

University of Groningen

Component analysis of multisubject multivariate longitudinal data

Timmerman, Marieke Engelen

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2001

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Timmerman, M. E. (2001). *Component analysis of multisubject multivariate longitudinal data*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

2. Component models for longitudinal three-way data

2.1. Introduction

In this chapter, a class of component models, that can be applied to longitudinal three-way data is discussed. Longitudinal three-way data are defined as multisubject multivariate data measured at various occasions that are comparable across all subjects (see Section 1.3). The models aim at providing a meaningful summary of the data. For that purpose, the data are decomposed into a number of interpretable matrices, so that the data are described parsimoniously. Before the models and some issues in empirical applications are discussed (Sections 2.3 through 2.6), some terminology and notation used throughout this study will be introduced. The interpretation of longitudinal three-way models is treated in Section 2.7.

2.2. Notational issues and some matrix algebraic properties

Element x_{ij} of the $I \times J$ two-way datamatrix \mathbf{X} is denoted by x_{ij} . The j^{th} column of \mathbf{X} is denoted by \mathbf{x}_j , and the i^{th} row by \mathbf{x}_i' . The symbol \mathbf{I}_Q denotes the $Q \times Q$ identity matrix. The $P \times Q$ matrix having each element equal to zero is denoted by $\mathbf{0}_{P \times Q}$. The $Q \times 1$ vectors with each element equal to zero and one are denoted by $\mathbf{0}_Q$ and $\mathbf{1}_Q$, respectively, or just $\mathbf{0}$ and $\mathbf{1}$ if the size of the vector is clear from the context.

Longitudinal three-way data consist of scores on J variables ($j=1, \dots, J$) collected from I subjects ($i=1, \dots, I$) at K measurement occasions ($k=1, \dots, K$). The data can be collected in a three-way data array $\underline{\mathbf{X}}$ ($I \times J \times K$). A graphical representation of such a data array $\underline{\mathbf{X}}$ is presented in Figure 2.1.

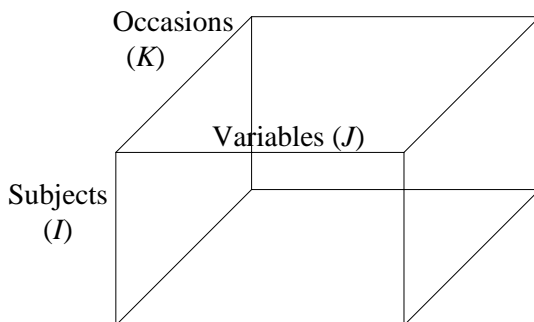


Figure 2.1. Graphical representation of a three-way data array $\underline{\mathbf{X}}$ of size $I \times J \times K$.

This type of three-way data is also called three-mode data (Carroll & Arabie, 1980). The number of ways refers to the geometrical dimension of the array, whereas the number of modes refers to the number of different sets of entities. Hence, a set of covariance matrices obtained at different occasions from scores on the same variables would make up a two-mode three-way data array.

The data in Figure 2.1 can also be presented in I , J or K separate submatrices. The k^{th} frontal slab of $\underline{\mathbf{X}}$ is denoted by \mathbf{X}_k ($I \times J$). Analogously, \mathbf{X}_i ($J \times K$) is the i^{th} horizontal slab, and \mathbf{X}_j ($K \times I$) the j^{th} lateral slab of $\underline{\mathbf{X}}$. Element x_{ijk} of the array $\underline{\mathbf{X}}$ ($I \times J \times K$) is denoted by x_{ijk} . Sometimes it is convenient to present the full three-way data set in ‘matricized’ form (see Kiers, 2000), that is, as a supermatrix with frontal, lateral or horizontal slabs next to each other. The size of the supermatrix and the subscript of the matrix symbol indicate the positioning of the submatrices in the sequel. Specifically, the matrix \mathbf{X}_a ($I \times JK$) contains K frontal slabs \mathbf{X}_k ($I \times J$), $k=1, \dots, K$, positioned next to each other. Analogously, the matrix \mathbf{X}_b ($J \times KI$) denotes a supermatrix in which the I horizontal slabs \mathbf{X}_i ($J \times K$), $i=1, \dots, I$, are positioned next to each other, and the matrix \mathbf{X}_c ($K \times IJ$) is the supermatrix that contains the J lateral slabs \mathbf{X}_j ($K \times I$) positioned next to each other. Another option, though rarely used, is to present the data in $\underline{\mathbf{X}}$ as a vectorized version of \mathbf{X}_a . The vectorized version of \mathbf{X}_a ($I \times JK$) is denoted by $\text{Vec}(\mathbf{X}_a)$, and it is obtained by stringing out \mathbf{X}_a column-wise to a column vector of length IJK .

