mathbf{V}$ of order $s_j$, which yields $\mathbf{UAV} = \mathbf{D}$ with diagonal $\mathbf{D}$. Then, setting $\mathbf{P} = \mathbf{U}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}$, one gets

$$\mathbf{P}^{\mathrm{T}}\mathbf{A} = \mathbf{VUA} = \mathbf{VUU}^{\mathrm{T}}\mathbf{DV}^{\mathrm{T}} = \mathbf{VDV}^{\mathrm{T}},$$

which is the desired symmetric matrix. A numerical example shall serve as an illustration of the algorithm. Consider the maximization of the variance-of-squares criterion var for the three-dimensional core array of order (2,2,2) given by

$$\mathbf{C} = \left(\begin{array}{cc|cc} C_{111} & C_{121} & C_{112} & C_{122} \\ C_{211} & C_{221} & C_{212} & C_{222} \end{array}\right) = \left(\begin{array}{cc|cc} 0 & 1 & 1 & 2 \\ 1 & 1 & 0 & 1 \end{array}\right)$$

The mean of squared entries is $\overline{C} = 1.125$ and the variance-of-squares criterion for the initial core is $\text{var}(\mathbf{C}) = 10.875$. In the first step of the algorithm, one has to compute the matrix $\mathbf{A}$ with general element

$$A_{kl} = \sum_{i_2=1}^{2} \sum_{i_3=1}^{2} \left( C^2_{l i_2 i_3} - \overline{C} \right) C_{l i_2 i_3} C_{k i_2 i_3} \qquad (k = 1,2;\ l = 1,2)$$

For instance, $A_{11} = 0 - 0.125 - 0.125 + 11.5 = 11.25$. For the whole matrix, one has

$$\mathbf{A} = \begin{pmatrix} 11.25 & -0.375 \\ 5.625 & -0.375 \end{pmatrix}$$

From singular value decomposition of this matrix, one finds that $\mathbf{P}^{\mathrm{T}}\mathbf{A}$ becomes symmetric for

$$\mathbf{P} = \begin{pmatrix} 0.911 & 0.412 \\ 0.412 & -0.911 \end{pmatrix}$$

Applying the transformation $(\mathbf{P} \otimes \mathbf{I}_2 \otimes \mathbf{I}_2)$ vec $\mathbf{C}$ to the original core array yields the new core array

$$\mathbf{C}^{\text{new}} = \left(\begin{array}{cc|cc} 0.412 & 1.323 & 0.911 & 2.234 \\ -0.911 & -0.500 & 0.412 & -0.088 \end{array}\right)$$

with a significantly increased variance-of-squares value of var($\mathbf{C}^{\text{new}}$) = 19.36. In the next iteration, the matrix $\mathbf{A}$ has to be considered with $j = 2$. Accordingly, its general element is

$$A_{kl} = \sum_{i_1=1}^{2} \sum_{i_3=1}^{2} \left( C^2_{i_1 l i_3} - \bar{C} \right) C_{i_1 l i_3} C_{i_1 k i_3} \qquad (k,l = 1,2)$$

where $\mathbf{C}$ refers to the previously obtained $\mathbf{C}^{\text{new}}$. Proceeding as before, one finds a transformation matrix $\mathbf{P}$ symmetrifying $\mathbf{P}^{\mathrm{T}}\mathbf{A}$ and a new core via vec $\mathbf{C}^{\text{new}} = (\mathbf{I}_2, \mathbf{P}, \mathbf{I}_2)$vec $\mathbf{C}$. The new variance-of-squares value then becomes var($\mathbf{C}^{\text{new}}$) = 27.64. Finally, after three main iterations (i.e., $3 \times 3 = 9$ single iterations), the var-value reaches a relative precision of 0.001 at var = 48.6. The resulting core is

$$\mathbf{C} = \left( \begin{array}{cc|cc} -0.12 & 0.29 & -0.15 & 2.76 \\ -0.83 & 0.63 & 0.08 & -0.41 \end{array} \right)$$

Although the sum of squares is the same as in the original core, there is practically only one significant entry left now (after squaring in mind all contributions).

