Degeneracy in Candecomp/Parafac explained for $5 \times 3 \times 3$ arrays of rank 6 or higher

Alwin Stegeman[1]

University of Groningen

The Netherlands

Draft: December 15, 2005

Abstract

The Candecomp/Parafac (CP) method decomposes a real-valued three-way array into a pre-specified number R of rank-1 arrays, by minimizing the sum of squares of the residual array. The practical use of CP is sometimes hampered by the occurrence of so-called degenerate solutions, in which several rank-1 arrays are highly correlated in all three modes and some elements of the rank-1 arrays become arbitrarily large. We consider the CP decomposition of $5 \times 3 \times 3$ arrays of three-way rank 6 or higher, with the number of components R equal to 5. We show that the CP objective function may not have a minimum but an infimum. In such cases, any sequence of feasible CP solutions of which the objective value approaches the infimum, will become degenerate. We illustrate this result by means of simulations.

Keywords: Candecomp, Parafac, three-way arrays, degenerate solutions.

[1]Author's address: Alwin Stegeman, Heijmans Institute of Psychological Research, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands, phone: ++31 50 363 6193, fax: ++31 50 363 6304, email: *a.w.stegeman@rug.nl*

1 Introduction

Carroll and Chang [1] and Harshman [2] independently proposed the same method for component analysis of three-way arrays, and named it Candecomp and Parafac, respectively. In the sequel, we will denote column vectors as $\mathbf{x}$, matrices as $\mathbf{X}$ and three-way arrays as $\underline{\mathbf{X}}$. Unless stated otherwise, we assume all vectors, matrices and three-way arrays to be real-valued. Candecomp/Parafac (CP) decomposes an $I \times J \times K$ array $\underline{\mathbf{X}}$ into a prespecified number of R components $\underline{\mathbf{Y}}^{(r)}$, $r = 1, \ldots, R$, and a residual term $\underline{\mathbf{E}}$, all of the same order as $\underline{\mathbf{X}}$, i.e.

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \underline{\mathbf{Y}}^{(r)} + \underline{\mathbf{E}} \,. \tag{1.1}$$

Each component $\underline{\mathbf{Y}}^{(r)}$ is defined as the outer product of three vectors $\mathbf{a}^{(r)}$, $\mathbf{b}^{(r)}$ and $\mathbf{c}^{(r)}$, i.e. $y_{ijk}^{(r)} = a_i^{(r)}\, b_j^{(r)}\, c_k^{(r)}$. For fixed R, the CP decomposition (1.1) is found by minimizing the sum of squares of $\underline{\mathbf{E}}$.

The three-way rank of $\underline{\mathbf{X}}$ is defined as the smallest number of rank-1 arrays whose sum equals $\underline{\mathbf{X}}$. A three-way array has rank 1 if it is the outer product of three vectors. Hence, the concept of rank is the same for matrices and three-way arrays. Notice that, in the CP decomposition (1.1), each of the R components $\underline{\mathbf{Y}}^{(r)}$ has rank 1. The three-way rank of $\underline{\mathbf{X}}$ is equal to the smallest number of components for which full CP decomposition exists, i.e. with an all-zero residual term $\underline{\mathbf{E}}$.

A CP solution is usually expressed in terms of the component matrices $\mathbf{A}$ $(I \times R)$, $\mathbf{B}$ $(J \times R)$ and $\mathbf{C}$ $(K \times R)$, which have as columns the vectors $\mathbf{a}^{(r)}$, $\mathbf{b}^{(r)}$ and $\mathbf{c}^{(r)}$, respectively. Let the k-th slices of $\underline{\mathbf{X}}$ and $\underline{\mathbf{E}}$ be denoted by $\mathbf{X}_k$ $(I \times J)$ and $\mathbf{E}_k$ $(I \times J)$, respectively. Then (1.1) can be written as

$$\mathbf{X}_k = \mathbf{A}\,\mathbf{C}_k\,\mathbf{B}^T + \mathbf{E}_k\,, \qquad k = 1, \ldots K\,, \tag{1.2}$$

where $\mathbf{C}_k$ is the diagonal matrix with the k-th row of $\mathbf{C}$ as its diagonal.

The uniqueness of a CP solution $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is usually studied for given residuals $\underline{\mathbf{E}}$. The fitted part of a CP decomposition, i.e. $\sum_{r=1}^{R} \underline{\mathbf{Y}}^{(r)}$, can only be unique up to rescaling/counterscaling and jointly permuting columns of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. Indeed, the residuals will be the same for the solution given by $\overline{\mathbf{A}} = \mathbf{A}\,\mathbf{\Pi}\,\mathbf{T}_a$, $\overline{\mathbf{B}} = \mathbf{B}\,\mathbf{\Pi}\,\mathbf{T}_b$ and $\overline{\mathbf{C}} = \mathbf{C}\,\mathbf{\Pi}\,\mathbf{T}_c$, for a permutation matrix $\mathbf{\Pi}$ and diagonal matrices $\mathbf{T}_a$, $\mathbf{T}_b$ and $\mathbf{T}_c$ with $\mathbf{T}_a\,\mathbf{T}_b\,\mathbf{T}_c = \mathbf{I}_R$. When, for given residuals, the CP solution $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is unique up to these indeterminacies, it is called *essentially unique*. The scaling indeterminacy can be avoided by norming the columns of two component matrices at unit length (in this way, the diagonal elements of the corresponding diagonal matrices $\mathbf{T}_x$ are only allowed to be -1 or 1). When these constraints have been imposed, we label each component matrix as either *restricted* (of which there are two) or *unrestricted* (of which there is one).

Kruskal [4] has shown that essential uniqueness of the CP solution holds under relatively mild conditions. Kruskal's condition relies on a particular concept of matrix rank that he introduced, which has been named k-rank after him. Specifically, the k-rank of a matrix is the largest number

x such that every subset of x columns of the matrix is linearly independent. We denote the k-rank of a matrix $\mathbf{A}$ as $k_{\mathbf{A}}$. For a CP solution $(\mathbf{A},\mathbf{B},\mathbf{C})$, Kruskal [4] proved that

$$k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2\,R + 2\,, \tag{1.3}$$

is a sufficient condition for essential uniqueness. In practice, the value of R is usually small enough to satisfy (1.3). This uniqueness property of CP is one of its most attractive features.

The practical use of CP, however, has been hampered by the occurrence of so-called *degenerate solutions*. In the majority of such cases, exactly two components, say $\underline{\mathbf{Y}}^{(s)}$ and $\underline{\mathbf{Y}}^{(t)}$, of the solution display the following pattern:

- In all three component matrices, the columns s and t are almost exactly equal up to a sign change, the product of these sign changes being -1.
- The magnitudes of the elements of columns s and t in the unrestricted component matrix become arbitrarily large.

This pattern is called a *two-factor degeneracy*, see Kruskal, Harshman and Lundy [5]. The contributions of $\underline{\mathbf{Y}}^{(s)}$ and $\underline{\mathbf{Y}}^{(t)}$ diverge in nearly opposite directions. However, their sum $\underline{\mathbf{Y}}^{(s)} + \underline{\mathbf{Y}}^{(t)}$ still contributes to a better fit of the CP decomposition. It is clear that such CP solutions are impossible to interpret and should be avoided, if possible.

Analogous to two-factor degeneracies, also *three-factor degeneracies* have been encountered, for which holds that:

- In all three component matrices, any two of the columns s, t and u are almost exactly equal up to a multiplicative constant. The product of these constants may be positive or negative. In the restricted component matrices, columns s, t and u are almost equal up to a sign change.
- The magnitudes of the elements of columns s, t and u in the unrestricted component matrix become arbitrarily large.

