Fresenius J Anal Chem (1998) 361 : 15–22

ORIGINAL PAPER

René Henrion

Simultaneous simplification of loading and core matrices in N-way PCA: application to chemometric data arrays

Received: 21 November 1997 / Revised: 11 February 1998 / Accepted: 15 February 1998

Abstract In the Tucker3 model of N-way principal components analysis (NPCA), a so-called core matrix describes the possible interactions between components from different modes. For an easy interpretation of solutions, it is necessary to have as few interactions as possible (in conventional PCA of data tables, such interactions can always be avoided). This goal may be realized by various approaches of core matrix transformations. At the same time, it is desirable to have simple component (or loading) matrices. Usually, the simplicity of the core conflicts to a certain degree with the simplicity of the components. The paper demonstrates how the conditional optimization of both goals can be used to find a compromise. For the purpose of illustration, the procedure is first applied to a small three-way data array from heavy metal analysis of tissues in different samples of game. Later, a data array of bigger size from a three-way interlaboratory study is considered.

1 Introduction

The data-analytic investigation of three- or higher dimensional data arrays becomes more and more natural in chemometrics. This is due to both the use of modern hyphenated instrumentation, which canonically provides multidimensional information, and to the increasingly complex design of research, e.g. in environmental studies. Formally, an N-way data array **A** of order $(n_1, \ldots, n_N)$ is identified with a mapping assigning to each N-tuple $(i_1, \ldots, i_N)$ of indices with $i_1 \le n_1, \ldots, i_N \le n_N$ some real number $a_{i_1, \ldots, i_N}$ (the measurement). For $N = 1$, one gets a measurement vector **A** of length n_1 and with general element a_{i_1}, while, for $N = 2$, one arrives at a matrix **A** of order (n_1, n_2) with general entry a_{i_1,i_2}. Physically, this matrix can be represented as a data table, where the rows (corresponding to the first index) may refer to certain objects under investigation whereas the columns (corresponding to the second index) usually relate to some variables measured at the objects. Similarly, a three-way data array can be represented as a so-called data cube as depicted in Fig. 1. Sometimes, the three index sets (or modes) of a three way array are referred to as objects, variables and conditions in order to emphasize that the array resulted from stacking several data tables with a fixed set of objects and variables but measured under different conditions. For instance, we shall consider below a data example which is a small part of an environmental study described by Lusky et al. [11]. The three-way data array in question comprises the measurements of 5 heavy metal concentrations (variables) in 3 organs ('conditions') for a set of 16 samples of game (objects).

In order to extract the basic information contained in higher dimensional (or N-way) arrays, one needs appropriate generalizations of well-known tools from the conventional data analysis of data tables, such as principal components (PCA), factor analysis (FA) or partial least squares (PLS). Here, appropriateness refers to taking ac-

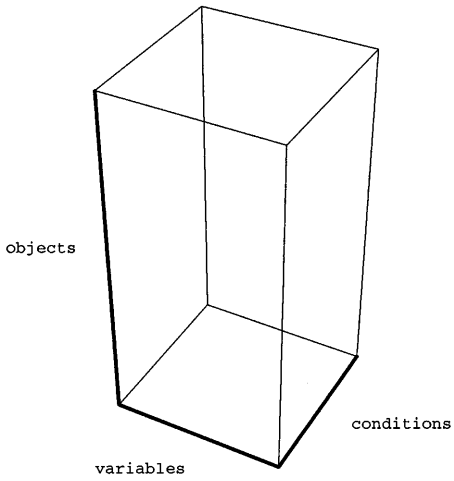

Fig. 1 Illustration of a three-way array

Dedicated to my father, Professor Dr. Günter Henrion, on the occasion of his 65th birthday

R. Henrion
Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, D-10117 Berlin, Germany

count of the complex structure of the given array rather than just reducing it to a table by some unfolding procedure, which usually results in some loss of information. The probably most prominent methods in the context of N-way PCA or FA are the so-called Tucker3 and Parallel Factor Analysis (PARAFAC) models introduced by Tucker [12] and Harshman [2], respectively. The Tucker3 model of decomposing an N-way array **A** of order ($n_1, \ldots, n_N$) into N component matrices $\mathbf{X}_i$ of orders (n_i, s_i) (where $s_i \leq n_i$) is defined via

$$\text{vec}\,\mathbf{A} \approx (\mathbf{X}_1 \otimes \ldots \otimes \mathbf{X}_N)\,\text{vec}\,\mathbf{C}, \qquad (1)$$

where **C** is the so-called core matrix (better: core array) of order ($s_1, \ldots, s_N$). The symbols 'vec' and '⊗' refer to the vectorization operator and Kronecker product, respectively. The approximation in (1) has to be understood in the least squares sense. Frequently, the component matrices are required to be orthonormal, i.e. $\mathbf{X}_i^T\mathbf{X}_i = \mathbf{I}_{s_i}$ (**I** = identity matrix). The aspect of data reduction is given by choosing the number of components s_i in each of the N modes of the array much smaller than the number n_i of original items in that mode. In particular, the choice $s_i = 2$ allows to generate (separate) diagram plots for the items of each mode (objects, variables, conditions, etc.) from the corresponding component matrix $\mathbf{X}_i$. In contrast to the well-known biplot technique of conventional PCA, it is no longer possible for higher dimensional arrays to make direct superpositions of diagram plots for different modes. This is due to the fact that the Tucker3 model allows for 'non-diagonal' interactions between components of different modes. For instance, there may be a joint contribution to the data variation by the first component of objects along with the second component of variables and the first component of conditions. PCA solutions of conventional data tables always provide diagonal contributions (first component of objects with first component of variables and, independently, second component of objects with second component of variables), so one might wonder what a consideration of non-diagonal interactions between components could be good for.

