A Constrained PARAFAC Method for Positive Manifold Data

Wim P. Krijnen and Jos M. F. Ten Berge
University of Groningen

A set of non-negatively correlated variables, referred to as positive manifold data, display a peculiar pattern of loadings in principal components analysis (PCA). If a small set of principal components is rotated to a simple structure, the variables correlate positively with all components, thus displaying positive manifold. However, this phenomenon is critically dependent on the freedom of rotation, as is evident from the unrotated loadings. That is, although the first principal component is without contrast (which means that all variables correlate either positively or negatively with the first component), subsequent components have mixtures of positive and negative loadings—which means that positive manifold is absent.

PARAFAC is a generalization of PCA that has unique components, which means that rotations are not allowed. This paper examines how PARAFAC behaves when applied to positive manifold data. It is shown that PARAFAC does not always produce positive manifold solutions. For cases in which PARAFAC does not produce a positive manifold solution, a constrained PARAFAC method is offered that restores positive manifold by introducing non-negativity constraints. Thus, noncontrast PARAFAC components can be found that explain only a negligible amount of variance less than the PARAFAC components. These noncontrast components cannot be degenerate and cannot be partially unique in the traditional sense.

The PARAFAC model (Harshman, 1970) can be used to represent a three-way array of data consisting of scores from n persons on m variables at p occasions. PARAFAC is a generalization of principal components analysis (PCA), because PARAFAC decomposes every frontal slab—a matrix of order $n \times m$ containing the scores for an occasion—in a factor scores matrix and in a factor loadings matrix. That is, PARAFAC performs simultaneous components analysis of several two-way matrices (Harshman & Lundy, 1984, p. 155). Some properties of PCA are lost with this generalization. In particular, there is no rotational freedom of the PARAFAC components (Kruskal, 1977). Whether this absence of rotational freedom is an advantage is unclear. It will be illustrated that, in PCA, rotational freedom can be used to find clearly interpretable components.

Using Rotational Freedom in PCA

If a correlation matrix has no negative elements, the variables show positive manifold, and the data are called positive manifold data. If positive manifold data are analyzed by PCA, the first component has no negative correlations with the variables before rotation. The other components correlate positively with some variables and negatively with other variables. This implies that the structure matrix has positive and negative elements in the same column; therefore, these components have contrasting signs and can be called contrast components. For example, positive manifold data are found for variables measuring intelligence. Table 1 shows a correlation matrix and the unrotated structure matrix from PCA for positive manifold data, based on persons who were measured on three intelligence variables (Wolters, Wolters-Hoff, & Liebrand, 1988).
Table 1
Correlation Matrix (R), Unrotated Principal Components, and Varimax Rotated Structure Matrix From PCA for Vocabulary Analogies (1), Verbal Analogies (2), Nonverbal Abstraction (3)

<table>
<thead>
<tr>
<th>Variable</th>
<th>R</th>
<th>PCA</th>
<th>Varimax</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1.00</td>
<td>.59</td>
<td>.30</td>
</tr>
<tr>
<td>2</td>
<td>.59</td>
<td>1.00</td>
<td>.43</td>
</tr>
<tr>
<td>3</td>
<td>.30</td>
<td>.43</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The unrotated first and second components can be labeled “General Intelligence” and “Nonverbal-Verbal Contrast,” respectively. By rotating the structure matrix according to the varimax criterion, the contrast components disappear. The rotated components can be interpreted as “Nonverbal Abstraction” and “Analogies.” The rotated components are noncontrast components that partially coincide with two groups of variables. This illustrates that rotating PCA components can produce more clearly interpretable components. Note that these rotated components explain the variables as well as the unrotated components.

The PARAFAC model (Harshman, 1970) can be used to represent data such as these, based on a three-way array X with scores from n persons on m variables that measure intelligence at p occasions. Do PARAFAC components derived from such data resemble principal components or rotated components? Before this question can be answered, the interpretation of PARAFAC components must be clarified.

