ON THE NON-EXISTENCE OF OPTIMAL SOLUTIONS AND THE OCCURRENCE OF "DEGENERACY" IN THE CANDECOMP/PARAFAC MODEL

WIM P. KRIJNEN

HANZE UNIVERSITY GRONINGEN

THEO K. DIJKSTRA AND ALWIN STEGEMAN

UNIVERSITY OF GRONINGEN

The CANDECOMP/PARAFAC (CP) model decomposes a three-way array into a prespecified number of R factors and a residual array by minimizing the sum of squares of the latter. It is well known that an optimal solution for CP need not exist. We show that if an optimal CP solution does not exist, then any sequence of CP factors monotonically decreasing the CP criterion value to its infimum will exhibit the features of a so-called "degeneracy". That is, the parameter matrices become nearly rank deficient and the Euclidean norm of some factors tends to infinity. We also show that the CP criterion function does attain its infimum if one of the parameter matrices is constrained to be column-wise orthonormal.

Key words: Candecomp, Parafac, level sets, bounded sequences, factor analysis.

1. Introduction

The Canonical Decomposition (CANDECOMP) (Carroll & Chang, 1970) and the Parallel Factor Analysis (PARAFAC) model (Harshman, 1970) are identical methods for component analysis of three-way arrays. The CANDECOMP/PARAFAC (CP) model assumes that a three-way array containing, for example, scores of cases on variables measured at several occasions is the sum of a systematic part and a residual part, where the former is the sum of R factors. The CP model has been applied in various disciplines such as linguistics (Harshman, Ladefoged, & Goldstein, 1977), psychology (Meyer, 1980; Krijnen & Ten Berge, 1992), marketing (Harshman & DeSarbo, 1984), chemometrics (Smilde, 1992; Leurgans & Ross, 1992), and neuroimaging (Andersen & Rayens, 2004; Beckmann & Smith, 2005).

Let $\circ$ denote the outer vector product, i.e., for vectors $\mathbf{x}$ and $\mathbf{y}$ we define $\mathbf{x} \circ \mathbf{y} = \mathbf{xy}'$. For three vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$, the product $\mathbf{x} \circ \mathbf{y} \circ \mathbf{z}$ is a three-way array with elements $x_i y_j z_k$. The CP model can be written as

$$\underline{X} = \sum_{r=1}^{R} \boldsymbol{a}_r \circ \boldsymbol{b}_r \circ \boldsymbol{c}_r + \underline{E}, \tag{1}$$

where $\underline{X}$ is the $I \times J \times K$ three-way data array; $\boldsymbol{a}_r$, $\boldsymbol{b}_r$, and $\boldsymbol{c}_r$ are the vectors of the rth factor in each of the three modes; and $\underline{E}$ is the residual array. The vectors $\boldsymbol{a}_r$, $\boldsymbol{b}_r$, and $\boldsymbol{c}_r$ are found by minimizing the sum of squares of $\underline{E}$. We refer to the latter as the CP criterion function. A CP solution is usually denoted by a triplet (A, B, C), where the parameter matrices contain the vectors $\boldsymbol{a}_r$, $\boldsymbol{b}_r$, and $\boldsymbol{c}_r$ as rth columns.

In this paper, we consider the real-valued CP model. The three-way rank of $\underline{X}$ is usually defined as the minimal number of rank-1 arrays whose sum equals $\underline{X}$, where a rank-1 array is the

outer product of three vectors. Hence, it follows from (1) that the CP model assumes $\underline{X}$ is the sum of R rank-1 arrays and a residual array. The smallest R for which $\underline{X}$ satisfies the CP model with residuals $\underline{E}$ equal to zero, is by definition equal to the three-way rank of $\underline{X}$. Moreover, CP tries to find the best three-way rank-R approximation of $\underline{X}$.

For estimating the CP parameters (A, B, C), several alternating least squares type of algorithms are available (Harshman, 1970; Carroll & Chang, 1970; Ten Berge, Kiers, & Krijnen, 1993; Krijnen & Ten Berge, 1992). Other CP algorithms can be found in Hopke, Paatero, Jia, Ross, and Harshman (1998) and Tomasi and Bro (2006). See also the Multilinear Engine of Paatero (1999).