Indeed, one can consider a restricted version of the Tucker3 model (1) where the core array **C** is required to have a so-called body-diagonal structure. This means that its entries are zero whenever not all indices coincide, i.e. $c_{i_1}, \ldots, {}_{i_N} = 0$ if not $i_1 = \ldots = i_N$. This reduced version of (1) is called the PARAFAC model. Due to its body-diagonal core, only equally-indexed component contributions have to be discussed as in classical PCA. In particular, diagram plots of different modes can be superposed in the sense of tri- or N-plots. PARAFAC is an important and extremely useful tool for qualitative and quantitative analysis of a set of samples characterized by bilinear data (e.g. excitation-emission matrices in fluorescence spectroscopy). The reason is that, up to scaling and permutation, PARAFAC yields unique component solutions. If, for instance, several mixtures of a certain set of pure components are characterized by bilinear responses, then PARAFAC will identify the right pairs of profiles (e.g. emission profile and excitation profile) of the pure components as well as the right concentration proportions of the mixtures. This is a big advantage (sometimes referred to as the 'third order advantage') over characterizations of samples by means of one-dimensional responses (e.g. single spectrum), where the PCA solutions always have to be after-treated by some appropriate rotation in order to get physically meaningful results.

On the other hand, allowing for non-diagonal component contributions may result in a significantly better fit to the data array by using only a few extra parameters (the core elements). Data arrays from environmental studies (e.g. samples/chemical parameters/sampling times, sites or conditions) are likely to meet this situation. Apart from the aspect of better fit, the consideration of (non-diagonal) core arrays comes as a natural consequence of simple-structure transformations for component matrices.

As an illustration of the Tucker3 and PARAFAC models, Fig. 2 shows the resulting diagram plots of component matrices for the data array introduced above. As one can

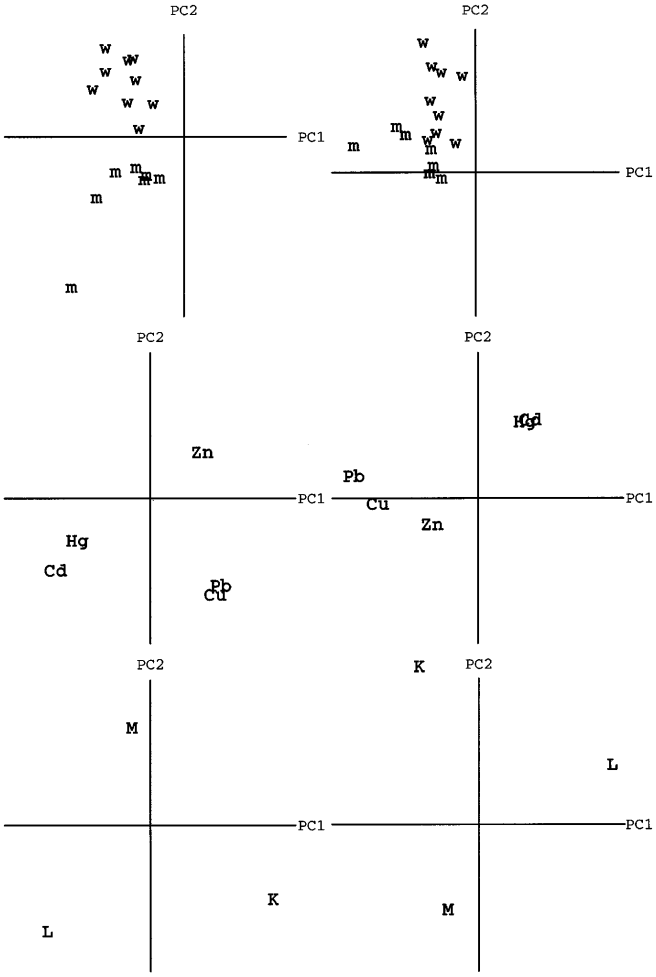

Fig. 2 Diagram plots of the Tucker3 (first column) and PARAFAC (second column) solutions for the data example of game samples. From top to bottom, the objects (game samples with 'm' referring to moufflon and 'w' to wild boar), variables (heavy metals) and 'conditions' (organs or tissues with 'L' = liver, 'K' = kidney and 'M' = muscle) are displayed

see, there are minor differences in the plots, for instance, the moufflon samples are better separated from the wild boar samples in the Tucker3 solution. Also, correlations among the variables are indicated in slightly different ways. An axis-wise interpretation can be carried out immediately for the PARAFAC solutions, whereas a similar way of argumentation is not possible for the Tucker3 solutions. Here one has to take into account the core array, which, in the example, is given by

$$\mathbf{C} = \left(\begin{array}{cc|cc} 1.188 & -0.069 & -0.075 & -0.917 \\ -0.089 & -0.620 & 0.413 & -0.101 \end{array}\right)$$