For an example of a three-factor degeneracy, see Stegeman [7]. Also degeneracies involving four or five components have been encountered. Kruskal et al. [5] have argued that degenerate solutions occur due to the fact that the CP objective function has no minimum, but an infimum. They reason that every sequence of CP solutions of which the objective value is approaching the infimum, must fail to converge and displays the pattern of degeneracy as stated above. Stegeman [7] has proven this statement for CP decompositions of $p \times p \times 2$ arrays of rank $p + 1$ or higher, with $R = p$. In particular, the main result of Stegeman [7] is the following. Let $\mathcal{R}_p$ denote the set of real-valued $p \times p \times 2$ arrays of which the first slice is invertible, i.e.

$$\mathcal{R}_p = \{\underline{\mathbf{Y}} : p \times p \times 2 \text{ such that } \mathbf{Y}_1^{-1} \text{ exists}\}\,. \tag{1.4}$$

Theorem 1.1 *Let $\underline{\mathbf{X}}$ be a real-valued $p \times p \times 2$ array with $p \times p$ slices $\mathbf{X}_1$ and $\mathbf{X}_2$. Suppose that $\mathbf{X}_1^{-1}$ exists. If the rank of $\underline{\mathbf{X}}$ is $p+1$ or higher, then:*

- *the CP objective function of the best approximation of $\underline{\mathbf{X}}$ by rank-p arrays in $\mathcal{R}_p$ does not have a minimum, but an infimum, and*
- *any sequence of rank-p arrays in $\mathcal{R}_p$ of which the objective value approaches the infimum, will become degenerate.* □

In this paper, we use the tools developed in Stegeman [7] to prove the analogue of Theorem 1.1 for $5 \times 3 \times 3$ arrays of rank 6 or higher. In Section 2, a sketch of the proof of Theorem 1.1 is presented, using a rank criterion for $p \times p \times 2$ arrays based on the eigendecomposition of $\mathbf{X}_2\mathbf{X}_1^{-1}$. A similar rank criterion for $5 \times 3 \times 3$ arrays is due to Ten Berge [10]. In Section 3, Ten Berge's rank criterion is presented and elaborated upon. Our analogue of Theorem 1.1 is proven in Section 4, using Ten Berge's rank criterion. Finally, Section 5 contains the results of rank-5 approximations to random $5 \times 3 \times 3$ arrays of rank 6. They show what type of degenerate CP solutions may be encountered in this context.

2 A sketch of the proof of Theorem 1.1

To explain the similarity between Theorem 1.1 and our main result for $5 \times 3 \times 3$ arrays, we give a brief sketch of the proof of Theorem 1.1. We need the rank criterium for $p \times p \times 2$ arrays, which can be found in Ja' Ja' [3]; see also Ten Berge [8].

Lemma 2.1 *Let $\underline{\mathbf{X}}$ be a real-valued $p \times p \times 2$ array with $p \times p$ slices $\mathbf{X}_1$ and $\mathbf{X}_2$. Suppose that $\mathbf{X}_1^{-1}$ exists. There holds:*

(*i*) *If $\mathbf{X}_2\mathbf{X}_1^{-1}$ has p real eigenvalues and p linearly dependent eigenvectors, then $\underline{\mathbf{X}}$ has rank p.*

(*ii*) *If $\mathbf{X}_2\mathbf{X}_1^{-1}$ has at least one pair of complex eigenvalues then $\underline{\mathbf{X}}$ has rank $p+1$ or higher.*

(*iii*) *If $\mathbf{X}_2\mathbf{X}_1^{-1}$ has p real eigenvalues but less than p linearly independent eigenvectors, then $\underline{\mathbf{X}}$ has rank $p+1$ or higher.* □

Ja' Ja' [3] has shown that the rank of a real-valued $p \times p \times 2$ array is at most $p + \text{floor}(p/2)$, where the upper bound can be attained. Let $\underline{\mathbf{X}}$ be as in Theorem 1.1. Then it satisfies either (*ii*) or (*iii*) of Lemma 2.1. We need to solve the following optimization problem:

$$\begin{aligned} \text{Minimize} \quad & \|\underline{\mathbf{X}} - \underline{\mathbf{Y}}\|^2 \\ \text{subject to} \quad & \underline{\mathbf{Y}} \in \mathcal{D}_p\,, \end{aligned} \tag{2.1}$$

where $\|.\|$ denotes the Frobenius norm and $\mathcal{D}_p$ is the set of $p \times p \times 2$ arrays in $\mathcal{R}_p$, which have rank p or less, i.e.

$$\mathcal{D}_p = \{\underline{\mathbf{Y}} \in \mathcal{R}_p : \ \mathbf{Y}_2\mathbf{Y}_1^{-1} \text{ has } p \text{ real eigenvalues and } p \text{ linearly independent eigenvectors}\}\,. \tag{2.2}$$

We will consider $\mathcal{D}_p$, and other subsets of the space of real-valued $p \times p \times 2$ arrays, as subsets of $\Re^{2p^2}$. Next, consider the optimization problem

$$\begin{aligned} \text{Minimize} \quad & \|\underline{\mathbf{X}} - \underline{\mathbf{Y}}\|^2 \\ \text{subject to} \quad & \underline{\mathbf{Y}} \in \overline{\mathcal{D}}_p\,, \end{aligned} \tag{2.3}$$

where $\overline{\mathcal{D}}_p$ is the closure of $\mathcal{D}_p$ in $\mathcal{R}_p$, i.e. the union of $\mathcal{D}_p$ and its boundary points in $\mathcal{R}_p$. Since $\overline{\mathcal{D}}_p$ is closed and $\underline{\mathbf{X}} \notin \overline{\mathcal{D}}_p$, the objective function in (2.3) has a minimum. Any optimal solution $\underline{\tilde{\mathbf{X}}}$ of (2.3) is a boundary point of $\mathcal{D}_p$.

There holds that the boundary $\mathcal{D}_p$ is formed by $p \times p \times 2$ arrays $\underline{\mathbf{Y}}$ for which $\mathbf{Y}_2\mathbf{Y}_1^{-1}$ has p real eigenvalues but is not diagonalizable (and a lower dimensional subset which is immaterial in practice). Since such arrays do not have a full rank-p decomposition, they do not lie in $\mathcal{D}_p$ and, hence, the set $\mathcal{D}_p$ is open. Since $\underline{\mathbf{X}}$ does not lie in $\mathcal{D}_p$, this implies that the CP objective function in (2.1) has no minimum, but an infimum. If $\underline{\mathbf{X}}$ satisfies (*iii*), then it is a boundary point of $\mathcal{D}_p$ and the CP objective function in (2.1) has an infimum of 0. If $\underline{\mathbf{X}}$ satisfies (*ii*), then the CP objective function in (2.1) has an infimum of $\|\underline{\mathbf{X}} - \underline{\tilde{\mathbf{X}}}\|^2$, where $\underline{\tilde{\mathbf{X}}}$ is an optimal solution of (2.3). This proves the first statement of Theorem 1.1.