## 5. Applications

Two applications shall illustrate the ideas discussed so far. The first application is a data example considered in more detail in Ref. [12]. It relates to a water quality study carried out in the course of 1 year in the Niger delta area. More precisely, 13 physicochemical parameters were measured 22 times at 10 sampling stations, thus, yielding a three-way data array of order (10, 13, 22). The data were scaled in a way to give all physicochemical parameters zero mean and unit variance (over all sampling stations and sampling times). The first column of diagrams in Fig. 2 provides the loading plots resulting from a Tucker3 decomposition of the array with two components considered for each mode. Thus, the diagrams correspond to the component matrices $\mathbf{A}_i$ in (1). The first diagram reveals a strong grouping among sampling stations ('a' and 'b') in accordance with their known degree of pollution. The second diagram refers to the physicochemical parameters among which a salinity-related group (conductivity 'co', chloride concentration 'Cl' and hardness 'hr') shows high loadings on the first factor and the chemical oxygen demand '$cO_2$' has a high loading on the second factor. In the third diagram, successive sampling times (1,2 = February, . . . ,21,22 = December) have a strong temporal trend along the first factor. For better visualization the loadings of this first factor are plotted vs. time in the diagram at the very bottom. The resulting curve indicates a clear temporal factor in the data. In order to detect how these factors of different modes relate to each other, one has to study the core array, which in unfolded form, is given by

$$\mathbf{C} = \left( \begin{array}{cc|cc} 1.36 & 0.48 & 0.35 & -0.35 \\ -0.37 & -0.11 & 1.02 & -0.57 \end{array} \right)$$

Accordingly, two major entries seem to be present, namely $c_{111} = 1.36$ and $c_{212} = 1.02$. The first one relates to the joint effect of all first factors in the three modes. Re-inspecting the diagrams one recognizes this factor as a seasonal change of salinity which is almost uniform for all stations (similar loadings of stations on the first factor). Indeed, the time curve reflects quite well the rainfall period (September to November) with low salinity. The second contribution relates to the combination of the second factors of sampling stations and times with the first factor of physicochemical parameters. Hence, again salinity is involved, but now with a geographical rather than seasonal meaning: the vertical arrangement of sampling stations corresponds quite well to their geographical positions with increasing distance to the shore resulting in decreasing salinity, while there is no systematic variation of the loadings of sampling times on the second factor.

Among the remaining entries of the core array there are five with comparable contributions, and it seems hard to decide whether all or which of these have additional importance in the explanation of data structure. To