The consideration of diagram plots corresponds to the choice $s_1 = s_2 = s_3 = 2$ for the number of components for each mode in the decomposition (1). Hence, the core array **C** is of the order (2,2,2), which means that it consists of two matrices of the order (2,2) stacked in front of each other. Above, we have given the unfolded notation of **C**, which means that the elements $c_{i,j,1}$ are located in the matrix left and the elements $c_{i,j,2}$ in the matrix right to the separating line. In the case of orthonormal component matrices, the squared values of the core elements indicate how much variance of the data array is explained by a combination of those components in the three modes which are indexed in the corresponding core entry. For instance, $c^2_{1,1,1} = 1.188^2$ gives the biggest part (around 28% of the total data variance) which is related to the first components (PC1) in all three modes. The next contribution is defined by $c^2_{1,2,2} = (-0.917)^2$ and relates PC1 of objects with PC2 of variables and PC2 of 'conditions' (organs). According to the squared values, one would have to include 4 component contributions into the interpretation of the Tucker3 model here. Compared to the PARAFAC solution, not only the number of contributions has increased but, even worse, the interaction between components has become much more complex. That is why simplification procedures as described in the following section are of extreme importance in the context of the Tucker3 model. The introduced small data array will continue to serve the purpose of illustration, whereas in the last section a more extensive data example will be considered. For a more detailed introduction to N-way PCA with applications in chemometrics, we refer to [5].

2 Simple-structure transformations for component and core matrices

The simplification of PCA solutions by exploiting a transformation degree of freedom is an attractive goal even in the context of data tables. Assuming that some data table **A** has been decomposed into component matrices **X** and **Y** for objects and variables (usually called score and loading matrix) according to $\mathbf{A} \approx \mathbf{X}\mathbf{Y}^T$, it is well known that, after passing to rotated component matrices $\tilde{\mathbf{X}} = \mathbf{X}\mathbf{P}$ and $\tilde{\mathbf{Y}} = \mathbf{Y}\mathbf{P}$, where **P** is some orthogonal matrix (i.e. $\mathbf{P}^T\mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I}$), the same decomposition and, hence, the same degree of fit to **A** is achieved by using the rotated component matrices due to $\tilde{\mathbf{X}}\tilde{\mathbf{Y}}^T = \mathbf{X}\mathbf{P}\mathbf{P}^T\mathbf{Y}^T = \mathbf{X}\mathbf{Y}^T$. Consequently, the same techniques (e.g. biplot of diagrams) as for the original PCA solution can be employed, but now with the advantage that the PC-axes may have become much easier to interpret: usually, the PC's represent latent factors, but using an appropriate rotation matrix **P**, the axes of the new diagrams generally may be identified with certain interpretable groups of objects or variables. This may result in a considerably easier discussion of the data structure. The right rotation matrix can be found according to different criteria, the most popular of which is Kaiser's varimax criterion [7].

Note, however, that in the derivation above, only one rotation matrix has been applied, so one can use it either to simplify the representation of objects or the one of variables, whereas the transformation of the respective complementary representation is automatically determined by the former one. As a consequence, one cannot get simple configurations of objects and variables simultaneously by the approach described above. In order to do so, one has to permit independent rotations, defined by orthogonal matrices **P** and **Q**, for the two component matrices. Then, denoting the transformed component matrices by $\tilde{\mathbf{X}} = \mathbf{X}\mathbf{P}$ and $\tilde{\mathbf{Y}} = \mathbf{Y}\mathbf{Q}$, respectively, the same decomposition is obtained according to

$$\mathbf{A} \approx \mathbf{X}\mathbf{Y}^T = \tilde{\mathbf{X}}\mathbf{P}^T\mathbf{Q}\tilde{\mathbf{Y}}^T = \tilde{\mathbf{X}}\mathbf{C}\tilde{\mathbf{Y}}^T,$$

where $\mathbf{C} = \mathbf{P}^T\mathbf{Q}$ can be understood as a core matrix linking the two (transformed) component matrices. Indeed, using the more complicated notation of Kronecker product and vectorization (which is indispensable when working with general N-way arrays), the same relation writes as

$$\text{vec}\,\mathbf{A} \approx \left(\tilde{\mathbf{X}} \otimes \tilde{\mathbf{Y}}\right) \text{vec}\,\mathbf{C},$$

which is immediately recognized as the Tucker3 model (1) in the special case of data tables ($N = 2$). It is important to note that **C** is not diagonal in general, which means that the consideration of non-diagonal contributions of components from different modes is a direct consequence of aiming at simultaneously simple component matrices. This gives a different, quite natural motivation for the use of core matrices, which in the Tucker3 model seemed to be somehow artificial at first. The stated phenomenon already encountered in the context of data tables is typical for all data arrays: either one has a simple core matrix (which could be formally set equal to the identity matrix in the PCA model of data tables) but complicated component matrices (latent factors), or one has simple component matrices (axes identifiable as subgroups of items) but a complicated (non-diagonal, perhaps completely filled with significant elements) core matrix. As will be seen later, the situation becomes even more difficult for three- or general N-way arrays: neither the component matrices $\mathbf{X}_\mathbf{i}$ nor the core matrix **C** in (1), which are obtained, for instance, by the ALS algorithm described in [10], need to be simple.

There is no problem, of course, to calculate suitably transformed component matrices with simple structure in the general N-way case, since one still deals with matrices then, hence, the same procedure (e.g. varimax) applies as

