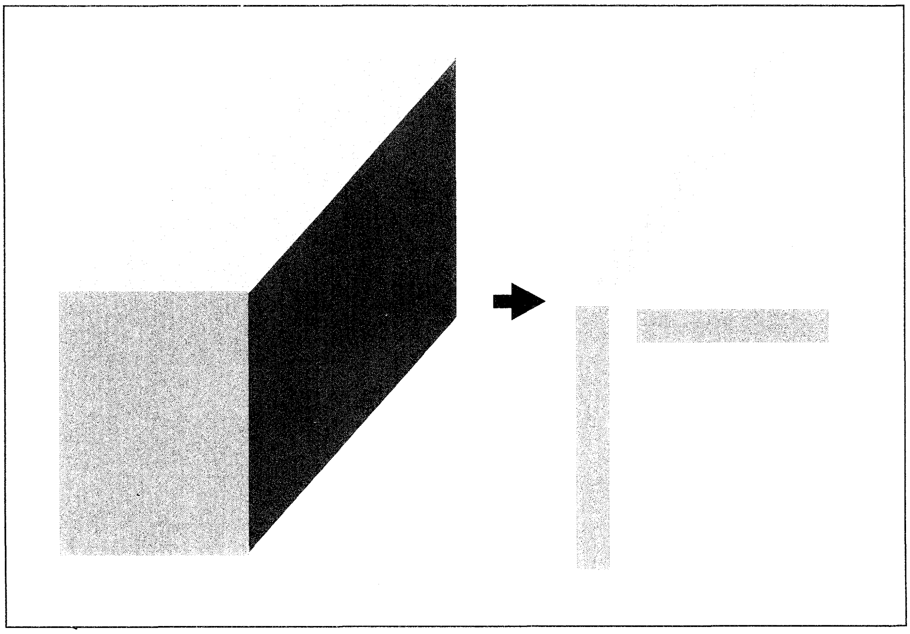
THE ANALYSIS OF THREE-WAY ARRAYS BY CONSTRAINED PARAFAC METHODS

Wim P. Krijnen



DSWO PRESS

RIJKSUNIVERSITEIT GRONINGEN

THE ANALYSIS OF THREE-WAY ARRAYS BY CONSTRAINED PARAFAC METHODS

Proefschrift

ter verkrijging van het doctoraat in de

Psychologische, Pedagogische en Sociologische Wetenschappen

aan de Rijksuniversiteit Groningen

op gezag van de

Rector Magnificus Dr. S. K. Kuipers

in het openbaar te verdedigen op

donderdag 21 januari 1993

des namiddags te 4.00 uur

door

Wilhelmus Petrus Krijnen

geboren op 29 juli 1960

te Stad Delden

Promotor: Prof. Dr. J.M.F. ten Berge

Referent: Dr. H.A.L. Kiers

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Krijnen, Wilhelmus Petrus

The analysis of three-way data by constrained PARAFAC methods /
Wilhelmus Petrus Krijnen. - [S.l. : s.n.]
Also publ. as commercial ed.: Leiden : DSWO Press, Leiden
University, 1993. - (M&T series ; 23). - Thesis Groningen. -
With ref. - With summary in Dutch.
ISBN 90-9005493-6
Subject headings: PARAFAC / psychotechnics.

Cover design: Wilhelmus Petrus Krijnen.
Printed by 'Reprodienst, Subfaculteit Psychologie, Rijksuniversiteit Leiden' and by printing office 'Karsten drukkers b.v.', Leiden.

ISBN 90-9005493-6

ACKNOWLEDGEMENTS

While working on this study I encountered a number of interesting research questions. Only a few of them have been answered. For a number of reasons I am indebted to Jos ten Berge and Henk Kiers. They always had an open ear for my new ideas. Whenever I had written a manuscript or a note I got it back with useful comments from either of them in due course.

During the research period I had the opportunity to sharpen my views by discussing them with Richard Harshman. His questions and remarks were a valuable contribution to the study.

I like to thank Richard Harshman, Ivo Molenaar and Willem Heiser for useful comments on an earlier version of the manuscript.

For kindly making their data available I like to thank Fred Wolters, Theo Nijsse, Richard Harshman, Pieter Kroonenberg, Gudrun Eckblad, Ria Neütel, Henk Lutje Spelberg, and Garmt Dijksterhuis.

For convenient working conditions during the research period I am grateful to the Department of Personality and Educational Psychology of the University of Groningen.

For personal support I am indebted to Siebina Slager.

CONTENTS

CHAPTER 1

A REVIEW OF THE PARAFAC METHOD

As a first exploratory step in the analysis of a data matrix with scores of n persons on m variables Principal Component Analysis (PCA) is often used. The main purpose of PCA is to summarize the most important information of the data. This is accomplished by representing the persons and the variables simultaneously in a limited number of optimal components. These components are optimal in the sense that they explain a maximal amount of variance. It is well-known that PCA allows for rotating the components without affecting the optimality of the components. Rotations are commonly used in order to find other components which allow, for instance, an easier interpretation.

In case one has scores from n persons on m variables measured at p occasions a generalization of PCA, called PARAFAC (Harshman, 1970), seems useful for the exploratory analysis of such data. PARAFAC represents the persons, the variables and the occasions simultaneously by a limited number of components. Unlike PCA, PARAFAC yields unique components. As a consequence, rotations cannot be used to find other components that can, for instance, be interpreted more easily.

The history of the PARAFAC method goes back to Cattell (1944). He reviewed seven principles for the choice of rotation in components analysis and advocated the principle of "parallel proportional profiles" as "the most fundamental principle" for rotation. Specifically, this principle means that two data matrices with the same variables should contain the same components and in the second matrix each component "should be accentuated or reduced in influence" (Cattell, 1944, p. 274). Harshman (1970) proposed using this principle as a constraint in a new method to analyze two or more data matrices that contain scores for the same persons on the same variables, and christened this method PARAFAC. The uniqueness of the PARAFAC components then obviates the difficult decision how to rotate the components.

For certain data sets the principle of proportional profiles seems to be inappropriate in the sense that there is hardly any differential accentuation or reduction in influence of the PARAFAC components. For such data, it does not make much sense to maintain the PARAFAC components, and alternative approaches seem indicated. Even if the PARAFAC components do have differential accentuations or reductions in influence, the PARAFAC solution may be undesirable from an interpretative point of view, and it may be worth considering alternative representations that fit the data almost as well and allow an easier interpretation. To verify whether such alternative representations can be found, it will be proposed here to consider certain constrained versions of PARAFAC. Constrained PARAFAC methods can be designed for various purposes. The present study will focus on constrained PARAFAC methods for assessing the nature of uniqueness, considering components with the same relative influences across occasions, analyzing positive manifold data, and determining PARAFAC components that correspond to non–overlapping clusters of variables. For each of these methods, algorithms will be constructed, and their usefulness will be demonstrated by means of analyses of empirical data. However, before treating any of these variants of PARAFAC, a review of the main features of PARAFAC will be given.

1.1 Definition of PARAFAC

Suppose that scores are available from, for instance, n persons on m variables measured at p occasions. Such data are called three–way data because per occasion scores are available of the same group of persons on the same group of variables. These three–way data can be collected in an $n \times m \times p$ three–way array X, which can be visualized as a box of scores with frontal slices containing the $n \times m$ data matrices for each of the p occasions. The PARAFAC model (Harshman, 1970) represents such a three–way array by

$$x_{ijk}=\sum_{r=1}^{q} a_{ir}b_{jr}c_{kr}+e_{ijk}, \tag{1.1}$$

where q is the number of components, x_{ijk} is the element (i,j,k) of X, a_{ir} is the element (i,r) of an $n \times q$ matrix A with coefficients of the persons on the q components, b_{jr} is the element (j,r) of an $m \times q$ matrix B with coefficients of the variables on the q components, c_{kr} is the element (k,r) of a $p \times q$ matrix C with coefficients of the occasions on the q

components, and e_{ijk} is the ijk^{th} residual element, i=1,...,n, j=1,...,m and k=1,...,p. In the sequel A, B and C will be called component matrix, pattern matrix and occasions matrix, respectively. In matrix notation the PARAFAC model can be written as

$$X_k = AD_kB' + E_k, \tag{1.2}$$

where X_k is frontal slice k of X, D_k is a diagonal matrix with the elements of row k of C on its diagonal, and E_k is the nxm residual matrix for occasion k, k=1,...,p.

From (1.1) it can be seen that the PARAFAC model treats its parameter matrices in a symmetric way. If X is sliced into n horizontal slices or into m lateral slices, then matrix equations equivalent to (1.2) can be formulated. In the sequel this symmetric role of the parameter matrices will be called the symmetry property of the PARAFAC model.

From (1.2) it can be seen that the columns of the matrices A and B can be scaled to arbitrary lengths, if this scaling is compensated for by D_k, k=1,...,p. From this and the symmetry property it follows that, without loss of generality, two of the matrices A, B and C can be scaled to unit length column-wise, as is done frequently.

From (1.2) it can also be seen that the representations AD_1B',...,AD_pB' of the frontal slices only differ through the diagonal matrices D_1,...,D_p. That is, the PARAFAC representations of the frontal slices consist of a common matrix A and a common matrix B. The diagonal matrices D_1,...,D_p rescale the columns of A. As a result, there are PARAllel FACtors (called components here) in AD_k, k=1,...,p, across the occasions, hence the name PARAFAC. In other words, these components are perfectly congruent, and according to the model, the frontal slices are represented by linear combinations of these components. The weights for these linear combinations are contained in B.

1.2 Fitting the PARAFAC model

For fitting the PARAFAC model Harshman (1970) defined a loss function that expresses the discrepancy between the data and the PARAFAC representation. In PARAFAC, the loss function to be minimized is the sum of squares of the elements of the residual matrices, which may be written as

$$\text{PARAFAC}(A,B,C)=\sum_{k=1}^{p}\|X_k-AD_kB'\|^2. \tag{1.3}$$

There is no method available to determine the minimizing A, B and C directly. Therefore, an Alternating Least Squares (ALS) algorithm to minimize PARAFAC(A,B,C) is used (Harshman, 1970; Carroll & Chang, 1970). Carroll and Chang (1970, p. 310) called this algorithm a "CANonical DECOMPosition" procedure of a three-way array, or, briefly, CANDECOMP. For the sake of simple terminology, it will be called PARAFAC algorithm in this study. The PARAFAC algorithm is based on alternatingly minimizing PARAFAC(A,B,C) over A for fixed B and C, over B for fixed A and C, and over C for fixed A and B. The updates for A, B, and C are

$$A=\sum_{k=1}^{p} X_k B D_k \left(\sum_{l=1}^{p} D_l B' B D_l\right)^{-1}, \tag{1.4}$$

$$B=\sum_{k=1}^{p} X_k' A D_k \left(\sum_{l=1}^{p} D_l A' A D_l\right)^{-1} \tag{1.5}$$

and

$$\mathbf{c}_k=(A'A*B'B)^{-1}[\mathrm{Diag}(A'X_kB)]\mathbf{1}, \tag{1.6}$$

k=1,...,p, respectively, where * denotes the element-wise (Hadamard) product and **1** the q-vector with unit elements (see, e.g., Kroonenberg, 1983, p. 116). The process of updating A, B, and C is continued until the value of PARAFAC(A,B,C) stabilizes up to an arbitrary constant, taken as 0.0001 in this study. The PARAFAC algorithm decreases the function value monotonically. Unfortunately, it is not guaranteed that the global minimum will be reached. In practice, this problem of local minima can be dealt with by taking a number (5 in the present study) of different runs of the PARAFAC algorithm from different starting configurations.

The above stopping criterion differs from the stopping criterion in the PARAFAC program (Lundy & Harshman, 1985) which is based upon the percentage of change in every column of a parameter matrix from one iteration to the next. The iterative process is continued until this percentage of change is lower than 0.1 for all parameters. It should be noted, however, that convergence of the parameter matrices cannot be guaranteed, see section 1.6. On the other hand, it is clear that, if the parameter matrices do converge, the above function value also converges.

Throughout this study, the iterative process was continued until both stopping criteria were satisfied, with a maximum of 1000 iterative cycles. In the sequel, in case the stopping criteria were not found to be satisfied or in case of non-convergence this will be mentioned.

As noted above, the number q denotes the number of components or the dimensionality of the PARAFAC solution. It has to be chosen by the user. It is well-known that solutions of different dimensionalities are not nested. That is, if a solution for PARAFAC(A,B,C) is globally optimal in q dimensions, then there need not be a subset of q–1 columns in these A, B and C that globally minimizes the PARAFAC loss function with dimensionality q–1.

PARAFAC usually finds oblique components. In case p=1, PARAFAC minimizes PARAFAC$(A,B,D_1)=\|X_1-AD_1B'\|^2$, which is equivalent to PCA. Hence PCA is a special case of PARAFAC, and PARAFAC can be seen as a straightforward generalization of PCA with oblique components. Often in PCA orthonormal components are determined. PCA can also be generalized to PARAFAC with orthonormal components (Harshman & Lundy, 1984a, p. 129). This requires minimizing PARAFAC(A,B,C) subject to the constraint $A'A=I_q$. The matrix A, subject to the constraint $A'A=I_q$, can be updated as follows (see, e.g., Kiers & Krijnen, 1991). Let $\sum_{k=1}^{p} X_kBD_k=PDQ'$ ($P'P=Q'Q=QQ'=I_q$) be the singular value decomposition (SVD) of $\sum_{k=1}^{p} X_kBD_k$. The solution for A is PQ' (Cliff, 1966). An ALS algorithm, that minimizes PARAFAC(A,B,C) subject to $A'A=I_q$, is constructed by taking, after an arbitrary start, PQ' as the update for A and taking updates for B and C according to (1.5) and (1.6). This method will be called PFORTA here, because it enables one to analyze a three-way array by ParaFac with a column-wise ORThonormal A. Note that, due to the orthonormality constraint, the residual sum of squares of PFORTA will be at least as large as the residual sum of squares of PARAFAC.

1.3 Partitioning the fit in PFORTA

From (1.2) it can be seen that the matrix D_k accentuates or reduces the importance of the components for the representation of X_k, as prescribed by Cattell's principle of proportional profiles. For this reason, Harshman and Lundy (1984a, p. 155) propose to interpret the diagonal elements of D_k as the relative importances of the q components for occasion k. For example, if the l^{th} diagonal element of D_k equals zero, then component l is of no importance for occasion k. If the components (columns of A) are orthonormal, and B has unit length column-wise, then, like the squared singular values in PCA, the squared diagonal elements of D_k express exactly the sum of squares accounted for by the components for occasion k, (see Harshman & Lundy, 1984, p.199). Specifically, if $A'A=I_q$ and $\text{Diag}(B'B)=I_q$, then, upon convergence of the PARAFAC algorithm,

$$\|X_k-AD_kB'\|^2=\|X_k\|^2-\text{tr}D_k^2=\|X_k\|^2-\sum_{l=1}^{q}c_{kl}^2. \tag{1.7}$$

Because $\|X_k\|^2$ is the sum of squares to be explained for occasion k, and $\|X_k\|^2-\text{tr}D_k^2=\|X_k\|^2-\sum_{l=1}^{q}c_{kl}^2$ is the unexplained part, it follows that $\sum_{l=1}^{q}c_{kl}^2$ denotes the explained part, called fit for occasion k. Therefore, c_{kl}^2

denotes the sum of squares that component l explains for occasion k.

Equation (1.7) can be proven as follows. For every stationary point of PARAFAC(A,B,C) equation (1.6) holds. From $A'A=\text{Diag}(B'B)=I_q$, it follows that $(A'A{*}B'B)=(I_q{*}B'B)=\text{Diag}(B'B)=I_q$, and hence $D_k=\text{Diag}(A'X_kB)$. Expanding $\|X_k-AD_kB'\|^2$ and substituting D_k for $\text{Diag}(A'X_kB)$ and I_q for $\text{Diag}(B'B)$ shows that

$$\|X_k-AD_kB'\|^2=\text{tr}X_k'X_k-2\text{tr}X_k'AD_kB'+\text{tr}D_k^2B'B$$

$$=\ \|X_k\|^2-2\text{tr}\text{Diag}(A'X_kB)D_k+\text{tr}D_k^2\text{Diag}(B'B)$$

$$=\ \|X_k\|^2-2\text{tr}D_k^2+\text{tr}D_k^2=\|X_k\|^2-\text{tr}D_k^2, \tag{1.8}$$

which completes the proof.

From (1.8) it follows immediately that, upon convergence, $\text{PARAFAC}(A,B,C)=\sum_{k=1}^{p}\|X_k\|^2-\|C\|^2$. Analogously, it can be proven that, if $A'A=I_q$, $\text{Diag}(C'C)=I_q$ and PARAFAC has converged, then b_{jl}^2 is the sum of squares that component l explains for variable j, and we have

$$\text{PARAFAC}(A,B,C)=\sum_{j=1}^{m}\left(\sum_{k=1}^{p}\|\mathbf{x}_{jk}\|^2-\sum_{l=1}^{q}b_{jl}^2\right)=\sum_{k=1}^{p}\|X_k\|^2-\|B\|^2, \tag{1.9}$$

where $\mathbf{x}_{jk}$ denotes variable j at occasion k.

1.4 Preprocessing the data before a PARAFAC analysis

Before one analyzes a three–way data array one will usually preprocess the data in order to remove, for instance, unwanted constants, see Harshman and Lundy (1984b, p.216) and Kruskal (1984). There are many methods to preprocess a three–way array, see Ten Berge (1989) for a taxonomy. In the present study, two preprocessing methods will be used. Both methods can be seen as generalizations of standardizing variables prior to PCA. The first method is to center within the occasions, that is, to center X_k, k=1,...,p, column–wise, and to rescale the variables to unit length across the occasions, that is, to rescale $\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ to unit length column–wise. The second method is to center across the occasions and rescale the variables to unit length across the occasions. These two preprocessing methods are simple and can be applied to different types of data.

There are two other preprocessing methods that can be seen as generalizations of standardizing variables prior to PCA. In a first method the variables are centered within the occasions and scaled to unit length within the occasions. In a second method the variables are centered across the occasions and scaled to unit length within the occasions. These preprocessing methods are considered as "not appropriate for the PARAFAC model", because the number of components tends to increase for such data (Harshman & Lundy, 1984b, p. 246). In addition, it seems highly unlikely, for such preprocessed data, that the components can have distinct accentuation and reduction in influence. For this reason, these preprocessing methods are not used in the present study.

1.5 Interpretation of the PARAFAC components

In PCA followed by an oblique rotation a structure matrix, which contains correlations between the variables and the components, is helpful for interpreting or naming the components (Mulaik, 1972, pp. 101–102; Harman, 1976, p. 22), because it enables one to see which variables are similar to the components. In definitions of the PARAFAC model a structure matrix is lacking. Therefore, a structure matrix will be defined here. A structure matrix may be defined per frontal slab as $S_k \equiv X_k' AD_k$. If X_k is centered column–wise, then the elements of S_k are covariances between the variables of occasion k and the weighted components. Consequently, the PARAFAC components can be interpreted directly on the basis of S_k, $k=1,...,p$, as in PCA. In case p is large, a possibly prohibitive number of structure elements must be considered. Fortunately, however, the components $AD_1,...,AD_p$ are perfectly congruent across the occasions, and, consequently, these components may be interpreted identically. Therefore, a pooled structure matrix is defined as $S \equiv \sum_{k=1}^{p} X_k' AD_k$. If the columns of X_k, $k=1,...,p$, are centered within the occasions, and scaled to unit length over the occasions, and A and C are scaled to unit length column–wise, and the PARAFAC algorithm has converged, then the elements of S are correlations between the variables over the occasions, collected in $\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$, and the perfectly congruent components, collected in $\begin{pmatrix} AD_1 \\ \vdots \\ AD_p \end{pmatrix}$, as can be shown as follows. From the fact that $\text{Diag}(A'A)=\text{Diag}(C'C)=I_q$ it follows that $\text{Diag}\left(\sum_{l=1}^{p} D_l A' AD_l\right)=I_q$. Hence, $\begin{pmatrix} AD_1 \\ \vdots \\ AD_p \end{pmatrix}$ has unit length column–wise. From the fact that A satisfies (1.4) and X_k, k=1,...,p, is centered column-wise, it follows that A is centered column-wise. Hence, $\begin{pmatrix} AD_1 \\ \vdots \\ AD_p \end{pmatrix}$ is centered column-wise. By writing $S=\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}' \begin{pmatrix} AD_1 \\ \vdots \\ AD_p \end{pmatrix}$ it can be seen that every element in S is the product of two vectors that are centered and have unit length, which completes the proof.

In case of orthogonal components, the structure matrix S and the pattern matrix B coincide. That is, if $A'A=I_q$ and $\text{Diag}(C'C)=I_q$, then from (1.5) it follows that

$$B=\sum_{k=1}^{p} X_k' AD_k \left(\sum_{l=1}^{p} D_l A' AD_l \right)^{-1} = S\left(\sum_{l=1}^{p} D_l^2 \right)^{-1} = S\left(\text{Diag}(C'C)\right)^{-1} = S. \tag{1.10}$$

This shows that, with the generalization of PCA to PFORTA, the property that structure and pattern are equal is retained.

Because the pattern matrix B has coefficients of the variables on the components one might want to use B as a basis for the interpretation of the components, as suggested by Harshman and DeSarbo (1984, p. 627). However, it can be argued that, analogously to the pattern matrix in PCA after an oblique rotation, the components can only be interpreted indirectly on the basis of B. This is because the elements of B are regression weights in the regression of the variables, collected in $\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$, on $\begin{pmatrix} AD_1 \\ \vdots \\ AD_p \end{pmatrix}$. So from the rows of B it can be seen how the variables are optimally reconstructed from the components. Therefore, if the interpretation of the components is known, and the interpretation of the variables is unknown, then the pattern matrix B may be used as a basis for the interpretation of the variables (cf. Brogden 1969; Gorsuch, 1983, p. 207). It can be concluded that the pattern matrix B has no direct relationship with the interpretation of the components. The author prefers S instead of B for the interpretation of the PARAFAC components. Because it is quite current to use B for the interpretation of the PARAFAC components, this matrix will be considered too. In the sequel, substantial differences between S and B will be reported, whenever they are encountered.