One of the most attractive features of the CP model is the rotational uniqueness of its solutions. Kruskal (1977, 1989) showed that, for a fixed residual array $\underline{E}$, a CP solution (A, B, C) is unique up to rescaling/counterscaling and jointly permuting columns of the three parameter matrices if

$$k_A + k_B + k_C \geq 2R + 2, \tag{2}$$

where k_A, k_B, and k_C denote the k-ranks of the component matrices. The k-rank of a matrix is the largest number x such that every subset of x columns of the matrix is linearly independent. For an accessible proof of (2), see Stegeman and Sidiropoulos (2007).

To avoid the scaling indeterminacy in a CP solution, the columns of two component matrices can be set to unit length. Throughout the paper, we impose this restriction on A and B.

It is well known that the practical use of the CP model is complicated by the occurrence of "degeneracies" while running a CP algorithm. In such cases, the CP criterion function decreases very slowly, some factor magnitudes seem to increase without bound, and the parameter matrices become nearly rank deficient (Harshman & Lundy, 1984, p. 271; Kruskal, Harshman, & Lundy, 1983, 1985, 1989; Mitchell & Burdick, 1994). Such degeneracies are a problem for the analysis of three-way arrays, since the obtained CP solution is hardly interpretable. Degeneracies can be avoided by imposing orthogonality and non-negativity restrictions on the parameter matrices; see Theorem 2 and Lim (2005).

Synthetic data for which degeneracies occur in the CP model were considered by Kruskal et al. (1983) and Paatero (2000). Stegeman (2006, 2008a, 2008b) analysed the structure of degeneracies for all $I \times J \times 2$ arrays and several $I \times J \times 3$ arrays. It is claimed but not generally proven that in case of degeneracy the CP criterion function does not have a global minimum, that is, does not attain its infimum (Kruskal et al., 1983, 1985). For a synthetic $2 \times 2 \times 2$ array it is shown that this is indeed true (Ten Berge, Kiers, & De Leeuw, 1988; Stegeman, 2006). De Silva and Lim (2006) showed that for $R = 1$ there always exists an optimal CP solution, while for $2 \leq R \leq \min(I, J, K)$ there always exists an array $\underline{X}$ of three-way rank $R + 1$ which has no optimal CP solution. Also, the same authors show that all $2 \times 2 \times 2$ arrays of three-way rank 3 have no optimal CP solution for $R = 2$.

Apart from the (unrestricted) CP model, degeneracies also occur in other component models (DeSarbo & Carroll, 1985; Krijnen & Ten Berge, 1992; Stegeman, 2008b). Zijlstra and Kiers (2002) showed that degeneracies do not occur in component models which yield rotationally indeterminate components.

Here we show that there is a close relation between the occurrence of CP degeneracies and the non-existence of an optimal CP solution. In Section 3, we investigate the situation where the CP criterion function does not attain its infimum. We show that any sequence of $(A, B, C)_n$ which monotonically decreases the CP criterion function to its infimum will exhibit the features of a degeneracy. This implies that any CP algorithm minimizing the CP criterion function will yield a degeneracy if the CP model does not have an optimal solution for a particular array $\underline{X}$. In Section 4, we consider orthogonality and non-negativity restrictions under which the CP criterion function attains its infimum. Hence, under these restrictions degeneracies do not occur in the CP

model and (we hope) an interpretable CP solution is obtained. Section 5 contains a discussion of our results. In the next section, we introduce some notation.