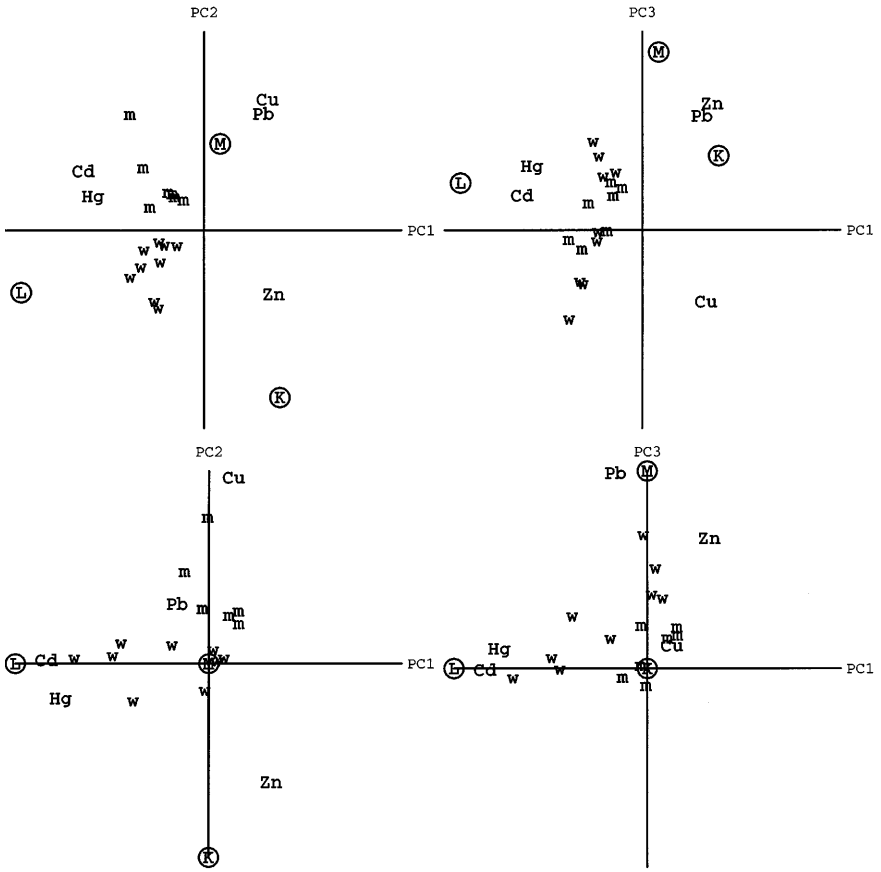

Fig. 3 Diagram plots of the Tucker3 solutions (all modes superposed, same legend as in Fig. 2) for the data example of game samples. 3 principal components have been taken into account. The two diagrams on top refer to the original PC solutions while the two diagrams at the bottom correspond to the varimax rotated solutions for all three modes

in the case of data tables. Figure 3 illustrates the effect of component simplification by the varimax method. In order to save space, the plots of all three modes are superposed although this does not mean that a triplot-like discussion would be possible. In contrast to Fig. 2, a three-component solution was considered here (i.e. $s_1 = s_2 = s_3 = 3$) and represented as PC1/PC2 and PC1/PC3-diagrams, respectively. The upper two diagrams refer to the original solutions as they are output of the ALS algorithm. Similarly to Fig. 2, one would have to interpret the PC axes here as latent factors. The lower two diagrams show the varimax rotated plots for the three modes. Since the original component matrices were required to be orthonormal, the maximization of varimax reduces to that of the so-called quartimax expression, which simply is the sum of fourth powers of all entries in the corresponding component matrix. The figure makes evident that the varimax rotated plots provide much better identification of the PC-axes with certain groups of real items: for instance, each of the PC's may be identified with one organ (PC1 = liver, PC2 = kidney, PC3 = muscle). This is not surprising, of course, since the third mode consists of three items only and since we have chosen a three-component solution here. Furthermore, among variables, PC1 may be identified with the (Cd,Hg)- group or PC2 with Cu and, among objects, PC1 represents wild boar whereas PC2 corresponds to moufflon.

In order to make an interpretation of the joint contribution of all three modes, one has to know the underlying

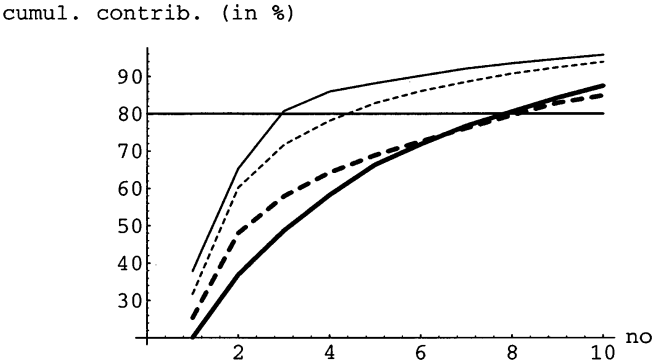

Fig. 4 Cumulative contributions of the first 10 out of 27 squared core entries for the original core (*thick line*), the core with all modes varimax-rotated (*thick line dashed*), the core with one mode (organs) varimax-rotated and the other two modes rotated to maximum variance of squared core entries (*thin line dashed*) and for the core with all three modes rotated to maximum variance of squared core entries (*thin line*)

core matrix. Unfortunately, neither the original nor the varimax solution has a simple core structure. This is easily seen from Fig. 4 where, in order to avoid printing all of the 27 elements for each core, the cumulative percentages of the leading 10 squared core entries (as compared to the sum of all 27 squared entries) are plotted. Obviously, in the original core as well as in that obtained after varimax rotation of all three modes, one needs 8 elements in order to arrive at a reasonable percentage like 80%. This, of

course, makes the interpretation rather arduous, and one might prefer to look for simpler core matrices instead. The possibility to do so relies on the fact that, after transforming the component matrices $\mathbf{X}_i$ in (1) to $\tilde{\mathbf{X}}_i = \mathbf{X}_i\mathbf{P}_i$ by means of orthogonal matrices $\mathbf{P}_i$ of order (s_i, s_i), the same approximation as in (1) is obtained when using the transformed component matrices $\tilde{\mathbf{X}}_i$ along with the counter-transformed core matrix

$$\operatorname{vec} \tilde{\mathbf{C}} = (\mathbf{P}_1^T \otimes \ldots \otimes \mathbf{P}_N^T) \operatorname{vec} \mathbf{C}. \quad (2)$$

Consequently, the rotational degree of freedom in the identification of the core matrix in (1) can be exploited in order to select among the infinite variety of congruent cores the one with highest simplicity.