1.6 Degenerate PARAFAC solutions and how they can be avoided

Harshman and Lundy (1984b, p. 271) and Kruskal, Harshman and Lundy (1989) report that the PARAFAC algorithm sometimes yields a degenerate solution, which they describe in terms of seven criteria. Instead of repeating these criteria, a definition will be provided that essentially covers these criteria. Let

$$\cos ABC = \cos(\mathbf{a}_l, \mathbf{a}_{l'}) \cos(\mathbf{b}_l, \mathbf{b}_{l'}) \cos(\mathbf{c}_l, \mathbf{c}_{l'}), \tag{1.11}$$

where $\cos(\mathbf{a}_l, \mathbf{a}_{l'})$ denotes the cosine between columns l and l' of A, and $\cos(\mathbf{b}_l, \mathbf{b}_{l'})$ and $\cos(\mathbf{c}_l, \mathbf{c}_{l'})$ denote the cosines for the two corresponding columns in B and C, respectively. A PARAFAC solution is degenerate if, for certain l and l', the limit of $\cos ABC$, as the number of iterative steps tends to infinity, is -1. So in case of degeneracy we can make $\cos ABC$ as close to −1 as we please, by increasing the number of iterative steps. This does not imply that cos*ABC* can reach −1, which can be proven as follows. Suppose that cos*ABC*=−1, and that the rank of the matrix $(X_1|...|X_p)$ is greater than q−1. It can always be arranged that $\text{Diag}(A'A)=\text{Diag}(C'C)=I_q$, l=1, l'=2, $\cos(\mathbf{a}_1,\mathbf{a}_2)=1$, $\cos(\mathbf{b}_1,\mathbf{b}_2)=1$ and $\cos(\mathbf{c}_1,\mathbf{c}_2)=-1$. It follows that $\mathbf{a}_1=\mathbf{a}_2$, $\mathbf{b}_1=\lambda\mathbf{b}_2$ for some scalar $\lambda>0$, and $\mathbf{c}_1=-\mathbf{c}_2$. Thus, for $k=1,...,p$, we have

$$AD_kB'=\sum_{r=1}^{q}\mathbf{a}_r c_{kr}\mathbf{b}_r'=\mathbf{a}_1c_{k1}\mathbf{b}_1'+\mathbf{a}_2c_{k2}\mathbf{b}_2'+\sum_{r=3}^{q}\mathbf{a}_r c_{kr}\mathbf{b}_r'=\mathbf{a}_2c_{k2}\mathbf{b}_2'-\mathbf{a}_2\lambda c_{k2}\mathbf{b}_2'+\sum_{r=3}^{q}\mathbf{a}_r c_{kr}\mathbf{b}_r'$$

$$=\mathbf{a}_2(c_{k2}-\lambda c_{k2})\mathbf{b}_2'+\sum_{r=3}^{q}\mathbf{a}_r c_{kr}\mathbf{b}_r'. \quad (1.12)$$

This shows that the PARAFAC representation of X_k has at most rank q−1, k=1,...,p. From this and the supposition that the rank of $(X_1|...|X_p)$ is greater than q−1 it follows that for these parameter matrices PARAFAC(A,B,C)>0. This is because PARAFAC(A,B,C)=0 is incompatible with having the rank of $(X_1|...|X_p)$ greater than q−1, as will be shown now. If PARAFAC(A,B,C)=0 then $X_k=AD_kB'$, k=1,...,p, and hence the rank of $(X_1|...|X_p)$ would equal the rank of $(AD_1B'|...|AD_pB')$. From (1.12) it follows immediately that the rank of $(AD_1B'|...|AD_pB')$ would be less than or equal to q−1, which is incompatible with having the rank of $(X_1|...|X_p)$ greater than q−1. Therefore, we have PARAFAC(A,B,C)>0.

From (1.12) and PARAFAC(A,B,C)>0 it follows that the PARAFAC loss function can be further decreased by minimizing

$$f(\mathbf{u},\mathbf{v},\mathbf{w})=\sum_{k=1}^{p}\left\| \left(X_k-\mathbf{a}_2(c_{k2}-\lambda c_{k2})\mathbf{b}_2'-\sum_{r=3}^{q}\mathbf{a}_r c_{kr}\mathbf{b}_r'\right)-\mathbf{u}w_{k1}\mathbf{v}'\right\|^2, \tag{1.13}$$

where $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are of the same order as $\mathbf{a}_1$, $\mathbf{b}_1$ and $\mathbf{c}_1$, respectively, and f is a PARAFAC loss function with dimensionality 1. By taking, as the first columns in A, B and C, the minimizing $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ of f, respectively, at least a lower value of the PARAFAC loss function can be found (for the same dimensionality) than the value of the PARAFAC loss function with cosABC=–1. Therefore, the value of the PARAFAC loss function is not globally minimal if cosABC=–1, which proves that cosABC can never reach –1.

The fact that cosABC cannot reach –1 implies that, in case of degeneracy, PARAFAC(A,B,C) has no minimum, and hence the iterative process does not converge to stable parameter matrices. The above definition of degenerate components is in accordance with results presented by Ten Berge, Kiers and De Leeuw (1988). They proved that PARAFAC(A,B,C) has no minimum for a certain contrived 2×2×2 array that yields degenerate PARAFAC components.

It can easily be seen that, in case of degeneracy, it is impossible to interpret the components consistently. That is, if cosABC tends to –1, then it can be arranged that cos($\mathbf{a}_l$,$\mathbf{a}_{l'}$) tends to 1, cos($\mathbf{b}_l$,$\mathbf{b}_{l'}$) tends to 1 and cos($\mathbf{c}_l$,$\mathbf{c}_{l'}$) tends to –1. So it may be assumed that cos($\mathbf{a}_l$,$\mathbf{a}_{l'}$) and cos($\mathbf{b}_l$,$\mathbf{b}_{l'}$) are close to 1, and that cos($\mathbf{c}_l$,$\mathbf{c}_{l'}$) is close to –1. But then on the basis of A and B component l and l' should have the same interpretations, whereas from C these components should have opposite interpretations.

Of course, in practice the question arises, for what value of cosABC can it be said that the PARAFAC solution is degenerate? In the sequel it will be said that the PARAFAC solution is degenerate, if it is found that cosABC<–.85 and one additional step of the iterative process further decreases cosABC. Of course, this is an arbitrary choice, but for practical purposes it can be justified as follows. If cosABC < –.85, then $|\cos(\mathbf{a}_l,\mathbf{a}_{l'})| > .85$, $|\cos(\mathbf{b}_l,\mathbf{b}_{l'})| > .85$ and $|\cos(\mathbf{c}_l,\mathbf{c}_{l'})| > .85$, where $|.|$ denotes that the absolute value of (.) is taken. Clearly, the cosine of the angle between two vectors equals Tucker's (1951) congruence coefficient for the same two vectors. Haven and Ten Berge (1977) have found that components are judged as being 'virtually equal' whenever Tucker's (1951) congruence coefficient for the loadings of two components is above .85. So, if cosABC < –.85, then from two parameter matrices one would conclude that the components concerned are virtually equal, whereas from the third parameter matrix one would conclude the opposite.

Table 1.1 *The value of cosABC after various numbers of iterative steps from PARAFAC analysis of a* 38x3x2 *data-array.*

Number of iterative steps	10	100	1000	10000
cosABC	–.72	–.78	–.93	–.99
Value of PARAFAC(A,B,C)	0.690	0.676	0.666	0.666

To illustrate a PARAFAC analysis where a degenerate solution appears, the results of a two-dimensional PARAFAC analysis of a three-way array* with

*The author is obliged to Fred Wolters who kindly made the data available.

scores of 38 persons on three variables measured at two occasions are reported, see Table 1.1. Prior to the analysis the data were centered within the occasions and scaled to unit length over the occasions.

From Table 1.1 it can be seen that cos*ABC* decreases, as the number of iterative steps increases. This suggests that cos*ABC* tends to −1 as the number of iterative steps tends to infinity and hence that this PARAFAC solution is degenerate.

To avoid a degenerate solution, Harshman and Lundy (1984b, p. 274) suggest constraining one parameter matrix, for example A, to be column-wise orthonormal. Clearly, PFORTA cannot yield a degenerate solution because having $A'A=I_q$ implies that the value of cos*ABC* is fixed to zero for all pairs of columns of the parameter matrices. It is clear that a degenerate solution can also be avoided by subjecting B or C to the column-wise orthonormality constraint. In the next chapters, it will be demonstrated that imposing other constraints than orthonormality constraints also has the effect of avoiding degenerate components.

1.7 Uniqueness of the PARAFAC components

The most salient property of the PARAFAC model is its uniqueness. Jennrich (see Harshman, 1970, pp. 61–62) has first established certain sufficient conditions for uniqueness. Harshman (1972) has relaxed these conditions. He proved that, under certain conditions, the PARAFAC representations of the frontal slices are unique up to an arbitrary simultaneous permutation of the columns in the parameter matrices and an arbitrary scaling of the columns in two of the parameter matrices. In particular, let the matrices A, B and C, all of the same order as before, be given such that

$$\hat{X}_k = AD_kB', \tag{1.14}$$

$k=1,\ldots,p$, where A and B have full column rank, D_j is non–singular, and for some matrix D_i all diagonal elements of $D_iD_j^{-1}$ are distinct. This condition will be called Harshman's condition. Let $\overset{*}{A}$, $\overset{*}{B}$ and $\overset{*}{C}$ also satisfy (1.14), where $\overset{*}{A}$, $\overset{*}{B}$ and $\overset{*}{C}$ are of the same order as A, B and C. If Harshman's condition is satisfied, then $A=\overset{*}{A}\Pi\Lambda_a$, $B=\overset{*}{B}\Pi\Lambda_b$ and $C=\overset{*}{C}\Pi\Lambda_c$ for certain diagonal matrices Λ_a, Λ_b and Λ_c such that $\Lambda_a\Lambda_b\Lambda_c=I_q$, and a certain permutation matrix Π. This equivalence result pinpoints uniqueness in PARAFAC. In case $q=1$ it can readily be verified that Harshman's (sufficient) condition for uniqueness is satisfied for all non–zero A, B and C, hence there is uniqueness in this case. For the other cases, with $q>1$, Kruskal (1977), see also Kruskal (1989), has proven more relaxed conditions for uniqueness based on the concept of k–rank. Specifically, an $n\times q$ matrix A is said to have a k–rank k_a if all sets of k_a columns in A are linearly independent and there is at least one set of k_a+1 columns in A that is linearly dependent. Let k_b and k_c denote the k–rank of B and C, respectively. Theorem 4a of Kruskal (1977) states that the condition $k_a+k_b+k_c \geq 2q+2$ is sufficient for uniqueness.

By noting that, for $k_a=k_b=q$ and $k_c>1$, satisfaction of Harshman's condition implies satisfaction of Kruskal's condition, it is clear that Kruskal's condition contains Harshman's condition as a special case if k_c>1. To clarify the above sufficient conditions for uniqueness, it is instructive to consider a few cases. From Harshman's condition it already follows that it is not necessary for uniqueness that all the parameter matrices have full column rank. For example, consider the case q=3, with A and B of rank 3, and $C=\begin{pmatrix}1&1&2\\1&2&3\\1&3&4\end{pmatrix}$ of rank 2. Now, $D_3D_1^{-1}=\begin{pmatrix}1&0&0\\0&3&0\\0&0&2\end{pmatrix}$ has distinct diagonal elements. Hence, these parameter matrices satisfy Harshman's sufficient condition for uniqueness.

From Kruskal's condition it can be seen that there can be uniqueness even if *none* of the parameters matrices has full rank. For example, in case q=5 with all parameter matrices having k–rank = 4 = q–1, it follows that $k_a+k_b+k_c=3(q-1)=2q+2$ and hence Kruskal's condition is satisfied. This exemplifies uniqueness in a case where none of the three parameter matrices has full rank. Clearly, in this case Harshman's condition is not satisfied whereas Kruskal's condition is satisfied. This illustrates that Kruskal's condition is more relaxed than Harshman's condition if q>1. For practical purposes it is convenient to realize that, in case the parameter matrices have full rank, there is uniqueness because $k_a=k_b=k_c=q$, and hence $k_a+k_b+k_c=3q\geq 2q+2$ if q>1.

Having clarified the above sufficient conditions for uniqueness the question arises: What are necessary conditions for uniqueness? In section 2.1 it will be shown that having k–rank>1 for all parameter matrices is necessary for uniqueness if q>1.

1.8 Problems with PARAFAC: The meaning of uniqueness

A first problem with PARAFAC to be discussed in the present study is related to uniqueness, as follows. Let A, B and C minimize PARAFAC(A,B,C) and let $\hat{X}_k=AD_kB'$ denote the representation of frontal slice k of X, k=1,...,p. All available sufficient conditions for uniqueness critically depend on the assumption of fixed representations $\hat{X}_k$, k=1,...,p, of the frontal slices. This means that uniqueness only pertains to uniqueness *given* the representations. So the uniqueness considered above does *not exclude* the existence of parameter matrices that minimize PARAFAC(A,B,C) but yield different representations. This will be illustrated by a first example.

Suppose that $X_1=P\Lambda_1Q'$ and $X_2=P\Lambda_2Q'$ $(P'P=Q'Q=QQ'=I_3)$, with $\Lambda_1=\begin{pmatrix}3&0&0\\0&2&0\\0&0&2\end{pmatrix}$ and $\Lambda_2=\begin{pmatrix}2&0&0\\0&1&0\\0&0&1\end{pmatrix}$. Note that $X_1=P\Lambda_1Q'$ and $X_2=P\Lambda_2Q'$ are the SVD's of X_1 and X_2, respectively, and that the second and third singular values are equal, both for X_1 and X_2. Now it can be demonstrated that there are at least two different solutions that minimize PARAFAC(A,B,C) in two dimensions. Let $\mathbf{p}_r$ and $\mathbf{q}_r$ denote column r of P and Q, respectively, r=1,2,3. We have a first solution $A=(\mathbf{p}_1,\mathbf{p}_2)$, $B=(\mathbf{q}_1,\mathbf{q}_2)$ and $C=\begin{pmatrix}3&2\\2&1\end{pmatrix}$, and a second solution $A=(\mathbf{p}_1,\mathbf{p}_3)$, $B=(\mathbf{q}_1,\mathbf{q}_3)$ and $C=\begin{pmatrix}3&2\\2&1\end{pmatrix}$. Both solutions approximate X_1 and X_2 as close as two separate PCA analyses, and hence optimally approximate the frontal slabs in two dimensions (Eckart & Young, 1936). Therefore, in both cases the PARAFAC loss function is globally minimal. In both solutions the parameter matrices have full column rank, so both solutions are unique. Furthermore, they cannot be transformed into each other by scalings and permutations. This demonstrates that uniqueness given the representations does *not exclude* the existence of alternative parameter matrices that minimize PARAFAC(A,B,C) but yield different representations. This case of non-uniqueness is very unlikely to occur in practice. What is more likely to happen in practice is that the uniqueness occurs up to some small constant. This will be illustrated by a second example.

Suppose that a PARAFAC analysis yields an A and a B with full column rank and $C=\begin{pmatrix}1 & 1\\ 1 & 1+\varepsilon\end{pmatrix}$, with ε close to zero. Then Harshman's conditions are satisfied, so there is uniqueness. In such a case it can be seen that there are PARAFAC parameter matrices that almost minimize PARAFAC(A,B,C), as follows. Clearly, $D_1=I_2$ and D_2 is almost equal to I_2. Therefore, by taking AB' to represent both X_1 and X_2, the data are represented almost as well as when X_1 and X_2 are represented by AD_1B' and AD_2B', respectively. For the representations AB' we have $AB'=AT(T)^{-1}B'$, which implies rotational freedom. In this case the uniqueness may be considered to be *weak*, as there are infinitely many representations that fit the data almost as well as PARAFAC does. On the other hand, the uniqueness may be called *strong* in case there are no alternative components that fit the data almost as well as PARAFAC does. Here a first problem that arises is, how it can be determined, to what extent, for any particular data set, the PARAFAC components are unique.

In the above, the concept of uniqueness is interpreted as a gradual concept in the sense that uniqueness exists to a certain extent. Clearly, this practical interpretation differs from the mathematical interpretation of uniqueness which is a matter of all or nothing: A certain PARAFAC representation is either unique or it is not.

In case of weak uniqueness a second problem arises. To see this, PCA will be reconsidered. In past decades, various rotations have been developed. For instance, rotation according to the VARIMAX criterion (Kaiser, 1958) is commonly used after a PCA, to find components that have structure elements (also called loadings) that have simple structure in the sense that the loadings are either close to zero or close to –1 or 1. In this way components are found which allow for an easier interpretation and at the same time provide a more parsimonious representation of the data. However, PARAFAC uniqueness implies that, up to arbitrary scalings and permutations, there are no rotations (or transformations) such that the residual matrices remain the same. In case uniqueness is weak, this absence of rotational freedom may detract from the usefulness of PARAFAC for the purpose of exploratory or confirmatory analysis of three-way data. Specifically, in such a case the question arises: Can other components be found that allow an easier interpretation and represent the data almost as well as the PARAFAC components? This question will be answered by using certain constrained versions of PARAFAC. In fact, the present study will be largely devoted to these constrained versions of PARAFAC.

1.9 The purpose and the organization of the present study

The present study has two main purposes. For many data sets it seems rather difficult or even impossible to know in advance whether or not the

PARAFAC model will be appropriate in the sense that the importance of the components will indeed be accentuated or reduced from occasion to occasion. In other words, it seems difficult for many data sets to know in advance whether or not the PARAFAC components will be strongly unique. The first purpose is to show how, for any particular data set, it can be determined to what extent the PARAFAC components are unique. The second purpose is to show how, in case of weak uniqueness, the PARAFAC method can be modified to the effect that a simpler representation of the three-way data can be found in the sense that its components allow an easier interpretation. Both purposes will be attained by subjecting the PARAFAC parameter matrices to certain constraints.

In general, it seems desirable that (constrained) PARAFAC components are stable, for instance, in the splithalf sense. Specifically, in case of weak uniqueness it can be expected that the PARAFAC components are not very stable. Likewise, the stability of constrained components is of interest. In all exemplary analyses, the splithalf stability of the PARAFAC components will be compared with that of the constrained PARAFAC components.

In case of weak uniqueness, it seems indicated to rotate the components (columns of A) to determine components that are easier to interpret. In the present study, rotation of the PARAFAC components will not be considered, for the following reason. Rotating the columns of A means that, in the search for components with simpler interpretations, we limit ourselves to the column space of the representations of the frontal slices. Fortunately, we can obviate the limitation by subjecting the

PARAFAC parameter matrices to certain constraints. Specifically, by using constraints it is possible to seek for easier interpretable components, not only within the column space of A, but over all possible solutions that satisfy the constraints under consideration.

The organization of the present study is as follows. The second chapter deals with the question how to determine, for any particular data set, to what extent the PARAFAC components are unique. It is proposed to answer this question by studying the discrepancy between the residual sum of squares of PARAFAC and the residual sum of squares of a constrained PARAFAC method that has two proportional columns in one of the parameter matrices. In case this discrepancy is small, it is concluded that the uniqueness is weak. This way of examining the uniqueness of the PARAFAC components will be illustrated by the analysis of several empirical data sets and it will be compared with splithalf analysis which can be seen as a different way of examining the uniqueness.

The third chapter deals with a question that is suggested by some of the results of Chapter 2: Can the data be represented satisfactorily by components with the same relative importances from occasion to occasion? In order to answer this question, a constrained PARAFAC method will be introduced, called Weighted PCA, that has proportional columns in the occasions matrix C. An efficient algorithm for Weighted PCA will be developed. This algorithm will be employed to illustrate that certain data sets can indeed be represented satisfactorily by components with the same relative importances from occasion to occasion. It will be shown that Weighted PCA has a number of properties, like, for instance, rotational freedom. It will be illustrated how this rotational freedom can be used in order to find components that allow for an easier interpretation than the PARAFAC components.

The fourth chapter starts with a demonstration of the role of rotational freedom when so-called positive manifold data (the term positive manifold was coined by Thurstone, 1947, p. 216) are analyzed by PCA. It is illustrated that VARIMAX rotation tends to yield components that have no elements in the structure matrix that contrast in sign. It is explained that such components without contrast allow for an easier interpretation than components with contrast. For three-way positive manifold data it is illustrated that PARAFAC may yield components that do have contrasting structure and/or pattern elements, and hence these components are more difficult to interpret than components without contrast. This raises the question whether or not there are constrained PARAFAC components without contrast that represent the data (almost) as well as the unconstrained PARAFAC components? To answer this question a constrained PARAFAC method is proposed that finds optimal PARAFAC components without contrast.

The fifth chapter deals with questions concerning components that correspond to non-overlapping clusters of variables. It is illustrated that PARAFAC components need not correspond to such clusters of variables. In practice, confirmatory and exploratory research questions arise about components that correspond to non-overlapping clusters of variables. That is, if a researcher has a hypothesis about a partitioning of the variables into non-overlapping clusters, then the question arises whether or not this hypothesis is sustained by the results of a PARAFAC analysis (confirmatory research question). On the other hand, an exploratory question arises if a researcher merely wants to see whether or not the variables can be clustered at all. Because PARAFAC components need not correspond to non-overlapping clusters of variables, PARAFAC is not an appropriate method for answering either of these questions. In order to answer these questions, two constrained PARAFAC methods are proposed, one for each of the research questions.

In the last chapter an overview is presented of a number of constrained PARAFAC methods and some general conclusions are drawn.

CHAPTER 2

CONFIRMATORY EVIDENCE FOR WEAK UNIQUENESS

In chapter 1 the sufficient condition $k_a+k_b+k_c \geq 2q+2$ for uniqueness (Kruskal, 1977) was explained. In practice, this sufficient condition is always fulfilled, because the probability that the PARAFAC algorithm yields (exactly) proportional columns is zero for empirical data. So in practice there is always uniqueness. In section 1.8 an example was given where the uniqueness of the PARAFAC components is weak. This diagnosis was based on the existence of alternative parameter matrices that do have rotational freedom and almost minimize the PARAFAC loss function. Now the question arises: How can it be determined to what extent, for any particular data set, the PARAFAC components are unique? In order to answer this question, we start with considering a necessary condition for uniqueness.

Suppose that a parameter matrix has k–rank 1, which means that there are at least two proportional columns in the matrix. Then there is in fact no uniqueness, hence having k–rank greater than 1 for the three parameter matrices is necessary for uniqueness. A proof follows in section 2.1. Accordingly, to verify the uniqueness of the PARAFAC solution in practice, one may reason as follows. If PARAFAC, constrained to have two proportional columns in A, B, or C, represents the data (almost) as well as unconstrained PARAFAC, then the PARAFAC components are *weakly* unique. On the other hand, the PARAFAC components are considered to be *strongly* unique if the discrepancy in fit between PARAFAC and PARAFAC with two proportional columns in one of the parameter matrices is large. In the present chapter such a constrained PARAFAC method will be described. First however, a necessary condition for uniqueness will be established.

2.1 A necessary condition for uniqueness

In chapter 1 it was explained that, if two columns in, for instance, the matrix C are proportional, then $k_C=1$ and therefore the above sufficient condition for uniqueness (Kruskal, 1977) is not satisfied. In fact, the condition that C does not have proportional columns is necessary for uniqueness, as will be proven now. No generality is lost by assuming that the first two columns in C are proportional. Let $A=(A_-|A_+)$, $B=(B_-|B_+)$, and $C=(C_-|C_+)$, where A_- is an $n\times 2$ matrix, A_+ is an $n\times(q-2)$ matrix, B_- is an $m\times 2$ matrix, B_+ is an $m\times(q-2)$ matrix, C_- is a $p\times 2$ matrix with two proportional columns, and C_+ is a $p\times(q-2)$ matrix, then $AD_kB'=A_-D_{-k}B_-'+A_+D_{+k}B_+'$. From the fact that the first two columns in C are proportional it follows that $C_-=\mathbf{cd}'$ for a p-vector $\mathbf{c}$ and a 2-vector $\mathbf{d}$. Therefore, $D_{-k}=\text{Diag}(c_k\mathbf{1d}')=c_kD_-$, where c_k is a scalar and $D_-=\text{Diag}(\mathbf{1d}')$. Hence, $A_-D_{-k}B_-'=c_kA_-D_-B_-'=\overset{*}{A}\overset{*}{D}_k\overset{*}{B}'$, $k=1,\ldots,p$, where $\overset{*}{A}=A_-T$, $\overset{*}{B}=B_-D_-(T')^{-1}$, $\overset{*}{D}_k=c_kI$, for any non-singular matrix T. This shows that the first two columns in A and B are determined up to a non-singular transformation, hence the first two components are not unique.

Note that the non-uniqueness only occurs in the first two columns of A and B. Therefore, in case $q>2$, having proportional columns in C implies rotational freedom only for the components that correspond to the proportional columns in C and can therefore be called partial rotational freedom.

From the symmetry property and the above result, it follows that the condition that two columns in A or B are proportional is also sufficient for partial rotational freedom. For this reason, the PARAFAC components may be considered to be weakly unique if the discrepancy in fit between PARAFAC and PARAFAC constrained to have two proportional columns in A, B, or C is small.