2. Notation

In matrix notation, the CP model is

$$X_k = A D_k B' + E_k, \quad \text{for } k = 1, \dots, K, \tag{3}$$

where X_k is the kth $I \times J$ frontal slice of $\underline{X}$, E_k the kth $I \times J$ residual matrix, and D_k the diagonal matrix with row k of the matrix C as diagonal. For our purposes, it is convenient to rewrite the model. Let $\otimes$ be the Kronecker product and "vec" the operator that stacks the columns of a matrix one underneath the other. Let $x = \text{vec}(\text{vec } X_1, \dots, \text{vec } X_K)$ contain the data, $e = \text{vec}(\text{vec } E_1, \dots, \text{vec } E_K)$ the residuals, and $\theta = \text{vec}(\text{vec } A, \text{vec } B, \text{vec } C)$ the $q = R(I + J + K)$ parameters, where θ lies in $\mathbb{R}^q$. We denote the CP criterion function by $f(\theta)$.

Let the Euclidean norm of a vector and the Frobenius norm of a matrix be denoted by the symbol $\| \cdot \|$. As mentioned earlier, we restrict A and B to have columns of length 1. Let $\widetilde{C}$ have the unit length columns $\tilde{c}_r = c_r \| c_r \|^{-1}$, $r = 1, \dots, R$. The factors $f_r = (\tilde{c}_r \otimes b_r \otimes a_r)$, $r = 1, \dots, R$, which have unit length, are collected as columns in the matrix F. The magnitude or, more specifically, the Euclidean length $d_r = \| c_r \|$, $r = 1, \dots, R$, of the factors is collected in $d = (d_1, \dots, d_R)'$. It follows that

$$\|\theta\|^2 = \|d\|^2 + 2R. \tag{4}$$

By (3) and $\text{vec}(a_r b_r') = b_r \otimes a_r$, we have $x = F d + e$. For the CP criterion function f, we have

$$f(\theta) = \|x - F d\|^2. \tag{5}$$

An optimal CP solution is defined as a vector $\widehat{\theta} \in \mathbb{R}^q$ which globally minimizes $f(\theta)$. Various alternating least squares algorithms have been constructed to minimize f. These yield a sequence $\{\theta_1, \theta_2, \dots\} = \{\theta_n\}$ of parameter vectors, which monotonically decreases the CP criterion function, i.e., $f(\theta_n) \geq f(\theta_{n+1})$. The monotonicity of the sequence $\{f(\theta_n)\}$ is assumed throughout this paper. This is guaranteed to hold for alternating least squares algorithms. But practical experience shows that many other CP algorithms also yield monotonically decreasing sequences $\{f(\theta_n)\}$. For an element θ_n of a sequence of CP updates, we denote the corresponding matrices as A_n, B_n, $\widetilde{C}_n$, and the factors and their lengths as F_n and d_n, respectively.

3. When an Optimal CP Solution Does not Exist

Here, we present a result for the case where the CP model does not have an optimal solution, that is, the CP criterion function f does not attain its infimum. We have

Theorem 1. *If the CP criterion function f does not attain its infimum and $f(\theta_n) \downarrow \inf f$, as $n \to \infty$, then $\|\theta_n\| \to \infty$.*

Proof: Suppose that f does not attain its infimum, $\{\theta_n\}$ is a sequence such that $f(\theta_n) \downarrow \inf f$, and $\{\theta_n\}$ has a bounded subsequence $\{\theta_{n_k}\}$. It follows that the latter has a further subsequence $\{\theta_{n_{k_i}}\}$ such that $\lim_{i \to \infty} \theta_{n_{k_i}} = \widehat{\theta}$ for a certain limit point $\widehat{\theta} \in \mathbb{R}^q$ (Rudin, 1976, p. 51). Hence, by continuity of f,

$$\lim_{i \to \infty} f(\theta_{n_{k_i}}) = f\left(\lim_{i \to \infty} \theta_{n_{k_i}}\right) = f(\widehat{\theta}) = \inf f. \tag{6}$$

That is, f attains its infimum, a contradiction. It follows that $\{\boldsymbol{\theta}_n\}$ does not have a bounded subsequence. Therefore, the infimum over all subsequential limits of $\|\boldsymbol{\theta}_n\|$ is infinite, so that $\|\boldsymbol{\theta}_n\| \to \infty$, as $n \to \infty$.   □