There are several well-known procedures for core simplification, the more classical among them being related with diagonality properties. For example, the following core matrices of the order (2,2,2), which are unfolded in the same way as in the example of section 1, have slice-wise (cf. [10]) and body diagonal (cf. [4, 8]) structure, respectively:

$$\left(\begin{array}{cc|cc} * & 0 & * & 0 \\ 0 & * & 0 & * \end{array}\right) \quad \left(\begin{array}{cc|cc} * & 0 & 0 & 0 \\ 0 & 0 & 0 & * \end{array}\right)$$

The theoretical and practical interest of these diagonality concepts is that they allow a reduction of the Tucker3 solution to a formally much simpler PARAFAC solution. If, however, an almost diagonal core cannot be achieved by suitable rotations of the component matrices, then a transformed core with the best approximation to diagonality needs not be a simple core (i.e. one with the least number of significant entries) as well. The reason is that diagonalization does not only aim at reducing the number of non-zero elements but simultaneously looks for a specific structure where these non-zero elements have to be placed. In the case of data tables, both goals are identical as it was shown by Henrion and Andersson in [6], but starting from three-way arrays, extreme differences may occur. An appropriate criterion for measuring the simplicity of cores is the so-called variance of squared core entries:

$$var = \sum_{i_1=1}^{s_1} \cdots \sum_{i_N=1}^{s_N} \left(C^2_{i_1 \ldots i_N} - \overline{C}\right)^2 \quad (3)$$

where

$$\overline{C} = \sum_{i_1=1}^{s_1} \cdots \sum_{i_N=1}^{s_N} C^2_{i_1 \ldots i_N} / (s_1 \cdots s_N) \quad (4)$$

The justification of this criterion relies on the fact, that the highest value of *var* is attained at cores with one single non-zero element. Although the rotational degree of freedom in (2) is not large enough to end up with such an extremely simple solution in general, it is clear that the *var*-criterion rewards cores with many zero or almost-zero elements. In complementary terms, solutions with a minimal number of significant entries are sought. This feature directly translates the goal of minimizing the interpreta-

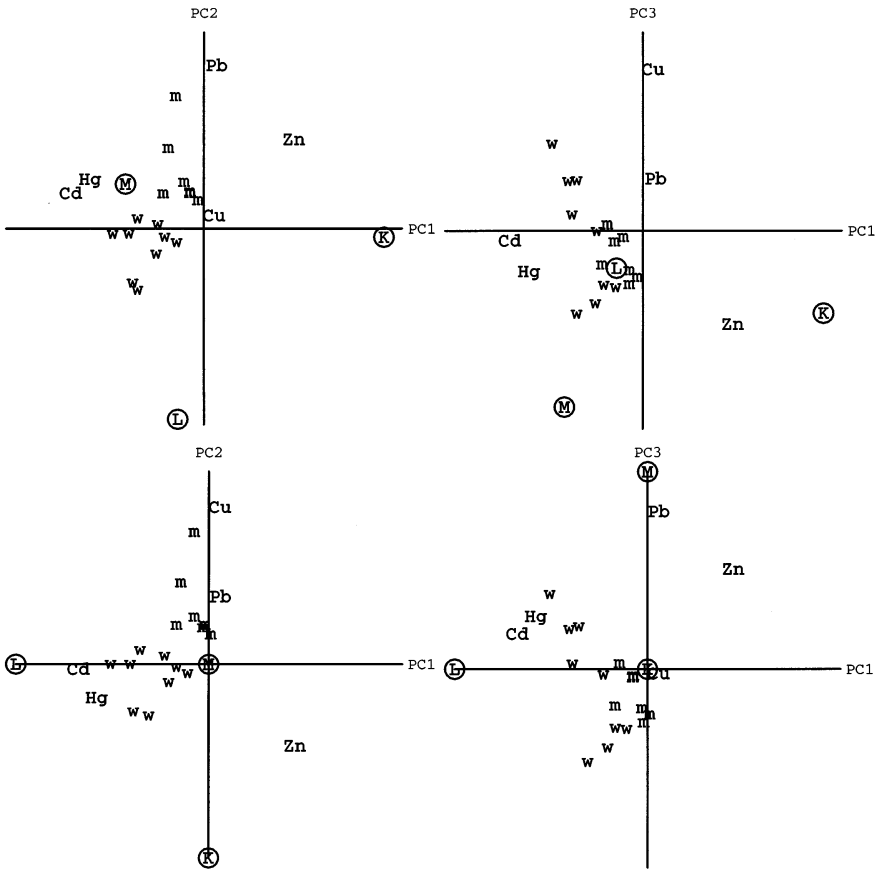

Fig. 5 Diagram plots of the Tucker3 solutions for the data example of game samples. The two diagrams on top refer to the transformation yielding a core with maximum variance of squared entries while the two diagrams at the bottom correspond to the third mode being varimax-rotated and the first two modes being rotated (with the third one held fixed) as to provide a core with maximum variance of squared entries

Table 1 Results of different simple-structure transformations in the data example of game samples (maximum values emphasized, for details see text)

Core no.	var	q_{obj}	q_{var}	q_{cond}
1)	1.057	0.388	0.897	1.955
2)	**3.074**	0.493	1.382	2.260
3)	2.371	**0.733**	1.299	2.374
4)	2.857	0.497	**1.460**	2.176
5)	2.445	0.546	1.333	**3.000**
6)	1.492	**0.733**	**1.460**	**3.000**

tional effort for the core matrix. The theoretical relations of the *var*- criterion with conventional diagonalization are investigated in [6]. An algorithm for finding the core with highest *var*- value and the correspondingly rotated component matrices is described in [1] as part of a general purpose algorithm, which allows the maximization of any differentiable criterion depending on a set of *N* orthogonal matrices. In particular, the conventional diagonality criteria are also included. Below we shall describe a variant where a conditional maximization of the *var*- criterion for a subset of modes is carried out.