2.2 An Algorithm for PARAFAC with two proportional columns in one of the parameter matrices

In order to examine the extent to which the PARAFAC components are uniquely determined it has been proposed to compare the discrepancy in fit between PARAFAC and PARAFAC with two proportional columns in, for instance, C. Therefore, the problem is to minimize the loss function

$$f(A,B,C)=\sum_{k=1}^{p}\|X_k-AD_kB'\|^2, \tag{2.1}$$

subject to the constraint that the first two columns in C are proportional.

In order to minimize f(A,B,C) an ALS algorithm will be employed, in which first A is updated while B and C are fixed, next B is updated while A and C are fixed, and finally C is updated while A and B are fixed. From (2.1) it can be seen that to update A and to update B the PARAFAC algorithm can be used. This leaves us with updating C for fixed A and B, subject to the constraint that the first two columns in C are proportional. Imposing the identification constraint $\text{Diag}(C'C)=I_q$ implies that the first two columns of C are constrained to be equal. Specifically, using the fact that $\text{Vec}(\mathbf{ab}')=\mathbf{a}\otimes\mathbf{b}$ where Vec(.) denotes the vector containing all the elements of the matrix strung out rowwise into a column vector, and $\otimes$ denotes the (right) Kronecker product, we have

$$\text{h}(C)=\sum_{k=1}^{p}\|X_k-AD_kB'\|^2=\sum_{k=1}^{p}\|X_k-\sum_{r=1}^{q}\mathbf{a}_r\mathbf{b}_r'c_{kr}\|^2=\sum_{k=1}^{p}\|\text{Vec}(X_k)-\sum_{r=1}^{q}\mathbf{a}_r\otimes\mathbf{b}_rc_{kr}\|^2$$

$$=\sum_{k=1}^{p}\|\text{Vec}(X_k)-(\mathbf{a}_1\otimes\mathbf{b}_1|....|\mathbf{a}_q\otimes\mathbf{b}_q)\mathbf{c}_k\|^2, \quad (2.2)$$

where $\mathbf{c}_k'$ is row k in C. Let $\mathbf{d}_k$ denote the q–1 vector with $\mathbf{d}_k'=(c_{k1},c_{k3},...,c_{kq})$. We may write

$$(\mathbf{a}_1\otimes\mathbf{b}_1|...|\mathbf{a}_q\otimes\mathbf{b}_q)\mathbf{c}_k=(\mathbf{a}_1\otimes\mathbf{b}_1+\mathbf{a}_2\otimes\mathbf{b}_2|\mathbf{a}_3\otimes\mathbf{b}_3|...|\mathbf{a}_q\otimes\mathbf{b}_q)\mathbf{d}_k=U\mathbf{d}_k, \quad (2.3)$$

for k=1,...,p, where $U\equiv(\mathbf{a}_1\otimes\mathbf{b}_1+\mathbf{a}_2\otimes\mathbf{b}_2|\mathbf{a}_3\otimes\mathbf{b}_3|...|\mathbf{a}_q\otimes\mathbf{b}_q)$. From (2.2) and (2.3) it follows that the problem is to minimize

$$h(\mathbf{d}_1,...,\mathbf{d}_p)=\sum_{k=1}^{p}\|\text{Vec}(X_k)-U\mathbf{d}_k\|^2. \tag{2.4}$$

From (2.4) it is clear that the minimizing $\mathbf{d}_k$ can be obtained as

$$\mathbf{d}_k=(U'U)^{-1}U'\text{Vec}(X_k) \tag{2.5}$$

for k=1,...,p. From (2.5) the solution for $\mathbf{c}_k'$ can be obtained as $\mathbf{c}_k'=(d_{k1},d_{k1},d_{k3},...,d_{kq})$, and D_k=Diag($\mathbf{1c}_k'$), for k=1,...,p. With the above procedures for updating the PARAFAC parameter matrices an ALS algorithm can be constructed which, after an arbitrary start, monotonically decreases f(A,B,C) until the function value stabilizes. A number of test trials revealed that the above algorithm is not sensitive to local minima. Because the danger of local minima cannot be ruled out completely, it is suggested to run more than one analysis, with different starting configurations for both A and B.

From the symmetry property it follows that analyzing the horizontal slices, or the lateral slices, by the above algorithm corresponds to PARAFAC subject to the constraint that the first two columns in the matrix A or B are proportional, respectively. So no additional algorithms need be developed to examine the strength of uniqueness of the PARAFAC components in other directions.

2.3 Examining uniqueness for empirical data

To illustrate how PARAFAC with proportional columns can be used to examine the uniqueness of the PARAFAC components, the results from analyzing three empirical data sets will be reported here. Prior to the analyses the data were centered and scaled to unit length over the occasions. The data were analyzed by PARAFAC and separately by PARAFAC with two proportional columns in A, B, or C, respectively, yielding four fit values per dimensionality. In three cases a degenerate PARAFAC solution was found. In these cases the data were analyzed by PFORTA in order to arrive at a non-degenerate solution. In these cases the uniqueness of the PFORTA components rather than the uniqueness of the PARAFAC components will be examined. In case of dimensionality 2, a non-degenerate solution is guaranteed for PARAFAC with two proportional columns in one of the parameter matrices, due to the freedom to orthonormalize one of the two other parameter matrices column-wise, see section 2.1. Unfortunately, the solutions for dimensionality 3, having one parameter matrix with two proportional columns, can be degenerate. Two such solutions were encountered. These two solutions will not be used for the examination of the degree of uniqueness of the PARAFAC or the PFORTA components. In Table 2.1 the percentages of the variance accounted for are reported for dimensionality 2 and 3.

Table 2.1 *The percentages of variance explained by PARAFAC with proportional columns in A, in B and in C, by unconstrained PARAFAC, and by PFORTA for three data sets.*

q=2	PARAFAC with prop. col. in			PARAFAC	PFORTA
Data set	*A*	*B*	*C*		
Affective Response	39.0	38.4	42.9	46.4	44.3
Tongue Shape	64.4	58.1	71.4	74.8	–
Triple Personality	41.0	57.4	42.8	60.9	–
q=3	PARAFAC with prop. col. in			PARAFAC	PFORTA
Data set	*A*	*B*	*C*		
Affective Response	48.1	53.8	50.5	54.0	51.7
Tongue Shape	82.2	79.2	84.8	86.6	–
Triple Personality	62.6	65.5	66.3	67.3	66.7

As a first data set the 32x6x8 Affective Response* data (see section 3.6) were analyzed. It was found that the two-dimensional PARAFAC solution is borderline degenerate (cosABC=–.78 after 2000 iterations, see section 3.6 for a more detailed description of these results), and that the three-dimensional solution is degenerate (cosABC=–.89 after 2000 iterations). For dimensionality 2, the discrepancy between PFORTA and PARAFAC with two proportional columns in C is 1.4% of variance explained, which indicates that the PFORTA components may be considered as weakly

*The author is obliged to Gudrun Eckblad who kindly made the data available.

unique. For dimensionality 3 of the PARAFAC solution having proportional columns in B, a value of –.93 of cosABC was encountered, hence this solution is degenerate. The discrepancy between PFORTA and PARAFAC with two proportional columns in C is 1.2%, which indicates that the PFORTA components for dimensionality 3 are also weakly unique.

The Tongue Shape data consist of heights of 13 tongue positions that were measured during the pronunciation of 10 vowels by five speakers, see Harshman, Ladefoged and Goldstein (1977) for a more detailed description of these data. The data were previously analyzed in two dimensions by Harshman, Ladefoged and Goldstein (1977), and Kruskal (1984), who called it a 'successful application'. For dimensionality 2, the discrepancy in fit between PARAFAC and PARAFAC with two proportional columns in C is 3.4% of variance explained, which indicates at least some degree of uniqueness. For dimensionality 3, the discrepancy in fit between PARAFAC and PARAFAC with two proportional columns in C is 1.8% of variance explained, which may be considered as some evidence for weak uniqueness of these PARAFAC components.

The Triple Personality data consist of scores from one person for 15 concepts on 10 variables (semantic differential scales) measured at six administrations, see Osgood and Luria (1954) or Kroonenberg (1983, pp. 227–242). For dimensionality 2, the discrepancy between PARAFAC and PARAFAC with two proportional columns in B is 3.5% of variance explained, which indicates at least some degree of uniqueness. For dimensionality 3, PARAFAC yielded a value of –.94 for cosABC, and PARAFAC with two proportional columns in B yielded a value of –1.00 for cosABC hence these

PARAFAC solutions are degenerate. For dimensionality 3 the discrepancy between PFORTA and PARAFAC with two proportional columns in C is 0.4% of variance explained, hence these PFORTA components are weakly unique.

2.4 Examining uniqueness by means of splithalf analysis

Harshman and Lundy (1984a, pp. 164–167) proposed to examine the uniqueness of the PARAFAC components by means of splithalf analysis. They reason that, if the data are split into two different halves and two separate PARAFAC analyses of these splithalves yield two sets of identical components, then the data contains enough systematic variation to determine components uniquely. To get an impression on how this approach works in practice, the results will be presented of a splithalf analysis of the so-called DAT* data. The DAT data consist of scores of 133 persons on nine variables that measure intelligence, at two occasions. The data were split into two sets by random assignment of the persons. The sets were separately centered within the occasions and scaled to unit length over the occasions. Each set was analyzed by an independent PARAFAC analysis with dimensionality 2. Next, Tucker's (1951) congruence coefficients for the corresponding columns of B and C across the two sets were computed as a measure for the proportionality of the components

*The author is obliged to Theo Nijsse who kindly made the data available.

across the two sets. In addition, for columns having unit length, say $\mathbf{b}_1$ and $\mathbf{b}_2$, the congruence coefficient equals $\mathbf{b}_1'\mathbf{b}_2$ and, because $\|\mathbf{b}_1 - \mathbf{b}_2\|^2 = 2(1 - \mathbf{b}_1'\mathbf{b}_2)$, it can also be seen as a measure of equality. Without loss of generality, the columns of B and C can be scaled to unit length. Hence, in case the congruence value, for the corresponding columns from PARAFAC analysis of each of the two splits, is close to 1, these columns are nearly equal. In case the congruence values encountered are greater than or equal to .85 (see Haven & Ten Berge, 1977), it will be said that the splithalf analysis indicates stable components. For the DAT data, five separate splithalf analyses were conducted. Two of the five splithalf analyses indicated stable components. The three lowest congruence values encountered for the unstable components were .31, .71, and .63. These results suggest that the PARAFAC components are not stable in the splithalf sense. In addition, it seems that it is hazardous to rely on only one splithalf analysis.

Now the question arises whether or not unstable components occur together with weak uniqueness. To shed a first light on this question, the same DAT data were analyzed by PARAFAC and by PARAFAC with two proportional columns in C. The data were preprocessed as before. The percentages of variance explained by PARAFAC and by PARAFAC with two proportional columns in C were 48.3 and 48.2, respectively. This discrepancy in fit can be regarded as negligible, and hence these PARAFAC components are weakly unique. This illustrates a case where the PARAFAC components are not stable in the splithalf sense and are weakly unique. For these data other components may exist, which, for instance, allow for an easier interpretation, and are stable in the splithalf sense.

Above, an example was presented where the PARAFAC components are unstable and the uniqueness is weak. To investigate the relation between stability and uniqueness more systematically, a simulation study was conducted.

2.5 A simulation study on PARAFAC uniqueness

It can be expected that, in case the PARAFAC components are stable in the splithalf sense, a characteristic of the population is responsible for this (see also Harshman & Lundy, 1984a, pp. 164–169). For instance, one might expect that stable sample components indicate unique PARAFAC components in the population. The question arises whether or not splithalf stability is sufficient to reveal such population characteristics. In order to answer this question a small simulation study was done. The data of this simulation study will also be used in chapter 4.

Three covariance matrices, Σ_1, Σ_2, and Σ_3, were constructed according to $\Sigma_i=\begin{pmatrix} BD_{1i} \\ BD_{2i} \\ BD_{3i} \end{pmatrix}(D_{1i}B' \,|\, D_{2i}B' \,|\, D_{3i}B')$, using one matrix $B=\begin{pmatrix} .85 & .15 \\ .75 & .25 \\ .15 & .85 \\ .25 & .75 \end{pmatrix}$, and three matrices $C_1=\begin{pmatrix} 1.50 & 0.50 \\ 1.00 & 1.00 \\ 0.50 & 1.50 \end{pmatrix}$, $C_2=\begin{pmatrix} 1.25 & 0.75 \\ 1.00 & 1.00 \\ 0.75 & 1.25 \end{pmatrix}$, and $C_3=\begin{pmatrix} 1.50 & 1.50 \\ 1.00 & 1.00 \\ 0.50 & 0.50 \end{pmatrix}$, where D_{ki} is row k of C_i, i=1,..,3. Such covariance matrices can be seen as the covariance matrices for data matrices of the form $(X_1 | X_2 | X_3)$, where the PARAFAC model holds for X_i, i=1,..,3, having orthonormal columns in A. By holding the matrix B constant and manipulating the matrices C_1, C_2, and

C_3, the PARAFAC components from Σ_1, Σ_2, and Σ_3, respectively, have different degrees of uniqueness. Accordingly, the covariance matrix Σ_1 is said to have 'strong' uniqueness, the covariance matrix Σ_2 is said to have 'medium' uniqueness and the covariance matrix Σ_3 is without uniqueness. These covariance matrices served as input for the SIMLIS program (Boomsma, 1983), which is a program for pseudo random sampling from a multivariate normal distribution. Samples of size n=40 and n=20 were drawn from these distributions. For each combination of degree of uniqueness and sample size, three samples were drawn, yielding 18 arrays of order n×4×3 in total.

Each of the 18 arrays was analyzed by the splithalf method followed by PARAFAC, using five splits. For the six samples drawn from the strongly unique population and the six samples drawn from the medium unique population all congruence values encountered were greater than .97. For the six samples drawn from the population without uniqueness one stable splithalf solution (out of 15 splithalf analyses) was encountered for n=20 and six stable splithalf solutions (out of 15) were encountered for n=40. The lowest congruence value encountered was .14 for n=20 and .25 for n=40. These results suggest that the PARAFAC components for samples drawn from populations having uniqueness are stable in the splithalf sense, even when the sample size is as low as 20. In addition, the PARAFAC components for samples from populations without uniqueness are not stable in the splithalf sense.

Next, to examine the degrees of uniqueness in the samples, the 18 arrays were analyzed by PARAFAC and by PARAFAC with two proportional columns in

C, with dimensionality 2. The discrepancies in percentage of fit between both methods for n=40 and n=20 were highly similar. For this reason, only the discrepancies for n=40 will be reported. The discrepancies encountered were, for strong uniqueness, 12.0, 13.6, and 15.9, for medium uniqueness, 3.6, 5.8, and 4.5 and, for non-uniqueness, 0.1, 0.1, and 0.1. It can be concluded that these discrepancies in the samples quite accurately reflect the extent of uniqueness in the population. It seems that both the proposed discrepancy in fit and an analysis of stability by splithalf can be used to make inferences about the degree of uniqueness of the components in the population.

2.6 Examining uniqueness in case the parameter matrices have disproportional columns

In case the PARAFAC solution of dimensionality 2 contains a matrix C with nearly proportional columns, it might be expected that the discrepancy between PARAFAC and PARAFAC constrained to have two proportional columns in C is small. On the other hand, it might be expected that in case clearly non-proportional columns in the parameter matrices are found, the uniqueness is strong. However, this expectation need not come true. This will be illustrated with the results of a PARAFAC analysis of a 42×4×2 empirical data set*. The data were centered within the occasions and

*The author is obliged to Fred Wolters who kindly made the data available.

scaled to unit length across the occasions. Tucker's (1951) congruence coefficients between the columns of the parameter matrices were computed as a measure for the proportionality of the columns in the parameter matrices. The values of the congruence coefficients are –.06, .17, and .66, respectively, for the columns of A, B, and C. Hence, the columns of the parameter matrices are clearly non-proportional. These parameter matrices have full rank, hence this PARAFAC solution is unique. However, the amount of explained variance by PARAFAC and by PARAFAC with proportional columns in C is 55.5 and 55.4, respectively. This demonstrates that these PARAFAC components are weakly unique. It can be concluded that clear non-proportionality of the columns in the PARAFAC parameter matrices is not sufficient for strong uniqueness of the PARAFAC components.

2.7 Discussion

It has been illustrated that splithalf analysis and PARAFAC with two proportional columns in one of the parameter matrices is useful to detect weak uniqueness of the PARAFAC and of the PFORTA components. In addition to this, one might want to use PARAFAC with two proportional columns in one of the parameter matrices for the purpose of representing a three-way array. This will be taken up in Chapter 3.

In case a splithalf analysis reveals that the PARAFAC components are not unique Harshman and Lundy (1984a, pp. 160–161) recommend to remove such non-unique components by appropriate preprocessing. Unfortunately, some preprocessing methods are rather difficult to understand from a substantive point of view. Instead of introducing different preprocessing methods in this study, alternative procedures for dealing with weakly unique PARAFAC solutions will be introduced. Specifically, in case the PARAFAC components are weakly unique, other components than the PARAFAC components exist that may be preferred, because, for instance, they have simple structure and allow therefore an easier interpretation. In the next three chapters, various constraints to impose on the PARAFAC parameter matrices will be considered in order to find (constrained) PARAFAC components that allow for an easier interpretation than the unconstrained PARAFAC components.

CHAPTER 3

ANALYSIS OF THREE-WAY DATA BY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS

A constrained variant of PARAFAC has been considered by Van der Kloot and Kroonenberg (1985, p. 483) and by Ten Berge, De Leeuw and Kroonenberg (1987, pp. 189–190). Recently, this constrained variant of PARAFAC has been applied by Van IJzendoorn and Kroonenberg (1990). In this variant of PARAFAC, the data are supposed to be based on the same components, but now these are to have the same relative importances from occasion to occasion; only the joint importance of the components differs from occasion to occasion. Accordingly, this method, called Weighted Principal Components Analysis (Weighted PCA) here, minimizes

$$\text{WPCA}(A,B,\mathbf{c})=\sum_{k=1}^{p}\|X_k-c_kAB'\|^2, \tag{3.1}$$

where the component matrix A and the pattern matrix B are of the same orders as in PARAFAC, $\mathbf{c}$ is a p–vector such that $\mathbf{c}=(c_1,\ldots,c_p)'$, and c_k denotes the weight for occasion k, $k=1,\ldots,p$. From (3.1) it is clear that Weighted PCA minimizes the PARAFAC loss function subject to the constraint that $D_k=c_kI$, $k=1,\ldots,p$. The main purpose of the present chapter is to compare this constrained variant of PARAFAC with unconstrained PARAFAC, and to give some special properties of Weighted PCA.

An obvious difference between PARAFAC and Weighted PCA is that PARAFAC will usually give the better fitting representation of the data, whereas Weighted PCA gives the more parsimonious representation (using fewer parameters than PARAFAC). As a result, Weighted PCA may be preferred over PARAFAC in cases where the discrepancy in fit is small. One such situation arises if the PARAFAC solution yields matrices $D_1,...,D_p$ that are exactly proportional, because then Weighted PCA will represent the data exactly as well as PARAFAC does, as is readily verified. In case the PARAFAC solution yields $D_1,...,D_p$ that are *nearly* proportional, it can be expected that the representation $c_1AB',...,c_pAB'$ fits the data almost as well as the unconstrained PARAFAC representation.

Another difference between Weighted PCA and PARAFAC is related to uniqueness. In Weighted PCA, which is PARAFAC with the matrices $D_1,...,D_p$ constrained to be proportional, Kruskal's (1977) sufficient condition $k_a+k_b+k_c \geq 2q+2$ for uniqueness is not satisfied. In fact, Weighted PCA does not have a unique solution, as follows at once from $c_kAB'=c_kATT^{-1}B'=c_k\overset{*}{A}\overset{*}{B}'$, $k=1,...,p$, where $\overset{*}{A}=AT$, $\overset{*}{B}=B(T')^{-1}$, for any non-singular T.

In the previous chapter, it has been proposed to use PARAFAC with two proportional columns in, for instance, C to detect weak uniqueness. This raises the question how PARAFAC with two proportional columns in C is related to Weighted PCA, which can be seen as PARAFAC with proportional columns throughout the matrix C. From (3.1) it is clear that, for dimensionality 2, Weighted PCA and PARAFAC with two proportional columns in C coincide and, for dimensionality greater than 2, Weighted PCA is a constrained variant of PARAFAC with two proportional columns in C. Apart from this relation of Weighted PCA with the constrained PARAFAC methods in the previous chapter, one may wonder how Weighted PCA is related to other methods to analyze three-way data. Kiers (1991) has presented a hierarchy for a number of three-way methods, where every method in the hierarchy is a constrained variant of its predecessor. Weighted PCA is not mentioned in this hierarchy. In the present chapter, it will be shown how Weighted PCA fits into this hierarchy.

Van IJzendoorn and Kroonenberg (1990) used the TUCKALS3 algorithm (Kroonenberg & De Leeuw, 1980; Kroonenberg, Ten Berge, Brouwer, & Kiers, 1989) for Weighted PCA. It will be shown in the present chapter that a simpler and more efficient algorithm than the TUCKALS3 algorithm can be used for Weighted PCA.

In the case of exploratory data analysis, one is interested in finding a representation for the data that fits the data satisfactorily and consists of components that allow an easy interpretation. In this chapter, it will be illustrated that Weighted PCA may fit the data almost as well as PARAFAC, and that the Weighted PCA representation may be preferred because its components allow an easier interpretation. In addition, it will be illustrated that, in case PARAFAC yields a degenerate solution, Weighted PCA may be used in addition to PFORTA, to avoid a degenerate solution. Specifically, it will be illustrated that from a PFORTA and a Weighted PCA representation different types of conclusions can be drawn. First, however, a reinterpretation of Weighted PCA will be given.

3.1 Weighted PCA interpreted as PCA of a weighted sum of data matrices

An important property of the Weighted PCA method is that, for fixed **c** scaled to unit length, Weighted PCA is equivalent to PCA of $\sum_{k=1}^{p} c_k X_k$. That is, for optimal A and B the product AB' optimally represents $\sum_{k=1}^{p} c_k X_k$, so Weighted PCA can be interpreted as PCA of a weighted sum of the frontal slabs, hence its name. Specifically, for fixed **c** with $\mathbf{c}'\mathbf{c}=1$, and from (3.1) it follows that

$$\text{WPCA}(A,B,\mathbf{c})=\sum_{k=1}^{p}\|X_k\|^2-2\text{tr}\sum_{k=1}^{p}c_kX_k'AB'+\|AB'\|^2,$$

$$=\|\sum_{k=1}^{p}c_kX_k-AB'\|^2+\sum_{k=1}^{p}\|X_k\|^2-\|\sum_{k=1}^{p}c_kX_k\|^2. \tag{3.2}$$

Because **c** is considered fixed, $\sum_{k=1}^{p}\|X_k\|^2-\|\sum_{k=1}^{p}c_kX_k\|^2$ is constant. So for the A and B that minimize WPCA(A,B,**c**), the product AB' must optimally represent $\sum_{k=1}^{p} c_k X_k$, and can therefore be interpreted as PCA of $\sum_{k=1}^{p} c_k X_k$.

3.2 How Weighted PCA fits in a hierarchy of three-way methods

Kiers (1991) has presented the following hierarchical relations between a number of three-way methods: SUMPCA, a method to be discussed shortly, is a constrained variant of PARAFAC subject to orthonormality constraints on A and B (called PFORTAB here); PFORTAB is a constrained variant of PARAFAC; PARAFAC can be seen as constrained TUCKALS3 (see Carroll & Chang, 1970, p. 312). We will now show that Weighted PCA fits in this hierarchy at a position in between SUMPCA and PFORTAB. Specifically, it will be shown that Weighted PCA is constrained PFORTAB and that SUMPCA is constrained Weighted PCA.