It remains to show that any sequence of CP solutions approaching the infimum, becomes degenerate. For any $\underline{\mathbf{Y}} \in \mathcal{D}_p$, the matrix $\mathbf{Y}_2\mathbf{Y}_1^{-1}$ has an eigendecomposition $\mathbf{K}\,\mathbf{\Lambda}\,\mathbf{K}^{-1}$, where the diagonal matrix $\mathbf{\Lambda}$ contains the eigenvalues and $\mathbf{K}$ has the associated eigenvectors as columns. It can be shown that a CP decomposition (1.2) of $\underline{\mathbf{Y}} \in \mathcal{D}_p$ is necessarily of the form $\mathbf{A} = \mathbf{K}$, $\mathbf{B}^T = \mathbf{K}^{-1}\mathbf{Y}_1$, $\mathbf{C}_1 = \mathbf{I}_p$ and $\mathbf{C}_2 = \mathbf{\Lambda}$. Suppose (*ii*) holds. Any sequence of arrays $\underline{\mathbf{Y}} \in \mathcal{D}_p$ of which the CP objective value in (2.1) approaching the infimum, will converge to an optimal solution $\underline{\tilde{\mathbf{X}}}$ of problem (2.3). Hence, $\mathbf{Y}_2\mathbf{Y}_1^{-1}$ will converge to $\tilde{\mathbf{X}}_2\tilde{\mathbf{X}}_1^{-1}$. Since $\underline{\tilde{\mathbf{X}}}$ satisfies (*iii*), the following holds for the eigendecomposition of $\mathbf{Y}_2\mathbf{Y}_1^{-1}$: the matrix $\mathbf{K}$ will converge to a singular matrix and $\mathbf{\Lambda}$ will converge to a matrix with not all diagonal elements distinct. If we assume that $\mathbf{A}$ and $\mathbf{C}$ are the restricted component matrices, we obtain the following for the CP decomposition of $\underline{\mathbf{Y}}$: matrices $\mathbf{A}$ and $\mathbf{C}$ will converge to matrices in which some columns are equal up to a sign change, and the elements of the corresponding columns of $\mathbf{B}$ will become arbitrarily large. Hence, the sequence of CP solutions indeed becomes degenerate. If (*iii*) holds, the same line of reasoning can be used with $\underline{\tilde{\mathbf{X}}}$ replaced by $\underline{\mathbf{X}}$ itself. This proves the second statement of Theorem 1.1.

The degenerate CP solutions in Theorem 1.1 are caused by the fact that the best approximation $\mathbf{Y}_2\mathbf{Y}_1^{-1}$ of $\mathbf{X}_2\mathbf{X}_1^{-1}$ is not diagonalizable. In case (*ii*), the characteristic polynomial of $\mathbf{X}_2\mathbf{X}_1^{-1}$ has at least one pair of complex roots and is approximated by characteristic polynomials of the same degree, which have only real roots. Such approximations may play a role in the occurrence of degenerate CP solutions in general. We will show that this is indeed the case for $5 \times 3 \times 3$ arrays, using a rank criterion similar to Lemma 2.1, which is due to Ten Berge [10]. In the next section, Ten Berge's rank criterion is presented.

3 A polynomial rank criterion for $5 \times 3 \times 3$ arrays

Here, we present a rank criterion for $5 \times 3 \times 3$ arrays due to Ten Berge [10], which is an analogue of Lemma 2.1. We give a summary of Ten Berge's analysis and extend it to suit our purposes.

Sometimes, we will refer to the situation where $5 \times 3 \times 3$ arrays are randomly sampled from a 45-dimensional continuous distribution. More formally, we assume the following in this case.

(A1) The $5 \times 3 \times 3$ array $\underline{\mathbf{X}}$ is randomly sampled from a 45-dimensional continuous distribution F with $F(S) = 0$ if and only if $L(S) = 0$, where L denotes the Lebesgue measure ans S is an arbitrary Borel set in $\Re^{45}$.

Notice that the requirement on F guarantees that $F(S) = 0$ if and only if the set S has dimensionality lower than 45. If the sequel, we will say that some property of $5 \times 3 \times 3$ arrays holds "with probablity 1". This means that the set of arrays for which the property does not hold, has dimensionality lower than 45. Analogously, a property holds "with probability zero" if the set of arrays for which the property holds has dimensionality lower than 45. For example, under (A1), each of the 5×3 slices $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ has rank 3 with probability 1 and $\underline{\mathbf{X}}$ has three-way rank 1 with probability zero.

Ten Berge [10] has proven that, under (A1), $5 \times 3 \times 3$ arrays have a three-way rank of 5 or 6 with probability 1, where both rank values occur with positive probability. To arrive at this result, Ten Berge developed a rank criterion for $5 \times 3 \times 3$ arrays involving a seventh degree polynomial P, the coefficients of which are functions of the elements of the array. If P has seven distinct real roots, then the array has rank 5, and if P has at least one pair of complex roots, then the array has rank 6 or higher. A summary of the analysis of Ten Berge [10] is as follows.

Ten Berge and Kiers [9] have shown that, with probability 1, there exist nonsingular matrices $\mathbf{S}$ (5×5) and $\mathbf{T}$ (3×3) such that

$$\mathbf{S}^T\mathbf{X}_1\,\mathbf{T} = \mathbf{Z}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{S}^T\mathbf{X}_2\,\mathbf{T} = \mathbf{Z}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{S}^T\mathbf{X}_3\,\mathbf{T} = \mathbf{Z}_3\,, \tag{3.1}$$

where, under (A1), the last slice can be treated as randomly sampled from a 15-dimensional continuous distribution. The matrix $\mathbf{S}$ is obtained as

$$\mathrm{Vec}(\mathbf{S}) = \left[\mathbf{I}_{25} - \mathbf{V}(\mathbf{V}^T\mathbf{V})^+\mathbf{V}^T\right]\,\mathrm{Vec}(\mathbf{I}_5)\,, \tag{3.2}$$

where $(\mathbf{V}^T\mathbf{V})^+$ denotes the Moore-Penrose inverse of $\mathbf{V}^T\mathbf{V}$ and

$$\mathbf{V} = [\mathbf{I}_5 \otimes \mathbf{X}_2 \,|\, \mathbf{O}] - [\mathbf{O} \,|\, \mathbf{I}_5 \otimes \mathbf{X}_1]\,, \tag{3.3}$$

with $\otimes$ denoting the Kronecker product and $\mathbf{O}$ denoting an all-zero matrix of order 25×6. The matrix $\mathbf{T}$ is then obtained as

$$\mathbf{T} = (\mathbf{X}_1^T\mathbf{S}\,\mathbf{S}^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T\mathbf{S}\,\mathbf{Z}_1\,. \tag{3.4}$$

The transformation in (3.1) is rank-preserving and simplifies the analysis. In the remaining part of this section, we consider only transformed arrays $\underline{\mathbf{Z}}$ of the form (3.1). Notice that, since the rank of the matrix $[\mathbf{Z}_1\,|\,\mathbf{Z}_2\,|\,\mathbf{Z}_3]$ equals 5, the three-way rank of $\underline{\mathbf{Z}}$ is at least 5. Next, we determine for which arrays $\underline{\mathbf{Z}}$ a rank-5 CP decomposition is possible.

First, we assume that a rank-5 CP decomposition $(\mathbf{A},\mathbf{B},\mathbf{C})$ exists in which the first row of the component matrix $\mathbf{C}$ does not contain any zeros. The CP decomposition in the form of (1.2) can be written as

$$\mathbf{Z}_1 = \mathbf{A}\,\mathbf{I}_5\,\mathbf{B}^T\,, \qquad \mathbf{Z}_2 = \mathbf{A}\,\tilde{\mathbf{C}}\,\mathbf{B}^T\,, \qquad \mathbf{Z}_3 = \mathbf{A}\,\tilde{\mathbf{D}}\,\mathbf{B}^T\,, \tag{3.5}$$

where $\mathbf{A}$ is 5×5, $\mathbf{B}$ is 3×5 and $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{D}}$ are 5×5 diagonal matrices containing, respectively, the second and third row of $\mathbf{C}$ as diagonal elements. Note that the elements of the first row of $\mathbf{C}$ have been set to 1 (this is possible due to the scaling indeterminacy of CP and because the first row of $\mathbf{C}$ does not contain any zeros). It follows from (3.5) that the columns of $\mathbf{Z}_1$ and $\mathbf{Z}_2$ lie in the column space of $\mathbf{A}$. Therefore, $\mathbf{A}$ must have rank 5 and its inverse exists. We have

$$\mathbf{A}^{-1}\mathbf{Z}_1 = \mathbf{B}^T\,, \qquad \mathbf{A}^{-1}\mathbf{Z}_2 = \tilde{\mathbf{C}}\,\mathbf{B}^T\,, \qquad \mathbf{A}^{-1}\mathbf{Z}_3 = \tilde{\mathbf{D}}\,\mathbf{B}^T\,, \tag{3.6}$$

which yields

$$\tilde{\mathbf{C}}\,\mathbf{A}^{-1}\mathbf{Z}_1 - \mathbf{A}^{-1}\mathbf{Z}_2 = \tilde{\mathbf{D}}\,\mathbf{A}^{-1}\mathbf{Z}_1 - \mathbf{A}^{-1}\mathbf{Z}_3 = \mathbf{O}\,. \tag{3.7}$$