Figure 5 (top) shows the PC1/PC2 and PC1/PC3 diagrams for the data example of game samples, where the original component matrices of Fig. 3 (top) are rotated as to provide a core matrix with a maximum *var*- value. In contrast to the original core (*var* = 1.057) and also to the core of varimax rotated components (*var* = 1.492) (compare Table 1), one obtains a much higher value of *var* = 3.074. As a consequence, the core matrix has got a much simpler structure: according to Fig. 4, the three most significant core entries are sufficient now, in order to explain 80% of the total sum of squares contained in the core. More precisely, these three entries are identified in the example as $c_{1,1,1}$, $c_{1,2,2,}$ and $c_{2,3,2}$. From Fig. 5, one derives the following rough assignment of PC-axes:

objects	PC1	common factor for all game samples
	PC2	discriminating factor between moufflon and wild boar
	PC3	non-interpretable specific factor
variables	PC1	(Cd, Hg)
	PC2	Pb
	PC3	Cu
'conditions'	PC1	kidney
	PC2	liver
	PC3	muscle

Then, the interpretation of the three core entries is as follows: the common factor for all game samples is characterized by dominating concentrations of cadmium and mercury in kidney ($c_{1,1,1}$) and, independently, by dominating concentration of lead in liver ($c_{1,2,2,}$). The factor discriminating the moufflon and wildboar group among game samples is determined by the different copper concentrations in liver ($c_{2,3,2}$). All the remaining core entries represent minor contributions of factor combinations with a considerably smaller amount of explained data variation. It is emphasized that a similarly simple interpretation was not possible in the case of the component matrices in Fig. 3. This demonstrates the gain obtained by the simple-structure transformation of the core matrix.

On the other hand, as to be expected from the remarks in the introduction, the simplicity of the core structure is bought by a stronger orientation towards latent factors rather than directly interpretable groups of items. In the example, this is true first of all for the PC's of objects. This rises the question if it is possible to make a reasonable compromise between the simplicity of the core and the simplicity of the component or loading matrices. One could formulate such a compromise in terms of Pareto or vector optimization, but it is not clear if this approach is really useful here. A simpler strategy was suggested by Kiers in [9]: a certain subset of modes is considered for simple-structure transformation of its component matrices, whereas the complementary modes are used in order to simplify the core matrix, but with the modes already having simple component matrices being held fixed. This approach of conditional optimization was discussed in [9] within the framework of a specific simple-structure transformation for core matrices, the so-called SIMPLIMAX method, which, similarly to the diagonalization methods, prescribes a certain structure of the ideal core (a fixed number of zero's has to be realized as close as possible). We follow the same idea here, but with the variance-of-squares criterion *var* instead, which allows for core transformations without any a priori restrictions on its structure.

Table 1 collects the results of several possible conditional optimizations in the data example of game samples. The cores considered there are: 1) = core of original solution, 2) = core with maximum simplicity, 3),4),5) = cores obtained after transforming the component matrices of objects, variables and 'conditions', respectively, to simple structure and using the respective complementary modes for conditionally maximizing the core simplicity, 6) = core obtained after the component matrices of all modes had been simplified. The columns contain the variance-of-squares criterion *var* for measuring core simplicity and the quartimax criterion *q* for measuring the simplicity of the component matrix for the respectively indexed mode. The extreme values in each column are emphasized. The core of original solutions gives poor results both in the degree of core simplicity and of component simplicity with respect to all modes. The core resulting from a varimax transformation of all three modes has, of course, maximum simplicity of components with respect to all modes, but the core is far from being simple at the same time. On the other hand, the core with maximum simplicity fails to have simple component matrices for objects and for 'conditions'. Among the three conditionally optimized cores (we do not discuss here the three remaining possibilities of combining pairs of modes for varimax) the one with components simplified for variables reaches a core with almost as simple structure as in the optimal case (*var* = 2.857). However, the component matrix for objects and conditions are then quite far from simplicity. The core with 'conditions' simplified (no. 5) realizes more or less reasonable values with respect to all criteria and therefore may be considered as a good compromise in this example.

The two diagrams at the bottom of Fig. 5 show the plots of component matrices for this solution. The corresponding curve in Fig. 4 (thin line dashed) indicates that now four core entries are necessary to explain around 80% of the amount of variance represented by the sum of all core entries. At this point, we do not go into the discussion of further details of this small data example.

At the end of this section, a possible algorithmic realization of the described compromise between simplicity of core and of loading matrices is presented. It is assumed that an N-way data array **A** has been decomposed into N component matrices and a core matrix according to (1) with the orders of arrays and matrices as introduced there. The first part of the algorithm (until step 4) realizes the maximization of the quartimax criterion (sum of fourth powers) for the simplification of the component matrices of a certain pre-selected set of modes. The second part realizes the conditional maximization of the variance-of-squares criterion (3) for core simplification with the simplified component matrices of the previously selected modes held fixed. In other terms, part of the modes gets simplified component matrices while the remaining degree of freedom in the complementary modes is used in order to simplify the core.

