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 $n \times m$ 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

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

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 $\text{PARAFAC}(A,B,D_1)=\|X_1-AD_1B'\|^2$, 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 $A'A$=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 Diag($A'X_kB$) and I_q for Diag($B'B$) shows that

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

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

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

which completes the proof.

From (1.8) it follows immediately that, upon convergence, PARAFAC(A,B,C)$=\sum_{k=1}^{p}\|X_k\|^2-\|C\|^2$. Analogously, it can be proven that, if $A'A=I_q$, 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

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

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, $k=1,\dots,p$, as in PCA. In case p is large, a possibly prohibitive number of structure elements must be considered. Fortunately, however, the components $AD_1,\dots,AD_p$ 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, $k=1,\dots,p$, 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 $\mathrm{Diag}(A'A) = \mathrm{Diag}(C'C) = I_q$ it follows that $\mathrm{Diag}\left(\sum_{l=1}^{p} D_l A'AD_l\right) = I_q$. 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 Diag$(C'C)=I_q$, then from (1.5) it follows that

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

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 cosABC can reach -1, which can be proven as follows. Suppose that cosABC=-1, and that the rank of the matrix $(X_1|\dots|X_p)$ 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 $k=1,\dots,p$, we have

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

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

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

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

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 $\cos ABC=-1$. Therefore, the value of the PARAFAC loss function is not globally minimal if $\cos ABC=-1$, which proves that $\cos ABC$ can never reach -1.

The fact that $\cos ABC$ 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 2x2x2 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 $\cos ABC$ 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 $\cos ABC$ 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 $\cos ABC < -.85$ and one additional step of the iterative process further decreases $\cos ABC$. Of course, this is an arbitrary choice, but for practical purposes it can be justified as follows. If $\cos ABC < -.85$, 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 $\cos ABC < -.85$, 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* 38×3×2 *data-array.*

Number of iterative steps	10	100	1000	10000
$\cos ABC$	−.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 cosABC decreases, as the number of iterative steps increases. This suggests that cosABC 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 cosABC 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

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

$k=1,...,p$, 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.