Interpretation of PARAFAC Components

The PARAFAC parameter matrices are defined as minimizing A, B, and C of the function

\[
CP(A,B,C) = \sum_{k=1}^{p} \|X_k - AD_kB'\|^2 ,
\]

where \(X_k\) is the \(k\)th \(n \times m\) frontal slab of \(X\),
- \(A\) is a \(n \times q\) matrix,
- \(B\) is a \(m \times q\) matrix, and
- \(D_k\) is a diagonal matrix with the elements of the \(k\)th row of a \(p \times q\) matrix \(C\) on its diagonal.

The value of the function in Equation 1 indicates the amount of variance not explained by the PARAFAC model. \(A\) contains the coefficients of \(n\) persons on \(q\) components and is called a components matrix. \(B\) contains the coefficients of \(m\) variables on the same \(q\) components, and is called a pattern matrix.

No method minimizes \(CP(A,B,C)\) directly. Therefore, an iterative algorithm, called the CANDECOMP algorithm (Carroll & Chang, 1970) (also called the CP algorithm), is used to minimize \(CP(A,B,C)\). The CP algorithm minimizes \(CP(A,B,C)\) over \(A\) for fixed \(B\) and \(C\); over \(B\) for fixed \(A\) and \(C\); and over \(C\) for fixed \(A\) and \(B\). After selecting an arbitrary starting point, the updates for \(A\), \(B\), and \(C\), respectively, can be computed as

\[
A = \sum_{k=1}^{p} X_kBD_k \left( \sum_{k=1}^{p} D_kB'BD_k \right)^{-1} ,
\]
\[ \mathbf{B} = \sum_{k=1}^{p} x_k \mathbf{A} \mathbf{D}_k \left( \mathbf{D}_k \mathbf{A}' \mathbf{A} \right)^{-1}, \]  

and

\[ c_k = (\mathbf{A}' \mathbf{A} \ast \mathbf{B}' \mathbf{B})^{-1} \mathbf{D}_k \mathbf{A}' \mathbf{A} \mathbf{X}_k, \]

where \( k = 1, \ldots, p \), \( c_k \) is row \( k \) of \( \mathbf{C} \), and \( \ast \) is the element-wise (Hadamard) product (Magnus & Neudecker, 1991, pp. 45-46). The process of updating \( \mathbf{A} \), \( \mathbf{B} \), and \( \mathbf{C} \) is continued until the percentage of explained variance from the last iterative cycle improves by only some small value.

Harshman & Desarbo (1984, p. 627) suggested using \( \mathbf{B} \) to interpret the components. However, analogously to the pattern matrix in PCA, after an oblique rotation the components can be interpreted only indirectly with the elements of \( \mathbf{B} \). This is because the elements of \( \mathbf{B} \) are regression weights in the regression of the variables on the PARAFAC components. Therefore, the rows of \( \mathbf{B} \) show how the variables can be reconstructed from the components. Consequently, in the PARAFAC model, the primary role of \( \mathbf{B} \) is to reconstruct the variables, rather than to serve as the basis for interpreting the components.

\( \mathbf{C} \) contains the coefficients of the \( p \) occasions on the \( q \) components. From the PARAFAC representation of \( \mathbf{X}_k \), the matrix \( \mathbf{D}_k \), which contains the elements from row \( k \) of \( \mathbf{C} \), stretches and contracts the components for occasion \( k \). For this reason, the diagonal elements of \( \mathbf{D}_k \) can be interpreted as the relative importance of the \( q \) components for the \( k \)th occasion (Harshman & Lundy, 1984, p. 155). If the components are orthonormal, then, similar to the singular values in PCA, the diagonal elements of \( \mathbf{D}_k \) express the importance of the columns in \( \mathbf{A} \) for occasion \( k \). Specifically, if \( \mathbf{A}' \mathbf{A} = \mathbf{I}_q \) and \( \mathbf{B} \) is scaled to unit length column-wise, then a squared diagonal element of \( \mathbf{D}_k \) equals the amount of variance in \( \mathbf{X}_k \) explained by the corresponding component (for a proof, see Appendix A).