1. Select a subset $J \subseteq \{1, \ldots, N\}$ of modes, the component matrices of which are supposed to be simplified. Set $\mathbf{C}^{new}$: = original core matrix (compare (1)). Select a first index $j \in J$.
2. Set $\mathbf{X}_j^{new}$: = original component matrix of mode j (compare (1)).
3. Set $\mathbf{X}_j^{old}$: = $\mathbf{X}_j^{new}$ and $\mathbf{C}^{old}$: = $\mathbf{C}^{new}$. Compute an orthogonal matrix $\mathbf{P}$ such that $\mathbf{P}^T\mathbf{Q}$ becomes a symmetric matrix, where

   $$Q_{kl} := \sum_{i=1}^{n_j} x_{il}^3 x_{ik} \quad (k,\ l\ =\ 1,\ \ldots,\ s_j).$$

   Set $\mathbf{X}_j^{new}$: = $\mathbf{X}_j^{old}$ $\mathbf{P}$ and update the core according to (cf. (2))

   $$\operatorname{vec} \mathbf{C}^{new} := \left(\mathbf{I}_{s_1} \otimes \cdots \otimes \mathbf{I}_{s_{j-1}} \otimes \mathbf{P}^T \otimes \mathbf{I}_{s_{j+1}} \otimes \mathbf{I}_{s_N}\right) \operatorname{vec} \mathbf{C}^{old}.$$

   Repeat step 3 until convergence of the transformed component matrix $\mathbf{X}_j^{new}$.
4. Select an index $j \in J$ which has not been considered yet (if any is left) and go to step 2.
5. Set $\mathbf{X}_j^{new}$: = $\mathbf{X}_j$ for all indices $j \in \{1, \ldots, N\} \setminus J$ of the complementary modes which have to be used for core simplification. Select a first complementary index $j \in \{1, \ldots, N\} \setminus J$
6. Set $\mathbf{X}_j^{old}$: = $\mathbf{X}_j^{new}$ and $\mathbf{C}^{old}$: = $\mathbf{C}^{new}$. Compute an orthogonal matrix $\mathbf{P}$ such that $\mathbf{P}^T\mathbf{Q}$ becomes a symmetric matrix, where

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

   and $\bar{C}$ refers to (4). Set $\mathbf{X}_j^{new}$: = $\mathbf{X}_j^{old}$ $\mathbf{P}$ and update the core $\mathbf{C}^{new}$ the same way as in step 3.
7. Select a complementary index $j \in \{1, \ldots, N\} \setminus J$, which has not been considered yet (if any is left) and go to step 6.
8. If convergence of the component matrices or of the core matrix has not been obtained yet, then restart by selecting a first complementary index $j \in \{1, \ldots, N\} \setminus J$ and go to step 6.

The decisive parts of the algorithm are contained in steps 3 and 6, respectively. The concrete shape of the occurring matrix $\mathbf{Q}$ is a consequence of the derivation of the given criterion (quartimax or variance-of-squares) with respect to the orthogonal matrix providing an appropriate transformation. The required symmetrification of the expression $\mathbf{P}^T\mathbf{Q}$ is easily realized: let $\mathbf{Q} = \mathbf{UDV}$ be a singular value decomposition of $\mathbf{Q}$, where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal (recall that $\mathbf{Q}$ is quadratic although not necessarily symmetric) and $\mathbf{D}$ is diagonal. Then $\mathbf{P}$: = $\mathbf{UV}$ is the desired orthogonal matrix since $\mathbf{P}^T\mathbf{Q} = \mathbf{V}^T\mathbf{U}^T\mathbf{UDV} = \mathbf{V}^T\mathbf{DV}$ is a symmetric matrix.

3 Application to three-way interlaboratory data

The described method shall now be applied to a more extensive data array, which is taken as an appropriate for three-way analysis part of a complex interlaboratory study discussed in [3]. In the original study, 102 laboratories participated in the trace element analysis of four different sediment samples. The whole spectrum of variables covered a range of 55 elements but, due to instrumental restrictions, practically all laboratories considered only a suitable subset of these elements. Each analysis was provided with four parallel estimations. In [3], conventional data analysis was applied. For our purposes, one may extract a data cube consisting of 88 objects (22 laboratories with 4 parallel estimations), 15 variables (trace concentrations of Hg, Cd, Pb, Cu, Ni, Cr, Zn, As, Fe, Mn, Al, Ca, Mg, Tl, Sn) and 4 ‘conditions’ (sediment samples). For simplicity, a two-component solution was considered for each mode.

It turns out that a conditional simplification of the core with the optimally simplified component matrices for objects and variables held fixed yields a value of *var* = 99.1 which is almost the same as the maximum value of *var* = 104.8 obtained for a completely free core simplification. Since this latter solution comes along with component matrices not being simple at all, the mentioned conditional solution is much more preferable. On the other hand, one has to restrict the component simplification to objects and variables, since further inclusion of ‘conditions’ (sediments) leads to a considerable increase of core complexity (*var* = 47.4).