From (3.7), we determine $\mathbf{A}$, $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{D}}$. The matrix $\mathbf{B}$ can then be obtained from $\mathbf{B}^T = \mathbf{A}^{-1}\mathbf{Z}_1$. Let $\mathbf{a}_j^T$ denote row j of $\mathbf{A}^{-1}$. Then we need to determine five linearly independent solutions to the vector equation $\mathbf{a}_j^T(c_j\,\mathbf{Z}_1 - \mathbf{Z}_2) = \mathbf{a}_j^T(d_j\,\mathbf{Z}_1 - \mathbf{Z}_3) = \mathbf{0}^T$, where c_j and d_j are the j-th diagonal elements of $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{D}}$, respectively, and $\mathbf{0}$ denotes the all-zero vector. Without loss of generality we set the first element of $\mathbf{a}_j$ to 1. From the form of $\mathbf{Z}_1$ and $\mathbf{Z}_2$, see (3.1), it follows that $\mathbf{a}_j^T(c_j\,\mathbf{Z}_1 - \mathbf{Z}_2) = \mathbf{0}^T$ is equivalent to

$$\mathbf{a}_j^T = (1\ \ e_j\ \ c_j\ \ c_je_j\ \ c_j^2)\,, \tag{3.8}$$

for some scalar e_j. Let $\mathbf{Z}_3 = [\mathbf{f}\,|\,\mathbf{g}\,|\,\mathbf{h}]$. It remains to satisfy $\mathbf{a}_j^T(d_j\,\mathbf{Z}_1 - \mathbf{Z}_3) = \mathbf{0}^T$. It can be seen that this is equivalent to the vector $(1\ \ e_j)^T$ being orthogonal to the columns of

$$\mathbf{W}_j = \begin{bmatrix} c_j^2 f_5 + c_j f_3 + f_1 - d_j & c_j^2 g_5 + c_j g_3 + g_1 & c_j^2 h_5 + c_j(h_3 - d_j) + h_1 \\ c_j f_4 + f_2 & c_j g_4 + g_2 - d_j & c_j h_4 + h_2 \end{bmatrix}. \tag{3.9}$$

To get such an e_j, the matrix $\mathbf{W}_j$ needs to have rank 1. Hence, we demand that the submatrix of columns 1 and 2 and that of columns 1 and 3 have determinant zero. However, this is not enough. We have to make sure that column 1 is not all-zero, which is the case for $c_j = -f_2/f_4$ and $d_j = c_j^2 f_5 + c_j f_3 + f_1$. The proportionality of columns 1 and 3 yields

$$(c_j^2 f_5 + c_j f_3 + f_1)(c_j h_4 + h_2) - (c_j f_4 + f_2)(c_j^2 h_5 + c_j h_3 + h_1) - d_j(c_j h_4 + h_2) + d_j(c_j^2 f_4 + c_j f_2) = 0\,, \tag{3.10}$$

from which d_j can be solved explicitly. When this solution for d_j is inserted in the equation for the proportionality of columns 1 and 2, we obtain

$$P(c_j) = z_7 c_j^7 + z_6 c_j^6 + z_5 c_j^5 + z_4 c_j^4 + z_3 c_j^3 + z_2 c_j^2 + z_1 c_j + z_0 = 0\,, \tag{3.11}$$

where the coeffcients z_j, which are functions of $\mathbf{f}$, $\mathbf{g}$ and $\mathbf{h}$, are given in the Appendix of Ten Berge [10]. The seventh degree polynomial P in (3.11) has one root $-f_2/f_4$ and six other roots, which we denote as $\lambda_1, \ldots, \lambda_6$. If these six roots are real, then five of them can be used as c_j and the corresponding d_j follows from (3.10). Moreover, if five *different* real roots are picked, then (3.8) guarantees five linearly independent rows of $\mathbf{A}^{-1}$ and a rank-5 decomposition is possible. If $\lambda_1, \ldots, \lambda_6$ are real and different, then six different rank-5 decomposisitons are possible in this way. Note that the polynomial P has all roots different with probability 1 and all roots real with positive probability. Also, complex roots of P occur with positive probability.

Next, we present an extension to the analysis in Ten Berge [10]. We would like to state that if P has at least one pair of complex roots, then a rank-5 decomposition is not possible. However, we only know that a rank-5 decomposition with no zeros in the first row of $\mathbf{C}$ is not possible then. Therefore, we also consider the possibility of having one or two zeros in the first row of $\mathbf{C}$. Notice that three or more zeros in the first row of $\mathbf{C}$ is not possible since $\mathbf{Z}_1$ in (3.5) has rank 3.

First, we set the first row of $\mathbf{C}$ to $(0\ \ 1\ \ 1\ \ 1\ \ 1)$ and see if a rank-5 decomposition is possible. Instead of $\mathbf{A}^{-1}\mathbf{Z}_1 = \mathbf{B}^T$, we then have

$$\mathbf{A}^{-1}\mathbf{Z}_1 = \begin{bmatrix} \mathbf{0}^T \\ \mathbf{b}_2^T \\ \vdots \\ \mathbf{b}_5^T \end{bmatrix}, \tag{3.12}$$

where $\mathbf{b}_j$ denotes the j-th column of $\mathbf{B}$. The equations $\mathbf{a}_j^T(c_j\,\mathbf{Z}_1 - \mathbf{Z}_2) = \mathbf{a}_j^T(d_j\,\mathbf{Z}_1 - \mathbf{Z}_3) = \mathbf{0}^T$ now only need to hold for $j = 2, 3, 4, 5$. Hence, for $c_2, \ldots, c_5$ we need to have four different real roots of P among $\lambda_1, \ldots, \lambda_6$. Then the corresponding d_j and $\mathbf{a}_j$ follow, as before, from (3.10) and (3.8), respectively. For the first columns of the component matrices, there must hold

$$\mathbf{a}_1^T\mathbf{Z}_1 = \mathbf{0}^T, \qquad \mathbf{a}_1^T\mathbf{Z}_2 = c_1\,\mathbf{b}_1^T, \qquad \mathbf{a}_1^T\mathbf{Z}_3 = d_1\,\mathbf{b}_1^T. \tag{3.13}$$

It can be verified that (3.13) yields $\mathbf{a}_1^T = (0\ \ 0\ \ 0\ \ x\ \ y)$, where x and y have to satisfy

$$\begin{bmatrix} f_4 & f_5 \\ c_1 g_4 - d_1 & c_1 g_5 \\ c_1 h_4 & c_1 h_5 - d_1 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{3.14}$$

The matrix in (3.14) must have rank 1. For nonzero f_4 and f_5, this implies that

$$\frac{g_4 f_5 - f_4 g_5}{f_5} = \frac{f_4 h_5 - h_4 f_5}{f_4}. \tag{3.15}$$

If (3.15) holds and

$$d_1 = \frac{c_1(f_4 h_5 - h_4 f_5)}{f_4}, \tag{3.16}$$

where c_1 can be set to 1, then the matrix in (3.14) has rank 1. When $\mathbf{a}_1$, c_1 and d_1 are known, $\mathbf{b}_1$ can be determined from (3.13). Since we only need four different real roots among $\lambda_1, \ldots, \lambda_6$ to construct a rank-5 decomposition, one may wonder whether P can have one pair of complex roots in this case. It turns out that the answer is no. The reason is that (3.15) is equivalent to the leading coefficient z_7 of P being equal to zero. Hence, in this case P is a sixth degree polynomial with one root $-f_2/f_4$, four different real roots and one other root, which is necessarily real. If P has five different real roots, besides $-f_2/f_4$, then five different rank-5 decompositions are possible with the first row of $\mathbf{C}$ equal to $(0\ \ 1\ \ 1\ \ 1\ \ 1)$. Notice that the requirement (3.15) implies that arrays permitting such a rank-5 decomposition occur with probability zero under the sampling regime (A1).