It will first be shown that Weighted PCA is a constrained variant of PFORTAB. For this purpose, the Weighted PCA loss function may be written as

$$\text{WPCA}(A,B,\mathbf{c},D)=\sum_{k=1}^{p}\|X_k-Ac_kDB'\|^2, \tag{3.3}$$

where D is a diagonal matrix. Analogously to the derivation of (3.2), it can be shown that in (3.3) A and B can be constrained to be column-wise orthonormal without loss of fit. It follows at once that Weighted PCA can be seen as PFORTAB with D_k constrained to be c_kD, $k=1,...,p$.

It will secondly be shown that SUMPCA is a constrained variant of Weighted PCA. From Kiers (1991), it can be seen that SUMPCA minimizes

$$\text{SUMPCA}(A,B,D)=\sum_{k=1}^{p}\|X_k-ADB'\|^2, \tag{3.4}$$

where A and B are of the same order as in (3.1) and D is a diagonal matrix. To show that SUMPCA is a constrained variant of Weighted PCA, it merely has to be noted that imposing the constraint $c_k=c$, $k=1,...,p$, in Weighted PCA and absorbing c in D gives the description of SUMPCA.

Clearly, SUMPCA consists of less parameters than Weighted PCA. Therefore, SUMPCA provides a more parsimonious representation of the three-way data than Weighted PCA and might be preferred if it represents the data (almost) as well as Weighted PCA. Similarly, Weighted PCA provides a more parsimonious representation of the three-way data than PFORTAB and PARAFAC and it might be preferred if it represents the data (almost) as well as PFORTAB and PARAFAC.

3.3 An Algorithm for Weighted PCA

Van IJzendoorn and Kroonenberg (1990) used the TUCKALS3 algorithm for Weighted PCA. The TUCKALS3 algorithm minimizes

$$\text{TUCKALS3}(A,B,C,G_1,\ldots,G_{q_3})=\sum_{k=1}^{p}\|X_k-A\sum_{l=1}^{q_3}c_{kl}G_lB'\|^2, \tag{3.5}$$

over column-wise orthonormal matrices $A(n\times q_1)$, $B(m\times q_2)$, $C(p\times q_3)$, and arbitrary matrices $G_1,\ldots,G_{q_3}$ of order $q_1\times q_2$, where q_1, q_2 and q_3 denote the dimensionality of the solution for A, B, and C, respectively. The so-called core array consists of the matrices $G_1,\ldots,G_{q_3}$, which contain interactions between the components collected in the columns of the matrices A, B, and C. As noted by Ten Berge et al. (1987, pp. 189–190), TUCKALS3 reduces to Weighted PCA in case $q_1=q_2=q$ and $q_3=1$. In the context of Weighted PCA the $q_1\times q_2$ core array has only one frontal slab, which will be denoted by G. It is proposed to use an alternative algorithm, which is more efficient. The alternative is based on considering Weighted PCA as a special case of TUCKALS3, with $q_1=q$, $q_2=m$ and $q_3=1$, so B is a square matrix and C is a vector, denoted by $\mathbf{c}$. Then, the TUCKALS3 algorithm minimizes

$$\text{TUCKALS3}(A,B,\mathbf{c},G)=\sum_{k=1}^{p}\|X_k-c_kAGB'\|^2. \tag{3.6}$$

Because B is a square orthonormal matrix, the rotational freedom of B in TUCKALS3 can be used, without loss of fit, to fix B to I_m. Then, the TUCKALS3 loss function coincides with the Weighted PCA loss function with G' fulfilling the role of B from Weighted PCA. For our special case, the TUCKALS3 algorithm can be elaborated as follows. According to Kiers, Kroonenberg and Ten Berge (1992) the TUCKALS3 update for G, using the fact that $B=I_m$, is

$$G=A'\sum_{k=1}^{p}c_kX_kB=A'\sum_{k=1}^{p}c_kX_k, \tag{3.7}$$

and the update for A is

$$A=\text{GS}\left(\sum_{k=1}^{p}c_kX_kBG'\right)=\text{GS}\left(\sum_{k=1}^{p}c_kX_kG'\right), \tag{3.8}$$

where $\text{GS}\left(\sum_{k=1}^{p}c_kX_kG'\right)$ denotes the Gram-Schmidt orthogonalization of $\sum_{k=1}^{p}c_kX_kG'$. Finally, the update for $\mathbf{c}$ is GS($\mathbf{d}$), where d_k, element k of $\mathbf{d}$, is equal to $\text{tr}A'X_kBG'=\text{tr}A'X_kG'$. In the TUCKALS3 algorithm, one iteration consists of updating G, A, G, B, G, and C, respectively (see Kiers et al., 1992). These iterative steps can be simplified without affecting the monotonical convergence of the TUCKALS3 algorithm. Note that B is constant, so the update for B and the consecutive update for G can be left out. Also note that, because C has rank 1, the above GS-update equals the optimal Bauer-Rutishauser update (see Kiers et al., 1992), hence updating G may be skipped, without affecting the monotonical convergence of the algorithm. In conclusion, it is proposed to update A for fixed G and $\mathbf{c}$, update G for fixed A and $\mathbf{c}$, and update $\mathbf{c}$ for fixed A and G, in each full cycle of the algorithm. This algorithm will be called the Weighted PCA algorithm. After the same stopping criteria of section 1.2 are satisfied, the optimal B in (3.1) is found as G' in the Weighted PCA algorithm.

In Appendix A it is proven that the above Weighted PCA algorithm and the TUCKALS3 algorithm yield the same series of function values. Weighted PCA is based on fewer (matrix) multiplications than TUCKALS3, as can be seen, for instance, by comparing the expressions for A^1 (see equation A.1 in Appendix A) and for E^1 in Appendix A. Therefore, Weighted PCA uses less computation time. In a few test runs on empirical data it was found that the Weighted PCA algorithm was about 3 times as fast as the TUCKALS3 algorithm.

A number of test trials on empirical data revealed that the Weighted PCA algorithm is not sensitive to local minima. Because the danger of local minima cannot be ruled out completely, it is suggested to run more than one Weighted PCA analysis on the same data with different starting configurations for the matrix A and the vector $\mathbf{c}$.

As a useful additional result, it is found that, if the A parameters have converged, the Weighted PCA algorithm gives the principal components of $\sum_{k=1}^{p} c_k X_k$. This can be seen as follows. Upon convergence of the A parameters in the algorithm, it follows that $A=\text{GS}\left(\sum_{k=1}^{p}\sum_{l=1}^{p} c_l c_k X_k X_l' A\right)=\sum_{k=1}^{p}\sum_{l=1}^{p} c_l c_k X_k X_l' A U$, for a certain upper triangular matrix U. Hence, $I_q=A'A=\left(A'\sum_{k=1}^{p}\sum_{l=1}^{p} c_l c_k X_k X_l' A\right)U$. For all practical purposes it may be assumed that $\left(A'\sum_{k=1}^{p}\sum_{l=1}^{p} c_l c_k X_k X_l' A\right)$ has full rank. Then, $U=\left(A'\sum_{k=1}^{p}\sum_{l=1}^{p} c_l c_k X_k X_l' A\right)^{-1}$, which is symmetric. Because U is also upper triangular, it follows that U

is diagonal. So, $\left(\sum_{k=1}^{p}\sum_{l=1}^{p} c_l c_k X_k X_l'\right)A=AU^{-1}$, and hence A contains eigenvectors of $\left(\sum_{k=1}^{p}\sum_{l=1}^{p} c_l c_k X_k X_l'\right)$ and therefore the principal components of $\left(\sum_{k=1}^{p} c_k X_k\right)$, which completes the proof.

Recently, Kiers et al. (1992) proposed a modification of the TUCKALS3 algorithm that handles three-way arrays of order $n \times m \times p$ for any n. The modified algorithm is efficient in the sense that, if $n > mp$, it needs less work space to store the data during the iterative process on a computer and it has a higher execution speed. Above, the Weighted PCA algorithm was derived as a special case of the TUCKALS3 algorithm, and hence the modified TUCKALS3 algorithm can also be used for Weighted PCA. This modified TUCKALS3 algorithm is based on updating A implicitly. For Weighted PCA, this modified TUCKALS3 algorithm can be further simplified in the sense that updating A can be skipped from the iterative process. This has the advantage over the modified TUCKALS3 algorithm that even less storage space is needed and that computational speed is increased. This can be seen as follows. Without loss of generality, for the TUCKALS3$(A,B,\mathbf{c},G)=\sum_{k=1}^{p}\|X_k-c_k AGB'\|^2$ loss function Weighted PCA may be subjected to the constraints $B=I_m$, $\mathbf{c}'\mathbf{c}=1$ and $G'G=I_q$. It can be verified that, for fixed $\mathbf{c}$ and G, it follows from regression theory (Draper & Smith, 1981) that the optimal A satisfies $A=\left(\sum_{k=1}^{p} c_k X_k BG'\right)=\left(\sum_{k=1}^{p} c_k X_k G'\right)$. Upon substituting $\left(\sum_{k=1}^{p} c_k X_k G'\right)$ for A into TUCKALS3$(A,B,\mathbf{c},G)$ and using the fact that $B=I_m$, it follows that minimizing TUCKALS3$(A,B,\mathbf{c},G)$ is equivalent to maximizing

$$g(G,\mathbf{c})=\mathrm{tr}G'\left(\sum_{k=1}^{p} c_k X_k\right)'\left(\sum_{l=1}^{p} c_l X_l\right)G=\sum_{k=1}^{p}\sum_{l=1}^{p} c_k c_l \mathrm{tr}G'X_k'X_l G=\mathbf{c}'V\mathbf{c}, \qquad (3.9)$$

where V is a positive semi-definite matrix with $v_{kl}=\text{tr}G'X_k'X_lG$ as element (k,l), subject to the constraints $\mathbf{c}'\mathbf{c}=1$ and $G'G=I_q$. From (3.9) it can be seen that G and $\mathbf{c}$ can be updated by using Bauer–Rutishauser or by using Gram–Schmidt in the same way as in the TUCKALS3 algorithm (Kroonenberg & De Leeuw, 1980; Kroonenberg et al., 1989). This shows that, in order to minimize the TUCKALS3 loss function for Weighted PCA, only G and $\mathbf{c}$ need to be updated, and only the matrix with $X_k'X_l$, $k=1,...,p$ and $l=1,...,p$ need to be stored by the computer.

3.4 Interpretation of the Weighted PCA components

In chapter 1 the structure matrix for PARAFAC was defined as $S=\sum_{k=1}^{p} X_k'AD_k$, and it was proposed to use S for the interpretation of the PARAFAC components. Analogously, a structure matrix for Weighted PCA can be defined as $S\equiv\sum_{k=1}^{p} c_kX_k'A$, which follows from substituting c_kI_q for D_k in $S=\sum_{k=1}^{p} X_k'AD_k$.

It is well–known that in PCA, when the components are orthonormal, the structure matrix and the pattern matrix are equal. With our definition of the structure matrix this property is retained in the Weighted PCA generalization of PCA. That is, if $A'A=I_q$ and $\mathbf{c}'\mathbf{c}=1$, which can always be arranged, then $B=S$. This can be proven by using the substitution c_kI_q for D_k analogously to the proof for $S=B$ in case of PFORTA, see section 1.5. Therefore, using S or B as a basis for interpretation of the Weighted PCA components leads to the same interpretation if $A'A=I_q$ and $\mathbf{c}'\mathbf{c}=1$.

The occasion parameters in $\mathbf{c}$ of Weighted PCA can be interpreted from the knowledge of the occasions. In addition to this, it may be important to know how well the Weighted PCA model fits the data for each occasion separately. For this reason, the constraint $\mathbf{c}'\mathbf{c}=1$ will be dropped and the constraint $\|B\|^2=1$ will be imposed instead, as can be arranged without loss of fit. If X_k is centered column-wise, and if c_k is optimal, then c_k^2 is the amount of variance in X_k that the Weighted PCA components explain, $k=1,\ldots,p$. This can be proven as follows. From the optimality of c_k, it follows that c_k minimizes $\mathrm{f}(c_k)=\|X_k-c_kAB'\|^2$. From this, $A'A=I_q$ and $\|B\|^2=1$, it follows that $c_k=\mathrm{tr}A'X_kB$, because

$$\mathrm{f}(c_k)=\|X_k-c_kAB'\|^2=\|X_k\|^2-2\mathrm{tr}A'X_kBc_k+c_k^2+(\mathrm{tr}A'X_kB)^2-(\mathrm{tr}A'X_kB)^2$$

$$=(\mathrm{tr}A'X_kB-c_k)^2+\|X_k\|^2-(\mathrm{tr}A'X_kB)^2. \quad (3.10)$$

This function is minimal for $c_k=\mathrm{tr}A'X_kB$. From the substitution of $\mathrm{tr}A'X_kB$ for c_k into (3.10) it follows that

$$\mathrm{f}(c_k)=\|X_k\|^2-c_k^2. \quad (3.11)$$

From (3.11), it can be seen that for occasion k the amount of residual variance, written as $\|X_k-c_kAB'\|^2$, is partitioned into the amount of variance to be explained $\|X_k\|^2$, and the squared element of the occasion parameter c_k^2. Therefore, c_k^2 is the amount of variance that Weighted PCA explains in X_k, $k=1,\ldots,p$, which completes the proof.

3.5 An empirical example

To illustrate the above hierarchical relations and Weighted PCA, a data set, denoted as GOS* data, with scores of 34 children of age 2–3 on 10 variables that measure intelligence on two occasions, was analyzed. The variables are divided into four groups: Simultaneous Processing (Magic Window, Face Recognition, and Gestalt Closure), Achievement (Expressive Vocabulary and Faces and Places), Sequential Processing (Hand Movements and Number Recall), and Motor Skills (Gross Motor Skills, Fine Motor Skills, and Figure Movement in Disc), see Van Eldik, Neutel, Van der Meulen, and Spelberg (1990). Prior to the analysis, the variables were centered column-wise within the occasions, and scaled to unit length over the occasions.

To illustrate the above hierarchical relations the GOS data were analyzed by SUMPCA, Weighted PCA and PARAFAC, all with dimensionality 2. The matrix C that resulted from the PARAFAC analysis, with A and B scaled to unit length column-wise, was $\begin{pmatrix} 1.14 & .61 \\ 1.61 & 1.27 \end{pmatrix}$. The value of Tucker's (1951) coefficient of congruence between the two columns of C is .99. So it can be concluded that the columns in the matrix C are nearly proportional. From this it can be expected that the discrepancy between PARAFAC and Weighted PCA is small. The percentage of variance explained by PARAFAC, Weighted PCA, and SUMPCA was 47.5, 47.2, and 45.5, respectively. The small discrepancy in fit between PARAFAC and Weighted PCA indicates that these

---

*The author is obliged to Ria Neutel and Henk Lutje Spelberg who kindly made the data available.

PARAFAC components are weakly unique. The SUMPCA model, or, equivalently, PCA of the average matrix, seems not very restrictive for these data.

To illustrate Weighted PCA and to compare Weighted PCA with PARAFAC, more detailed results of the analyses are given now. The occasion parameters that resulted from Weighted PCA (with $\|B\|^2=1$) for occasion 1 and 2 were 1.18 and 1.82, respectively. By simply squaring these values, see section 3.4, it can be found that the amount of variance explained for the first occasion is 1.40 and for the second occasion 3.32. The amounts of variance to be explained were 4.01 and 5.99, for occasion 1 and 2, respectively. So the Weighted PCA components explain 34.6% and 55.4% of the variance for occasion 1 and 2, respectively. For PARAFAC, the corresponding percentages are 34.6% and 56.2%, respectively. From this it can be concluded that for both methods the data for occasion 2 are represented better than those for occasion 1. The structure matrices resulting from PARAFAC and Weighted PCA are depicted in Table 3.1. Before the structure matrices were computed A, C, and $\mathbf{c}$ were scaled to unit length column-wise.

From Table 3.1 it can be seen that the structure matrices from Weighted PCA (without rotation) and PARAFAC are quite similar. The variables are assumed to belong to two clusters, but the PARAFAC components, as well as the Weighted PCA components, do not correspond to these clusters of variables. Now the fact that Weighted PCA can be seen as PCA of $\sum_{k=1}^{p} c_k X_k$ allows one to use all available rotation techniques for PCA in Weighted PCA analysis, as if $\sum_{k=1}^{p} c_k X_k$ were analyzed by PCA. It was chosen to rotate the components to oblique components that approximate simple structure, by using the independent cluster rotation proposed by Harris and Kaiser (1964). This resulted in rotated components that correlate 0.34. In Table

3.1 the structure matrix from rotated Weighted PCA analysis is depicted with elements greater than .40 in bold face. It can be seen from the rotated Weighted PCA structure matrix, that the variables of the first two groups load mainly on the first rotated component, and the variables of the second two groups load mainly on the second rotated component. So the components may be interpreted as Simultaneous Processing and Sequential Processing, respectively. Apparently, by using the rotational freedom that Weighted PCA has, components can be identified that correspond better to the clusters of variables and therefore allow for a simpler interpretation than the PARAFAC components.

Table 3.1 *The structure matrices from PARAFAC, and unrotated and rotated Weighted PCA analysis of the GOS data.*

	PARAFAC		WPCA		ROTATED	
Variable	I	II	I	II	I	II
Magic Window	.76	.48	.68	.40	**.78**	.19
Face Recognition	.62	.40	.56	.29	**.63**	.19
Gestalt Closure	.63	.36	.58	.27	**.64**	.21
Expressive Vocabulary	.78	.37	.73	.27	**.77**	.31
Faces and Places	.75	.40	.70	.29	**.76**	.28
Hand Movements	.49	–.33	.63	–.41	.42	**.73**
Number Recall	.31	–.53	.48	–.60	.20	**.76**
Gross Motor Skills	.28	–.54	.46	–.62	.17	**.77**
Fine Motor Skills	.32	–.21	.41	–.29	.26	**.49**
Figure Movement in Disc	.27	–.02	.31	–.13	.23	.30

In order to assess the stability of the PARAFAC and the rotated Weighted PCA components the data were subjected to five different splithalf analyses. After determining the two sets randomly the data were preprocessed per set as before. For PARAFAC only one component in one splithalf analysis was stable. The lowest congruence values encountered in the five splithalf analyses were .62, .57, .56, .67, and .26. For Weighted PCA, Tucker's congruence coefficient was computed between the corresponding **c** vectors and between the corresponding columns of the rotated matrix *BT*, were *T* is the rotation matrix, proposed by Harris and Kaiser (1964). For Weighted PCA with rotation, the five **c** vectors and the five first columns of *B* were stable and two of the second columns of *B* were stable. For the unstable columns of *B* the congruence values encountered were .65, .81, and .75. These results suggest that neither the PARAFAC nor the rotated Weighted PCA components are stable in the splithalf sense, and that the rotated Weighted PCA components are less unstable than the PARAFAC components.

3.6 Using Weighted PCA to arrive at a non-degenerate solution

Here it will be illustrated, by reporting the results of various analyses of the 32x6x8 Affective Response data (Eckblad, 1981), that PARAFAC may yield degenerate components and that for the same data Weighted PCA may be preferred over PFORTA. The Affective Response data consist of scores of 32 students who judged eight tasks of increasing complexity on six scales: Simple–Complex, Monotone–Varied, Boring–Interesting, Unpleasant–Pleasant, Uncomfortable–Comfortable and Disorderly–Clear. The tasks are manipulated such that they should only differ with respect to complexity, which is one–dimensional. So in advance it is at least doubtful that the tasks show distinct stretching and contraction for two PARAFAC components. Eckblad (1981, p. 3–6) predicted and found, by analyzing $\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ by PCA, that the variables are represented (more or less) along a semi–circle. In addition, Eckblad (1981, pp. 3–4) predicted that the rank order of the tasks is the same for all the subjects. As was done by Eckblad (1981), the data array X was centered and scaled to unit length over the persons and the occasions. The Affective Response data were analyzed by PARAFAC in two dimensions. After 2000 iterative steps the value –.78 was found for cosABC. By taking more iterative steps until machine precision broke down the monotonical convergence the value of cosABC decreased further, but did not reach –.85. After 2000 iterative steps the cosines between the columns of A, B, and C are .96, .84, and –.98, respectively. This solution can be seen as borderline degenerate.

To avoid a degenerate solution, Harshman and Lundy (1984b, p. 274) suggest imposing column–wise orthonormality on one of the parameter matrices. It will now be illustrated that another way to arrive at a non–degenerate solution is to apply Weighted PCA. The Weighted PCA solution will be compared to that of PARAFAC with column–wise orthonormality on A. The Affective Response data were analyzed by PFORTA, Weighted PCA (without rotation), and SUMPCA. The percentage of explained variance by PFORTA, Weighted PCA, and SUMPCA is 44.3, 42.9, and 8.0, respectively. Clearly, SUMPCA is overly restrictive for the Affective Response data, whereas the Weighted PCA fit to the data is 1.4% of variance explained less than that of PFORTA, see also section 2.3. This rather small discrepancy between PFORTA and Weighted PCA suggests that these data can be fitted reasonably well by the more parsimonious Weighted PCA model. The structure matrices that resulted from PFORTA and from Weighted PCA (without rotation) are depicted in Table 3.2. Before the structure matrices were computed C and $\mathbf{c}$ were scaled to unit length column-wise.

From the structure matrices in Table 3.2, it can be seen that the variables are indeed (more or less) represented on a semi-circle. On the basis of the structure matrices, the PFORTA and the Weighted PCA components have almost identical interpretations, with the first component interpreted as Complexity and the second component as Pleasantness.

Table 3.2 *The structure matrices from PFORTA and Weighted PCA analysis of the Affective Response data.*

	Structure Matrices			
	PFORTA		Weighted PCA	
Variable	I	II	I	II
Simple–Complex	.80	–.10	.80	.03
Monotone–Varied	.70	.05	.69	.16
Boring–Interesting	.29	.40	.27	.43
Unpleasant–Pleasant	–.09	.48	–.12	.36
Uncomfortable–Comfortable	–.46	.35	–.48	.33
Disorderly–Clear	–.82	.14	–.83	.06

The matrix C (containing the task parameters) that resulted from PFORTA, and the vector $\mathbf{c}$ that resulted from Weighted PCA are reported in Table

3.3. For the task parameters in PFORTA, the matrix B was scaled to unit length column-wise and in Weighted PCA B was scaled such that $\|B\|^2=1$. As was explained in sections 1.3 and 3.4, by squaring these (task) parameter values it can be seen how much variance the corresponding component(s) explain(s) for the corresponding task.

From Table 3.3 it can be seen that the task parameters of Weighted PCA increase, except for one task, with the complexity of the tasks. Hence, Weighted PCA recovered, up to one exception, the rank order of the tasks consistently with the one-dimensional manipulation of the tasks.

Table 3.3 *The Task parameters from PFORTA and Weighted PCA analysis of the Affective Response data.*

	Task Parameters		
	PFORTA		Weighted PCA
Task	I	II	
1	−.72	−.17	−.76
2	−.59	−.04	−.64
3	−.44	−.10	−.48
4	−.20	.13	−.20
5	.27	.45	.38
6	.53	.29	.61
7	.41	.29	.46
8	.70	.35	.76

The first column of the task parameter matrix of PFORTA resembles the task parameters of Weighted PCA and can therefore be interpreted accordingly.