Traditionally, if a PCA is followed by an oblique rotation, then a structure matrix is used to interpret the components; but in the PARAFAC model, a structure matrix is not defined. Therefore, a structure matrix will be defined here as \( \mathbf{S}_k = \mathbf{X}_k \mathbf{A} \mathbf{D}_k \) for each frontal slab. If \( \mathbf{X}_k \) is centered column-wise, then the elements of \( \mathbf{S}_k \) are covariances between the variables of the \( k \)th occasion and the PARAFAC components (weighted with \( \mathbf{D}_k \)). Consequently, the PARAFAC components can be interpreted using the elements of \( \mathbf{S}_k, k = 1, \ldots, p \). However, if \( p \) is large, then a prohibitively large number of elements must be considered for the interpretation of the PARAFAC components. To overcome this, an overall structure matrix will be defined. The structure matrix in PARAFAC is defined as

\[ \mathbf{S} = \sum_{k=1}^{p} \mathbf{X}_k \mathbf{A} \mathbf{D}_k. \]
and from

\[ B = \sum_{k=1}^{p} X_k A D_k \left( \sum_{l=1}^{p} D_l A' D_l \right)^{-1}, \tag{7} \]

\[ S = BA \left( \sum_{l=1}^{p} D_l \right). \tag{8} \]

Because \( A \left( \sum_{l=1}^{p} D_l \right) \) is diagonal and positive definite, every column in \( S \) is proportional to the corresponding column in \( B \).

Contrast components for the PARAFAC parameter matrices now can be defined. A PARAFAC component is a contrast component if at least one of the matrices \( S, B, \) or \( C \) has at least one column that contains both negative and positive elements. This definition covers many types of contrasts. For instance, if \( S \) has a column with contrasting elements, then the components are contrast components (in the same sense as contrast components defined for PCA). If \( C \) has a column with contrasting elements, then contrasting interpretations of the components arise from at least two structure matrices that are defined for each frontal slab. If \( B \) has a column with contrasting elements, then there is no positive manifold if the components are interpreted on the basis of \( B \).

### PARAFAC Representations of Some Empirical Datasets

Do PARAFAC components resemble principal components or rotated components? The results from the analysis of three datasets are presented. The three data arrays all consisted of scores of persons on variables that measured intelligence on two occasions. To compare the CP method with PCA, the persons (per frontal slab) with the maximum amount of person variance were removed until a rotated PCA structure matrix from PCA (with rank 2) of every frontal slice was found without negative elements. This ensured that the rotated components from PCA on every frontal slab were noncontrast components, and that if the PARAFAC components were contrast components they would not resemble the rotated components from PCA.

Prior to the CP analysis, the variables were centered column-wise within the occasions and scaled to unit length over the occasions. The CP rank parameter (number of components) was fixed to 1, 2, and 3, respectively. The CP process was terminated when the percentage of explained variance improved by less than .00001 from the last iterative cycle. When using the CP algorithm, if components are found without contrast, then after a suitable reflection of the columns of \( A, B, \) and \( C \), the matrices \( S, B, \) and \( C \) do not have negative elements. From the definition of contrast components, if the lowest elements found in \( S, B, \) and \( C \) are all positive, then the PARAFAC components are noncontrast components. If a negative element in \( S, B, \) or \( C \) is found and the largest element found in the corresponding column is positive, then the corresponding component is a contrast component.

Table 2 shows results for the CP method applied to three datasets of variables measuring intelligence. All datasets were based on two-occasion data. The Drenth (Drenth & Hoolwerf, 1977) dataset (Wolters et al., 1988) was based on 33 persons and 3 variables, the Groninger Intellientie Test (GIT; Luteijn & Van Der Ploeg, 1983) dataset (Wolters et al., 1988) was based on 20 persons and 4 variables, and the Differentiele Aanleg Test (DAT; Evers & Lucassen, 1984) dataset (Nijse, 1988) comprised 87 persons and 9 variables.