Suppose an optimal CP solution does not exist. If a CP algorithm is used which is designed to minimize the CP criterion function (and does terminate with a suboptimal solution), then the Euclidean norm of the parameter vector $\boldsymbol{\theta}_n$ diverges to infinity as the iterations of this CP algorithm increase without bound. It follows from (4) that $\|\boldsymbol{d}_n\| \to \infty$, as $n \to \infty$. Hence, there are factor magnitude(s) which diverge to infinity as the number of iterative steps increases without bound. Equivalently, this means that given any fixed large number M, there exists a finite number of iterative steps N such that $\|\boldsymbol{d}_n\| > M$ for all $n \geq N$. This is also observed when degeneracies occur while running a CP algorithm and was proven analytically for degeneracies occurring for $I \times J \times 2$ arrays and some $I \times J \times 3$ arrays by Stegeman (2006, 2008a, 2008b).

Next, we relate the non-existence of an optimal CP solution to near linear dependency of the factors and near rank deficiency of the individual parameter matrices. Let $\mathrm{ev}_{\min}(\boldsymbol{F}'_n \boldsymbol{F}_n)$ be the smallest, $\mathrm{ev}_{\max}(\boldsymbol{F}'_n \boldsymbol{F}_n)$ be the largest eigenvalue, and $\kappa(\boldsymbol{F}_n) = \mathrm{ev}_{\max}^{1/2}(\boldsymbol{F}'_n \boldsymbol{F}_n)/\mathrm{ev}_{\min}^{1/2}(\boldsymbol{F}'_n \boldsymbol{F}_n)$ be the condition number of $\boldsymbol{F}_n$ (Ortega & Rheinboldt, 1970, p. 42). We have the following corollary to Theorem 1.

Corollary 1. *If the CP criterion function f does not attain its infimum and $f(\boldsymbol{\theta}_n) \downarrow \inf f$, as $n \to \infty$, then $\mathrm{ev}_{\min}(\boldsymbol{F}'_n \boldsymbol{F}_n) \to 0$, as $n \to \infty$.*

Proof: From the triangle inequality and $f(\boldsymbol{\theta}_n) \downarrow \inf f$, it follows that

$$\|\boldsymbol{F}_n \boldsymbol{d}_n\| \leq \|\boldsymbol{x} - \boldsymbol{F}_n \boldsymbol{d}_n\| + \|\boldsymbol{x}\| = f(\boldsymbol{\theta}_n)^{\frac{1}{2}} + \|\boldsymbol{x}\| \downarrow (\inf f)^{\frac{1}{2}} + \|\boldsymbol{x}\|. \tag{7}$$

Hence, the sequence $\{\|\boldsymbol{F}_n \boldsymbol{d}_n\|^2\}$ is bounded. That is, there exists a positive number M such that $\|\boldsymbol{F}_n \boldsymbol{d}_n\|^2 \leq M$ for all n. From

$$0 \leq \mathrm{ev}_{\min}\big(\boldsymbol{F}'_n \boldsymbol{F}_n\big) \|\boldsymbol{d}_n\|^2 \leq \|\boldsymbol{F}_n \boldsymbol{d}_n\|^2 \leq M, \tag{8}$$

and (4) it follows that

$$0 \leq \mathrm{ev}_{\min}\big(\boldsymbol{F}'_n \boldsymbol{F}_n\big) \leq \frac{M}{\|\boldsymbol{d}_n\|^2} = \frac{M}{\|\boldsymbol{\theta}_n\|^2 - 2R}. \tag{9}$$

By Theorem 1, $\|\boldsymbol{\theta}_n\|^2 \to \infty$, so that $\mathrm{ev}_{\min}(\boldsymbol{F}'_n \boldsymbol{F}_n) \to 0$, as $n \to \infty$.   □

Since $\boldsymbol{F}'_n \boldsymbol{F}_n$ is positive semidefinite with unit diagonal elements, $1 \leq \mathrm{ev}_{\max}(\boldsymbol{F}'_n \boldsymbol{F}_n) \leq R$. Hence, by Corollary 1, $\kappa(\boldsymbol{F}_n) \to \infty$. Furthermore, Corollary 1 implies the following corollary.