Next, we set the first row of $\mathbf{C}$ equal to $(0\ \ 0\ \ 1\ \ 1\ \ 1)$ and examine under which conditions a rank-5 decomposition is possible. Instead of (3.12), we now have

$$\mathbf{A}^{-1}\mathbf{Z}_1 = \begin{bmatrix} \mathbf{0}^T \\ \mathbf{0}^T \\ \mathbf{b}_3^T \\ \mathbf{b}_4^T \\ \mathbf{b}_5^T \end{bmatrix} . \tag{3.17}$$

The equations $\mathbf{a}_j^T(c_j\,\mathbf{Z}_1 - \mathbf{Z}_2) = \mathbf{a}_j^T(d_j\,\mathbf{Z}_1 - \mathbf{Z}_3) = \mathbf{0}^T$ only need to hold for $j = 3, 4, 5$. Hence, we need to have three different real roots among $\lambda_1, \ldots, \lambda_6$ to use as c_3, c_4, c_5. The corresponding d_j and $\mathbf{a}_j$ are obtained as before. The equations for the first columns of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are (3.13) as above. Analogously, the second columns of the component matrices need to satisfy

$$\mathbf{a}_2^T\mathbf{Z}_1 = \mathbf{0}^T , \qquad \mathbf{a}_2^T\mathbf{Z}_2 = c_2\,\mathbf{b}_2^T , \qquad \mathbf{a}_2^T\mathbf{Z}_3 = d_2\,\mathbf{b}_2^T . \tag{3.18}$$

Again, $\mathbf{a}_1^T = (0\ \ 0\ \ 0\ \ x\ \ y)$, where x and y have to satisfy (3.14). But also $\mathbf{a}_2^T = (0\ \ 0\ \ 0\ \ v\ \ w)$, where v and w have to satisfy

$$\begin{bmatrix} f_4 & f_5 \\ c_2 g_4 - d_2 & c_2 g_5 \\ c_2 h_4 & c_2 h_5 - d_2 \end{bmatrix} \begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} . \tag{3.19}$$

Since both $(x\ \ y)^T$ and $(v\ \ w)^T$ need to be orthogonal to $(f_4\ \ f_5)^T$, they are linearly dependent unless $f_4 = f_5 = 0$. If the latter is not the case, then $\mathbf{a}_1$ and $\mathbf{a}_2$ are linearly dependent and a rank-5 decomposition does not exist. Hence, $f_4 = f_5 = 0$ has to hold. Under this condition, the matrices in (3.14) and (3.19) have rank 1 if

$$(g_4 h_5)c_j^2 - (g_4 + h_5)c_j d_j + (g_5 - h_4)c_j + d_j^2 = 0 , \qquad j = 1, 2 . \tag{3.20}$$

Hence, we need a real solution (c_1, c_2, d_1, d_2) to (3.20). Then $(x\ \ y)^T$ and $(v\ \ w)^T$ follow from (3.14) and (3.19), respectively. The columns $\mathbf{b}_1$ and $\mathbf{b}_2$ can be determined from (3.13) and (3.18), respectively. This time, we only need three different real roots among $\lambda_1, \ldots, \lambda_6$ to construct a rank-5 decomposition. However, the constraint $f_4 = f_5 = 0$ yields that the coefficients z_7, z_6 and z_5 of the polynomial P are zero. Hence, P is a fourth degree polynomial with one root $-f_2/f_4$ and three other roots, which need to be real and distinct in order to obtain a rank-5 decomposition with the first row of $\mathbf{C}$ equal to $(0\ \ 0\ \ 1\ \ 1\ \ 1)$. Also in this case, P cannot have any complex roots if a rank-5 decomposition is possible. Notice that, as above, a rank-5 solution with the first row of $\mathbf{C}$ equal to $(0\ \ 0\ \ 1\ \ 1\ \ 1)$ is possible with probability zero under the sampling regime (A1).

The analysis above results in the following analogue of Lemma 2.1.

Lemma 3.1 *Let* $\underline{\mathbf{X}}$ *be a* $5 \times 3 \times 3$ *array for which the matrices* $\mathbf{S}$ *in* (3.2) *and* $\mathbf{T}$ *in* (3.4) *are both nonsingular. Then the transformation to the form* (3.1) *is possible. Assume the polynomial* P *in* (3.11) *has degree seven and roots* $-f_2/f_4$ *and* $\lambda_1, \ldots, \lambda_6$. *There holds:*

(*i*) *If there are five distinct real roots among* $\lambda_1, \ldots, \lambda_6$, *then* $\underline{\mathbf{X}}$ *has rank* 5.

(*ii*) *If there is at least one pair of complex roots among* $\lambda_1, \ldots, \lambda_6$, *then* $\underline{\mathbf{X}}$ *has rank* 6 *or higher.*

(*iii*) *If the roots* $\lambda_1, \ldots, \lambda_6$ *are real, but* (*i*) *does not hold, then* $\underline{\mathbf{X}}$ *has rank* 6 *or higher.* □

Recall that, under (A1), the matrices $\mathbf{S}$ and $\mathbf{T}$ are nonsingular with probability 1 and P has degree seven with probability 1. In Lemma 3.1, cases (*i*) and (*ii*) occur with positive probability, while (*iii*) occurs with probability zero. Analogous to Lemma 2.1, the rank of $\underline{\mathbf{X}}$ depends on whether the roots of a unique polynomial associated with the array are real or complex. The coeffcients of the polynomial are functions of the elements of the array.

4 Degenerate CP solutions for $5 \times 3 \times 3$ arrays of rank 6 or higher

When a $5 \times 3 \times 3$ array of rank 6 (or higher) is decomposed in CP with $R = 5$ components, the solution often becomes degenerate. Here, we show why and how that happens. We consider the following set of $5 \times 3 \times 3$ arrays:

$$\mathcal{R} = \{\underline{\mathbf{Y}} : \ \ \mathbf{S} \text{ in (3.2) and } \mathbf{T} \text{ in (3.4) are nonsingular, and } P \text{ in (3.11) has degree seven}\,\}\,. \qquad (4.1)$$

Note that the restriction to arrays in the set $\mathcal{R}$ is only virtual. This can be seen as follows. Under (A1), any array lies in $\mathcal{R}$ with probability 1. This implies that any array not in $\mathcal{R}$ can be approximated arbitrarily close by arrays in $\mathcal{R}$. Indeed, if $\underline{\mathbf{X}} \notin \mathcal{R}$, then the set

$$B(\underline{\mathbf{X}},\, \epsilon) = \{\underline{\mathbf{Y}} : \ \|\underline{\mathbf{X}} - \underline{\mathbf{Y}}\|^2 < \epsilon\}\,, \qquad (4.2)$$

has positive Lebesgue measure (i.e. positive 45-dimensional volume) for any $\epsilon > 0$. Under (A1), this is equivalent to the set $B(\underline{\mathbf{X}},\, \epsilon)$ having positive probability. Since the set $\mathcal{R}$ has probability 1, the set $B(\underline{\mathbf{X}},\, \epsilon)$ contains an array which lies in $\mathcal{R}$. Hence, for any $\underline{\mathbf{X}} \notin \mathcal{R}$ and any $\epsilon > 0$, there exists an $\underline{\mathbf{Y}} \in \mathcal{R}$ such that $\|\underline{\mathbf{X}} - \underline{\mathbf{Y}}\|^2 < \epsilon$. Although we cannot consider any CP algorithm as a generator of random arrays, we can safely assume, based on the observations above and the results from our simulations in the next section, that CP solutions not lying in the set $\mathcal{R}$ will not be encountered in practice.