The results in Table 2 were based on the best of five starts for each dataset, with a maximum of 1,000 iterations. The matrices \( A \) and \( B \) were scaled to unit length column-wise. Table 2 shows that CP yielded both negative and positive elements in \( S \) or \( B \) (or both) when the rank parameter was fixed at 2 or 3. Thus, if PCA per frontal slab followed by varimax rotation of the structure matrix
yields components without contrast, then CP may yield contrast components. These PARAFAC components do not resemble rotated PCA components, because the PARAFAC components are contrast components. Therefore, the absence of rotational freedom when analyzing this type of data is a disadvantage of the PARAFAC model. Table 2 also shows that when the rank parameter was fixed at 1, no contrast component was found. Under mild conditions, contrast components cannot occur (see Result 1a and Result 1b below).

Let \( \mathbf{a}, \mathbf{b}, \) and \( \mathbf{c} \) be the parameter vectors of the CP solution for rank 1. Let \( \mathbf{X}_{ij} (j, k = 1, \ldots, p) \) have no negative elements.

Result 1a: If \( \mathbf{X}_{ij}a (k = 1, \ldots, p) \) and \( \mathbf{b} \) have no negative elements, then the vector \( \mathbf{c} \) has no negative elements (for a proof, see Appendix B).

Result 1b: If the vector \( \mathbf{c} \) has no negative elements and \( \mathbf{b} \) has one positive element, then \( \mathbf{X}_{ij}a (k = 1, \ldots, p) \) and \( \mathbf{b} \) have no negative elements (for a proof, see Appendix B).

Table 2 illustrates that the CP algorithm can yield contrast components for rank greater than 1. However, can an algorithm be constructed that finds optimal components without contrast?

**An ALS Algorithm for Optimal Noncontrast PARAFAC Components**

If \( \text{CP}(\mathbf{A}, \mathbf{B}, \mathbf{C}) \) is minimized subject to the constraints that \( \mathbf{X}_{ij}a, k = 1, \ldots, p, \) and \( \mathbf{B} \) and \( \mathbf{C} \) do not have negative elements, then the PARAFAC components are noncontrast components by definition. In order to find optimal noncontrast PARAFAC components, an alternating least squares (ALS) algorithm can be derived that minimizes \( \text{CP}(\mathbf{A}, \mathbf{B}, \mathbf{C}) \) subject to these non-negativity constraints.

In general, a non-negativity constraint can be imposed directly on each row of the parameter matrix \( \mathbf{B} \). With the non-negative least squares (NNLS) algorithm (Lawson & Hanson, 1974; Tenenhaus, 1988), \( \text{CP}(\mathbf{A}, \mathbf{B}, \mathbf{C}) \) can be minimized globally over \( \mathbf{B} \) for fixed \( \mathbf{A} \) and \( \mathbf{C} \) subject to the non-negativity of the elements in \( \mathbf{B} \). Analogously, with the NNLS algorithm, \( \text{CP}(\mathbf{A}, \mathbf{B}, \mathbf{C}) \) can be minimized globally over \( \mathbf{C} \) for fixed \( \mathbf{A} \) and \( \mathbf{B} \) subject to the non-negativity of \( \mathbf{C} \). Indirectly, a non-negativity constraint also can be imposed on the columns of the parameter matrix \( \mathbf{A} \), such that the matrix \( \mathbf{X}_{ij}a (k = 1, \ldots, p) \) has no negative elements. By using the Vec operator (Magnus & Neudecker, 1991, pp.