Corollary 2. *If the CP criterion function f does not attain its infimum and $f(\boldsymbol{\theta}_n) \downarrow \inf f$, as $n \to \infty$, then the smallest singular value of each column-wise normalized parameter matrix tends to zero, as $n \to \infty$.*

Proof: Suppose that f does not attain its infimum. From the definition of $\boldsymbol{F}'_n \boldsymbol{F}_n$ and elementary properties of the Kronecker product, $\boldsymbol{F}'_n \boldsymbol{F}_n = \boldsymbol{A}'_n \boldsymbol{A}_n * \boldsymbol{B}'_n \boldsymbol{B}_n * \widetilde{\boldsymbol{C}}'_n \widetilde{\boldsymbol{C}}_n$ follows, where $*$ is the element-wise Hadamard product. Since $\boldsymbol{A}'_n \boldsymbol{A}_n, \boldsymbol{B}'_n \boldsymbol{B}_n$ and $\widetilde{\boldsymbol{C}}'_n \widetilde{\boldsymbol{C}}_n$ are positive semidefinite with

unit diagonal elements, it follows for all n that

$$\max\{\text{ev}_{\min}(\boldsymbol{A}'_n\boldsymbol{A}_n),\ \text{ev}_{\min}(\boldsymbol{B}'_n\boldsymbol{B}_n),\ \text{ev}_{\min}(\widetilde{\boldsymbol{C}}'_n\widetilde{\boldsymbol{C}}_n)\} \le \text{ev}_{\min}(\boldsymbol{A}'_n\boldsymbol{A}_n * \boldsymbol{B}'_n\boldsymbol{B}_n * \widetilde{\boldsymbol{C}}'_n\widetilde{\boldsymbol{C}}_n) \quad (10)$$

(Schur, 1911; Styan, 1973). An application of Corollary 1 completes the proof. $\square$

Since the smallest singular value of each normalized parameter matrix tends to zero, it follows that the smallest singular value is arbitrarily small if the number of iterations of the CP algorithm are sufficiently large. In this sense the normalized parameter matrices are nearly rank deficient for n sufficiently large. Since the largest singular value of each column-wise normalized parameter matrix lies between 1 and $R^{1/2}$, Corollary 2 implies that $\kappa(\boldsymbol{A}_n) \to \infty$, $\kappa(\boldsymbol{B}_n) \to \infty$, and $\kappa(\widetilde{\boldsymbol{C}}_n) \to \infty$. That is, the condition number of each normalized parameter matrix tends to infinity as the number of iterative steps increases without bound. These phenomena were proven analytically for degeneracies occurring for $I \times J \times 2$ arrays and some $I \times J \times 3$ arrays by Stegeman (2006, 2008a, 2008b).

If $R = 2$, a geometric interpretation of Corollary 1 can be given as follows. Since we may assume $d_{1n} > 0$ and $d_{2n} > 0$ without loss of generality, and $\boldsymbol{F}_n$ has unit column length, it follows that

$$\|\boldsymbol{F}_n\boldsymbol{d}_n\|^2 = \|\boldsymbol{d}_n\|^2 + 2d_1 d_2 \cos(\boldsymbol{f}_{1n}, \boldsymbol{f}_{2n}). \quad (11)$$

Therefore, $\{\|\boldsymbol{F}_n\boldsymbol{d}_n\|^2\}$ bounded, $\{\|\boldsymbol{d}_n\|^2\}$ unbounded, $d_{1n} > 0$, $d_{2n} > 0$, $\text{ev}_{\min}(\boldsymbol{F}'_n\boldsymbol{F}_n) = 1 - |\cos(\boldsymbol{f}_{1n}, \boldsymbol{f}_{2n})| \to 0$, as $n \to \infty$, implies that $\cos(\boldsymbol{f}_{1n}, \boldsymbol{f}_{2n}) \to -1$, as $n \to \infty$. (If $\cos(\boldsymbol{f}_{1n}, \boldsymbol{f}_{2n}) \to 1$, as $n \to \infty$, it would follow that $\|\boldsymbol{F}_n\boldsymbol{d}_n\|^2 \to \infty$, which is contradictory.) Hence, the angle between the factors tends to $180°$. More specifically, the two factors may be represented as two vectors on the boundary of a unit "ball" in $\mathbb{R}^q$ of which the end points tend to positions on a straight line that contains the center as well. This is in line with what is usually observed when a degeneracy occurs while running a CP algorithm: the factors involved nearly cancel out but still contribute to a better fit of the CP model.