Recall that the polynomial P in (3.11) has roots $-f_2/f_4$ and $\lambda_1, \ldots, \lambda_6$. We define the following subsets of $\mathcal{R}$ in (4.1). Let

$$\mathcal{S} = \{\underline{\mathbf{Y}} \in \mathcal{R} : \; P \text{ in (3.11) has seven real roots}\}\,, \tag{4.3}$$

and

$$\mathcal{D} = \{\underline{\mathbf{Y}} \in \mathcal{R} : \; P \text{ in (3.11) has six distinct real roots } \lambda_1, \ldots, \lambda_6\}\,. \tag{4.4}$$

It is clear that $\mathcal{D} \subset \mathcal{S} \subset \mathcal{R}$. Since the roots of P are continuous functions of the elements of the array, it follows that $\mathcal{D}$ is an open subset of $\mathcal{S}$, which is a closed subset of $\mathcal{R}$. Moreover, any array in $\mathcal{S}\backslash\mathcal{D}$ can be approximated arbitrarily close by arrays in $\mathcal{D}$. Hence, the interior of $\mathcal{S}$ equals the set $\mathcal{D}$ and $\mathcal{S}$ is the closure of $\mathcal{D}$ in $\mathcal{R}$. By Lemma 3.1, all arrays in $\mathcal{D}$ have rank 5. The arrays in $\mathcal{S}\backslash\mathcal{D}$ may have rank 5 or rank 6 or higher. All arrays in $\mathcal{R}\backslash\mathcal{S}$ have rank 6 or higher.

We consider a $5 \times 3 \times 3$ array $\underline{\mathbf{X}} \in \mathcal{R}$, which has rank 6 or higher. Hence, $\underline{\mathbf{X}}$ satisfies either *(ii)* or *(iii)* of Lemma 3.1. We approximate $\underline{\mathbf{X}}$ by rank-5 arrays in $\mathcal{R}$. Any rank-5 array in $\mathcal{R}$ either lies in $\mathcal{D}$ or in $\mathcal{S}\backslash\mathcal{D}$. Therefore, we consider the following optimization problems:

$$\begin{aligned} \text{Minimize} \quad & \|\underline{\mathbf{X}} - \underline{\mathbf{Y}}\|^2 \\ \text{subject to} \quad & \underline{\mathbf{Y}} \in \mathcal{D}\,, \end{aligned} \tag{4.5}$$

and

$$\begin{aligned} \text{Minimize} \quad & \|\underline{\mathbf{X}} - \underline{\mathbf{Y}}\|^2 \\ \text{subject to} \quad & \underline{\mathbf{Y}} \in \mathcal{S}\,. \end{aligned} \tag{4.6}$$

It is clear that if an optimal solution of problem (4.6) has rank 5, then this is also a best rank-5 approximation (in $\mathcal{R}$) of $\underline{\mathbf{X}}$. Our main result is the following, which is the analogue of Theorem 1.1.

Theorem 4.1 *Let $\underline{\mathbf{X}}$ be a real-valued $5 \times 3 \times 3$ array which lies in $\mathcal{R}$ and has rank 6 or higher. If all optimal solutions of problem (4.6) have rank 6 or higher, then:*

- *the CP objective function of the best approximation of $\underline{\mathbf{X}}$ by rank-5 arrays in $\mathcal{R}$ does not have a minimum, but an infimum, and*
- *any sequence of rank-5 arrays in $\mathcal{R}$ of which the objective value approaches the infimum, will become degenerate.* □

The remaining part of this section contains the proof of Theorem 4.1. It will be seen that the structure of the proof of Theorem 4.1 is similar to the proof of Theorem 1.1 described in Section 2.

Our proof of Theorem 4.1 is as follows. Suppose $\underline{\mathbf{X}}$ satisfies *(ii)* of Lemma 3.1. Since $\mathcal{S}$ is a closed subset of $\mathcal{R}$ and $\underline{\mathbf{X}} \notin \mathcal{S}$, it follows that any optimal solution of problem (4.6) will be a boundary point of $\mathcal{S}$. The assumption of Theorem 4.1 is that all optimal solutions of problem (4.6) have rank 6 or higher. Since the interior of $\mathcal{S}$ equals $\mathcal{D}$, this implies that the best approximation of $\underline{\mathbf{X}}$ by rank-5 arrays in $\mathcal{R}$ is obtained by solving problem (4.5).

From the above, we know that the objective value in (4.6) of any interior point of $\mathcal{S}$ can be decreased. The interior of $\mathcal{S}$ equals $\mathcal{D}$, which yields that, in problem (4.5), the objective value of any feasible solution can be decreased. Hence, the objective function of problem (4.5) has no minimum, but an infimum. The value of this infimum is equal to $\|\underline{\mathbf{X}} - \tilde{\underline{\mathbf{X}}}\|^2 > 0$, where $\tilde{\underline{\mathbf{X}}}$ is an optimal solution of (4.6).

If $\underline{\mathbf{X}}$ satisfies (iii) of Lemma 3.1, then $\underline{\mathbf{X}} \in \mathcal{S}\backslash\mathcal{D}$ can be approximated arbitrarily close by arrays in $\mathcal{D}$. Hence, the objective function of problem (4.5) has no minimum, but an infimum, and the value of the infimum is zero. This proves the first statement of Theorem 4.1.

Next, we prove the second statement of Theorem 4.1. It is assumed that all optimal solutions of problem (4.6) have rank 6 or higher, i.e. they satisfy (iii) of Lemma 3.1. From the observations above, it follows that any sequence of feasible solutions to problem (4.5), of which the objective value approaches the infimum, will approximate (arbitrarily close) a boundary point $\tilde{\underline{\mathbf{X}}}$ of $\mathcal{S}$, which is an optimal solution of problem (4.6). Since $\tilde{\underline{\mathbf{X}}}$ satisfies (iii) of Lemma 3.1, it has less than five distinct roots among the real $\lambda_1, \ldots, \lambda_6$ for its associated polynomial P.

Let $(\underline{\mathbf{Y}}_n)$ be a sequence of feasible solutions to problem (4.5), which converges to $\tilde{\underline{\mathbf{X}}}$. Since $\underline{\mathbf{Y}}_n \in \mathcal{D}$ for all n, the possible rank-5 decompositions for each $\underline{\mathbf{Y}}_n$ can be found as described in Section 3. First, the array is transformed to the form (3.1), using matrices $\mathbf{S}_n$ in (3.2) and $\mathbf{T}_n$ in (3.4). Any rank-5 decomposition $(\mathbf{A}_n, \mathbf{B}_n, \mathbf{C}_n)$ of the transformed array then yields a rank-5 decomposition of $\underline{\mathbf{Y}}_n$ of the form $((\mathbf{S}_n^{-1})^T\mathbf{A}_n, (\mathbf{T}_n^{-1})^T\mathbf{B}_n, \mathbf{C}_n)$.

Since the sequence $(\underline{\mathbf{Y}}_n)$ converges to $\tilde{\underline{\mathbf{X}}}$, the corresponding polynomials P_n converge to the polynomial P associated with $\tilde{\underline{\mathbf{X}}}$. This implies that the roots $\lambda_1^{(n)}, \ldots, \lambda_6^{(n)}$ of P_n will converge to the roots $\lambda_1, \ldots, \lambda_6$ of P. Hence, for $\underline{\mathbf{Y}}_n$ close to $\tilde{\underline{\mathbf{X}}}$, some roots of P_n will become more and more alike such that when any five of $\lambda_1^{(n)}, \ldots, \lambda_6^{(n)}$ are picked, at least two of them are nearly identical. In the lemma below, it is shown that this yields a degenerate rank-5 decomposition of $\underline{\mathbf{Y}}_n$.