---

**Table 2**

Percent of Explained Variance and Lowest Matrix Elements for the CP Method Applied to Three Datasets (Entries in Parentheses Are the Largest Elements in the Matrix When a Negative Lowest Element Was Found)

<table>
<thead>
<tr>
<th>Dataset and Rank</th>
<th>% Explained Variance</th>
<th>Lowest Element</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>Drenth 1</td>
<td>54.01</td>
<td>.46</td>
</tr>
<tr>
<td>Drenth 2</td>
<td>75.10</td>
<td>.21</td>
</tr>
<tr>
<td>Drenth 3</td>
<td>87.39</td>
<td>-.28 (.34)</td>
</tr>
<tr>
<td>GIT 1</td>
<td>53.96</td>
<td>.68</td>
</tr>
<tr>
<td>GIT 2</td>
<td>68.29</td>
<td>-.09 (.76)</td>
</tr>
<tr>
<td>GIT 3</td>
<td>80.79</td>
<td>-.03 (.87)</td>
</tr>
<tr>
<td>DAT 1</td>
<td>39.17</td>
<td>.30</td>
</tr>
<tr>
<td>DAT 2</td>
<td>50.64</td>
<td>-.43 (.53)</td>
</tr>
<tr>
<td>DAT 3</td>
<td>60.00</td>
<td>-.40 (.54)</td>
</tr>
</tbody>
</table>
30–31), it follows that $\text{CP}(A, B, C)$ can be minimized over $A$ for fixed $B$ and $C$, subject to the non-negativity of the elements in the matrix $X_k A$ ($k = 1, \ldots, p$), by using the least squares with inequality constraints (LSI) algorithm proposed by Lawson & Hanson (1974). Therefore, an ALS algorithm can be constructed by updating $A$ with the LSI algorithm and updating $B$ and $C$ with the NNLS algorithm; this can be called the CPNC algorithm, because it minimizes the CP loss function subject to the requirement of noncontrast components.

Practical experience with the CPNC algorithm has been quite satisfactory. In terms of local minima, CPNC behaves like the CP algorithm, but it is slightly slower. Because the danger of local minima cannot be ruled out completely, more than one CPNC analysis should be run on the same data with different starting configurations for $B$ and $C$.

**Using the CPNC Algorithm for Analyzing Three-way Data**

The three datasets described above were analyzed using the CPNC algorithm, with the rank parameter fixed to 2 and 3. The preprocessing method used prior to the CP analysis was also used here. Table 3 shows the matrices $S$ and $B$ that were obtained from the CP and the CPNC analyses of the Drenth data. The matrix $C$ that was obtained from the CPNC analysis had no negative elements.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$S$</th>
<th>$B$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CP</td>
<td>CPNC</td>
</tr>
<tr>
<td></td>
<td>I II</td>
<td>I II</td>
</tr>
<tr>
<td>1</td>
<td>.76</td>
<td>.21</td>
</tr>
<tr>
<td>2</td>
<td>.84</td>
<td>.47</td>
</tr>
<tr>
<td>3</td>
<td>.56</td>
<td>.85</td>
</tr>
</tbody>
</table>

The components resulting from the CPNC analysis are easier to interpret than the PARAFAC components resulting from the CP analysis (Table 3). That is, the CPNC components for the Drenth data resemble PCA components that are rotated according to the varimax criterion, whereas the PARAFAC components are contrast components. Additionally, the difference that arises from interpreting the PARAFAC components using $S$ or $B$ has disappeared. This is due to the fact that both $S$ and $B$ had no negative elements, which is a desirable property of the CPNC method. No confusion can arise in interpreting the components with $S$ or $B$.

Whether the CPNC components are preferable to the PARAFAC components primarily depends on the amount of variance that is explained by the CPNC components. The percentages of explained variance found using the CP algorithm and the CPNC algorithm are given in Table 4. The difference in the amount of explained variance using the PARAFAC method and the CPNC components is negligible. No essential information was lost by abandoning the PARAFAC components in favor of the CPNC components. After applying both the CP and the CPNC method, the amount of variance explained by the CP components that was not explained by the CPNC components can be seen. The gains in terms of interpretation from the CPNC components and the loss in explained variance can be weighted against each other. Therefore, the CPNC method, as well as the CP method, can be useful for finding PARAFAC components.