In case the CP criterion function f does not attain its infimum, Theorem 1 and its corollaries show the "degenerate" behavior of any sequence $\{\boldsymbol{\theta}_n\}$ such that $f(\boldsymbol{\theta}_n) \downarrow \inf f$. These results seem to provide a mathematical basis for the detection of cases where the criterion function does not attain its infimum. Indeed, if, for a large number of runs of a CP algorithm, the magnitudes of some factors and the condition numbers of the parameter matrices increase to arbitrary large values, then the conclusion that the CP criterion function does not attain its infimum seems inevitable. However, such reasoning need not be valid for a small number of such sequences. For example, cases are known where these phenomena occur in locally optimal neighborhoods while the CP model does have an optimal solution. See, for example, Paatero (2000) who showed this for a class of $2 \times 2 \times 2$ arrays and Stegeman (2008b) who showed this for $5 \times 3 \times 3$ arrays of rank 5. In these cases, running the CP algorithm several times with different starting values results in both degeneracies and optimal CP solutions.

4. Restrictions under which an Optimal CP Solution Exists

As mentioned earlier, De Silva and Lim (2006) showed that for $2 \le R \le \min(I, J, K)$ there always exists an array $\underline{\boldsymbol{X}}$ of three-way rank $R + 1$ which has no optimal CP solution. In this section we consider orthogonality and non-negativity restrictions under which the CP model always has an optimal solution. Such constraints have been included in alternating least squares CP algorithms (Kiers, 1989a, 1989b, 1991; Krijnen & Kiers, 1993; Bro & De Jong, 1997) and can be included in the Multilinear Engine (Paatero, 1999).

To show the existence of an optimal CP solution, we make use of level sets. Let $L(\gamma) = \{\boldsymbol{\theta} \in \mathbb{R}^q : f(\boldsymbol{\theta}) \leq \gamma\}$ be a level set of the CP criterion function f (Ortega & Rheinboldt, 1970, p. 98). Theorem 1 gives a condition under which f does not have a bounded level set. We need the following lemmas, proven by Ortega and Rheinboldt (1970, p. 104).

Lemma 1. *Let $g : D \subset \mathbb{R}^q \to \mathbb{R}^1$, where D is unbounded. Then all level sets of g are bounded if and only if $\lim_{n\to\infty} g(\boldsymbol{\theta}_n) = \infty$ whenever $\{\boldsymbol{\theta}_n\} \subset D$ and $\lim_{n\to\infty} \|\boldsymbol{\theta}_n\| = \infty$.*

Lemma 2. *Let $g : D \subset \mathbb{R}^q \to \mathbb{R}^1$ be continuous on the closed set D. Then g has a bounded level set if and only if the set of global minimizers of g is nonempty and bounded.*

Note that the CP criterion function f is continuous with an unbounded domain. The continuity of f implies that the level sets $L(\gamma)$ are closed. We have the following result.