The resulting PC plots are given in Fig. 6. The plot of objects reveals a strong outlying position of laboratory ‘j’, whereas the averages of the remaining ones are more or less distributed along the PC1 axis with quite small scat-

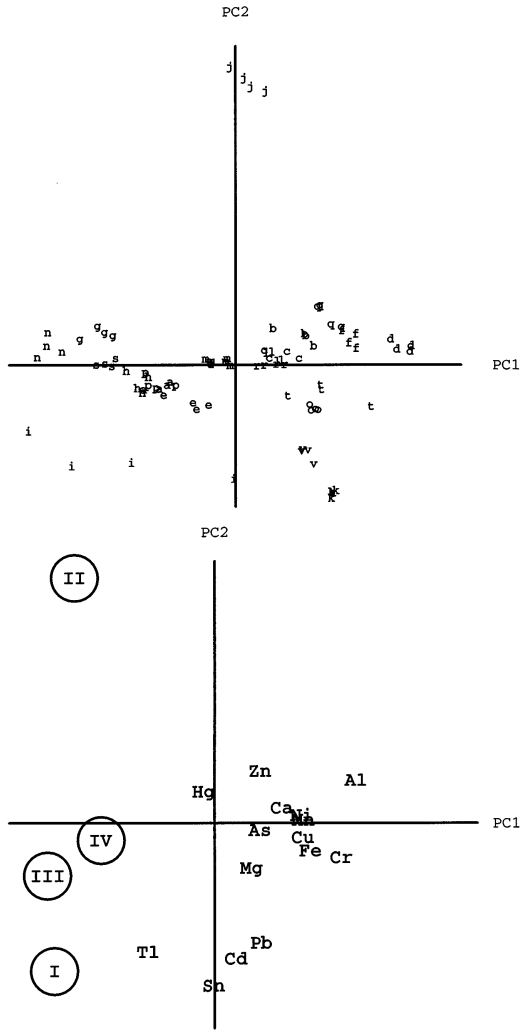

Fig. 6 Diagram plots of the Tucker3 solutions for the example of three-way interlaboratory data. The upper diagram refers to the plot of 4 parallel estimations of 22 laboratories. The lower diagram joins for space reasons the plots of the 15 elements and the 4 sediments (encircled) considered in the analysis

ter ranges for parallel estimations around these averages. That is why PC1 may be interpreted as a direction of systematic error and PC2 as a direction of a specific gross error. Among the variables, two subgroups are recognized and each of them identified with one of the two PC axes. Finally, the four sediments have almost uniform loadings on PC1 but sediment no. 2 gets a particularly high loading on PC2. A joint discussion requires the core matrix which is given here as

$$\mathbf{C} = \left(\begin{array}{cc|cc} 2.87 & -0.12 & -0.14 & -0.24 \\ -0.35 & 2.85 & 0.16 & 0.14 \end{array}\right).$$

Evidently, there are only two major contributions present which account for 98% of the total sum of squares in the core, namely $c_{1,1,1}$ and $c_{2,2,1}$. Both entries refer to an equal third index of 1 which relates to PC1 of the third mode (sediments). This means that the main error variation in this study is uniform with respect to all four sediments, hence the occurring errors are stable and reproducible for different samples. In other terms, there is no substantial error due to instable working which is sometimes observed in similar studies (high overestimations in one sample and extreme underestimations in the other). The outlying position of sediment 2 along PC2 is connected with a negligible source of error variation. Now, $c_{1,1,1,}$ combines the direction of systematic error with the following group of elements having high loadings on PC1 and the correlation of which is not surprising: Al, Cr, Fe, Ni, Mn, Cu. Consequently, laboratories ‘d’ on the one end of PC1 and ‘n’ on the other are mainly distinguished by significant over- or underestimations (compared to the average) concerning the mentioned group of elements. Laboratory ‘m’ takes a medium position and meets the average determinations in more or less all cases. In contrast, $c_{2,2,1}$ explains the outlying position of laboratory ‘j’ and relates it with extreme deviations in the analysis of Tl, Sn, Cd and Pb. To a smaller degree, such deviations become visible for laboratories, ‘i’, ‘k’, and ‘v’ too, but with the opposite sign of deviation. Similar to the systematic error along PC1, this specific gross error is stable over different sediments.

This short discussion demonstrates how a quick graphically oriented overview of the main error sources in a three-way interlaboratory study can be obtained. More detailed information would be available after passing to a bigger number of components in each mode as it was done in the example of game samples.

Acknowledgements The author thanks Prof. K. Lusky and Dr. M. Stoyke (Institut für Veterinär-Pharmakologie and Toxikologie GmbH Bernau) as well as Dr. P. Heininger and Dr. J. Pelzer (Bundesanstalt für Gewässerkunde Berlin) for providing data sets used in this paper.

References

1. Andersson CA, Henrion R (1998) Computational Statistics and Data Analysis (in press)
2. Harshman RA (1970) UCLA working papers in phonetics 16: 1–84
3. Heininger P, Pelzer J, Henrion G, Henrion R (1998) Fresenius J Anal Chem 360: 344–347
4. Henrion R (1993) J Chemometrics 7: 477–494
5. Henrion R (1994) Chemom Intell Lab Syst 25: 1–23
6. Henrion R, Andersson CA (1998) Chemom Intell Lab Syst (in press)
7. Kaiser HF (1958) Psychometrika 23: 187–200
8. Kiers HAL (1992) Statistica Applicata 4: 659–667
9. Kiers HAL (1998) Comp Stat Data Anal (in press)
10. Kroonenberg PM, de Leeuw J (1980) Psychometrika 45: 69–97
11. Lusky K, Lippert A, Stoyke M, Bohm D, Hecht H, Luthardt M (1994) Fleichwirtschaft 74: 189–191
12. Tucker LR (1963) In: Harris CW (ed) Problems of measuring change. Univ. of Wisconsin Press, Madison, p 122