Can the difference in explained variance between the CP and the CPNC method be large if $X_k X_{k'}$, $k, k' = 1, \ldots, p$, has no negative elements? To answer this question various simulation studies were conducted. No substantial differences in explained variance between the CP and the CPNC method
were found for data in which $X'kX_k$ had no negative elements. This indicates that, for all practical purposes, the non-negativity of $X'kX_k$ is likely to be sufficient for a small difference in explained variance between the CP and the CPNC method.

The CPNC Algorithm Cannot Yield Degenerate Components

Kruskal, Harshman, & Lundy (1989) noted that degenerate components sometimes can be found with the CP algorithm. These components can be defined as follows: Let $\cos(a, \alpha')$ denote the cosine between columns $l$ and $l'$ of $A$, and $\cos(b, \beta')$ and $\cos(c, \gamma')$ denote the cosines for the two corresponding columns in $B$ and $C$. Components are called degenerate if the limit of $\cos(a, \alpha') \cos(b, \beta') \cos(c, \gamma')$ (hereafter denoted as $\cos$), as the number of iterative steps tends toward infinity, is equal to $-1$. For rank 1 there are no degenerate components. If two components are degenerate, then $\cos$ approaches $-1$ as the number of iterative steps increases. However, $\cos$ cannot reach $-1$, because $\cos = -1$ implies a rank $q - 1$ fit, and a rank $q$ fit of the PARAFAC model to the data is obviously better than a rank $q - 1$ fit. Therefore, in the case of degeneracy, the limit of the CP function does not exist. To illustrate this, the Drenth $38 \times 3 \times 2$ data were reanalyzed, with the rank parameter fixed to 2. The same preprocessing method was used. Table 5 shows the value of $\cos$ at several stages of the iterative process. Table 5 illustrates that for the Drenth data, $\cos$ approached $-1$ as the number of iterations approached infinity; therefore, these components are degenerate.

In general, the fact that $\cos$ approaches $-1$ as the number of iterative steps increases implies that two of the corresponding columns of $B$ and $C$ approach perfect congruence and that the two corresponding columns of $A$ approach a value of $-1$ of the congruence coefficient (Tucker, 1951). Therefore, degenerate components cannot be interpreted consistently from the parameter matrices, which is why such components are called degenerate. It is clear that degenerate components are not acceptable. In addition, degenerate components are frequently contrast components. To illustrate that degenerate components can be contrast components and that a consistent interpretation of degenerate components

<table>
<thead>
<tr>
<th>Dataset</th>
<th>CP</th>
<th>CPNC</th>
<th>CP</th>
<th>CPNC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drenth</td>
<td>75.10</td>
<td>74.83</td>
<td>87.39</td>
<td>86.39</td>
</tr>
<tr>
<td>GIT</td>
<td>68.29</td>
<td>67.95</td>
<td>80.79</td>
<td>80.47</td>
</tr>
<tr>
<td>DAT</td>
<td>50.84</td>
<td>50.71</td>
<td>60.00</td>
<td>59.08</td>
</tr>
</tbody>
</table>

Table 4
Percent of Variance Explained by PARAFAC and CPNC Components for Three Datasets
is not possible, the S and B matrices resulting from the CP analysis of the complete Drenth data are depicted in Table 6. The resulting C matrix had no negative elements. From examination of B, the two components should have the same interpretation, but S shows that they should have opposite interpretations. Therefore, a consistent interpretation is impossible for these data. Also, the second column of S shows that these degenerate components are contrast components.

Table 6
Matrices S and B Resulting From CP Analysis of the Drenth Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>S</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>1</td>
<td>.77</td>
<td>-.72</td>
</tr>
<tr>
<td>2</td>
<td>.31</td>
<td>-.22</td>
</tr>
<tr>
<td>3</td>
<td>.55</td>
<td>.62</td>
</tr>
</tbody>
</table>

It is important to know whether the CPNC components can be degenerate. If $X'_iA$, B, and C have no negative elements, then the CPNC components cannot be degenerate. As proof of this, if B and C have no negative elements, then $\cos(b_{l'}b_l) \geq 0$ and $\cos(c_{l'}c_l) \geq 0$, where $b_l$, $b_{l'}$, $c_l$, and $c_{l'}$ are the $l$th and the $l'$th column of B and C, respectively, for $l, l' = 1, \ldots, q$. For $\cos$ to approximate $-1$, it is necessary to have $(l,l')$ such that $\cos(a_{l'},a_l)$ approximates $-1$. However, this contradicts the fact that $X'_iA$ has no negative elements.