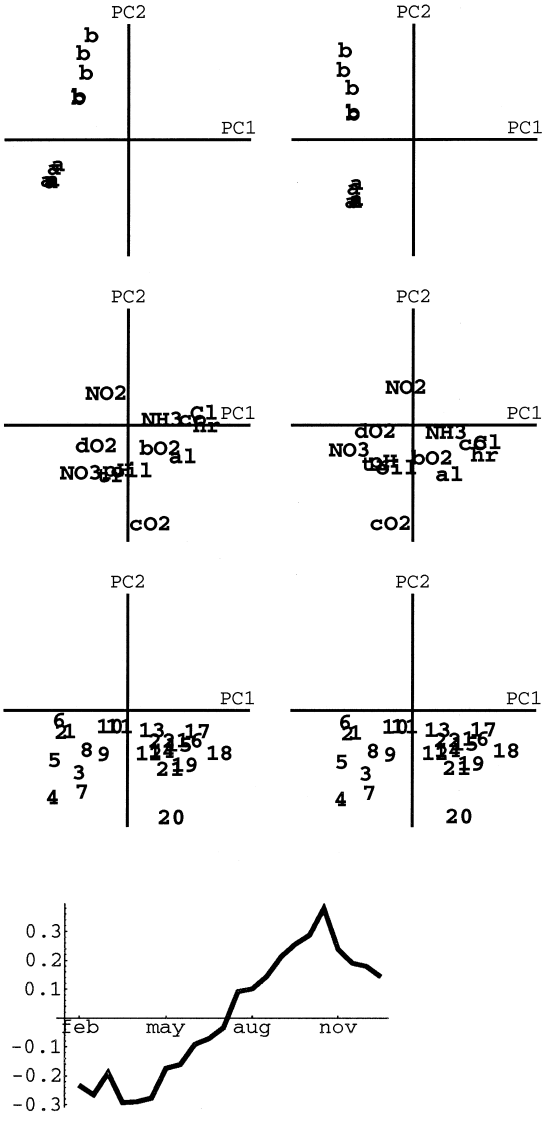


Fig. 2. Original (first column) and rotated (second column) loading plots for sampling stations, physicochemical parameters and sampling times in the water chemistry example. The loadings with respect to the first factor of sampling times are plotted as a curve over time (the same curve refers to both original and rotated loading plots for sampling times).

answer this question, a simple-structure transformation was realized according to the variance-of-squares criterion. The optimally transformed core array turns out to be

$$\mathbf{C} = \left(\begin{array}{cc|cc} 1.48 & 0.11 & 0.05 & -0.04 \\ 0.00 & 0.16 & 0.72 & 1.02 \end{array}\right)$$

In contrast to the original core array, a distinction between significant and nonsignificant contributions is much more evident now. This fact is also supported by the increase of the variance-of-squares criterion from 2.83 to 4.41. Obviously, three relevant factor combinations have to be taken into account. The corresponding rotated

loading diagrams leading to this core array are plotted in the second column of Fig. 2. It is remarkable that only a slight change takes place in the component matrices, nevertheless providing a much clearer core structure. In particular, all sampling stations get even more equal weights on the first factor. Minor changes take place for the parameters, too, whereas the loadings of sampling times remain practically unchanged (in particular the seasonal curve is the same as before). Apart from the two factor combinations $c_{111}$, $c_{212}$ already discussed before (but now with changed importance), a third combination $c_{222} = 1.02$—namely the one of all second factors—is found to be significant. According to the vertical axes in the diagrams, this relates to a distinction of sampling stations into groups 'a' and 'b' mainly due to differing values of chemical oxygen demand uniformly over time. Some effect of pollution is likely to be hidden in this factor combination. However, we do not go into further details of possible interpretations since the main objective of the example is to illustrate the Tucker3 model and the effect of core simplifications.

In order to emphasize the benefits of core rotation, we shall give another example dealing with the differences between cores with optimum variance-of-squares and optimum diagonality. To keep the discussion focused and aimed at core rotation, no explicit chemical interpretation of the factors will be given. The data to be analyzed are derived from fluorescence intensity measurements of 13 *thick juice* samples. Thick juice is an intermediary product in the production of sugar and ongoing projects aim at obtaining means to control and decrease the unwanted formation of colour during the process, see Refs. [5,24].

The 13 thick juice samples have been separated into 28 fractions (each of 700 μl) on a 200 mm Sephadex G25M column that separates components according to molecular size in the approximate range of 1000 to 5000 MW. A sample volume of 300 μl was introduced into the isocratic and aqueous 0.01 w/w% NaCl carrier running with a flow of 0.8 ml/min. For each fraction, six preselected combinations of excitation and emission wavelengths have been measured using spectrofluorometry. The filter combinations were found in earlier investigations [25]. The six combinations of excitation and emission wavelengths cover the range 270 nm to 390 nm of the excitation range and 280 to 420 nm of the emission wavelength range. The collected three-way data array has dimensions (13,28,6) where the respective modes refer to sample number, fraction number and combination of excitation–emission wavelengths.

For exploration of the data, we have chosen to analyze the data by *N*-way PCA, whereby the significant variation of the data is condensed into a few factors allowing for easy interpretation. In order to illustrate the benefits of core rotation, we will compare two cores derived from rotation of the initial core according to the maximum variance-of-squares measure and the maximum diagonality measure. From the cores discussed in the sequel, it will appear that the PARAFAC model is inadequate of handling the data in question due to severe non-diagonality of the core.