Theorem 2. *If one of the parameter matrices $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ is constrained to be column-wise orthonormal, then all level sets of f are bounded and the CP model has an optimal solution.*

Proof: Suppose that $\boldsymbol{A}'_n \boldsymbol{A}_n = \boldsymbol{I}$ for all n. Then it follows that $\boldsymbol{F}'_n \boldsymbol{F}_n = \boldsymbol{A}'_n \boldsymbol{A}_n * \boldsymbol{B}'_n \boldsymbol{B}_n * \widetilde{\boldsymbol{C}}'_n \widetilde{\boldsymbol{C}}_n = \boldsymbol{I}$. Hence, by the triangle inequality (Luenberger, 1969, p. 22),

$$f^{\frac{1}{2}}(\boldsymbol{\theta}_n) = \|\boldsymbol{x} - \boldsymbol{F}_n \boldsymbol{d}_n\| \geq \big| \|\boldsymbol{x}\| - \|\boldsymbol{F}_n \boldsymbol{d}_n\| \big| = \big| \|\boldsymbol{x}\| - \|\boldsymbol{d}_n\| \big| \to \infty \tag{12}$$

whenever $\|\boldsymbol{d}_n\| \to \infty$, which is equivalent to $\|\boldsymbol{\theta}_n\| \to \infty$ by (4). By Lemma 1 all level sets of f are bounded. Let $L(\gamma)$ be a nonempty level set of f. Then $L(\gamma)$ is bounded and closed and, hence, compact. Restricting f to $L(\gamma)$, it follows from Lemma 2 that f attains its infimum on $L(\gamma)$. $\qquad\square$

Also using level sets, Lim (2005) showed that the CP model has an optimal solution if the parameter matrices are constrained to have non-negative elements. We will state it here without a proof.

Theorem 3. *If each of the parameter matrices $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ is constrained to have non-negative elements, then all level sets of f are bounded and the CP model has an optimal solution.*

Using the theory of level sets, Krijnen (2006) analyzed the existence of optimal solutions for various factor models related to the CP model.

5. Conclusions and Discussion

We have analyzed the situation where the CP model does not have an optimal solution, i.e., the CP criterion function f does not attain its infimum. We showed that for any sequence of CP updates $\{\boldsymbol{\theta}_n\}$ such that $f(\boldsymbol{\theta}_n) \downarrow \inf f$, it holds that $\|\boldsymbol{\theta}_n\| \to \infty$, $\|\boldsymbol{d}_n\| \to \infty$, $\kappa(\boldsymbol{F}_n) \to \infty$, and, by Corollary 2, $\kappa(\boldsymbol{A}_n) \to \infty$, $\kappa(\boldsymbol{B}_n) \to \infty$, and $\kappa(\widetilde{\boldsymbol{C}}'_n) \to \infty$. Hence, the sequence of parameter vectors diverges to a CP "degeneracy", i.e. the factors become nearly linearly dependent and the individual parameter matrices become nearly rank deficient. Our result provides a general proof of the claim by Kruskal et al. (1983, 1985) that degeneracies occur when no optimal CP solution exists. Hence, any CP algorithm minimizing the CP criterion function will yield a degeneracy if the CP model does not have an optimal solution for this particular array $\underline{X}$. Moreover, our result

can be used to detect a degeneracy while running a CP algorithm, e.g. by monitoring the smallest singular values of the parameter matrices together with the factor lengths.

For $I \times J \times 2$ arrays and some $I \times J \times 3$ arrays, the occurrence of degeneracies while running a CP algorithm was mathematically described by Stegeman (2006, 2008a, 2008b). Here, the number of factors involved in the degeneracy and the type of rank deficiencies in the parameter matrices follow from the characteristics of the limit point of the sequence of $(A, B, C)_n$.

Apart from the work of Stegeman (2006, 2008a, 2008b) and De Silva and Lim (2006) no general criteria are known to indicate whether degeneracies will occur while running a CP algorithm for the CP model for a particular array $\underline{X}$. However, our result may help research in this direction, since it shows the importance of determining whether a CP model has an optimal solution or not.