It is not sufficient for nondegenerate components to impose only the non-negativity of B and $X'_iA$ on CP($A,B,C$). For example, only these constraints were imposed on the CP function and the Drenth data were analyzed with a corresponding algorithm, with $q = 3$. After 1,000 iterations, $\cos = -.99$ and a contrast was found in C.

Discussion

It has been shown that the CPNC method obviates the problem of degeneracy for three-way intelligence data. Harshman & Lundy (1984, p. 274) proposed overcoming degenerate components by imposing an orthonormality constraint on the columns of one of the parameter matrices. In the present context, orthonormal components have at least two disadvantages. First, degeneracy can be avoided by using orthonormal components, but contrast components cannot. Second, if the variables over all frontal slabs are all positively correlated and the variables fall into $q$ groups, then orthonormal components cannot coincide with the $q$ centroids of these groups of variables. Therefore, in the present context, orthonormal components are not a satisfactory solution for degenerate components.

Kruskal (1977) proved several sufficient conditions for the uniqueness of the PARAFAC components. A sufficient uniqueness condition, which can be derived from Kruskal (1977), is $\text{rank}(A) = \text{rank}(B) = \text{rank}(C) = q$. Kruskal’s uniqueness conditions are also sufficient for the uniqueness of the CPNC components. These sufficient conditions for uniqueness are fulfilled in practice. Therefore, both the PARAFAC components and the CPNC components usually are unique and, consequently, both sets of components lack rotational freedom. However, because contrast components are excluded by the CPNC method, the lack of rotational freedom is no longer a problem.

A similar difference between PARAFAC and CPNC components is related to partial uniqueness (Harshman & Lundy, 1984, p. 160), which comes from proportional columns in C, for instance. This partial uniqueness is paired with partial rotational freedom. Harshman & Lundy (1984, p.160) considered this problematic. They argued that nonunique components are “producing uninterpretable results that discourage attempts at further analysis” (Harshman & Lundy, 1984, p. 160). With the
CPNC method, proportional columns in C do not imply partial rotational freedom in the traditional sense. This is because the CPNC components are defined as noncontrast components. Therefore, rotational freedom is restricted to rotations that retain the CPNC components as noncontrast components. To obtain an interpretable PARAFAC representation of the three-way array with data from variables measuring intelligence, the CPNC method can be applied without concern about partial uniqueness.

Appendix A

It will be proven that if \( A' \) = \( I_d \) and \( \text{Diag}(B'B) = I_q \), then at every stationary point of \( \text{CP}(A, B, C) \), a squared diagonal element of \( D_k \) equals the amount of variance explained in \( X_k \) by the corresponding component.

Proof. For every stationary point of \( \text{CP}(A, B, C) \), \( D_k = (A'X_kB)B_k \). From the fact that \( A' = I_d \) and \( \text{Diag}(B'B) = I_q \), \( (A'X_kB)B_k = \text{Diag}(B'B) = I_q \), but then \( D_k = \text{Diag}(A'X_kB) \). For every stationary point of \( \text{CP}(A, B, C) \) the matrix \( D_k \) minimizes

\[
f_k(D_k) = \| X_k - AD_kB' \|^2
\]

\[
= \| X_k \|^2 - 2\text{tr} X_k'AD_kB' + \text{tr} BD_kA'AD_kB',
\]

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

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

Substituting \( D_k \) for \( \text{Diag}(A'X_kB) \) into Equation 9 results in

\[
f_k(D_k) = \| X_k \|^2 - \text{tr} D_k^2,
\]

and, therefore, \( c_{i_l}^2 \) equals the amount of variance explained by the \( l \)th component in the \( k \)th slab.