The Khatri-Rao product, or the column-wise Kronecker product, is denoted by \odot (Rao & Mitra, 1971). The Khatri-Rao product of two matrices \mathbf{A} and \mathbf{B} , each with R columns, is defined as

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \quad \mathbf{a}_2 \otimes \mathbf{b}_2 \quad \dots \quad \mathbf{a}_R \otimes \mathbf{b}_R], \quad (2.1)$$

where \otimes denotes the Kronecker product.

The elements of a matrix product of the form \mathbf{ABC} can be expressed as a function of $\text{Vec}(\mathbf{B})$ as

$$\text{Vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A}) \text{Vec}(\mathbf{B}), \quad (2.2)$$

see, for instance, Magnus and Neudecker (1991).

2.3. Component models for three-way data

As the component models described below are generalizations of the well-know principal component model, this model for two-way data is discussed first. A principal component analysis (PCA) aims at describing the scores on a number of variables as a weighted sum of scores on a smaller number of components. Specifically, in PCA one decomposes a two-way data matrix \mathbf{X} ($I \times J$) as

$$\mathbf{X} = \mathbf{A}\mathbf{B}' + \mathbf{E}, \quad (2.3)$$

where \mathbf{A} ($I \times Q$) denotes the subject component matrix, \mathbf{B} ($J \times Q$) the matrix with so-called variable loadings, \mathbf{E} ($I \times J$) the matrix with residuals and Q , $q=1, \dots, Q$, the number of components. The matrix with residuals is that part of the data that is not covered by the model. The data matrix is decomposed in a way that minimizes the sum of squared residuals. The part of the data that is covered by the structural part of the model ($\mathbf{A}\mathbf{B}'$ in the case of a PCA), hence without the residual part, is denoted by $\hat{\mathbf{X}}$ in the sequel.

Standard two-way principal component analysis leads to solutions that can be rotated without loss of fit. To avoid the rotational indeterminacy, Cattell (1944), and Cattell and Cattell (1955) proposed looking for 'Parallel Proportional Profiles' (PPP) in two PCA solutions of two two-way data matrices, where the same variables on the same observation units have been measured twice. Harshman (1970) generalized Cattell's idea from two to K 'parallel occasions', and showed that PPP entail the 'PARAFAC' model, which is a three-mode generalization of the principal component model. In the PARAFAC model, only proportional differences exist between the subjects, variables and occasions with respect to each component. Carroll and Chang (1970) independently proposed their CANDECOMP model, which is equivalent to the PARAFAC model. We will refer to the CANDECOMP/PARAFAC model as the CP model. The CP model is defined as

$$x_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr} + e_{ijk} \quad (2.4)$$

where x_{ijk} denotes element i, j, k of a three-way data array $\underline{\mathbf{X}}$, a_{ir} , b_{jr} , c_{kr} denote elements of the component matrices \mathbf{A} ($I \times R$), \mathbf{B} ($J \times R$), \mathbf{C} ($K \times R$), respectively, and e_{ijk} denotes element i, j, k of the error array $\underline{\mathbf{E}}$ ($I \times J \times K$), $i=1, \dots, I$, $j=1, \dots, J$, $k=1, \dots, K$, $r=1, \dots, R$.

The CP model can be written in matrix notation as

$$\mathbf{X}_a = \mathbf{A}(\mathbf{C} \odot \mathbf{B})' + \mathbf{E}_a, \quad (2.5)$$

where \mathbf{X}_a is the $I \times JK$ matricized data array $\underline{\mathbf{X}}$, \odot denotes the Khatri-Rao product, the matrix \mathbf{C} denotes the occasion component matrix, and, analogously to the PCA model, the matrix \mathbf{A} denotes the subject component matrix and \mathbf{B} the variable component matrix, and \mathbf{E}_a ($I \times JK$) is the matricized error array $\underline{\mathbf{E}}$. Note that the variable component matrix \mathbf{B} has an equivalent role and interpretation as the loading matrix \mathbf{B} has in the PCA model (see (2.3)). To stress the fact that the CP model is completely symmetric, we name the matrices \mathbf{A} , \mathbf{B} , and \mathbf{C} all 'component matrices', instead of using different names for the three.

An alternative three-way generalization of PCA is the Tucker3 model (Tucker, 1966a; Kroonenberg & De Leeuw, 1980; Kroonenberg, 1983). Just as in the CP

model, the three-way array is decomposed into three component matrices, and the model is symmetric. However, in the Tucker3 model, all components of \mathbf{A} , \mathbf{B} , and \mathbf{C} are related to each other via the so-called core array. Also, the numbers of components of \mathbf{A} , \mathbf{B} , and \mathbf{C} may differ. The Tucker3 model is defined as

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk}, \quad (2.6)$$

where x_{ijk} denotes element i, j, k of the $I \times J \times K$ three-way array $\underline{\mathbf{X}}$, a_{ip} , b_{jq} , and c_{kr} denote the elements of the component matrices \mathbf{A} ($I \times P$), \mathbf{B} ($J \times Q$), and \mathbf{C} ($K \times R$), respectively; g_{pqr} denotes the elements of the core array $\underline{\mathbf{G}}$ ($P \times Q \times R$), and e_{ijk} denotes element i, j, k of the error array $\underline{\mathbf{E}}$ ($I \times J \times K$), $i=1, \dots, I$, $j=1, \dots, J$, $k=1, \dots, K$; $p=1, \dots, P$, $q=1, \dots, Q$, $r=1, \dots, R$. The elements of the core array describe the weights of the interactions of the components in \mathbf{A} , \mathbf{B} , and \mathbf{C} .

The Tucker3 model can be written in matrix notation as

$$\mathbf{X}_a = \mathbf{A} \mathbf{G}_a (\mathbf{C}' \otimes \mathbf{B}') + \mathbf{E}_a, \quad (2.7)$$

where \mathbf{X}_a is the $I \times JK$ matricized data array $\underline{\mathbf{X}}$, \mathbf{G}_a ($P \times QR$) denotes the supermatrix containing the frontal slices of the core array $\underline{\mathbf{G}}$ ($P \times Q \times R$), \otimes the Kronecker product, and \mathbf{E}_a ($I \times JK$) the matricized error array $\underline{\mathbf{E}}$.

The CP model can be notated equivalently to the Tucker3 model (2.7), namely as

$$\mathbf{X}_a = \mathbf{A} \mathbf{H} (\mathbf{C}' \otimes \mathbf{B}') + \mathbf{E}_a, \quad (2.8)$$

where \mathbf{X}_a is $I \times JK$, the fixed matrix \mathbf{H} is the $Q \times Q^2$ two-way version of the 'superidentity' three-way array $\underline{\mathbf{H}}$, that is, an array with $h_{pqr}=1$ if $p=q=r$, and $h_{pqr}=0$ otherwise. This notation shows at once that the CP model is a constrained version of the Tucker3 model, a result given earlier by Carroll and Chang (1970, p. 312).

The Tucker2 model (Kroonenberg & De Leeuw, 1980) is less restricted than the Tucker3 and CP models. In the Tucker3 and CP models, each of the three modes are reduced into components. In the Tucker2 model, two modes are reduced and one is left unreduced. Consequently, the Tucker2 model is not symmetric, whereas the Tucker3 and CP models are. Here, we present the Tucker2 model for the situation where the subject mode is unreduced, but one can equally well choose to leave the variable or occasion mode unreduced. The Tucker2 model is given by

$$\mathbf{X}_a = \tilde{\mathbf{G}}_a (\mathbf{C}' \otimes \mathbf{B}') + \mathbf{E}_a, \quad (2.9)$$

where $\tilde{\mathbf{G}}_{\mathbf{a}}$ ($I \times QR$) denotes the ‘extended core matrix’. The elements of the extended core matrix describe the weights of the interaction between the combinations of the Q variable components and the R occasion components for each of the I subjects.

The relationship between the Tucker2 and Tucker3 models can be seen as follows. If in the Tucker3 model (2.7) the number of subject components is chosen to be I , then the estimated component matrix \mathbf{A} is a basis for the full subject space. In this case, the Tucker3 model boils down to the Tucker2 model, because then $\tilde{\mathbf{G}}_{\mathbf{a}}$ can be written as $\tilde{\mathbf{G}}_{\mathbf{a}} = \mathbf{A}\mathbf{G}_{\mathbf{a}}$.

The Tucker1 model is the most weakly constrained of the models for longitudinal three-way data that are discussed here. In the Tucker1 model, only one mode is reduced. Here, we treat the case for which the subject mode is reduced, and the variable and occasion modes are left unreduced. The Tucker1 model with reduced subject mode is given by

$$\mathbf{X}_{\mathbf{a}} = \mathbf{A}\mathbf{F}' + \mathbf{E}_{\mathbf{a}}, \quad (2.10)$$

where \mathbf{F} ($JK \times P$) denotes a component matrix containing the loadings of all combinations of variables and occasions on the P components. Note that the Tucker1 model equals a PCA model of a matricized three-way array. In practical applications of the Tucker1 model, it is often natural to reduce the variable mode, and to leave the subject and occasion modes unreduced. This can be done by applying the Tucker1 model to the matrix $\mathbf{X}_{\mathbf{b}}$ ($J \times KI$). With the exception of the Tucker2 model, the hierarchical relations between the three-way models mentioned above are also discussed by Kiers (1991). The hierarchical relation between the Tucker3 and Tucker2 models is treated by Kroonenberg and de Leeuw (1980).

2.4. Fitting the component models for three-way data

The Tucker1, Tucker2, Tucker3, and CP models are fitted to observed data by minimizing the sum of squared residuals $\|\mathbf{X}_{\mathbf{a}} - \hat{\mathbf{X}}_{\mathbf{a}}\|^2$, where $\hat{\mathbf{X}}_{\mathbf{a}}$ denotes the part of $\mathbf{X}_{\mathbf{a}}$ that is covered by the structural part of the model, and $\|\cdot\|^2$ denotes the squared Euclidean norm. The degree to which the estimated model describes the data is expressed by the proportion of sum of squares explained by the model, which we call the ‘fit’ in the sequel. The fit is defined as

$$1 - \frac{\|\mathbf{X}_{\mathbf{a}} - \hat{\mathbf{X}}_{\mathbf{a}}\|^2}{\|\mathbf{X}_{\mathbf{a}}\|^2}. \quad (2.11)$$

The fit is often expressed as a percentage by multiplying it by 100.

Because the Tucker1 model equals a PCA model of a matricized three-way array, estimates of the parameters of the model are obtained by applying a standard PCA procedure. Kroonenberg and de Leeuw (1980) gave algorithms for fitting the Tucker2 and Tucker3 models to data. Other ways of fitting the Tucker3 model to data have been proposed by Weesie and Van Houwelingen (1983), Kiers, Kroonenberg and Ten Berge (1992), Andersson and Bro (1998), and Paatero and Andersson (1999). Harshman (1970), Carroll and Chang (1970), Kiers (1998a) and Paatero (1999) gave algorithms to estimate the parameters of the CP model. Typically, the algorithms are Alternating Least Squares (ALS) algorithms. In ALS algorithms, the matrices over which the function has to be minimized are alternately updated until convergence. The algorithm is said to have converged, if from one cycle (i.e., update of all parameters) to another, the residual sum of squares decreases less than a prespecified tolerance value. The latter tolerance value is sometimes denoted by the 'convergence criterion'. As the residual sum of squares decreases monotonically, and is bounded below by zero, convergence in terms of this criterion is guaranteed. However, it is not guaranteed that the global minimum of the function is reached, hence the algorithm may land in a local minimum. In practice, it is recommended to use several differently started runs in order to decrease the chance of missing the global minimum of the function.

2.5. Transformational freedom within the component models for three-way data

The Tucker3, Tucker2 and Tucker1 models have transformational freedom. In the Tucker3 model, this means that estimates of the data array $\underline{\mathbf{X}}$ are insensitive to orthogonal and oblique transformations of the component matrices \mathbf{A} , \mathbf{B} , and \mathbf{C} , provided that such transformations are compensated in the core array $\underline{\mathbf{G}}$ (Tucker, 1966a). The same holds for the two component matrices in the Tucker2 model, provided that the transformations are compensated in the extended core matrix. The component matrix of the reduced mode in the Tucker1 model can also be transformed, provided that this transformation is compensated in the component matrix for the two unreduced modes. This case is identical to ordinary two-way PCA, whose transformational freedom is well known.

Standard rotational procedures aiming at a simple structure (e.g., Varimax; Kaiser, 1958) can be used to obtain more easily interpretable solutions for the subject and variable related component matrices. Those procedures are generally not useful in transforming the occasion component matrix, as will be illustrated later. In the Tucker2 and Tucker3 models, one can also transform the (extended) core array (over all two and three modes, respectively), provided that the transformations are compensated in the component matrices. Kiers (1997a) proposed a procedure that aims at obtaining simplicity of the core array in the Tucker3 model. Kiers (1998b)

discusses a method for *jointly* transforming the core, and the component matrices to simplicity in the Tucker3 model.

As Harshman (1972) proved, the parameter estimates of a CP model are ‘essentially unique’, if there is at least one pair $\mathbf{D}_k, \mathbf{D}_l$ ($k \neq l; k, l = 1, \dots, K$), so that $\mathbf{D}_k \mathbf{D}_l^{-1}$ has no equal pair of diagonal elements. Essential uniqueness of a CP solution means that CP estimates are unique up to trivial permutation, reflection and/or rescaling. An even more relaxed sufficient condition for the uniqueness of the estimates of the CP model was given by Kruskal (1977).

2.6. Issues in the application of the component models to three-way data

2.6.1. Preprocessing three-way data

In applying a PCA to a standard two-way matrix, one usually decomposes the centered and standardized data matrix, rather than the raw data. Centering and scaling serve different purposes, in two-way as well as multi-way component analysis (Harshman & Lundy, 1984b; Bro & Smilde, in preparation). Centering aims at removing constant terms in the data in order to make the data compatible with the model. That is, it causes interval scale data to behave as ratio scale data, which is required in component models. Scaling aims at eliminating artificial scale differences, and does not affect the model. It only influences the weights attached to certain elements in the least squares loss function.

Preprocessing in multi-way data is more complicated than it is in the two-way case. Harshman and Lundy (1984b) and Bro and Smilde (in preparation) extensively discuss centering and scaling of multi-way data. A centering is considered proper, if constants are eliminated from the data without introducing artificial variation. That is, if data (without error) consist of a structural part that is covered by the model, plus constant(s), then the centering should indeed remove those constants, without introducing new ones. Harshman and Lundy, and Bro and Smilde conclude that proper centering is obtained by centering *across one* mode. Several such centerings across one mode may be performed sequentially.

Harshman and Lundy (1984b), and Bro and Smilde (in preparation) showed that scaling is unproblematic only when performed *within one* mode (for example, per variable over occasions and subjects jointly). Combinations of centering and scaling do not conflict if scaling within one mode is combined with centering across (one of) the other modes.

2.6.2. Choice of a specific model and the numbers of components

In Section 2.3, four component models for three-way data were discussed. The models were ordered from severely constrained (CP model) to weakly constrained (Tucker1 model). A constrained version of a certain model implies that a restriction is imposed on certain parameters of that model. As a result, the fit of a constrained

model to the observed data set is at most equal to, but generally lower than the fit of its unconstrained version, all other things being equal. If one wishes to select a particular model and the number of components, one should take the interpretability and the degree of modeling error into account.

The interpretability of a model is important because in fitting models one aims at understanding underlying processes. The selection of an interpretable model is by no means a simple task. The interpretation of a model depends partly on the theoretical ideas and knowledge a researcher has. As a result, the criterion of interpretability adds a certain degree of subjectivity to the model selection process, and can therefore be subject to debate.

The second point of concern in model selection is the degree of modeling error. In fitting a model to data, one usually aims at obtaining a model that is more generally applicable than just to the current data. In the current context, this would imply that one has a three-way data set of which the individuals are thought to be drawn from a certain population. Then, the estimated model parameters are wished to be valid for the whole population, and not just for the current sample. In this context, it is useful to distinguish three types of error involved in fitting a particular model, namely the error of approximation, the error of estimation and the overall error (Linhart & Zucchini, 1986; Browne & Cudeck, 1992). Suppose one has chosen a particular model. The *error of approximation* is the lack of fit of the current model to the population data, if the parameters were optimally chosen. The error of approximation is not dependent on the data set at hand, on the sample size or the method of estimation. The *error of estimation* is the lack of fit of the current model fitted to the current data set, to the model of the population data with optimally chosen parameters. The error of estimation depends on the sample size, the current data set, and the method of estimation. Both types of error cannot be computed directly, because the population data, and the optimally chosen parameters of the model for the population data are unknown. The *overall error* refers to the lack of fit of the current model that is fitted to the current data set, to the population data. Usually, the overall error is the sum of the error of approximation and the error of estimation. Generally, as the number of parameters of a model increases, the error of approximation reduces, whereas the error of estimation increases.

In choosing a particular model, one should aim at finding an *interpretable* model with a small degree of overall error. A way to investigate the degree of overall error is using cross-validation, which assesses the predictive validity of the estimated model parameters. A cross-validation method, the ‘expectation maximization cross-validation’ (EM-CV), is proposed by Louwarse, Smilde and Kiers (1999) for application in the three-way context. The degree of overall error can also be examined using split-half analysis. Examples of split-half analysis applied to three-way models are presented by Harshman and Lundy (1984a), and Kiers and Van Mechelen (2001). In the references, the methods are applied to either the CP or Tucker3 model, but they can be adapted easily for application to other component models as well. In Chapter 6, the EM-CV method (Louwarse, Smilde & Kiers, 1999) and a variant of split-half

analysis (Harshman & Lundy, 1984a) are adapted, and applied to models for multivariate multisubject time series.

Besides modeling error, the interpretability of the model is important, because one aims at understanding processes. The requirement of interpretability may conflict with the desire to reduce the error of approximation, which is one part of the overall error. It is not useful to add parameters which are not interpretable at all to a model, only to reduce the error of approximation. Hence, it is more precise to say that it is desirable to minimize the error of approximation, provided that the fitted model is interpretable, and reflects underlying processes. Therefore, the decision on which model to take cannot be based solely on indices of the degree of overall error.

In certain cases, one may assume that the population three-way data consist of 'true scores', that follow the model concerned, and measurement error. In the framework discussed above, the measurement error is the error of approximation. The parameters of the models discussed here are obtained by least squares fitting of the model to observed three-way data. The estimated three-way array obtained this way will usually not be exactly equal to the 'true' three-way array, and the fitted model is partially based on measurement error. This phenomenon will be indicated by 'error fitting' in the sequel. Obviously, a high degree of error fitting is undesirable, since it may lead to errors in interpretation.

A special case, which is sometimes encountered in component modeling, occurs if one aims at meaningfully summarizing the current three-way data set only. One can view this case as having available the population data. Then, the overall error is fully or almost fully (depending on the discrepancy measure and the estimation procedure) determined by the error of approximation. Therefore, the fit of the estimated model to the observed data is a good indicator of the degree of error of approximation. One should find a balance between interpretability and degree of modeling error. One could try to select a parsimonious model that still covers the most important or most salient aspects of the data, thereby ignoring aspects of little importance (e.g., because they pertain to a small number of subjects, variables, or occasions) by using the fit to compare different estimated models. It is hoped that the model that covers the most important aspects is also a well interpretable model. Using this approach requires to choose the most constrained model with the smallest numbers of components that explains relatively much of the sum of squares. Specifically, for choosing the numbers of components in the Tucker3 model, Timmerman and Kiers (2000) proposed a method that aims at finding an optimal balance between the fit of solutions for the Tucker3 model and the numbers of components. They compared the results obtained by this method, called DIFFIT, with results from two other methods, by means of a simulation study. It was found that DIFFIT performed considerably better than the other methods in indicating the numbers of components.

2.7. Interpretation of three-way component models applied to longitudinal three-way data

First, the interpretation of the three parameter matrices of a CP model of longitudinal three-way data is discussed. The CP model (2.4) is the most constrained symmetrical three-way component model. The three modes are reduced into three component matrices that summarize the linear relationships between subjects, variables and occasions. As an empirical example, one could think of the scores on a number of variables that measure arithmetic and language skills. The scores are collected from a number of pupils receiving education in arithmetic and language at several time points. Suppose the variables are related to only two variable components, namely an arithmetic and a language component. The variable components can be viewed as referring to two ‘latent variables’, hence to variables underlying the observed variables. The relative size of a variable component score indicates the degree to which that (observed) variable measures the particular latent variable. In the CP model, the component matrices are usually not (columnwise) orthogonal, which implies that the components are linearly related to each other, to a smaller or larger extent. From a theoretical point of view, it is conceivable that latent variables are linearly related. For example, the degree of arithmetic and language skills are likely to be partly influenced by a general degree of intelligence. In the example, the two subject component scores indicate the relative position of the pupil to that of the other pupils on arithmetic and language. The occasion component scores at the time points are indicators of the overall level in arithmetic and language for the whole group at each particular time point. The scores at successive time points are interpreted as indicating the development over time. It may be helpful to plot the scores against time to get a notion of the development of the level over time. One interprets those scores as if they are evaluations of a certain latent curve. Note that the relative position of the subjects on each component is constant across time in the CP model. Supposing the time component scores to be positive, the model implies, for example, that if a pupil scores poorly in arithmetic at one time point, he will continue to score poorly at all subsequent time points.