Prior to analysis, the data were mean-centred across the third mode since the filter combinations of the apparatus result in quite different levels of signals. This pretreatment ensures that the arbitrary differences between the response levels are removed from the modelling step in accordance with the aim of the investigation.

A three-way model with three factors in each mode was chosen as a compromise between having a small number of factors and a close fit to the data. The SVD-based algorithm used to calculate the PCA model is described in Ref. [26]. The model explains 71.7% of the variation of data (i.e., sum of squares) and the initial core was found to be

$$\mathbf{C} = \left(\begin{array}{ccc|ccc|ccc} \mathbf{3383} & \mathbf{2805} & 493 & \mathbf{2600} & \mathbf{-2300} & -215 & -220 & -477 & 110 \\ \mathbf{-2037} & \mathbf{2116} & 124 & 1096 & 888 & -1484 & 16 & 530 & 213 \\ -284 & 526 & -1208 & 246 & 4 & 1353 & -251 & 554 & -454 \end{array}\right)$$

where the elements have been rounded to the nearest integer to provide a clear view of the significant elements. The variance-of-squares of **C** is $2.26 \times 10^{14}$ and the degree of diagonality is 25.2%. Apparently, there are up to six significant elements in the unrotated core. The sum of squares of the three largest squared elements explains 52.8% of the total sum of squares in **C**.

Upon maximization of the variance-of-squares, the core takes the form of

$$\tilde{\mathbf{C}} = \left(\begin{array}{ccc|ccc|ccc} \mathbf{4516} & -87 & 4 & 122 & \mathbf{-3377} & -152 & -114 & 94 & 167 \\ -194 & \mathbf{2837} & -229 & 1388 & -441 & -1495 & 112 & 744 & 318 \\ -33 & 320 & -912 & 263 & -78 & 1546 & -19 & 607 & -744 \end{array}\right)$$

where the variance-of-squares measure increases to $5.37 \times 10^{14}$ and the degree of diagonality is 42.8%. Now, the three largest squared elements are responsible for 80.7% of the total sum of squares. As seen directly from $\tilde{\mathbf{C}}$, the interpretation has become easier since much variation accounted for by several less significant factor combinations has been condensed into a lower number of more significant ones.

When optimizing the diagonality of $\mathbf{C}$ instead, the core transforms into

$$\hat{\mathbf{C}} = \left(\begin{array}{ccc|ccc|ccc} \mathbf{4424} & -484 & 12 & -407 & -1156 & \mathbf{2857} & 444 & -285 & 562 \\ -705 & -667 & 1747 & -794 & \mathbf{2700} & -67 & -322 & -396 & 87 \\ -139 & 225 & -1796 & 1285 & -35 & 978 & -48 & 165 & -1093 \end{array}\right)$$

with a degree of diagonality at 56.8%. The variance-of-squares measure becomes $4.40 \times 10^{14}$. For comparison with the core shown above, the three largest elements account for 70.9% of the total variation in the model, this is approximately 10%—points less than the core that is optimal in the variance-of-squares sense. This means that the analyst, by using the variance-of-squares optimized core $\tilde{\mathbf{C}}$ rather than the diagonalized core $\hat{\mathbf{C}}$, will include what corresponds to 10%-points more variation of data in his interpretation. It is noteworthy that the structure in the data does not conform with the PARAFAC model, since diagonality of the core $\hat{\mathbf{C}}$ cannot be obtained.

In Fig. 3, the 15 largest squared elements from the cores $\mathbf{C}$, $\tilde{\mathbf{C}}$ and $\hat{\mathbf{C}}$ are plotted. The line denoted by (a) represents the largest squared elements from $\mathbf{C}$. The differences between successive squared core elements are small, leading to a rather flat line that indicates the low variance of the core elements. Without core rotation, the analyst has to interpret, perhaps, five factor combinations in order to give a detailed picture of data. Line (b) describing $\tilde{\mathbf{C}}$ depicts a much higher variation in the core elements. We see that the three largest elements are all much higher than the fourth. This allows the analyst to focus on three factor combinations. Note, that the *three* largest elements from the rotated core explain the same amount of variation (80.7%) as the *five* largest elements from the initial core (80.9%). Line (c) describes the elements in the core with optimal diagonality, i.e., $\hat{\mathbf{C}}$. The indication of the presence of three significant factor combinations is more clear than with the unrotated core, but