Therefore, if \( A' = I_d \) and \( \text{Diag}(B'B) = I_q \), then examination of the element \( c_{i_l}^2, l = 1, \ldots, q \) shows how important the \( l \)th component is for the \( k \)th frontal slab and that the amount of variance explained in \( X_k \) equals \( \text{tr} D_k^2 \).

This proof is based on the sufficiency of \( A' = I_d \) and \( \text{Diag}(B'B) = I_q \) for \( (A'X_kB)B_k = I_q \). Analogously, it can be proven that \( \text{Diag}(A'X_kB) = I_q \) and \( B'B = I_q \) are sufficient for \( (A'X_kB)B_k = I_q \). Therefore, if \( \text{Diag}(A'X_kB) = I_q \) and \( B'B = I_q \), then a squared diagonal element of \( D_k \) equals the amount of variance explained in \( X_k \) by the corresponding component.

Appendix B

Let \( X_{j,k} \), \( j, k = 1, \ldots, p \), have no negative elements, and let \( a, b, c \) be the globally minimizing parameter vectors of \( \text{CP} \) for rank 1. It will be proven that if \( X_k \) is \( X_k \), then \( c \) has no negative elements (Result 1a).

Proof. Without loss of generality, it may be assumed that \( a'a = b'b = 1 \). For \( c \) to be a minimizing \( c \) of the \( \text{CP} \) function, it must minimize \( f_2(c) = \| X_k - c \Sigma \| \). From elementary algebra, it follows that the minimizing \( c \) is \( a'X_kb \). From this and the fact that \( X_k \) is \( X_k \) and \( b \) have no negative elements, it follows that \( c \) has no negative elements.

It also will be proven that if \( c \) has no negative elements and \( b \) has at least one positive element, then \( X_k \) is \( X_k \) and \( b \) have no negative elements (Result 1b).

Proof. Without loss of generality, it may be assumed that \( a'a = b'b = 1 \). Minimizing \( \text{CP}(a, b, c) \) subject to \( a'a = b'b = 1 \) over \( a \) for fixed \( b \) and \( c \) is equivalent to minimizing \( f_1(a) = a'X_kcX_k \) subject to \( a'a = 1 \). From the Cauchy-Schwarz inequality (Magnus & Neudecker, 1991, p. 199), the maximizing vector \( a \) of \( f_1 \) subject to \( a'a = 1 \) is \( \lambda \Sigma cX_kb \), for some \( \lambda \geq 0 \). Minimizing \( \text{CP}(a, b, c) \) subject to \( a'a = b'b = 1 \) over \( b \) is fixed \( a \) and \( c \) is equivalent to maximizing \( f_2(b) = a'X_kcX_kb \) subject to \( b'b = 1 \). If \( \lambda \Sigma cX_kb \) is substituted for \( a \) in \( f_2 \), then...
From the fact that the elements of $X'X_k$ and $c$ are non-negative, $c_cX'X_k$ has no negative elements. Therefore, the matrix $(\Sigma c_cX'X_k)'(\Sigma c_cX'X_k)$ is gramian and has no negative elements. The first eigenvector of $(\Sigma c_cX'X_k)'(\Sigma c_cX'X_k)$ is the maximizing $b$ of $f_b$ subject to $b'b = 1$. From the Perron-Frobenius theorem (Gifi, 1990, p. 309) and the fact that $b$ has at least one positive element, it follows that this eigenvector has no negative elements. Finally, it must be shown that $X'X_k$ has no negative elements. This follows from the non-negativity of the elements in $b$ and in $X'X_k$, and the substitution of $\sum c_cX'X_kb$ for $a$ in $X'X_k$.

References


Acknowledgments

The authors are obliged to Henk Kiers for his comments on an earlier draft of this paper. The authors thank Theo Nijsse and Fred Wolters for making the data available.

Author's Address

Send requests for reprints or further information to Wim P. Krijnen, Department of Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands.