CHAPTER 4

A CONSTRAINED PARAFAC METHOD FOR POSITIVE MANIFOLD DATA

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

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

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

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

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

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

4.1 PARAFAC representations of some empirical data sets

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

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

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

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

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

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

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

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

4.2 An ALS algorithm for optimal non-contrast PARAFAC components

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

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

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

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

Practical experience with the PFNC algorithm has been quite satisfactory.

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

4.3 Using PFNC for analyzing three-way positive manifold data

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

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

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

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

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

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

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

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

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

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

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

4.4 A simulation study on non-contrast components

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

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

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

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

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

4.5 Discussion

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

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