As explained in section 3.4, by squaring the elements of the second column of the task parameter matrix, it can be computed that the second component explains 0.1 of the sum of squares for the first four and 0.5 of the sum of squares for the last four tasks. Apparently, the second component accounts mainly for tasks 5 through 8, which are the more complex tasks.

In order to assess the stability of the PFORTA and the Weighted PCA components, the data were subjected to five different splithalf analyses. After determining the two sets randomly the data were preprocessed per set as before. For Weighted PCA, Tucker's congruence coefficient was computed between the corresponding **c** vectors and between the corresponding columns of the matrix B. The first PFORTA component was stable but the second was unstable over the five analyses. The lowest congruence values over B and C encountered for PFORTA in each of the five splits were .07, .34, .41, .60, and .76. The lowest congruence value encountered for Weighted PCA was .90. These results suggest that the PFORTA components are not stable, whereas the Weighted PCA components are stable in the splithalf sense.

Several conclusions can be drawn from the SUMPCA, the Weighted PCA, PFORTA and PARAFAC analysis of the Affective Response data. The SUMPCA model is overly restrictive for these data. The PARAFAC solution is borderline degenerate and is therefore unacceptable. PFORTA and Weighted PCA represent the Affective Response data almost equally well and can both be used to arrive at a non-degenerate solution. However, only the Weighted PCA components are stable in the splithalf sense. Up to one exception, the Weighted PCA task parameters recovered the rank order of the tasks.

3.7 Conclusions

A great discrepancy in fit between Weighted PCA and PARAFAC implies that the proportional columns constraint in Weighted PCA is overly restrictive. In such a case it seems appropriate for the components to have distinct relative importances from occasion to occasion. On the other hand, in case of a small discrepancy in fit between Weighted PCA and PARAFAC it seems appropriate for the components to have the same relative importances. In this sense, the Weighted PCA solution provides a more parsimonious representation of the three-way data. An even more parsimonious representation of the three-way data can be found by using the rotational freedom in Weighted PCA to approximate simple structure, as has been illustrated. It can be concluded that, in order to analyze three-way data, Weighted PCA can be useful in addition to PARAFAC.

In case one's goal is to find components that allow an easier interpretation, imposing other constraints than the proportional columns constraint on the occasion parameters, as is the case in Weighted PCA, may be useful. For instance, in case of three-way positive manifold data, components without contrasting signs allow an easier interpretation, as was explained in section 1.9. The subject of non-contrast components will be treated in the next chapter.

CHAPTER 4

A CONSTRAINED PARAFAC METHOD FOR POSITIVE MANIFOLD DATA

In case a correlation matrix has no negative elements, the variables show 'positive manifold'. Such data will be called positive manifold data. Positive manifold data arise in various research areas. For example, positive manifold data are often found if the variables measure intelligence. If positive manifold data are analyzed by PCA, then, before rotation, the first component has no negative correlations with the variables. The other components correlate positively with some variables and negatively with some other variables. This implies that the structure matrix has positive and negative elements in the same column. Accordingly, these components can be called contrast components. In Table 4.1 an example is given of a correlation matrix for positive manifold data (computed from the first frontal slice of the Drenth data in Table 4.2) and the unrotated structure matrix from PCA.

The unrotated first and second component can be interpreted as General Intelligence and Nonverbal-Verbal Contrast, respectively. By rotating the structure matrix according to the VARIMAX criterion the contrast components disappear, see Table 4.1. The rotated components can be interpreted as Analogies and Non-Verbal Abstraction, respectively. It can be seen that the rotated components are non-contrast components that coincide more or less with two groups of variables. This illustrates that rotation of the PCA components can be useful to find components that allow an easier interpretation. Note that these rotated components explain the variables equally well as the unrotated components.

Table 4.1 *A correlation matrix* (R), *the unrotated* (PCA) *and the rotated* (VARIMAX) *structure matrix that are found with PCA.*

	R			PCA		VARIMAX	
Variable				I	II	I	II
Vocabulary Analogies	1.00	.59	.30	.81	−.45	.92	.10
Verbal Analogies	.59	1.00	.43	.87	−.15	.80	.38
Non−verbal Abstraction	.30	.43	1.00	.69	.71	.16	.98

In case one has a three−way array X of positive manifold data, for example, with scores from n persons on m variables that measure intelligence at p occasions, PARAFAC can be used to represent such data. The question arises whether or not the PARAFAC components display contrast.

Before this question can be answered, contrast components for PARAFAC need to be defined. A PARAFAC component is a contrast component if at least one of the matrices S, B, and C has at least one column which contains both negative and positive elements. This definition covers various sorts of contrasts. Specifically, if the matrix S has a column with contrasting elements, then the components are contrast components in the same sense as contrast components defined for PCA. If the matrix C has a column with contrasting elements, then the contributions of that component to the representation of the variables have different signs across the occasions. If the pattern matrix B has a column with contrasting elements, then there is no positive manifold for those researchers who interpret the components on the basis of B.

4.1 PARAFAC representations of some empirical data sets

We will now turn to the question whether or not PARAFAC can yield contrast components if data without negative elements of $X_k'X_l$, $k,l=1,\ldots,p$ are analyzed. This question will be answered by presenting the results from analyzing three empirical data sets, called Drenth, GIT and DAT. The three data arrays all consist of scores of persons on variables that measure intelligence on two occasions. Specifically, the above question will be answered by comparing the PARAFAC components with the rotated PCA components per frontal slice. In order to study the performance of PARAFAC in cases where the rotated structure matrix from PCA per frontal slice indicates non-contrast components, the data were slightly modified to ensure that per frontal slice such non-contrast components were found: Per frontal slab the person with a maximal amount of person variance was removed until a VARIMAX rotated PCA structure matrix (with dimensionality 2) for every frontal slice was found without negative elements.

Prior to the PARAFAC analysis, the variables were centered column-wise within the occasions, and scaled to unit length over the occasions. The dimensionality was fixed to 1, 2, and 3, respectively. If PARAFAC finds components without contrast, then after a suitable reflection of the columns of the matrices A, B, and C, the matrices S, B, and C do not have negative elements. Note that from the definition of contrast components it follows that, if the lowest elements that are found in S, B, and C, respectively, are all non-negative, then the PARAFAC components are non-contrast components. But in case a negative element in S, B, or C is found, and the largest element that is found in the corresponding column is positive, then the corresponding component is a contrast component. For dimensionality 3, the PARAFAC analysis of the Drenth data yielded a value of −.89 of cosABC after 2000 iterations, hence this PARAFAC solution is degenerate. The results that are found by PARAFAC analysis of the three data sets are reported in Table 4.2. The matrices A and B are scaled to unit length column-wise.

From Table 4.2 it can seen that PARAFAC yields both negative and positive elements in S or B (or both) in case the dimensionality is fixed to 2 or 3. It can be concluded that PARAFAC may yield contrast components where PCA per frontal slab followed by VARIMAX rotation of the structure matrix yields components without contrast. Therefore, the absence of rotational freedom can be considered as a disadvantage of the PARAFAC model, for these data.

It should be noted from Table 4.2 that in case of dimensionality 1, no contrast component is found. In fact, it can be proven that, with dimensionality 1, no contrast components can occur for positive manifold data. Specifically, let **a**, **b**, and **c** be the parameter vectors of the PARAFAC solution, for dimensionality 1. If $X_k'X_l$, $k,l=1,\ldots,p$, have no negative elements, then $X_k'\mathbf{a}$, $k=1,\ldots,p$, **b** and **c** have no negative elements. The proof is given in appendix B. The theorem given here is stronger than the one derived by Krijnen and Ten Berge (1992), and it is a generalization of the Perron–Frobenius theorem.

Table 4.2 *Results of the PARAFAC analysis of three different sets of variables measuring intelligence, in terms of percentage of explained variance (Fit%), lowest element of S* (low$S$), *B* (low$B$) *and C* (low$C$). *In case a negative lowest element is found, the largest element of the corresponding column is recorded between parentheses.*

Data Set	order	Dim.	*Fit%*	low$S$	low$B$	low$C$
Drenth	33×3×2	1	54.0	.46	.37	.87
		2	75.1	.15	**−.52**(.84)	.62
		3	87.4[1]	**−.38**(.21)	**−.58**(.78)	.70
GIT	20×4×2	1	54.0	.68	.46	.98
		2	68.3	**−.10**(.75)	**−.66**(.56)	.49
		3	80.8	**−.04**(.86)	**−.49**(.79)	.47
DAT	87×9×2	1	39.2	.30	.16	1.24
		2	50.8	**−.41**(.54)	**−.27**(.59)	.74
		3	60.0[2]	**−.40**(.54)	**−.62**(.59)	.67

[1]PARAFAC yielded degenerate components. [2]After 1000 iterative cycles the parameters still did not converge.

It has been illustrated that PARAFAC may yield contrast components for dimensionality greater than 1. It has also been illustrated that contrast components are less easy to interpret. In case PARAFAC finds contrast components, the question arises how non–contrast components can be determined. A way to do so is by subjecting the PARAFAC parameter matrices to certain constraints. Specifically, in case the matrices $X_k'A$, k=1,...,p, B, and C are subjected to the constraints that these matrices have no negative elements, then such PARAFAC components are non-contrast components by definition. In order to find optimal non-contrast PARAFAC components, it is proposed to minimize PARAFAC(A,B,C) subject to the constraints that the matrices $X_k'A$, k=1,...,p, B, and C do not have negative elements. To solve this minimization problem, an ALS algorithm will be derived in the next section.

4.2 An ALS algorithm for optimal non-contrast PARAFAC components

In general, a non-negativity constraint can be imposed directly on each row of the parameter matrix B. With the Non-Negative Least Squares (NNLS) algorithm (Lawson & Hanson, 1974, pp. 158–165; Tenenhaus, 1988) PARAFAC(A,B,C) can be globally minimized over B for fixed A and C, subject to non-negativity of the elements in B. Analogously, with the NNLS algorithm PARAFAC(A,B,C) can be globally minimized over C for fixed A and B, subject to non-negativity of C. This seems a more efficient manner of imposing a non-negativity constraint on C than the one proposed by Ten Berge (1986). Indirectly, a non-negativity constraint can also be imposed on the columns of the parameter matrix A such that the matrix $X_k'A$, k=1,...,p, has no negative elements. By using the Vec operator it can be shown that, to minimize PARAFAC(A,B,C) over A for fixed B and C subject to non-negativity of the elements in the matrix $X_k'A$, k=1,...,p, is a constrained regression problem, called a Least Squares with Inequality constraints (LSI) problem by Lawson and Hanson (1974, pp. 165–169). With the LSI algorithm, PARAFAC(A,B,C) can be minimized over A for fixed B and

C subject to non–negativity of the elements in the matrix $X_k'A$, k=1,...,p. Hence, an ALS algorithm can be constructed by updating A with the LSI algorithm and updating B and C separately with the NNLS algorithm. This algorithm will be called PFNC algorithm, since it can be used to minimize the PARAFAC loss function subject to the requirement of Non–Contrast components.

It is of importance to know whether or not a PFNC solution can be degenerate. It will now be shown that if $X_k'A$, k=1,...,p, B and C have no negative elements, then the PFNC solution cannot be degenerate. From the fact that B and C have no negative elements it follows that $\cos(\mathbf{b}_l,\mathbf{b}_{l'}) \geq 0$ and $\cos(\mathbf{c}_l,\mathbf{c}_{l'}) \geq 0$, where $\mathbf{b}_l$, $\mathbf{b}_{l'}$, $\mathbf{c}_l$, and $\mathbf{c}_{l'}$ are column l and column l' of B and C, respectively, for l,l'=1,...,q. If cosABC tends to –1, then there must be a pair (l,l') such that $\cos(\mathbf{a}_l,\mathbf{a}_{l'})$ tends to –1. The fact that $\cos(\mathbf{a}_l,\mathbf{a}_{l'})$ tends to –1 implies that we can make $\cos(\mathbf{a}_l,\mathbf{a}_{l'})$ as close to –1 as we please by increasing the number of iterative steps. Therefore, we can make $\cos(\mathbf{a}_l,\mathbf{a}_{l'})$ so close to –1 that $X_k'A$, k=1,...,p, has at least one negative element, where it is assumed that the columns of X are not all proportional. This assumption is always fulfilled in practice. But a negative element in $X_k'A$, k=1,...,p, would contradict the fact that $X_k'A$, k=1,...,p, has no negative elements. This completes the proof.

Interestingly, it is not sufficient for non–degenerate components to impose only the non–negativity constraint on B and on $X_k'A$, k=1,...,p. If only these constraints are imposed and the Drenth data from Table 4.2 are analyzed with a corresponding algorithm (with dimensionality 3), then after 1000 iterations the value –.99 for cosABC is found.

Practical experience with the PFNC algorithm has been quite satisfactory.

In terms of local minima, PFNC behaves like the PARAFAC algorithm, and it is only a little slower than the PARAFAC algorithm. Because the danger of local minima cannot be ruled out completely, it is suggested to run more than one PFNC analysis on the same data with different non-negative starting configurations for B and C.

4.3 Using PFNC for analyzing three-way positive manifold data

It has been illustrated that a PARAFAC analysis may yield a degenerate solution. In order to avoid degeneracy, the data can be analyzed by PFORTA. But in case of positive manifold data, PFORTA might yield contrast components as will now be exemplified. In addition to this, PFNC will be illustrated by the results from analyzing the 33x3x2 Drenth data by PFORTA and PFNC with dimensionality 2. The resulting matrices S and B from PFORTA and PFNC are depicted in Table 4.3. Both analyses revealed a C matrix with positive elements.

Table 4.3 *The matrices S resulting from PFORTA and from PFNC and the matrix B from PFNC analysis of the Drenth* 33x3x2 *data.*

	S				B	
	PFORTA		PFNC		PFNC	
Variable	I	II	I	II	I	II
Vocabulary Analogies	.86	–.20	.88	.14	.75	.00
Verbal Analogies	.86	.09	.82	.40	.66	.30
Non–verbal Abstraction	.34	.78	.17	.85	.02	.96

From Table 4.3 it can be seen that the components resulting from PFORTA are contrast components. It can be concluded that degeneracy is overcome by PFORTA but contrast is not. Because the PFORTA components are contrast components, these components are more difficult to interpret than the PFNC components. Specifically, the PFNC components for the Drenth data resemble the PCA components rotated according to the VARIMAX criterion, reported in Table 4.1.

Whether or not the PFNC components are to be preferred over the PARAFAC components mainly depends on the amount of variance that is explained by the PFNC components. The three data sets described above were analyzed by PARAFAC and by PFNC with dimensionality 2 and 3. The percentages of explained variance found by PARAFAC and by PFNC are reported in Table 4.4

Table 4.4 *The percentages of variance explained by PARAFAC and by PFNC for three sets of data.*

	Dimensionality 2		Dimensionality 3	
Data Set	PARAFAC	PFNC	PARAFAC	PFNC
Drenth	75.1	74.8	87.4	86.2
GIT	68.3	68.0	80.8	80.5
DAT	50.8	50.7	60.0	59.1

From Table 4.4 it can be seen that, in these six cases, the discrepancy in amount of variance explained by PARAFAC and PFNC is negligible. Although PARAFAC yielded a degenerate solution in one case, there is no need to analyze the data by, for instance, PFORTA and report the percentage of variance explained by PFORTA, because PFORTA explains at most the same amount of variance as PARAFAC. It can be concluded that, in these examples, no essential information is lost by abandoning the PARAFAC components in favor of the PFNC components. In general, after applying PARAFAC and PFNC it is clear how much variance, that is explained by PARAFAC, is not explained by PFNC. Therefore, it is clear how the gain in terms of interpretation from the PFNC components and the loss in explained variance can be weighted against each other.

It might be asked whether or not the discrepancy in explained variance between PARAFAC and PFNC can be large if $X_k'X_l$, $k,l=1,...,p$, has no negative elements. To answer this question, various simulation studies were done. No substantial discrepancies in terms of explained variance between PARAFAC and PFNC were found with data that fulfill the condition that $X_k'X_l$, $k,l=1,...,p$, has no negative elements. This indicates that, for all practical purposes, the non-negativity of $X_k'X_l$, $k,l=1,...,p$ is likely to be sufficient for a small discrepancy in explained variance between PARAFAC and PFNC.

To conclude this section, the stability of the PFNC solution will be examined in comparison with two alternative methods. First, in order to assess the stability of the PFORTA and the PFNC solution the 38×3×2 Drenth data from Table 1.1 were subjected to five different splithalf analyses. Congruence coefficients were computed between corresponding columns of B and of C. For PFORTA, one splithalf analysis yielded two stable components and three yielded a stable first component and an unstable second component. The lowest congruence values encountered for these three unstable components were .68, .72, and .77. One splithalf analysis yielded two unstable components, where the lowest congruence values encountered for each dimension were .78 and .75. The lowest congruence value encountered for PFNC was .93. These results suggest that the PFORTA components are not stable whereas the PFNC components are stable in the splithalf sense.

As a second example of a PFNC analysis, the DAT data (section 2.4) were reanalyzed. For these data, only two splithalf solutions (out of five) were stable, so the PARAFAC solution can be considered unstable. The original PARAFAC solution as well as all the PARAFAC solutions of the splithalf analyses contained contrast components, although the variables are measuring intelligence. So it seems worthwhile to apply PFNC to these data. To assess the stability of the PFNC components, these data were subjected to the same five splithalf analyses as in chapter 2. The lowest congruence value encountered was .95, hence the PFNC solution is stable in the splithalf sense. Interestingly, the PFNC solution explained 48.2 percent of the variance, which is almost equal to the percentage of variance explained by PARAFAC (48.3). Thus, in this example PFNC yields stable components which allow for an easier interpretation and fit the data almost equally well as PARAFAC. For these reasons PFNC seems a useful alternative for representing these data.

4.4 A simulation study on non-contrast components

Above, it was found for three empirical data sets, that the PARAFAC components are contrast components and that the non-contrast components found by the PFNC method explain the data almost equally well. The existence of alternative components leads to the conclusion that the PARAFAC uniqueness is weak in these cases. For positive manifold data it can be expected that PARAFAC yields contrast components in case the uniqueness is weak, and non–contrast components in case the uniqueness is strong. Also, it can be expected, because of sampling bias, that if the sample size is small PARAFAC may yield contrast components. In order to get an impression whether or not these expectations come true, a small simulation study has been conducted. Clearly, the three covariance matrices in section 2.5 have positive manifold and different degrees of uniqueness. The 18 samples from section 2.5 were analyzed by PARAFAC and by PFNC. In case the population had no uniqueness, the PARAFAC solutions for the samples all consisted of contrast components; in all solutions an element lower than –.22 was encountered in the S (where A and C where scaled to unit length column–wise) or the B matrices (where B was scaled to unit length column–wise). For the population with medium uniqueness a negative lowest element (–.15) in S (n=40) was encountered in only one sample. For the population with strong uniqueness again a negative lowest element (–.08) in S (n=20) was encountered in only one sample. Hence, as expected, the PARAFAC solutions of the samples from populations with uniqueness showed hardly any contrast at all, whereas the PARAFAC solutions of the samples from populations without uniqueness did show substantial contrast.

It can be expected that the PARAFAC solutions from populations with uniqueness resemble the population more closely than solutions of samples without uniqueness. Congruence coefficients were computed between columns of B and C from PARAFAC and the corresponding columns in the population.

These congruence coefficients were also computed for PFNC, to see whether PFNC retrieves the population components better than PARAFAC. In case the four congruence coefficients are above .85 it will be said that PARAFAC retrieved the population components. For the samples from populations with strong and medium uniqueness, the smallest congruence value encountered for PARAFAC and PFNC was .98. For the three samples with n=20 drawn from the populations without uniqueness, PARAFAC never retrieved the second component and retrieved the first component only in two samples. The lowest congruence values encountered for the latter two cases were .68 and .68. The lowest congruence values encountered for the third sample were per component .78 and .51. For n=40 PARAFAC retrieved both components in the population without uniqueness in one sample and only retrieved the first component in the other two samples. The lowest congruence values encountered for the two cases where PARAFAC failed to retrieve the second population component were .81 and .82. For the six samples from the population without uniqueness the lowest congruence value encountered for PFNC was .98, hence PFNC did retrieve the population parameters in these cases.

In section 2.5 the stability of the PARAFAC components, for each of the 18 PARAFAC arrays, was assessed by conducting five separate splithalf analyses for each of the PARAFAC solutions. It was found that, in case of uniqueness in the population, the PARAFAC components are stable, and in case of non–uniqueness, the PARAFAC components are not stable. The samples drawn from the populations without uniqueness, yielding unstable PARAFAC components, were re–analyzed by the splithalf method in order to assess the stability of the PFNC components. Per splithalf analysis the same sets as in section 2.5 were used. The lowest congruence value encountered for the samples having n=40 was .94 and for the samples having n=20 was .86. Hence, the PFNC components are stable in the splithalf sense.

The above results suggest the following inferences. In case of uniqueness and positive manifold in the population, PARAFAC analysis of a sample yields a non-contrast solution which has parameters (nearly) equal to the population parameters even when the sample size is as low 20. However, in case of positive manifold and non-uniqueness in the population, PARAFAC may yield contrast components which are not stable and have parameters not equal to those of the population. For these cases, PFNC can be used as an alternative method yielding components which have no contrast, are stable in the splithalf sense, and correspond to parameters nearly equal to those of the population. In the latter case, PFNC can be seen as a method to correct for a sampling bias of PARAFAC.

4.5 Discussion

In chapter 1, Kruskal's sufficient condition for uniqueness was given and it has been argued that, in practice, this sufficient condition is always fulfilled. In practice, this condition is also always fulfilled for the PFNC components, and hence both sets of components are unique. In chapter 2 the degree of uniqueness was examined by comparing the discrepancy in fit of PARAFAC and PARAFAC with two proportional columns in one of the parameter matrices. Analogously, in order to examine the uniqueness of the PFNC components one may subject the matrix C, that is already constrained to have non–negative elements, to the additional constraint that the first two columns are proportional. An ALS algorithm to fit such a constrained PFNC method can easily be constructed because the PFNC algorithm can be used to update A and B and the NNLS algorithm can be used to update C by solving the regression problem in equation 2.4 subject to the non–negativity of C. It should be noted, however, that the rotational freedom that arises because there are two proportional columns in C is limited, because the non–negativity of S and B has to be maintained.

It has been illustrated that PARAFAC may yield degenerate solutions and that PFORTA may be used to avoid this. In addition, it has been illustrated that, in case of positive manifold data, PFORTA may yield contrast components, and it has been argued that such components are more difficult to interpret. For this reason, PFNC may be preferred over PFORTA. There is another reason for preferring PFNC over PFORTA in case one wants to represent positive manifold data by components that do allow an easy interpretation. Specifically, if all the variables over all frontal slabs are positively correlated and the variables fall apart into q groups, the orthonormal PFORTA components cannot coincide with the q 'centroids' of these groups of variables. Apart from PFNC, the methods proposed in the next chapter are particularly useful for such data sets.

CHAPTER 5

PARAFAC COMPONENTS THAT CORRESPOND TO NON-OVERLAPPING CLUSTERS OF VARIABLES

In many empirical studies, where n scores on m variables are obtained, the researcher has some idea about a possible clustering of the variables. This idea either comes from a study of the correlation matrix or from theoretical knowledge on the variables. A common situation is that the researcher has a hypothesis about the partitioning of the variables into certain non–overlapping clusters. In such a situation the data can be analyzed by the Multiple Group method (Guttman, 1952; Nunnally, 1978, pp. 394–400) or by a confirmatory factor analysis method (Jöreskog, 1969). In the Multiple Group method, the components are constructed by simply summing the variables that belong to a certain cluster. Next, the correlations between components and variables (structure), and the amount of explained variance are computed, in order to verify whether or not the data are represented satisfactorily. In confirmatory factor analysis, the researcher specifies which variables are to be associated with which factors. In case it is hypothesized that there are non–overlapping clusters of variables, each variable corresponds to exactly one factor. Next, the resulting model is fitted to the observed covariance matrix. Both methods can be seen as confirmatory methods.

A different situation is that where the researcher wants to see to what extent the variables can be clustered at all. To answer such an exploratory question, it is common practice to analyze the data by PCA or by Common Factor Analysis and to rotate the initial solution to simple structure.

In case of three-way data, the same two situations, one with a confirmatory and one with an exploratory research question, can be distinguished. Examples of three-way data where the researcher has an idea about the grouping of the variables, a confirmatory case, can be found in Kroonenberg (1983, p. 229) or Harshman and DeSarbo (1984, p. 605). The exploratory case does not seem to have been dealt with yet.

In case the variables are partitioned into certain clusters of variables, the PARAFAC method may identify components that do not correspond to these clusters of variables, as will be illustrated below. One way to obtain PARAFAC components that correspond to clusters of variables is by the following three-step procedure. In the first step, the data are analyzed by PARAFAC. In the second step, the pattern matrix B is rotated to simple structure according to the VARIMAX criterion. After rotating the pattern matrix B, the matrices A and C are not optimal given B. Therefore, in the third step, the rotated pattern matrix B is fixed and the PARAFAC loss function is minimized over A and C by using the corresponding steps of the PARAFAC algorithm. This three-step procedure may yield components that are simpler than the PARAFAC components but still do not exactly correspond to non-overlapping clusters of variables as is the case in Multiple Group Analysis and in Common Factor Analysis. In order to find PARAFAC components that do correspond to non-overlapping clusters of variables, a different procedure is proposed.

A way to obtain PARAFAC components that correspond to non-overlapping clusters of variables is by simply imposing this in a similar manner as in Multiple Group Analysis and confirmatory factor analysis. That is, if a hypothesis about the partitioning of the variables into non-overlapping clusters is available, then the PARAFAC parameters can be constrained such that the PARAFAC components correspond to these hypothesized clusters of variables. A first purpose of this chapter is to show how these constraints can be imposed on the PARAFAC parameter matrices. This yields a restricted PARAFAC method which will be called ParaFac with Clustered Variables (PFCV).

A second purpose of this chapter is to show how the PARAFAC parameters can be constrained such that PARAFAC can be used as an exploratory method for finding components that have simple structure. This restricted PARAFAC method is closely related to PFCV. That is, it finds those PARAFAC components that correspond to an optimal partitioning of the variables into non-overlapping clusters. This method will be called ParaFac with Optimally Clustered Variables (PFOCV).

5.1 The PFCV method and an algorithm

First, it will be explained how the PARAFAC parameters can be constrained such that PARAFAC identifies components that correspond to a hypothesized partitioning of the variables into non-overlapping clusters. In the first chapter it has been explained that the $m \times q$ pattern matrix B contains coefficients that are used to obtain linear combinations of the perfectly congruent components in order to optimally represent the variables. Like in PCA, the elements in B indicate how the variables are built from the components. To find simple components, it is proposed to constrain B such that a certain variable is represented only by the component to which it belongs according to the hypothesis of the researcher. Let an $m \times q$ matrix W contain zeroes and ones only, where $w_{jl}=1$ if variable j is assigned to cluster l, and else $w_{jl}=0$. Then the binary matrix W is an indicator matrix for the hypothesized partitioning of the variables into non-overlapping clusters. If the pattern matrix B is constrained to have zeroes in the cells where W has zeroes, that is, if $B=W*B$, where $*$ is the element-wise (Hadamard) product, then the elements of B will also correspond to the hypothesized partitioning of the variables into non-overlapping clusters. PFCV will be defined as the method that minimizes PARAFAC(A,B,C) subject to the constraint $B=W*B$, where W is considered fixed. PFCV(A,B,C) denotes the corresponding loss function.

One way to obtain the PFCV solution is by using the PARAFAC algorithm, as follows. First the data are partitioned into q three-way data sets corresponding to the q clusters of variables. Next, each one of these clusters is analyzed by means of PARAFAC with dimensionality 1. This yields the PFCV solution, because PFCV is equivalent to q separate one-dimensional PARAFAC analyses, which can be proven as follows. For convenience, the order of the columns in X_k, $k=1,...,p$, can be changed such that $X_k=(X_k^{(1)}|...|X_k^{(q)})$, where $X_k^{(l)}$ is an n by m_l matrix, $k=1,...,p$, containing the m_l variables of cluster l according to the hypothesis, $l=1,...,q$. We may partition the columns of B into q subvectors $\mathbf{b}_{rl}$ of order m_l, $r=1,...,q$, $l=1,...,q$. Then, from $B=W*B$ it follows that

$$B=\begin{pmatrix}\mathbf{b}_{11} & \mathbf{b}_{12} & .. & \mathbf{b}_{1q}\\ \mathbf{b}_{21} & \mathbf{b}_{22} & .. & \mathbf{b}_{2q}\\ \vdots & \vdots & & \vdots\\ \mathbf{b}_{q1} & \mathbf{b}_{q2} & .. & \mathbf{b}_{qq}\end{pmatrix}=\begin{pmatrix}\mathbf{b}_{11} & \mathbf{0} & .. & \mathbf{0}\\ \mathbf{0} & \mathbf{b}_{22} & .. & \mathbf{0}\\ \vdots & \vdots & & \vdots\\ \mathbf{0} & \mathbf{0} & .. & \mathbf{b}_{qq}\end{pmatrix}. \quad (5.1)$$

Substituting (5.1) for B into the PARAFAC loss function, it follows that

$$\text{PFCV}(A,B,C)=\sum_{k=1}^{p}\left\| X_k-(\mathbf{a}_1|...|\mathbf{a}_q)\begin{pmatrix}c_{k1} & 0 & .. & 0\\ 0 & c_{k2} & .. & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & & c_{kq}\end{pmatrix}\begin{pmatrix}\mathbf{b}'_{11} & \mathbf{0}' & ..\mathbf{0}'\\ \mathbf{0}' & \mathbf{b}'_{22} & ..\mathbf{0}'\\ \vdots & \vdots & \vdots\\ \mathbf{0}' & \mathbf{0}' & \mathbf{b}'_{qq}\end{pmatrix}\right\|^2$$

$$=\sum_{k=1}^{p}\|(X_k^{(1)}|...|X_k^{(q)})-(\mathbf{a}_1c_{k1}\mathbf{b}'_{11}|...|\mathbf{a}_qc_{kq}\mathbf{b}'_{qq})\|^2$$

$$=\sum_{l=1}^{q}\sum_{k=1}^{p}\|X_k^{(l)}-\mathbf{a}_lc_{kl}\mathbf{b}'_{ll}\|^2. \quad (5.2)$$

Clearly, minimizing (5.2) comes down to doing q one-dimensional PARAFAC analyses, one for each cluster of variables. Note that this implies that the PFCV components are unique even if the columns in A or C are proportional because a one dimensional PARAFAC solution is unique.

As explained above, the PFCV loss function could be minimized by applying the PARAFAC algorithm to the q subsets of three-way data. However, it is better to use a different algorithm that is more straightforward and uses less computation time. Specifically, PFCV will be minimized by alternatingly minimizing PFCV over A and C for given B, and over B for given A and C. For fixed B, the matrices A and C can be updated simultaneously, as follows. Note that, without loss of generality, the constraint $\text{Diag}(A'A)=\text{Diag}(C'C)=I_q$ may be imposed. From the PARAFAC loss function it can be seen that the problem is to minimize

$$f(A,C)=\sum_{k=1}^{p}\|X_k-AD_kB'\|^2=\left(\sum_{k=1}^{p}\mathrm{tr}X_k'X_k\right)-2\left(\sum_{k=1}^{p}\mathrm{tr}X_k'AD_kB'\right)+\left(\sum_{k=1}^{p}\mathrm{tr}BD_kA'AD_kB'\right), \quad (5.3)$$

subject to the constraint $\mathrm{Diag}(A'A)=\mathrm{Diag}(C'C)=I_q$. It should be noted that if $B=W^*B$, then B is orthogonal. For the third term on the right-hand side of (5.3), using the orthogonality of B, we have

$$\sum_{k=1}^{p}\mathrm{tr}BD_kA'AD_kB'=\mathrm{tr}\mathrm{Diag}(B'B)\sum_{k=1}^{p}D_kA'AD_k=\mathrm{tr}\mathrm{Diag}(B'B)\sum_{k=1}^{p}D_k\mathrm{Diag}(A'A)D_k$$

$$=\mathrm{tr}\mathrm{Diag}(B'B)\sum_{k=1}^{p}D_k^2=\mathrm{tr}\mathrm{Diag}(B'B)\mathrm{Diag}(C'C)=\|B\|^2. \quad (5.4)$$

From (5.4) it can be seen that the third term on the right of (5.3) is constant with respect to A and C. Hence, minimizing $f(A,C)$ subject to the constraint $\mathrm{Diag}(A'A)=\mathrm{Diag}(C'C)=I_q$ is equivalent to maximizing

$$g(A,C)=\mathrm{tr}\sum_{k=1}^{p}X_k'AD_kB'=\mathrm{tr}A'\sum_{k=1}^{p}X_kBD_k, \quad (5.5)$$

subject to the constraint $\mathrm{Diag}(A'A)=\mathrm{Diag}(C'C)=I_q$. From (5.5) it follows that

$$g(A,C)=\sum_{r=1}^{q}\mathbf{a}_r'\sum_{k=1}^{p}X_k\mathbf{b}_rc_{kr}=\sum_{r=1}^{q}\mathbf{a}_r'(X_1\mathbf{b}_r|\dots|X_p\mathbf{b}_r)\mathbf{c}_r, \quad (5.6)$$

where $\mathbf{a}_r$, $\mathbf{b}_r$ and $\mathbf{c}_r$ denote column r of A, B, and C, respectively. From (5.6) it can be seen that $g(A,C)$ is maximal if $\mathbf{a}_r$ and $\mathbf{c}_r$ are the first left and right hand singular vectors of the matrix $(X_1\mathbf{b}_r|\dots|X_p\mathbf{b}_r)$, respectively, for $r=1,\dots,q$. An efficient way to find these singular vectors is to define $Y_r\equiv(X_1\mathbf{b}_r|\dots|X_p\mathbf{b}_r)$ and find the first eigenvalue and eigenvector of $Y_r'Y_r$ by Hotelling's vector iteration method (Hotelling, 1933). Then $\mathbf{c}_r$ is the first eigenvector of $Y_r'Y_r$ and $\mathbf{a}_r=Y_r\mathbf{c}_r/\lambda_r$, where λ_r is the square root of the first eigenvalue of $Y_r'Y_r$.

For given A and C, PFCV can be minimized over B row by row. From $B=W*B$ it follows that $\mathbf{b}_j'$, row j of B, has only one non–zero element, say b_{jl}. From (5.1) it can be seen that, in order to update row j of B, the problem is to minimize

$$\mathrm{f}(\mathbf{b}_j)=\sum_{k=1}^{p}\|\mathbf{x}_{jk}-AD_k\mathbf{b}_j\|^2=\sum_{k=1}^{p}\|\mathbf{x}_{jk}-\mathbf{a}_l c_{kl} b_{jl}\|^2, \tag{5.7}$$

where $\mathbf{x}_{jk}$ is variable j at occasion k, and c_{kl} is element (k,l) of C. From $\mathbf{a}_l'\mathbf{a}_l=\mathbf{c}_l'\mathbf{c}_l=1$ and regression theory (Draper & Smith, 1981) it follows that the minimum of (5.7) over b_{jl} is attained for $b_{jl}=\sum_{k=1}^{p}c_{kl}\mathbf{x}_{jk}'\mathbf{a}_l$. Using this updating procedure for all rows of B amounts to updating B according to $B=\left(\sum_{k=1}^{p}X_k'AD_k\right)*W$.

An ALS algorithm that monotonically decreases PFCV(A,B,C) can be constructed as follows. After starting with arbitrary elements in B, it is proposed to iteratively perform two steps, that is, firstly, updating A and C simultaneously, and secondly, updating B, until a stable function value has been reached. This algorithm will be called PFCV algorithm. Various analyses of empirical data by the PFCV method revealed that the PFCV algorithm misses the global minimum only sporadically.

5.2 Partitioning the fit in PFCV

In (unconstrained) PARAFAC, it is not clear how much a certain component contributes to the fit of a certain variable, because the contributions of different components are mutually dependent. For the PFCV method, one can compute how much each component contributes to the fit of the data for a person i, a variable j, or an occasion k, as follows. Assume that A, B, and C minimize PFCV(A,B,C) and that A and C are scaled to unit length column-wise. First, consider the partitioning of the variables. The residual sums of squares (unexplained part) for variable j is given by (5.7). From the optimality of B it follows that $b_{jl}=\sum_{k=1}^{p} c_{kl}\mathbf{x}'_{jk}\mathbf{a}_l$. Upon substituting b_{jl} for $\sum_{k=1}^{p} c_{kl}\mathbf{x}'_{jk}\mathbf{a}_l$ in (5.7) and using $\mathbf{a}'_l\mathbf{a}_l=\mathbf{c}'_l\mathbf{c}_l=1$, it follows that $f(\mathbf{b}_j)=\sum_{k=1}^{p}\|\mathbf{x}_{jk}\|^2-b_{jl}^2$. Because $\sum_{k=1}^{p}\|\mathbf{x}_{jk}\|^2$ is the sum of squares to be explained for variable j, and $\sum_{k=1}^{p}\|\mathbf{x}_{jk}\|^2-b_{jl}^2$ is the unexplained part, it follows that b_{jl}^2 denotes the explained part, called the fit for variable j. Only component l contributes to the fit for variable j, given that variable j is assigned to component l.

Second, consider the fit for person i. If the columns of B and C are scaled to unit length column-wise, then a_{il}^2 denotes the sum of squares of person i that is explained by component l, where a_{il} denotes element (i,l) of A. This can be proven as follows. Let $\mathbf{a}'_i$ denote row i of A. From the fact that A is optimal it follows that $\mathbf{a}_i$ minimizes

$$g(\mathbf{a}_i)=\sum_{k=1}^{p}\|\mathbf{x}'_{ik}-\mathbf{a}'_iD_kB'\|^2=\sum_{k=1}^{p}\|\mathbf{x}_{ik}-BD_k\mathbf{a}_i\|^2=\left\|\begin{pmatrix}\mathbf{x}_{i1}\\ \vdots\\ \mathbf{x}_{ip}\end{pmatrix}-\begin{pmatrix}BD_1\\ \vdots\\ BD_p\end{pmatrix}\mathbf{a}_i\right\|^2, \tag{5.8}$$

where $\mathbf{x}'_{ik}$ denotes row i of X_k, k=1,...,p. From the orthogonality of B and the fact that B and C are scaled to unit length column-wise, it follows that $\sum_{k=1}^{p} D_k B'BD_k = I_q$. From this and the optimality of A, see equation (1.4), it follows that $\mathbf{a}_i = \sum_{k=1}^{p} D_k B'\mathbf{x}_{ik}$. It can be verified that

$$g(\mathbf{a}_i) = \sum_{k=1}^{p} \|\mathbf{x}_{ik}\|^2 - \mathbf{a}_i'\mathbf{a}_i = \sum_{k=1}^{p} \|\mathbf{x}_{ik}\|^2 - \sum_{l=1}^{q} a_{il}^2. \tag{5.9}$$

Hence, a_{il}^2 is the sum of squares for person i that is explained by component l, which had to be proven. Analogously, it can be derived that if A and B are scaled to unit length column-wise, then c_{kl}^2 denotes the contribution of component l to the fit for occasion k.

5.3 Interpretation of the PFCV components

As a basis for the interpretation of the PFCV components in terms of the variables, the pattern matrix B may be used. From the constraint $B = B*W$ it follows that the elements of component l correspond to the l^{th} hypothesized cluster of variables. Moreover, component l is a linear combination, over the occasions, of precisely those variables that belong to the l^{th} hypothesized cluster, which can be proven as follows. We are free to scale B and C to unit length column-wise. At the minimum of PFCV(A,B,C) A must be optimal given B and C, therefore $A = \sum_{k=1}^{p} X_k BD_k \left(\sum_{l=1}^{p} D_l B'BD_l \right)^{-1}$, see equation 1.4. From substituting $W*B$ for B in this equation and using the fact that $B'B = I_q$ and $\text{Diag}(C'C) = I_q$, we have

$$A = \sum_{k=1}^{p} X_k (W*B) D_k \left(\sum_{l=1}^{p} D_l^2 \text{Diag}(B'B) \right)^{-1} = \sum_{k=1}^{p} X_k (W*B) D_k. \tag{5.10}$$

From (5.10) and the fact that W is binary, it follows that the columns of A are linear combinations of non-overlapping clusters of variables.

In order to interpret the PFCV components, it is proposed to use the structure matrix $S=\sum_{k=1}^{p} X_k'AD_k$, in addition to the pattern matrix. The matrix B is related to S in a simple way. Assuming that A and C have unit length column-wise it follows that $B=\sum_{k=1}^{p} X_k'AD_k{*}W=S{*}W$, from which it is clear that the non-zero elements of B are equal to the corresponding elements in S. From the structure matrix it can be seen how each component covaries with all the variables over the occasions. The structure matrix in PFCV can be used to find disconfirmatory evidence regarding the existence of simple factors in the same way as the structure matrix in the Multiple Group method (Nunnally, 1978, pp. 394–400), as follows. First, the partitioning will be *incorrect* if the maximum of the squared elements of a row of S corresponds to a different component (cluster) than the component that corresponds to the non-zero element in B. Second, the partitioning will be *suspect* if a row of S has a squared element that is relatively high and corresponds to a different component (cluster) than the component that corresponds to the non-zero element in B.

5.4 A PFCV analysis of an empirical data set

To illustrate the PFCV method, the results of an analysis of the GOS data (see chapter 3) will be presented. Remember that the variables are divided into four groups: Simultaneous Processing (Magic Window, Face Recognition and Gestalt Closure), Achievement (Expressive Vocabulary and Faces and

Places), Sequential Processing (Hand Movements and Number Recall), and Motor Skills (Gross Motor Skills, Fine Motor Skills and Figure Movement in Disc). To study the relation between these four groups of variables, the W matrix was chosen such that the PFCV components correspond to these four groups of variables. Prior to the analysis, the variables were centered within the occasions and scaled to unit length over the occasions. These data were also analyzed by PARAFAC. The percentage of variance explained by the PARAFAC components and by the PFCV components was 61.4 and 55.9, respectively. This difference in percentages of explained variance by the two methods illustrates that the four-dimensional PARAFAC solution cannot be replaced by a simple representation in which each group of the variables belongs to one component. To study how the groups overlap, the structure matrix S and the component correlation matrix $(A'A)$, that resulted from the PFCV analysis, are depicted in Table 5.1. The elements in bold face correspond to non-zero elements of the matrix B.

From the structure matrix in Table 5.1 it can be seen that the variables of the first two groups have the highest correlations with the first two components and that the variables of the last two groups have the highest correlations with the last two components. Nearly all rows of the structure matrix have a relatively high element that corresponds to a different cluster. For this reason, this partitioning of the variables can be called suspect. In addition, it can be seen from the component correlations that both the first two components and the last two components correlate substantially. For these two reasons, the first two groups and the second two groups of variables were merged, and the GOS data were analyzed again by PARAFAC and PFCV, both with dimensionality 2.

The structure matrices that were found are depicted in Table 5.2. The elements in bold face correspond to non-zero elements of the matrix B.

Table 5.1 *The structure matrix and the component correlation matrix from PFCV analysis of the GOS data.*

	Structure Matrix			
Variable	I	II	III	IV
Magic Window	**.82**	.59	.25	.17
Face Recognition	**.66**	.50	.15	.24
Gestalt Closure	**.71**	.45	.17	.26
Expressive Vocabulary	.59	**.87**	.32	.26
Faces and Places	.59	**.83**	.33	.19
Hand Movements	.34	.39	**.80**	.52
Number Recall	.11	.28	**.83**	.47
Gross Motor Skills	.16	.15	.59	**.77**
Fine Motor Skills	.22	.22	.36	**.57**
Figure Movement in Disc	.21	.15	.14	**.52**

	Component Correlations			
	I	II	III	IV
I	1.00			
II	.71	1.00		
III	.26	.40	1.00	
IV	.30	.27	.61	1.00

From the structure matrices in Table 5.2 it can be seen that both the

PARAFAC and the PFCV components identify two clusters of variables. However, the PFCV components are purely composed of the two clusters of variables, whereas the PARAFAC components are based on *all* the variables. In addition, the second component from PARAFAC is a contrast component. For these two reasons, the PFCV components have a simpler interpretation than the PARAFAC components. The percentage of variance explained by PARAFAC and by PFCV was 47.5 and 46.6, respectively. It can be concluded that, for these variables, the PARAFAC representation can be replaced, at little cost, by a simple representation in which each group of the variables belongs to one component.

Table 5.2 *The structure matrices from PARAFAC and from PFCV analysis of the GOS data.*

	PARAFAC		PFCV	
Variable	I	II	I	II
Magic Window	.76	.48	**.78**	.23
Face Recognition	.61	.40	**.63**	.21
Gestalt Closure	.62	.36	**.63**	.23
Expressive Vocabulary	.78	.37	**.79**	.32
Faces and Places	.75	.40	**.76**	.29
Hand Movements	.49	−.33	.39	**.75**
Number Recall	.31	−.53	.21	**.75**
Gross Motor Skills	.28	−.54	.17	**.76**
Fine Motor Skills	.32	−.21	.23	**.51**
Figure Movement in Disc	.28	−.02	.19	**.31**

To assess the stability of the PFCV components, PFCV was subjected to the same 5 splithalf analyses as PARAFAC, see section 3.5. Three splithalf analyses revealed stable PFCV components, and two analyses revealed a stable first component only. The lowest congruence values for the unstable second components were .82 and .84. These results suggest that the PFCV components are almost stable in the splithalf sense. A comparison of the splithalf results for PFCV with those of PARAFAC in section 3.5 suggests that the PFCV components are less unstable than the PARAFAC components.
Until now, the variables were partitioned into q clusters on the basis of a theory derived hypothesis. In case such a hypothesis is absent, one might still want to see whether or not a simple structure of the variables can be obtained. An exploratory method to identify PARAFAC components that have simple structure will be presented in the next section.

5.5 PARAFAC with optimally clustered variables

PFOCV is defined as the method that finds the best PARAFAC components that are associated with a priori unknown non-overlapping clusters of variables. Specifically, for PFOCV, the problem is to minimize PARAFAC(A,B,C) over all matrices B that have exactly one non-zero element per row. In other words, the problem is to minimize PARAFAC(A,B,C) subject to the constraint that B=W*B, where W is to be identified such that W*B is optimal, and W is a matrix with one unit element per row and zeroes elsewhere. Let PFOCV(A,B,C) denote the corresponding loss function. In order to construct an algorithm for PFOCV, the problem is to find an update for B and W, for fixed A and C. If the matrices A and C are scaled to unit length column-wise, then

$$\text{PFOCV}(A,B,C)=\sum_{k=1}^{p}\|X_k-AD_k(B*W)'\|^2=\sum_{j=1}^{m}\sum_{k=1}^{p}\|\mathbf{x}_{jk}-\sum_{l=1}^{q}\mathbf{a}_l c_{kl} b_{jl} w_{jl}\|^2$$

$$=\sum_{j=1}^{m}\left(\sum_{k=1}^{p}\|\mathbf{x}_{jk}\|^2-2\sum_{l=1}^{q}\sum_{k=1}^{p}c_{kl}\mathbf{x}_{jk}'\mathbf{a}_l b_{jl} w_{jl}+\sum_{l=1}^{q}b_{jl}^2 w_{jl}\right). \qquad (5.11)$$

In case W is fixed and $w_{jl}=1$, it has been shown in section 5.1 that the optimal choice of b_{jl} is $b_{jl}=\sum_{k=1}^{p}c_{kl}\mathbf{x}_{jk}'\mathbf{a}_l$. If $w_{jl}=0$ then b_{jl} may be taken arbitrary. Upon substituting b_{jl} for $\sum_{k=1}^{p}c_{kl}\mathbf{x}_{jk}'\mathbf{a}_l$ in (5.11) it follows that

$$\text{PFOCV}(A,B,C)=\sum_{j=1}^{m}\left(\sum_{k=1}^{p}\|\mathbf{x}_{jk}\|^2-\sum_{l=1}^{q}b_{jl}^2 w_{jl}\right). \qquad (5.12)$$

This shows that, to minimize (5.12) over row j of B (and of W), $w_{jl}=1$ must be taken for the l such that b_{jl}^2 is the maximum of $\{b_{j1}^2,\ldots,b_{jq}^2\}$. So B can be updated first as $B=\sum_{k=1}^{p}X_k'AD_k$ (which corresponds to $b_{jl}=\sum_{k=1}^{p}c_{kl}\mathbf{x}_{jk}'\mathbf{a}_l$) and next all elements can be set equal to zero, except those that have the row-wise highest squared value (which corresponds to $W*B$). By starting with arbitrary elements in B, and next alternatingly updating A and C simultaneously according to the PFCV algorithm, and B according to the above, an algorithm has been constructed that monotonically decreases the PFOCV function, until the function value stabilizes.

This algorithm may yield, at some stage of the iterative process, a zero column in the matrix B, which would yield a suboptimal solution. In such a case, we can restore the full rank of the matrix B and at the same time decrease PFOCV(A,B,C), as follows. If b_{jl}^2 is the smallest squared element from the matrix B and $\mathbf{b}_r$ is the zero column in B then it is proposed to take $b_{jl}=0$, $\mathbf{a}_r$ and $\mathbf{c}_r$ equal to the first left and right singular vectors (say $\mathbf{u}$ and $\mathbf{v}$, respectively), of $(\mathbf{x}_{j1}|..|\mathbf{x}_{jp})$, and $b_{jr}=\mathbf{a}_r'(\mathbf{x}_{j1}|..|\mathbf{x}_{jp})\mathbf{c}_r$. It will be shown now that this procedure decreases the loss function. It will be assumed, without loss of generality, that $b_{jl}=b_{11}$, and $r=q$. First, note that changing $\mathbf{a}_q$ and $\mathbf{c}_q$ does not alter the value of PFOCV(A,B,C), since $\mathbf{b}_q$ is zero. By changing $\mathbf{a}_q$, $\mathbf{c}_q$, b_{11} and b_{1q}, into (5.3) and using that $\mathbf{b}_q=0$, we find

$$\text{PFOCV}(A,B,C)=\sum_{l=1}^{q}\sum_{k=1}^{p}\|X_k^{(l)}-\mathbf{a}_l c_{kl}\mathbf{b}_{ll}'\|^2=\sum_{l=1}^{q-1}\sum_{k=1}^{p}\|X_k^{(l)}-\mathbf{a}_l c_{kl}\mathbf{b}_{ll}'\|^2$$

$$=\sum_{l=2}^{q-1}\sum_{k=1}^{p}\|X_k^{(l)}-\mathbf{a}_l c_{kl}\mathbf{b}_{ll}'\|^2+\sum_{k=1}^{p}\|X_k^{(1)}-\mathbf{a}_1 c_{k1}\mathbf{b}_{11}'\|^2$$

$$=\sum_{l=2}^{q-1}\sum_{k=1}^{p}\|X_k^{(l)}-\mathbf{a}_l c_{kl}\mathbf{b}_{ll}'\|^2+\sum_{j=2}^{m}\sum_{k=1}^{p}\|\mathbf{x}_{jk}-\mathbf{a}_1 c_{k1}b_{j1}\|^2+\sum_{k=1}^{p}\|\mathbf{x}_{1k}-\mathbf{a}_1 c_{k1}b_{11}\|^2. \qquad (5.13)$$

Note that b_{11} occurs only in the last term on the right-hand side of (5.13), and that $\mathbf{a}_q$, $\mathbf{c}_q$ and b_{1q} do not play any role at all in (5.13). After replacing $\mathbf{a}_q$, $\mathbf{c}_q$, b_{11} and b_{1q} by $\mathbf{a}_q=\mathbf{u}$, $\mathbf{c}_q=\mathbf{v}$, and $b_{1q}=\mathbf{u}'(\mathbf{x}_{11}|..|\mathbf{x}_{1p})\mathbf{v}$ and $b_{11}=0$, it follows that

$$\text{PFOCV}(\tilde{A},\tilde{B},\tilde{C})=\sum_{l=2}^{q-1}\sum_{k=1}^{p}\|X_k^{(l)}-\mathbf{a}_l c_{kl}\mathbf{b}_{ll}'\|^2+\sum_{j=2}^{m}\sum_{k=1}^{p}\|\mathbf{x}_{jk}-\mathbf{a}_1 c_{k1}b_{j1}\|^2$$

$$+\sum_{k=1}^{p}\|\mathbf{x}_{1k}-\mathbf{a}_q c_{kq}b_{1q}\|^2, \qquad (5.14)$$

where $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ denote the adjusted values for A, B and C. The first two terms in (5.13) and (5.14) are equal, and for the third term in (5.13) and (5.14) we have

$$\sum_{k=1}^{p} \|\mathbf{x}_{1k}-\mathbf{a}_q c_{kq} b_{1q}\|^2 = \|(\mathbf{x}_{11}|..|\mathbf{x}_{1p})-\mathbf{a}_q b_{1q}\mathbf{c}_q'\|^2 \leq \|(\mathbf{x}_{11}|..|\mathbf{x}_{1p})-\mathbf{a}_1 b_{1q}\mathbf{c}_1'\|^2 = \sum_{k=1}^{p} \|\mathbf{x}_{1k}-\mathbf{a}_1 c_{k1} b_{11}\|^2, \quad (5.15)$$

where the inequality follows at once from the fact that $\mathbf{a}_q b_{1q}\mathbf{c}_q'$ is the best possible rank-one approximation for $(\mathbf{x}_{11}|..|\mathbf{x}_{1p})$, see Eckart and Young (1936). This proves that PFOCV$(\tilde{A},\tilde{B},\tilde{C}) \leq$ PFOCV(A,B,C), and hence that $\mathbf{a}_r$, $\mathbf{c}_r$ and $\mathbf{b}_j'$ can be updated according to the above without affecting monotonicity. This procedure has been implemented in the PFOCV algorithm.

5.6 Two examples of PFOCV of empirical data

The 34×10×2 GOS data were analyzed by the PFOCV method with dimensionality 2. The data were preprocessed as before. The PFOCV algorithm was run 20 times. This resulted in 12 local optima. The optimal run out of 20 resulted in exactly the same partitioning of the variables into non-overlapping clusters as the partitioning that was imposed by the second PFCV analysis (see Table 5.2). This illustrates that the partitioning which was found on the basis of theory is the best possible clustering. In order to assess the stability of the PFOCV components, the GOS data were subjected to the same splithalf analyses as in section 3.5 and 5.4. Two analyses revealed stable components whereas the other three analyses revealed a stable first component only. The lowest congruence values over B and C encountered for the unstable three second components were .82, .80 and .81. In five out of 10 splithalf analyses PFOCV yielded a different partition of the variables compared to the partition in Table 5.4. These results suggest that the PFOCV components are not stable in the splithalf sense.

To further illustrate the PFOCV method, an additional data set was analyzed by PARAFAC and PFOCV in two dimensions. The data consist of scores of 40 soups on five variables generated by 19 persons (Van der Burg, De Leeuw & Dijksterhuis, 1992). The variables are Spiciness, Saltiness, Amount of Vegetables, Firmness of Vegetables, and Thickness. Prior to the analysis, the variables were centered and scaled to unit length over the occasions. The PFOCV algorithm was run 20 times. No local optima were found. These data were also analyzed by PARAFAC. It was found that PARAFAC explains 24.0% and PFOCV explains 23.6% of the total variance. From this, it can be concluded that these data may be represented almost equally well by the PFOCV method. The structure and pattern matrices that both analyses revealed, are depicted in Table 5.3. The matrices A and C were scaled to unit length column-wise, and the elements in bold face correspond to non-zero elements of the matrix B.

From Table 5.3 it can be seen that the pattern matrix as well as the structure matrix found by PFOCV are simpler than the corresponding matrices found by PARAFAC. Typically, the non-zero elements in B from PFOCV are larger than the corresponding elements in B from PARAFAC. Specifically, the PFOCV analysis indicates that the variable Thickness does not belong to the first component, whereas the pattern matrix and the structure matrix from PARAFAC indicate that Thickness belongs to the first component almost as much as Saltiness belongs to the first component. For this reason, Thickness should play a part in the interpretation of the first PARAFAC component. However, the small discrepancy in fit between PARAFAC and PFOCV suggests that the PFOCV solution, with Thickness only in the second component, provides a satisfactory representation of the data as well.

Table 5.3 *The matrices S and B from PARAFAC and PFOCV analysis of the Soup data.*

	PARAFAC				PFOCV			
	S		*B*		*S*		*B*	
Variable	I	II	I	II	I	II	I	II
Spiciness	.44	−.17	.43	−.16	**.46**	−.05	.46	.00
Amount of Vegetables	.35	−.09	.35	−.08	**.36**	.00	.36	.00
Saltiness	.25	.01	.25	.01	**.25**	.04	.25	.00
Thickness	.20	.75	.23	.75	.02	**.78**	.00	.78
Firmness of vegetables	.40	−.14	.40	−.13	**.42**	−.02	.42	.00

Both analyses revealed that all but one of the person coefficients (in the matrix *C*) are positive. Therefore, all but one of the persons agree in the 'direction' to weight the components and only disagree in the amount by which they weight the components.

From the small discrepancy in fit between PARAFAC and PFOCV it can be concluded that PFOCV provides a satisfactory representation of the Soup data. From the structure matrix of PFOCV it can be concluded that the optimal partitioning of the variables into non-overlapping clusters of variables is not suspect.

To assess the stability of the PARAFAC and the PFOCV components, five separate splithalf analyses were conducted. In one splithalf analysis stable PARAFAC components were found. The lowest congruence values encountered for the four other splithalf analyses were .43, .74, .80, and .16. For PFOCV, using 20 runs per analysis, three splithalf analyses yielded stable components, and the other two had one stable component. For the unstable components the congruence values encountered were .78 and .81. For two sets, one from a sample suggesting stable components and one from a sample suggesting unstable components, the PFOCV solution yielded a different partitioning of the variables in non-overlapping clusters. These results suggest that neither the PARAFAC nor the PFOCV components are stable in the splithalf sense and that the PFOCV components are less unstable than the PARAFAC components.

5.7 Relations of PFCV and PFOCV with other methods

It can easily be seen how PFCV and PFOCV fit in the hierarchy of three-way methods given by Kiers (1991). PFCV can be seen as a constrained variant of PFOCV because the only difference between these two methods is that the binary matrix W is fixed in PFCV and free in PFOCV. In addition, the PFOCV method can be seen as a constrained variant of PARAFAC constrained to have an orthonormal B, denoted by PFORTB, which itself is a constrained variant of PARAFAC.

From (5.11) it follows that, in case p=1, the PFOCV method equals the method that Braverman (1970) and Escoufier (1988) proposed independently for components analysis of a single matrix. So PFOCV can be seen as a generalization of Braverman's and Escoufier's method.

5.8 Discussion

As noted in chapter 1, a degenerate solution can be avoided by analyzing the data by PFORTB. Due to the orthogonality of B, neither PFCV nor PFOCV can yield degenerate components.

In this chapter, two methods were presented to answer confirmatory and exploratory research questions pertaining to non-overlapping clusters of variables. These questions can also be raised for overlapping clusters of variables. A confirmatory and an exploratory method have been developed for PARAFAC analysis where the components correspond to overlapping clusters of variables. In the confirmatory method it is assumed that a hypothesis is available which states that certain pattern elements are zero. In the exploratory approach only the number of non-zero pattern elements needs to be specified by the user in advance, and next the optimal PARAFAC components that correspond to this number of non-zero pattern elements are determined. Both methods have been applied to various data sets, but examples where the corresponding components differ substantially from the Weighted PCA, PFNC and PFORTA components have not been found. For this reason, these methods will not be explained in detail.

CHAPTER 6

CONCLUSIONS

A first purpose of this chapter is to give some recommendations for the analysis of three-way data. In particular, an overview will be given of the methods to analyze three-way data that have been the subject of this study. It will be assumed that the data consist of scores of persons on variables at various occasions.
The second purpose of this chapter is to draw some conclusions. Specifically, degenerate components and uniqueness will be revisited.

6.1 Exploratory analysis of three-way data

After a preprocessing method is chosen, a three-way array can be analyzed. The hierarchical relations between (in order of increasing fit) SUMPCA, Weighted PCA, PFORTA, PARAFAC and TUCKALS3, can be used as follows. In case it is found that the fit of a more constrained variant is considerably lower than a less constrained variant, the more constrained variant is not appropriate to fit the data satisfactorily. For instance, in case the PARAFAC fit to the data is considerably lower than the TUCKALS3 fit, it can be concluded that PARAFAC is overly restrictive, and that certain interactions between the components, represented in the core array, are necessary to fit the data satisfactorily. For such a case one is referred to Kroonenberg (1983). In case the discrepancy in fit between PARAFAC and TUCKALS3 is negligible and the Weighted PCA fit is considerably less than the PARAFAC fit, one may represent the data by PARAFAC or by one of the constrained PARAFAC methods that have been the object of this study. On the other hand, in case Weighted PCA, PFORTA, PARAFAC and TUCKALS3 fit the data almost equally well, one may represent the data by Weighted PCA.

To illustrate the use of the above hierarchical relations, the Affective response data were analyzed by SUMPCA, Weighted PCA, PFORTA, PARAFAC and TUCKALS3 with dimensionality 2. The percentages of variance explained are: 8.0, 42.9, 44.3, 46.5, 46.5, respectively. In Chapter 2 and 3 it has been reported that PARAFAC yields a degenerate solution for these data. Two clear conclusions can be drawn. First, the SUMPCA method is overly restrictive, and second, there is no need to consider certain interactions between the components revealed in the TUCKALS3 core array. These fit values suggest to represent the Affective response data by Weighted PCA or by PFORTA, see section 3.6.

As noted above, in case the discrepancy in fit between PARAFAC and TUCKALS3 is negligible and the Weighted PCA fit is considerably lower than the PARAFAC fit, one may want to represent the data by a constrained PARAFAC method. A number of constrained PARAFAC methods, including those discussed in the present study, are listed in Table 6.1. The rows correspond to the methods, and the columns correspond to the parameter matrices. In each cell the constraint that is imposed on a parameter matrix is depicted. An empty cell means that the corresponding method is not constrained in terms of the corresponding parameter matrix.

Table 6.1 *A schematic representation of the methods discussed in this study.*

Methods	A	B	C
PFORTA	Orthonormal		
PFORTB		Orthonormal	
PFORTAB	Orthonormal	Orthonormal	
PARAFAC prop. col.			Two Proportional columns
SUMPCA			Equal rows
Weighted PCA			All columns proportional
PFNC	$X_k'A$ non-neg.	Non-negative	Non-negative
PFCV		$W*B=B$, W fixed	
PFOCV		$W*B=B$, W free	

Table 6.1 summarizes the constrained methods that have been mentioned in the previous chapters. It can be seen that four types of constraints were used: orthonormality (see PFORTA), lower rank (see Weighted PCA), non-negativity (see PFNC) and constraining elements to zero (see PFOCV). In some methods, more than one parameter matrix is subjected to one (type of) constraint, see, for example, PFORTAB or PFNC. By simply putting other combinations of constraints on the parameter matrices, still other PARAFAC methods with different properties can be developed. An example of this is PARAFAC subject to the constraints that A, B, and C are all non-negative, see Carroll, De Soete and Pruzansky (1989). One may also want to combine certain constraints on one parameter matrix. For example, in PFCV one may

subject B to the additional constraint that it has no negative elements. Another example of two different constraints, resting on one of the parameter matrices, is to subject A in PFNC to the additional constraint that it has orthonormal columns. Such constrained PARAFAC methods may turn out useful in the future.

6.2 Various approaches to cope with degenerate solutions

In Chapter 1 the degeneracy problem was described and Harshman and Lundy's (1984b, p. 274) solution was mentioned: Imposing a column-wise orthonormality constraint on one of the parameter matrices. In addition to these solutions, it has been demonstrated that Weighted PCA, PFNC, PFCV and PFOCV yield non-degenerate solutions as well.

6.3 Examining the degree of uniqueness by constrained PARAFAC

In chapter 1 the uniqueness of the PARAFAC components was introduced as the most salient property of the PARAFAC method. In chapter 2 it was explained how the degree of PARAFAC uniqueness can be examined by constrained PARAFAC. Specifically, it appeared that by analyzing samples from populations having increasing degrees of uniqueness, increasing amounts of discrepancy were encountered between the PARAFAC fit and the Weighted PCA fit. In chapters 3, 4, and 5 it has been found for various empirical data sets that the fit of constrained PARAFAC is close to that of unconstrained PARAFAC, and that the constrained components allow for an easier (and thus a different) interpretation. It is clear that the existence of such an alternative PARAFAC representation indicates that the uniqueness of the unconstrained PARAFAC components is weak. So each of the constrained PARAFAC methods can detect weak uniqueness of the unconstrained PARAFAC components.

6.4 Representing three-way data by unconstrained and constrained PARAFAC

In case of uniqueness in the population, PARAFAC retrieved the population parameters in the sample, and its components were stable in the splithalf sense, even with relatively small sample sizes. Therefore, it seems that, for empirical data where PARAFAC shows strong uniqueness, PARAFAC is a useful method for the exploratory analysis of three-way data.

On the other hand, in case of weak uniqueness in the sample, the PARAFAC components were unstable, and PARAFAC failed to retrieve the population parameters, and its components contained contrast, whereas the population showed positive manifold. Therefore, it seems that, for empirical data where PARAFAC shows weak uniqueness, PARAFAC is less useful. In case of weak uniqueness, PFNC did retrieve the population parameters and its components were stable in the splithalf sense. By the results of analyzing various empirical three-way arrays it was demonstrated that, in case of weak uniqueness, alternative (constrained) PARAFAC components exist, which fit the data almost equally well as PARAFAC, allow for an easier interpretation, and are less unstable or even stable in the splithalf sense. It seems that for empirical data, where PARAFAC shows weak uniqueness, constrained PARAFAC is useful for the exploratory analysis of three-way data.

APPENDIX A

Van IJzendoorn and Kroonenberg (1990) used the TUCKALS3 algorithm for Weighted PCA with $q_1=q_2=q$ and $q_3=1$. In the TUCKALS3 algorithm the order of updating the matrices in one iterative cycle implicitly is: G, A, G, B, G, C (see Kiers et al., 1992). It has to be proven that the TUCKALS3 algorithm produces the same series of function values as the Weighted PCA algorithm, if the same starting configurations are used.

Let A^i, B^i, $\mathbf{c}^i$, and G^i denote the i^{th} updates in TUCKALS3 for A, B, $\mathbf{c}$, and G, respectively, and let A^0, B^0, and $\mathbf{c}^0$ denote starting values for A, B and $\mathbf{c}$. According to Kiers et al. (1992) updates are found as follows. The first update of G is $G^1=A^{0\prime}\sum_{k=1}^{p}c_k^0X_kB^0$, the first update of A is

$$A^1=\mathrm{GS}\left(\sum_{k=1}^{p}c_k^0X_kB^0G^{1\prime}\right), \tag{A.1}$$

and the next update for G is $G^2=A^{1\prime}\sum_{k=1}^{p}c_k^0X_kB^0$. Next, by updating B it is found that

$$B^1=\mathrm{GS}\left(\sum_{k=1}^{p}c_k^0X_k'A^1G^2\right). \tag{A.2}$$

For convenience, let $Z^1\equiv\left(\sum_{k=1}^{p}c_k^0X_k'A^1\right)$ and $W^1\equiv\mathrm{GS}(Z^1)$. Then B^1 and Z^1 span the same column space, hence $B^1=W^1V^1$ for a certain rotation matrix V^1. Next, the third update of G satisfies $G^3=A^{1\prime}\sum_{k=1}^{p}c_k^0X_kB^1$. Using that $B^1=W^1V^1$, it is found that $G^3=Z^{1\prime}W^1V^1$. Hence, for the TUCKALS3 representation of X_k, it follows that

$$\hat{X}_k \equiv c_k^0 A^1 G^3 B^{1\prime} = c_k^0 A^1 Z^{1\prime} W^1 V^1 V^{1\prime} W^{1\prime} = c_k^0 A^1 Z^{1\prime} W^1 W^{1\prime}, \tag{A.3}$$

k=1,..,p. We may simplify this representation by noting that W^1 is a projector for the column space of Z^1, so that $Z^1 = W^1 W^{1\prime} Z^1$. By substituting Z^1 for $W^1 W^{1\prime} Z^1$ into (A.3) it is found that

$$\hat{X}_k = c_k^0 A^1 G^3 B^{1\prime} = c_k^0 A^1 Z^{1\prime}. \tag{A.4}$$

The first update of $\mathbf{c}$ satisfies $\mathbf{c}^1 = \text{GS}(\mathbf{d}^1)$, with $d_k^1 = \text{tr} A^{1\prime} X_k B^1 G^{3\prime}$, where d_k^1 denotes element k of $\mathbf{d}^1$. From (A.4) it follows that

$$d_k^1 = \text{tr} A^{1\prime} X_k B^1 G^{3\prime} = \text{tr} X_k^{\prime} A^1 G^3 B^{1\prime} = \text{tr} X_k^{\prime} A^1 Z^1, \tag{A.5}$$

which completes the first full cycle of updatings in TUCKALS3.

In order to show that the Weighted PCA algorithm produces the same function value after one complete cycle, it is convenient to express the Weighted PCA loss function as

$$\text{WPCA}(E,F,\mathbf{h}) = \sum_{k=1}^{p} \| X_k - h_k E F' \|^2, \tag{A.6}$$

where E plays the role of A, F the role of B and $\mathbf{h}$ the role of $\mathbf{c}$. Let E^i, F^i and $\mathbf{h}^i$ denote the i^{th} update of E, F and $\mathbf{h}$, respectively, and let starting configurations be given such that $F^0 = B^0 T^0$ and $\mathbf{h}^0 = \mathbf{c}^0$, where T^0 denotes an arbitrary non-singular matrix. Specifically, it will be proven that, after updating E and F, the representation $\hat{X}_k \equiv h_k^1 E^1 F^{1\prime}$ for X_k from the Weighted PCA algorithm equals the representation $\hat{X}_k = c_k^1 A^1 G^3 B^{1\prime}$ from one cycle of the TUCKALS3 algorithm, and hence $\text{TUCKALS}(A^1,B^1,\mathbf{c}^1,G^3)=\text{WPCA}(E^1,F^1,\mathbf{h}^1)$.