A graphical representation of the CP model is given in

Figure 2.2. The figure shows that the three-way array $\hat{\mathbf{X}}$ is a sum of R ($r=1, \dots, R$) rank one three-way arrays, of which each element x_{ijk} is the triple product $a_{ir}b_{jr}c_{kr}$. The vectors with component scores can be collected into three component matrices for each of the three modes.

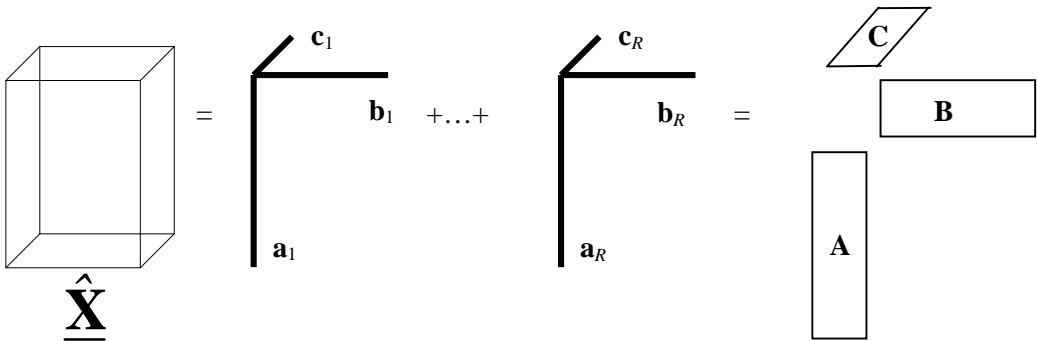


Figure 2.2. Graphical representation of the CP model.

As the Tucker3 model (2.7) is more flexible than the CP model, the interpretation is usually somewhat more complicated. The three component matrices summarize the properties of the subjects, variables and occasions. The weights of the combinations of the components are reflected in the core array. The variable component matrix is interpreted in the same way as in the CP model, that is in terms of latent variables. The subject components can be interpreted as ‘idealized’ or ‘prototype’ subjects (Kroonenberg, 1983). Real subjects are considered weighted combinations of the idealized subjects. The occasion components can be viewed as ‘idealized’ or ‘prototype’ occasions. The occasion component scores at the various time points reflect the level of the prototype occasion at the time points, and just as in the CP model, they can be interpreted as evaluations of a latent curve. Simplicity of the component matrices as well as the core array is needed to facilitate interpretability. A graphical representation of the Tucker3 model is given in Figure 2.3, which shows the decomposition of the three-way array into the three component matrices and the core array.

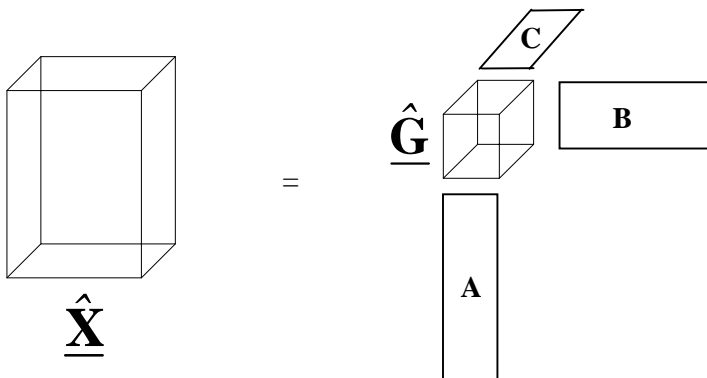


Figure 2.3. Graphical representation of the Tucker3 model.

The two component matrices of the two reduced modes in the Tucker2 model (2.9) are interpreted in the same way as their counterparts in the Tucker3 model. The extended core matrix provides the weights for all component combinations (of the reduced modes) per entity of the unreduced mode.

The component matrix of the reduced mode in the Tucker1 model (2.10) is interpreted in the same way as its counterpart in the Tucker2 or Tucker3 model. The component matrix contains the loadings for the components of the reduced mode for all combinations of entities of the two unreduced modes.