Lemma 4.2 *Let $\tilde{\underline{\mathbf{X}}}$ be a boundary point of $\mathcal{S}$ which satisfies (iii) of Lemma 3.1. Let $(\underline{\mathbf{Y}}_n)$ be a sequence of arrays in $\mathcal{D}$, which converges to $\tilde{\underline{\mathbf{X}}}$. Consider five roots $\lambda_1^{(n)}, \ldots, \lambda_5^{(n)}$ (not equal to $-f_2/f_4$) of the polynomial P_n associated with $\underline{\mathbf{Y}}_n$. For these roots, let $(\mathbf{A}_n, \mathbf{B}_n, \mathbf{C}_n)$ be a rank-5 decomposition of the transformed form (3.1) of $\underline{\mathbf{Y}}_n$. Suppose the roots $\lambda_1^{(n)}, \ldots, \lambda_m^{(n)}$, $2 \le m \le 5$, converge to λ and m is maximal. If matrices $\mathbf{B}_n$ and $\mathbf{C}_n$ are restricted to have columns of unit lenght, then the following holds for $(\mathbf{A}_n, \mathbf{B}_n, \mathbf{C}_n)$ when $(\underline{\mathbf{Y}}_n)$ is close to $\tilde{\underline{\mathbf{X}}}$.*

- *The first m columns of $\mathbf{B}_n$ are almost exactly equal up to a sign change.*
- *The first m columns of $\mathbf{C}_n$ are almost exactly equal up to a sign change.*
- *The magnitudes of the elements of the first m columns of $\mathbf{A}_n$ become arbitrarily large, while the sum of these m columns converges to a nonzero vector.*

Proof. For the roots $\lambda_1^{(n)}, \ldots, \lambda_5^{(n)}$ of the polynomial P_n associated with $\underline{\mathbf{Y}}_n$, the rank-5 decomposition $(\mathbf{A}_n, \mathbf{B}_n, \mathbf{C}_n)$ of the transformed form (3.1) of $\underline{\mathbf{Y}}_n$ is constructed as described in Section 3. From (3.8), it follows that the first m rows of $\mathbf{A}_n^{-1}$ are nearly identical. From the restriction on

$\mathbf{B}_n$ and $\mathbf{B}_n^T = \mathbf{A}_n^{-1}\mathbf{Y}_1$, where $\mathbf{Y}_1$ is as in (3.1), it then follows that the first m columns of $\mathbf{B}_n$ are almost exactly equal. Column $j \leq m$ of the unrestricted matrix $\mathbf{C}_n$ is of the form $(1 \;\; \lambda_j^{(n)} \;\; d_j^{(n)})^T$, where $d_j^{(n)}$ depends on $\lambda_j^{(n)}$ through (3.10) (replace c_j in (3.10) by $\lambda_j^{(n)}$). Hence, it is clear that the first m columns of the restricted matrix $\mathbf{C}_n$ will become almost exactly equal. The sign changes in the first m columns of $\mathbf{B}_n$ and $\mathbf{C}_n$ occur due to the inherent scaling indeterminacy of CP.

It remains to consider the matrix $\mathbf{A}_n$. The (i,j)-th element of $\mathbf{A}_n$ equals

$$a_{ij}^{(n)} = \frac{(-1)^{i+j}\det(\mathbf{M}_n^{(i,j)})}{\det(\mathbf{A}_n^{-1})}\,, \tag{4.7}$$

where $\mathbf{M}_n^{(i,j)}$ is the 4×4 submatrix of $\mathbf{A}_n^{-1}$ that is obtained by deleting row j and column i. For $j \leq m$, the submatrix $\mathbf{M}_n^{(i,j)}$ contains $m-1$ nearly identical rows. Since $\mathbf{A}_n^{-1}$ contains m nearly identical rows, $\det(\mathbf{A}_n^{-1})$ will converge to zero at a faster rate than $\det(\mathbf{M}_n^{(i,j)})$, when $(\underline{\mathbf{Y}}_n)$ converges to $\tilde{\underline{\mathbf{X}}}$. Hence, the absolute value of $a_{ij}^{(n)}$ in (4.7) will become arbitrarily large for $j \leq m$, when $(\underline{\mathbf{Y}}_n)$ converges to $\tilde{\underline{\mathbf{X}}}$. For $j > m$, both the submatrix $\mathbf{M}_n^{(i,j)}$ and $\mathbf{A}_n^{-1}$ contain m nearly identical rows. This implies that the value of $a_{ij}^{(n)}$ in (4.7) will converge to a nonzero constant when $(\underline{\mathbf{Y}}_n)$ converges to $\tilde{\underline{\mathbf{X}}}$.

It remains to show that the sum of the first m columns of $\mathbf{A}_n$ converges to a nonzero vector. We consider the i-th element of the said sum and assume i is odd. The proof for i even is completely analogous. We introduce the following notation. Let $\mathbf{A}_{n,i}^{-1}$ denote the submatrix of $\mathbf{A}_n^{-1}$ obtained by deleting column i. For $m = 3$ (for example), we use

$$\begin{pmatrix} 1-2 \\ 3 \\ * \end{pmatrix}, \tag{4.8}$$

to denote the 4×4 matrix which has: the first row equal to the difference between the first and second row of $\mathbf{A}_{n,i}^{-1}$; the second row equal to the third row of $\mathbf{A}_{n,i}^{-1}$; and the last two rows equal to the last two rows of $\mathbf{A}_{n,i}^{-1}$.

First, consider $m = 2$. We have

$$\begin{aligned} \det(\mathbf{A}_n^{-1})\,(a_{i,1}^{(n)} + a_{i,2}^{(n)}) &= \det(\mathbf{M}_n^{(i,1)}) - \det(\mathbf{M}_n^{(i,2)}) \\ &= \det\begin{pmatrix} 2 \\ * \end{pmatrix} - \det\begin{pmatrix} 1 \\ * \end{pmatrix} \\ &= \det\begin{pmatrix} 2-1 \\ * \end{pmatrix}. \end{aligned} \tag{4.9}$$

It can be shown that the determinant in (4.9) converges to zero at the same rate as $\det(\mathbf{A}_n^{-1})$. Hence, it follows that $a_{i,1}^{(n)} + a_{i,2}^{(n)}$ converges to a nonzero constant.

Next, consider $m = 3$. We have

$$\det(\mathbf{A}_n^{-1})\,(a_{i,1}^{(n)} + a_{i,2}^{(n)} + a_{i,3}^{(n)}) = \det(\mathbf{M}_n^{(i,1)}) - \det(\mathbf{M}_n^{(i,2)}) + \det(\mathbf{M}_n^{(i,3)})$$

$$= \det\begin{pmatrix} 2 \\ 3 \\ * \end{pmatrix} - \det\begin{pmatrix} 1 \\ 3 \\ * \end{pmatrix} + \det\begin{pmatrix} 1 \\ 2 \\ * \end{pmatrix}$$

$$= \det\begin{pmatrix} 2-1 \\ 3 \\ * \end{pmatrix} - \det\begin{pmatrix} 2-1 \\ 1 \\ * \end{pmatrix}$$

$$= \det\begin{pmatrix} 2-1 \\ 3-1 \\ * \end{pmatrix} . \tag{4.10}$$

As before, it can be shown that the determinant in (4.10) converges to zero at the same rate as $\det(\mathbf{A}_n^{-1})$. Hence, it follows that $a_{i,1}^{(n)} + a_{i,2}^{(n)} + a_{i,3}^{(n)}$ converges to a nonzero constant. We omit the proof for $m = 4$ and $m = 5$, since it is similar to the cases above. □

As explained above, a rank-5 decomposition of $\underline{\mathbf{Y}}_n$ itself is of the form $(\mathbf{S}_n^T \mathbf{A}_n, \mathbf{T}_n^T \mathbf{B}_n, \mathbf{C}_n)$, where $\mathbf{S}_n$ is as in (3.2) and $\mathbf{T}_n$ is as in (3.4). If $(\mathbf{A}_n, \mathbf{B}_n, \mathbf{C}_n)$ is degenerate as described in Lemma 4.2, then this also holds for $(\mathbf{S}_n^T \mathbf{A}_n, \mathbf{T}_n^T \mathbf{B}_n, \mathbf{C}_n)$. Hereby, the second statement of Theorem 4.1 is proven.