In the Weighted PCA algorithm (see section 3.3), the first update of E satisfies $E^1=\text{GS}\left(\sum_{k=1}^{p} h_k^0 X_k F^0\right)$. By substitution of B^0T^0 for F^0, and $\mathbf{c}^0$ for $\mathbf{h}^0$ into the first update for E and by using the fact that $\left(\sum_{k=1}^{p} c_k^0 X_k B^0 T^0\right)$ and A^1 span the same column space, we find

$$E^1=\text{GS}\left(\sum_{k=1}^{p} c_k^0 X_k B^0 T^0\right)=A^1S^1, \tag{A.7}$$

where S^1 denotes a certain rotation matrix. The first update of F satisfies $F^1=\sum_{k=1}^{p} h_k^0 X_k' E^1$, see section 3.3. Now substituting A^1S^1 for E^1, $\mathbf{c}^0$ for $\mathbf{h}^0$ and Z^1 for $\left(\sum_{k=1}^{p} c_k^0 X_k' A^1\right)$ into the first update of F, yields

$$F^1=\sum_{k=1}^{p} c_k^0 X_k' A^1 S^1=Z^1S^1=B^1T^1, \tag{A.8}$$

for a certain non-singular matrix T^1. Hence, it follows from (A.7), (A.8) and the substitution of $\mathbf{c}^0$ for $\mathbf{h}^0$ into $h_k^0E^1F^{1\prime}$ that $h_k^0E^1F^{1\prime}=c_k^0A^1S^1S^{1\prime}Z^{1\prime}=c_k^0A^1Z^{1\prime}$, which equals (A.4).

Now it will be shown that $\mathbf{c}^1=\mathbf{h}^1$. The first update of $\mathbf{h}$ satisfies $\mathbf{h}^1=\text{GS}(\mathbf{g}^1)$, with $g_k^1=\text{tr}E^{1\prime}X_kF^1=\text{tr}X_k'E^1F^{1\prime}$, where g_k^1 denotes element k of $\mathbf{g}^1$. From substituting $A^1Z^{1\prime}$ for $E^1F^{1\prime}$ into $g_k^1=\text{tr}X_k'E^1F^{1\prime}$, it follows that $g_k^1=\text{tr}X_k'A^1Z^{1\prime}$, which equals (A.5). This proves that $\mathbf{c}^1=\mathbf{h}^1$, and therefore $h_k^1E^1F^{1\prime}=c_k^1A^1Z^{1\prime}=c_k^1A^1G^3B^{1\prime}$. Thus, after the first complete cycle, the same representations are found by the two algorithms and hence the TUCKALS3 algorithm and the Weighted PCA algorithm produce the same function values after one complete cycle. In passing, it has been shown that after the first iterative cycle it is true that $F^1=B^1T^1$ (A.8) and $\mathbf{h}^1=\mathbf{c}^1$. Hence, using exactly the same reasoning as above, it can be shown that, after the second iterative cycle, TUCKALS(A^2,B^2,$\mathbf{c}^2$,G^6)=WPCA(E^2,F^2,$\mathbf{h}^2$), and, more generally, that after the i^{th} iterative cycle TUCKALS(A^i,B^i,$\mathbf{c}^i$,G^{3i})=WPCA(E^i,F^i,$\mathbf{h}^i$), which completes the proof.

APPENDIX B

Generalized Perron–Frobenius Theorem: Let $\mathbf{x}_{jt}$ denote column j of the $n{\times}m$ frontal slice t of the $n{\times}m{\times}p$ three-way array X. Let it be given that

$$\mathbf{x}_{jt}'\mathbf{x}_{is} \geq 0, \text{ for } i=1,...,m, \ j=1,...,m, \ s=1,...,p, \text{ and } t=1,...,p. \tag{B.1}$$

Let $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ denote the parameter vectors that minimize the PARAFAC loss function for dimensionality 1. Then the elements of $X_s'\mathbf{a}$, $s=1,...,p$, $\mathbf{b}$ and $\mathbf{c}$ can be taken to have no negative elements.

Proof: It will first be proven that $\mathbf{b}$ and $\mathbf{c}$ can be taken non-negative. It can always be arranged that $\mathbf{b}'\mathbf{b}=\mathbf{c}'\mathbf{c}=1$. Clearly, $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ minimize

$$\mathrm{f}(\mathbf{a},\mathbf{b},\mathbf{c})=\sum_{s=1}^{p}\|X_s-\mathbf{a}c_s\mathbf{b}'\|^2. \tag{B.2}$$

From this and the optimality of $\mathbf{a}$, it follows that $\mathbf{a}=\sum_{t=1}^{p}c_tX_t\mathbf{b}$, see equation (1.4). Upon substituting $\sum_{t=1}^{p}c_tX_t\mathbf{b}$ for $\mathbf{a}$ in (B.2), it is found that

$$\mathrm{g}(\mathbf{b},\mathbf{c})=\mathbf{b}'\left(\sum_{s=1}^{p}c_sX_s\right)'\left(\sum_{t=1}^{p}c_tX_t\right)\mathbf{b}, \tag{B.3}$$

is maximal with $\mathbf{b}'\mathbf{b}=\mathbf{c}'\mathbf{c}=1$. From (B.3) and (B.1) it follows that

$$\mathrm{g}(\mathbf{b},\mathbf{c})=\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{s=1}^{p}\sum_{t=1}^{p}b_ib_jc_sc_t\mathbf{x}_{jt}'\mathbf{x}_{is}\leq\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{s=1}^{p}\sum_{t=1}^{p}|b_i|\,|b_j|\,|c_s|\,|c_t|\,\mathbf{x}_{jt}'\mathbf{x}_{is}, \tag{B.4}$$

where $|.|$ denotes the absolute value of (.). From (B.4) it can be seen that reflection of negative elements in **b** or **c** does not decrease the value of g(**b**,**c**); hence **b** and **c** can be chosen to have non-negative values only.
Next, after substituting $\sum_{t=1}^{p} c_t X_t \mathbf{b}$ for **a** into $X_s'\mathbf{a} = \sum_{t=1}^{p} c_t X_s' X_t \mathbf{b}$ we find that $X_s'\mathbf{a}$, $s=1,\ldots,p$, is non-negative, because (B.1) is fulfilled and **b** and **c** can be taken non-negative, which completes the proof. It seems worthwhile noting that the above result generalizes the Perron-Frobenius theorem and that it can be generalized to the analysis of a so-called N-way array.

REFERENCES

Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality.* Doctoral dissertation, University of Groningen.

Braverman, E.M. (1970). Methods for the extremal grouping of parameters and the problem of determining essential factors. *Automation and Remote Control*, 1, 108–116.

Brogden, H.E. (1969). Pattern, structure, and the interpretation of factors. *Psychological Bulletin*, 72, 375–378.

Carroll, J.D., & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35, 283–319.

Carroll, J.D., De Soete, G, & Pruzansky, S (1989). Fitting of the latent class model via iteratively reweighted least squares CANDECOMP with non-negativity constraints (463–472). In R. Coppi and S. Bolasco (Eds.), *Analysis of multiway data matrices* (pp. 463–472). Amsterdam: North Holland.

Cattell, R.B. (1944). "Parallel proportional profiles" and other principles for determining the choice of factors by rotation. *Psychometrika*, 9, 267–283.

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, 31, 33–42.

Draper, N.R., & Smith, H. (1981). *Applied regression analysis.* New York: Wiley.

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1, 211–218.

Eckblad, G. (1981). Assimilation resistance and affective response in problem solving. *Scandinavian Journal of Psychology*, 1981, 22, 1–16.

Escoufier, Y. (1988). Beyond correspondence analysis. In: H.H. Bock (Ed.) *Classification and related methods of data analysis* (pp. 505–514). Amsterdam: North Holland.

Gorsuch, R,L. (1983). *Factor analysis*. Philadelphia: Saunders.

Guttman, L. (1952). Multiple group methods for common-factor analysis: Their basis, computation, and interpretation. *Psychometrika*, 17, 209–222.

Harman, H.H. (1976). *Modern factor analysis* (3rd. ed.). Chicago: University of Chicago Press.

Harris, C.W., & Kaiser, H.F. (1964). Oblique factor analytic solutions by orthogonal transformations. *Psychometrika*, 29, 347–362.

Harshman, R.A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "exploratory" multi-modal factor analysis. *UCLA working papers in Phonetics* 16, 1–84.

Harshman, R.A. (1972). Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA working papers in Phonetics* 22, 111–117.

Harshman, R.A., & DeSarbo, W.S. (1984). An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques. In H.G. Law, C.W. Snyder, J.A. Hattie, & R.P. McDonald (Eds.), *Research methods for multi-mode data analysis* (pp. 602–642). New York: Praeger.

Harshman, R.A., Ladefoged, P. & Goldstein, L. (1977). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America*, 62, 693–707.

Harshman, R.A., & Lundy, M.E. (1984a). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H.G. Law, C.W. Snyder, J.A. Hattie, & R.P. McDonald (Eds.), *Research methods for multi-mode data analysis* (pp. 122–215). New York: Praeger.

Harshman, R.A., & Lundy, M.E. (1984b). Data preprocessing and the extended PARAFAC model. In H.G. Law, C.W. Snyder, J.A. Hattie, & R.P. McDonald (Eds.), *Research methods for multi-mode data analysis* (pp. 216–284). New York: Praeger.

Haven, S., & Ten Berge, J.M.F. (1977). Tucker's coefficient of congruence as a measure of factorial invariance: An empirical study. *Heymans Bulletin* 290 EX. Department of Psychology, University of Groningen.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417–441.

Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202.

Kaiser, H.F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187–200.

Kiers, H.A.L. (1991). Hierarchical relations among three-way methods. *Psychometrika*, 56, 449–470.

Kiers, H.A.L., & Krijnen, W.P. (1991). An efficient algorithm for PARAFAC of three-way data with large numbers of observation units. *Psychometrika*, 56, 1, 147–152.

Kiers, H.A.L., Kroonenberg, P.M. & Ten Berge, J.M.F. (1992). An efficient algorithm for TUCKALS3 on data with large numbers of observation units. *Psychometrika*, 57, xxx–xxx.

Krijnen, W.P., & Ten Berge, J.M.F. (1992). A constrained PARAFAC method for positive manifold data. *Applied Psychological Measurement*, 16, xxx–xxx.

Kroonenberg, P.M. (1983). *Three-mode principal component analysis: Theory and applications*. Leiden: DSWO Press.

Kroonenberg, P.M. & De Leeuw, J. (1980). Principal components analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45, 69–97.

Kroonenberg, P.M., Ten Berge, J.M.F., Brouwer, P., & Kiers, H.A.L. (1989). Gram-Schmidt versus Bauer-Rutishauser in alternating least-squares algorithms for three-mode principal component analysis. *Computational Statistics Quarterly* 2, 81–87.

Kruskal, J.B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Linear Algebra and its Applications*, 18, 95–138.

Kruskal, J.B. (1984). Multilinear methods. In H.G. Law, C.W. Snyder, J.A. Hattie, & R.P. McDonald (Eds.), *Research methods for multi-mode data analysis* (pp. 36–62). New York: Praeger.

Kruskal, J.B. (1989). Rank, decomposition, and uniqueness for 3-way and n-way arrays. In R. Coppi and S. Bolasco (Eds.), *Analysis of multiway data matrices* (pp. 7–18). Amsterdam: North Holland.

Kruskal, J.B., Harshman, R.A., & Lundy, M.E. (1989). How 3–MFA data can cause degenerate PARAFAC solutions, among other relationships. In R. Coppi and S. Bolasco (Eds.), *Analysis of multiway data matrices* (pp. 115–122). Amsterdam: North Holland.

Lawson, C.L., & Hanson, R.J. (1974). *Solving least squares problems*. New Jersey: Prentice–Hall.

Lundy, M.E., & Harshman, R.A. (1985). *Reference manual for the PARAFAC analysis package*. London, Ontario: Scientific Software Associates.

Mulaik, S.A. (1972). *The foundations of factor analysis*. New York: McGraw–Hill.

Nunnally, J.C. (1978). *Psychometric theory*. New York: McGraw–Hill.

Osgood, C.E. & Luria, Z. (1954). A blind analysis of a case of multiple personality. *Journal of Abnormal and Social Psychology*, 49, 579–791.

Ten Berge, J.M.F. (1986). *Three notes on three-way analysis*. Paper presented at the workshop on TUCKALS and PARAFAC, University of Leiden, July 2.

Ten Berge, J.M.F. (1989). Convergence of Parafac preprocessing procedures and the Deming–Stephan method of iterative proportional fitting. In R. Coppi and S. Bolasco (Eds.), *Analysis of multiway data matrices* (pp. 53–64). Amsterdam: North Holland.

Ten Berge, J.M.F., De Leeuw, J. & Kroonenberg, P.M. (1987). Some additional results on principal components analysis of three–mode data by means of alternating least squares algorithms. *Psychometrika*, 52, 183–191.

Ten Berge, J.M.F., Kiers, H.A.L., & De Leeuw, J. (1988). Explicit CANDECOMP/PARAFAC solutions for a contrived 2x2x2 array of rank three. *Psychometrika*, 53, 579–584.

Tenenhaus, M. (1988). Canonical analysis of two polyhedral cones and applications. *Psychometrika*, 53, 503–524.

Thurstone, L.L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of mind.* Chicago: The University of Chicago Press.

Tucker, L.R. (1951). *A method for synthesis of factor analysis studies* (Personnel Research Section Report No. 984). Washington, D.C.: Department of the Army.

Van der Burg, E., De Leeuw, J. & Dijksterhuis, G.B. (1992). OVERALS: Nonlinear canonical correlation with k sets of variables. To appear in: R. Coppi & A. Di Ciaccio (Eds.), *Special Issue of Computational Statistics and Data Analysis devoted to Multiway Data analysis, software and applications.*

Van der Kloot, W.A. & Kroonenberg, P.M. (1985). External analysis with three-mode principal component models. *Psychometrika*, 50, 479–494.

Van Eldik, M.C.M., Neutel, R.J., Van der Meulen, B.F. & Spelberg, H.C.L. (1990). *Groningse Ontwikkelingsschalen. Constructiefase: Opzet ́ en resultaten.* Internal Report, Department of Special Education, University of Groningen.

Van IJzendoorn, M.H. & Kroonenberg, P.M. (1990). Cross-cultural consistency of coding the strange situation. *Infant Behavior And Development*, 13, 469–485.

DE ANALYSE VAN DRIE-WEG GEGEVENS DOOR GERESTRICTEERDE PARAFAC METHODEN

Voor de exploratieve analyse van tweeweg-gegevens, bijvoorbeeld scores van personen op variabelen, is Principale Componenten Analyse (PCA) een nuttige methode. Met PCA kunnen dergelijke gegevens samengevat (gerepresenteerd) worden in een aantal componenten die een maximale hoeveelheid variantie verklaren. PCA komt neer op het ontbinden van de matrix met gegevens in twee matrixen: de matrix met coëfficiënten van de personen op de componenten en de matrix met coëfficiënten van de variabelen op de componenten (patroonmatrix). Indien drieweg-gegevens, bijvoorbeeld scores van personen op variabelen gemeten op een aantal tijdstippen (tijdstippen-matrix), beschikbaar zijn, lijkt PARAFAC (Harshman, 1970), een afkorting voor PARAllele FACtor-analyse, een nuttige methode voor de exploratieve analyse van dergelijke gegevens. In dit proefschrift worden gerestricteerde PARAFAC-methoden ontwikkeld om enerzijds de mate van uniciteit vast te stellen en anderzijds een fraaiere PARAFAC-representatie van de gegevens te vinden.

In hoofdstuk 1 wordt een overzicht gegeven van de belangrijkste eigenschappen van PARAFAC. Er wordt uitgelegd dat PARAFAC opgevat kan worden als een generalisatie van PCA en dat PARAFAC neerkomt op het ontbinden van het array met drieweg-gegevens in drie matrixen: de matrix met coëfficiënten van de personen op de componenten (componentenmatrix), de matrix met coëfficiënten van de variabelen op de componenten (patroonmatrix) en de matrix met coëfficiënten van de tijdstippen op de componenten (tijdstippen–matrix). De PARAFAC–representatie van de gegevens heeft als eigenschap dat de componenten, over de tijdstippen heen, variëren in belangrijkheid. Een van de opmerkelijke eigenschappen van PARAFAC is de uniciteit van de PARAFAC–componenten. Hiermee wordt bedoeld dat rotatie van de PARAFAC–componenten tot een slechtere fit leidt en om deze reden onwenselijk is. Er worden voldoende voorwaarden voor uniciteit besproken en er wordt uitgelegd dat deze bij de analyse van empirische drieweg–gegevens altijd vervuld zijn. Tevens wordt er met een aantal theoretische voorbeelden aangetoond dat de mate van uniciteit van de PARAFAC componenten gering kan zijn. In een dergelijk geval wordt van zwakke uniciteit gesproken. Als de uniciteit van de PARAFAC–componenten zwak is dan leidt rotatie van, op zijn minst, twee PARAFAC–componenten nauwelijks tot een slechtere fit. Afgezien van het geval van zwakke uniciteit, kan men zich in het algemeen afvragen of er PARAFAC–componenten bestaan die eenvoudiger te interpreteren zijn en daarmee een fraaiere representatie van de gegevens vormen. In dit proefschrift wordt, zowel voor het vastellen van zwakke uniciteit als voor het vinden van een fraaiere PARAFAC–representatie van de gegevens, een oplossing gezocht in het opleggen van bepaalde randvoorwaarden aan de PARAFAC–methode. Tevens wordt via split–half analyse onderzocht in welke mate de (gerestricteerde) PARAFAC–componenten stabiel zijn.

Hoofdstuk 2 heeft als doel het vaststellen van zwakke uniciteit. Er wordt aangetoond dat het hebben van twee proportionele kolommen in één van de drie PARAFAC–matrixen voldoende is voor rotatievrijheid, en dus voldoende is voor de afwezigheid van uniciteit. Er wordt een algoritme afgeleid dat het PARAFAC model fit aan de gegevens onder de restrictie dat één van de drie matrixen een proportioneel kolommenpaar heeft. Aan de hand van analyses van empirische drieweg-gegevens wordt geïllustreerd dat de fit van deze gerestricteerde variant van PARAFAC vrijwel even goed kan zijn als de fit van (ongerestricteerde) PARAFAC. In een dergelijk geval wordt geconcludeerd dat de uniciteit van de PARAFAC componenten zwak is.

In Hoofdstuk 3 wordt nagegaan of componenten met dezelfde relatieve belangrijkheid over de tijdstippen een fraaie representatie van de gegevens kunnen vormen. Hiervoor wordt het idee van proportionele kolommen in de tijdstippen-matrix verder uitgewerkt. Er wordt aangetoond dat PARAFAC onder de restrictie van proportionele kolommen in de tijdstippen-matrix neerkomt op PCA van een gewogen som van gegevens-matrixen per tijdstip (Gewogen PCA). Gewogen PCA past in een hiërarchie van methoden om drieweg-gegevens te analyseren. Er wordt een efficiënt algoritme voor Gewogen PCA afgeleid. Aan de hand van een analyse van empirische gegevens wordt geïllustreerd hoe rotatievrijheid bij Gewogen PCA benut kan worden om simple structure te benaderen.

Hoofdstuk 4 heeft als doel het analyseren van zogenaamde positive manifold gegevens. Hiermee worden (drieweg-) gegevens bedoeld waarbij de variabelen onderling positief correleren. Het hoofdstuk begint met een demonstratie van de rol van rotatie-vrijheid bij PCA van dergelijke gegevens. Na rotatie ontstaan er componenten die contrastvrije componenten genoemd kunnen worden, omdat de matrix met correlaties tussen de variabelen en de componenten per component geen teken-contrast bevat. Vervolgens wordt met een voorbeeld geïllustreerd dat contrastvrije componenten qua interpretatie fraaier zijn dan contrast-componenten. De definitie van contrast-componenten bij PCA wordt gegeneraliseerd naar PARAFAC. Van deze definitie wordt gebruik gemaakt om aan te tonen dat PARAFAC contrast–componenten kan leveren. Dit roept de vraag op of er contrastvrije PARAFAC–componenten gevonden kunnen worden waarmee de gegevens even goed gefit kunnen worden als met de PARAFAC–componenten. Er wordt een algoritme afgeleid om het PARAFAC model met contrast–vrije componenten aan de gegevens te fitten. Met deze gerestricteerde variant van PARAFAC wordt geïllustreerd, dat de fit van PARAFAC met contrastvrije componenten vrijwel even goed kan zijn als die van PARAFAC. In een dergelijk geval wordt de voorkeur gegeven aan de contrastvrije componenten vanwege de fraaiere interpretatie.

Hoofdstuk 5 heeft als doel het bepalen van PARAFAC–componenten die overeenkomen met niet–overlappende clusters van variabelen. In sociaal–wetenschappelijk onderzoek ontstaan regelmatig confirmatieve en exploratieve onderzoeksvragen met betrekking tot dergelijke clusters. De PARAFAC–componenten komen niet noodzakelijk overeen met niet–overlappende clusters van variabelen. Derhalve worden twee gerestricteerde PARAFAC–methoden geïntroduceerd. Bij de eerste methode wordt ervan uitgegaan dat de onderzoeker een hypothese heeft omtrent de manier waarop de variabelen in niet–overlappende clusters zijn ingedeeld. Deze hypothese wordt vertaald in restricties op de patroonmatrix van PARAFAC. Er wordt een algoritme afgeleid om deze gerestricteerde PARAFAC–variant aan de gegevens te fitten en er worden voorwaarden gegeven waaraan de componenten moeten voldoen om te kunnen spreken van een bevestiging van de hypothese (confirmatie). Bij de tweede methode wordt ervan uitgegaan dat de onderzoeker wil weten of de variabelen zinvol in niet–overlappende clusters zijn op te delen. Deze exploratieve vraag wordt beantwoord door de restrictie aan de patroonmatrix van PARAFAC op te leggen, dat elke variable slechts door middel van één component wordt samengevat. Er wordt een algoritme afgeleid om deze gerestricteerde PARAFAC-variant aan de gegevens te fitten. Beide methoden worden geïllustreerd aan de hand van analyses van empirische gegevens.

In hoofdstuk 6 wordt een schematisch overzicht gepresenteerd van de gerestricteerde PARAFAC-methoden die in dit onderzoek besproken zijn. Er wordt geconcludeerd dat, in geval van zwakke uniciteit, de voorgestelde gerestricteerde PARAFAC-methoden een nuttige aanvulling zijn op de (ongerestricteerde) PARAFAC-methode, mede omdat ze stabielere componenten plegen te leveren.

THE ANALYSIS OF THREE-WAY ARRAYS BY CONSTRAINED PARAFAC METHODS

In three-way data, scores are available from a number of cases on a number of variables at a number of occasions. Among the earliest methods for the exploratory analysis of three-way data is a model and technique called PARAFAC, which was initiated by Richard Harshman, and which can be seen as a generalization of Principal Components Analysis. Some properties of Principal Components Analysis are retained under this generalization, while others are lost. PARAFAC yields a decomposition of the three-way data array into three component matrices: one for the persons, one for the variables, and one for the occasions.
In this book – a companion volume to Kroonenberg's *Three-mode Principal Component Analysis* that also appeared in the DSWO Press M&T series – much attention is directed to the most salient property of the PARAFAC model, the uniqueness of its components. It turns out that the theoretical property of uniqueness does not exclude the existence of alternative representations that fit the data almost as well as the standard PARAFAC solution. In such cases, the uniqueness is called weak. A constrained PARAFAC variant is developed specifically to determine the degree of uniqueness of PARAFAC components. For data having weak uniqueness, three new constrained PARAFAC methods are introduced that allow for easier interpretations. The first one employs constant relative importances of the components across occasions, the second one determines contrast-free components, and the third one determines components that correspond to non-overlapping clusters of variables. The usefulness of the constrained PARAFAC methods is illustrated by various analyses of empirical three-way data, and by simulations.

DSWO PRESS

ISBN 90-9005493-6