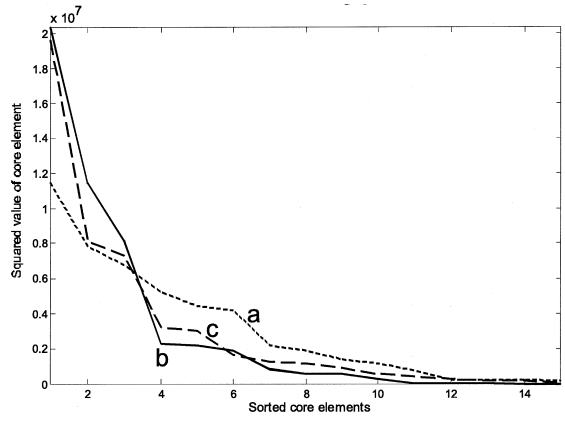


Fig. 3. The 15 largest squared core elements of the three cores are sorted and plotted. Line (a) represents the 15 largest squared core elements of the untreated core. Line (b) are the 15 largest squared elements of the core with optimal variance-of-squares measure, and (c) shows the 15 largest squared elements of the core with optimal diagonality.

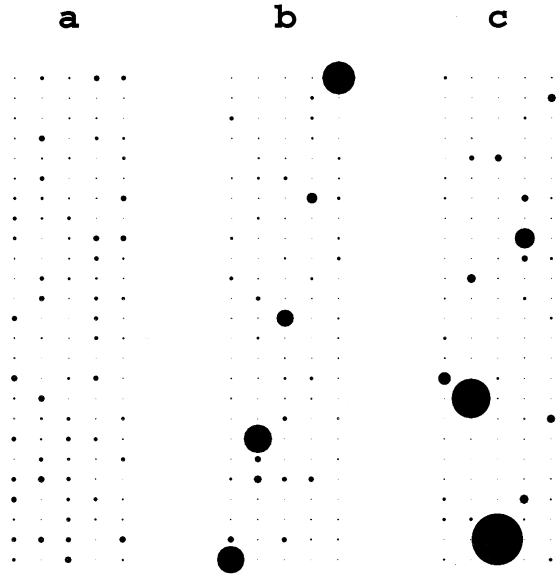


Fig. 4. Symbolic plots of core arrays of order (5,5,5) unfolded to matrices of order (25,5). The left array refers to the original core, the medium one to the transformed core with maximum body diagonality and the right one to the transformed core with maximum variance-of-squares. The squared entries of the cores are translated to diameters of filled circles.

not as clear as with the core with optimal variance-of-squares. The core $\hat{\mathbf{C}}$ suffers from the fact that the core could not be diagonalized, since this structure is not present in data. Furthermore, we see that the levels of the remaining squared core elements remain high for $\mathbf{C}$ and $\hat{\mathbf{C}}$. This is sub-optimal, because the size of the elements reflects what is not included in the interpretation of the model. The low level of all elements but the significant ones for the line (b) is a direct consequence of maximizing the variance-of-squares measure in $\tilde{\mathbf{C}}$.

Finally, we want to indicate the potentials of maximizing the variance-of-squares criterion in a higher-dimensional setting. To this aim, a random three-way core array of order (5,5,5) was created with entries uniformly distributed between $-1$ and 1. This original core was transformed both to maximum body diagonality and maximum variance-of-squares. The results for the unfolded cores (= matrices of order (25,5)) are shown in Fig. 4. For better visualization, the squared values of the entries are translated into diameters of filled circles. Due to the random nature of the original core, there are many positions of medium relevance in the first array (a). In contrast, the two transformed cores show a clear distinction between significant and nonsignificant elements. Not surprisingly, in the core with maximum body diagonality (b), the five major entries are distributed along the diagonal of the unfolded core. Such diagonal structure is lost in the array with maximum variance-of-squares (c), but this loss is in favour of a yet smaller number of significant elements. Comparing (c) with (b) on a rough scale, two rather than three entries are found to be significant. On a finer level, three rather than five entries are clearly distinguished from the rest.

## 6. Conclusion

The proposed variance-of-squares criterion has a great potential for simplifying the structure of core arrays in *N*-way PCA and, hence, for facilitating the interpretation of solutions obtained. Its main advantage over the well-established method of body diagonalization is directly to aim at a reduction of the number of significant entries. Moreover, its application is not restricted to cubic cores. The maximization of the criterion can be car-

ried out by an iterative solution algorithm providing reliable results in a short period of time, thus, higher dimensional arrays (e.g., $N = 7$) may be easily treated as well. Some theoretical results giving insight into the relations between the variance-of-squares and body diagonality criteria have been derived. A convergence proof for the algorithm is given in Ref. [23].

## Acknowledgements

The authors wish to thank Prof. L. Munck (Royal Veterinary and Agricultural University Copenhagen) and Prof. G. Henrion (Humboldt University Berlin) for facilitating the present cooperation. The second author wishes to thank the EU Affluence AIR project and the National Danish fund FØTEK2.

## References

[1] C.J. Appellof, E.R. Davidson, Anal. Chem. 53 (1981) 2053–2056.
[2] C. de Ligny, M. Spanjer, J. van Houwelingen, H. Weesie, J. Chromatogr. 301 (1984) 311–324.
[3] P. Geladi, Chemom. Intell. Lab. Syst. 7 (1989) 11–30.
[4] A.K. Smilde, Chemom. Intell. Lab. Syst. 15 (1992) 143–157.
[5] C.A. Andersson, Master Thesis, Royal Veterinary and Agricultural University, Copenhagen, Denmark, 1995.
[6] R.A. Harshman, UCLA Working Papers in Phonetics 16 (1970) 1–84.
[7] J.D. Carrol, C.C. Chang, Psychometrika 35 (1970) 283–319.
[8] L.R. Tucker, in: C.W. Harris (Ed.), Problems of Measuring Change, Univ. of Wisconsin Press, Madison, 1963, p. 122.
[9] L. Ståhle, J. Pharm. Biomed. Anal. 9 (1991) 671–678.
[10] Y.L. Xie, J.J. Baezabaeza, G. Ramisramos, Chemom. Intell. Lab. Syst. 27 (1995) 211–220.
[11] R. Bro, J. Chemometrics 10 (1996) 47–61.
[12] R. Henrion, Chemom. Intell. Lab. Syst. 25 (1994) 1–23.
[13] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, Wiley, Chichester, 1988.
[14] P.M. Kroonenberg, J. de Leeuw, Psychometrika 45 (1980) 69–97.
[15] H.A.L. Kiers, Statistica Applicata 4 (1992) 659–667.
[16] A.K. Smilde, R. Tauler, J.M. Henshaw, L.W. Burgess, B.R. Kowalski, Anal. Chem. 66 (1994) 3345–3351.
[17] P.M. Kroonenberg, Three-Mode Principal Component Analysis: Theory and Applications, DSWO Press, Leiden, 1983.
[18] H.A.L. Kiers, Statistica Applicata 4 (1992) 7–15.
[19] R. Henrion, J. Chemometrics 7 (1993) 477–494.
[20] J.D. Jobson, Applied Multivariate Data Analysis, Springer, New York, 1992.
[21] H.A.L. Kiers, Psychometrika 62 (1997) 579–598.
[22] H.A.L. Kiers, J. Classification 15 (1998) 245–268.
[23] C.A. Andersson, R. Henrion, submitted to Comp. Stat. Data Anal.
[24] L. Nørgaard, Zuckerindustrie 120 (1995) 970–981.
[25] C.A. Andersson, L. Munck, R. Henrion, G. Henrion, Fresenius J. Anal. Chem. 359 (1997) 138–142.
[26] C.A. Andersson, R. Bro, Chemom. Intell. Lab. Syst. 42 (1998) 93–105.