From the proof of Theorem 4.1 above, it follows that the best rank-5 approximation (in $\mathcal{R}$) of $\underline{\mathbf{X}}$ always results in degeneracy when all optimal solutions of problem (4.6) have rank 6 or higher. If an optimal solution of problem (4.6) has rank 5, then its polynomial P has two identical roots among $\lambda_1, \ldots, \lambda_6$, and five distinct roots can be picked to form a nondegenerate rank-5 CP decomposition. Notice that degeneracies may still occur in this case, due to picking the two roots which converge to the same limit. However, the problem can be overcome by picking the right set of five roots.

Notice that a degenerate solution of the type described in Lemma 4.2 is equal to a two-factor degeneracy for $m = 2$, but not to an m-factor degeneracy, as described in Section 1, for $m \geq 3$.

5 Simulation results

Here, we illustrate the result of Theorem 4.1 by calculating the CP decomposition (with $R = 5$ components) of random $5 \times 3 \times 3$ arrays of rank 6. For this, we use the Multilinear Engine by Paatero [6].

Under (A1), random $5 \times 3 \times 3$ arrays $\underline{\mathbf{X}}$ of rank 6 have a polynomial P with at least one pair of complex roots (with probability 1). We consider three categories of such arrays $\underline{\mathbf{X}}$, namely for which P has two, four or six complex roots. We calculate the rank-5 approximation of 10 arrays of each category. For each array $\underline{\mathbf{X}}$ we use 10 different (random) starting values for the component matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. After the algorithm terminates, we use the optimal component matrices for each run to calculate the rank-5 arrays $\underline{\mathbf{Y}}^*$ closest to $\underline{\mathbf{X}}$. Since a $5 \times 3 \times 3$ array of rank 5 has six possible rank-5 decompositions, see Ten Berge [10] and Section 3, we focus on the optimal arrays $\underline{\mathbf{Y}}^*$ rather than on the component matrices.

Category $\underline{\mathbf{X}}$	pattern of nearly identical roots of P^*						
	2	**3**	**4**	**2+2**	**2+3**	**2+4**	**2+2+2**
P has 2 complex roots	6	4	0	0	0	0	0
P has 4 complex roots	2	0	5	2	1	0	0
P has 6 complex roots	0	0	0	3	5	1	1

Table 1: Frequencies of different type of CP solutions resulting from rank-5 approximations of random $5 \times 3 \times 3$ arrays of rank 6. Of each category, 10 different arrays are considered.

In a majority of cases (254 out of 300) all 10 runs for one array $\underline{\mathbf{X}}$ yield approximately the same solution $\underline{\mathbf{Y}}^*$. In the other 46 cases the algorithm terminates with a suboptimal solution. We discard the outcomes of these 46 runs and will speak of *the* CP solution $\underline{\mathbf{Y}}^*$ for a certain array $\underline{\mathbf{X}}$. Notice that this indicates that, usually, problem (4.6) has a unique optimal solution $\underline{\tilde{\mathbf{X}}}$, which is a bounday point of $\mathcal{S}$. This $\underline{\tilde{\mathbf{X}}}$ is then approximated arbitrarily close (depending on the stopping criterion of the CP algorithm) by arrays in $\mathcal{D}$.

As we expected, all solution arrays $\underline{\mathbf{Y}}^*$ lie in $\mathcal{R}$, i.e. the transformation (3.1) is possible and the polynomial P^* associated with $\underline{\mathbf{Y}}^*$ has degree seven. All polynomials P^* have some nearly identical roots, which is in agreement with our analysis in Section 4. If P^* has only two nearly identical roots (apart from $-f_2/f_4$), then both nondegenerate and degenerate rank-5 decompositions of $\underline{\mathbf{Y}}^*$ are possible. This corresponds to the case where $\underline{\tilde{\mathbf{X}}}$ has rank 5. If P^* has more nearly identical roots, then only degenerate rank-5 decompositions of $\underline{\mathbf{Y}}^*$ exist. This corresponds to the case where $\underline{\tilde{\mathbf{X}}}$ has rank 6. For all degenerate solutions obtained, the component matrices display the pattern described in Lemma 4.2.

In Table 1, the frequencies of the different types of solutions can be found for each category of arrays $\underline{\mathbf{X}}$. If P^* has two nearly identical roots and a group of three other nearly identical roots, this is denoted as type "**2+3**". As can be seen, a variety of different types of solutions occur. In cases where P^* has only two nearly identical roots, both nondegenerate and degenerate solutions occur, where the latter occur about twice as often as the former.

It should be noted that degenerate CP solutions are an asymptotic phenomenon, i.e. a CP solution "becomes more and more degenerate" as the CP algorithm runs longer. In some runs (mostly for arrays $\underline{\mathbf{X}}$ for which P has four or six complex roots), the convergence of the solution array $\underline{\mathbf{Y}}^*$ to the (unknown) boundary point $\underline{\tilde{\mathbf{X}}}$ is very slow. In such cases, the type of solution (i.e. the number of nearly identical roots of P^*) is determined by extrapolation and the inspection of the component matrices, where we use Lemma 4.2 in the latter case.

References

[1] J.D. Carroll and J.J. Chang (1970) Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. *Psychometrika*, **35**, 283–319.

[2] R.L. Harshman (1970) Foundations of the Parafac procedure: models and conditions for an "explanatory" multimodal factor analysis. *UCLA Working Papers in Phonetics*, **16**, 1–84.

[3] J. Ja' Ja' (1979) Optimal evaluation of pairs of bilinear forms. *SIAM Journal on Computing*, **8**, 443–462.

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

[5] J.B. Kruskal, R.A. Harshman and M.E. Lundy (1989) How 3-MFA data can cause degenerate Parafac solutions, among other relationships. In: *Multiway Data Analysis*, R. Coppi and S. Bolasco (editors), North-Holland, 115–121.

[6] P. Paatero (1999) The Multilinear Engine - a table-driven least squares program for solving multilinear problems, including the n-way Parafac model. *Journal of Computational and Graphical Statistics*, **8**, 1–35.

[7] A.W. Stegeman (2004) Degeneracy in Candecomp/Parafac explained for $p \times p \times 2$ arrays of rank $p+1$ or higher. *Technical Report*, Heijmans Institute of Psychological Research, University of Groningen, The Netherlands. Submitted

[8] J.M.F. Ten Berge (1991) Kruskal's polynomial for $2 \times 2 \times 2$ arrays and a generalization to $2 \times n \times n$ arrays. *Psychometrika*, **56**, 631–636.

[9] J.M.F. Ten Berge and H.A.L. Kiers (1999) Simplicity of core arrays in three-way principal component analysis and the typical rank of $p \times q \times 2$ arrays. *Linear Algebra and its Applications*, **294**, 169–179.

[10] J.M.F. Ten Berge (2004) Partial uniqueness in Candecomp/Parafac. *Journal of Chemometrics*, **18**, 12–16.