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

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

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

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

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

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

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

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

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

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

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

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 $(X_1\mathbf{b}_r|\ldots|X_p\mathbf{b}_r)$, respectively, for $r=1,\ldots,q$. An efficient way to find these singular vectors is to define $Y_r\equiv(X_1\mathbf{b}_r|\ldots|X_p\mathbf{b}_r)$ 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

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

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

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

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 $A = \sum_{k=1}^{p} X_k B D_k \left(\sum_{l=1}^{p} D_l B' B D_l \right)^{-1}$, see equation 1.4. From substituting W^*B for B in this equation and using the fact that $B'B=I_q$ and $\mathrm{Diag}(C'C)=I_q$, we have

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

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 $S=\sum_{k=1}^{p} X_k' A D_k$, 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 $B=\sum_{k=1}^{p} X_k' A D_k * W = S * W$, 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

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

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

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

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

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 $B=\sum_{k=1}^{p}X'_kAD_k$ (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

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

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

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

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

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

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

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

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

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.