For the general situation considered in this paper, we may distinguish the following four cases with respect to Corollary 2 and the k-ranks k_A, k_B, and k_C in Kruskal's condition (2). Case 1, $k_A = k_B = k_C = R$ and Case 2, $k_A = k_B = R$, $k_C < R$, are frequently encountered in empirical applications of CANDECOMP/PARAFAC. In Case 1, all parameter matrices are nearly rank deficient for large n by Corollary 2. In Case 2, A_n and B_n are nearly rank deficient for large n by Corollary 2. Case 3, $k_A = R$, $k_B < R$, $k_C < R$, is less common. In this case Corollary 2 is nontrivial in the sense that it implies that A_n is nearly rank deficient for large n. To our best knowledge Case 4, $k_A < R$, $k_B < R$, $k_C < R$, has not been encountered in an empirical setting, but it was considered numerically (Harshman, 1972) and algebraically (Kruskal, 1976). In this case it is obvious that the parameter matrices are singular and that the conclusion of Corollary 2 is trivial.

Note that by multiplications of orthonormal commutation matrices it can be arranged that the order in the Kronecker product in (5) is altered (Magnus & Neudecker, 1979), so that the role played by A_n, B_n, or C_n is interchanged. Hence, the four cases above cover all possibilities.

In order to guarantee that CP has an optimal solution, one can impose orthogonality or non-negativity constraints on the parameter matrices (see Theorem 2 and Lim, 2005). Also, leaving out one data slice in one of the modes or changing the preprocessing scheme may overcome the problems of degeneracies.

References

Andersen, A.H., & Rayens, W.S. (2004). Structure-seeking multilinear methods for the analysis of fMRI data. *NeuroImage*, *22*, 728–739.

Beckmann, C.F., & Smith, S.M. (2005). Tensorial extensions of independent component analysis for multisubject FMRI analysis. *NeuroImage*, *25*, 294–311.

Bro, R., & De Jong, S. (1997). A Fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics*, *11*, 393–401.

Carroll, J.D., & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart–Young' decomposition. *Psychometrika*, *35*, 283–319.

DeSarbo, W.S., & Carroll, J.D. (1985). Three-way metric unfolding via alternating weighted least squares. *Psychometrika*, *50*, 275–300.

De Silva, V., & Lim, L.-H. (2006). *Tensor rank and the ill-posedness of the best low-rank approximation problem*. (SCCM Technical Report, 06-06). Preprint.

Harshman, R.A. (1970). Foundations of the PARAFAC procedure: models and conditions for an "exploratory" multimodal factor analysis. *UCLA Working Papers in Phonetics*, *16*, 1–84.

Harshman, R.A. (1972). Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA Working Papers in Phonetics*, *22*, 111–117.

Harshman, R.A., & DeSarbo, W.S. (1984). An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques. In H.G. Law, C.W. Snyder, J.A. Hattie, & R.P. McDonald (Eds.), *Research methods for multi-mode data analysis* (pp. 602–642). New York: Praeger.

Harshman, R.A., & Lundy, M.E. (1984). Data preprocessing and the Extended PARAFAC model. In H.G. Law, C.W. Snyder, J.A. Hattie, & R.P. McDonald (Eds.), *Research methods for multi-mode data analysis* (pp. 216–284). New York: Praeger.

Harshman, R.A., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America*, *62*, 693–707.

Hopke, P.K., Paatero, P., Jia, H., Ross, R.T., & Harshman, R.A. (1998). Three-way (Parafac) factor analysis: examination and comparison of alternative computational methods as applied to ill-conditioned data. *Chemometrics and Intelligent Laboratory Systems*, *43*, 25–42.

Kiers, H.A.L. (1989a). A computational short-cut for INDSCAL with orthonormality constraints on positive semi-definite matrices of lower rank. *Computational Statistics Quarterly*, *2*, 119–135.

Kiers, H.A.L. (1989b). An alternating least squares algorithm for fitting the two- and three-way DEDICOM model and the INDIOSCAL model. *Psychometrika*, *54*, 515–521.

Kiers, H.A.L. (1991). Simple structure in component analysis techniques for mixtures of qualitative and quantitative variables. *Psychometrika*, *56*, 197–212.

Krijnen, W.P. (2006). Convergence of the sequence of parameters generated by alternating least squares algorithms. *Computational Statistics and Data Analysis*, *51*, 481–489.

Krijnen, W.P., & Ten Berge, J.M.F. (1992). A constrained PARAFAC method for positive manifold data. *Applied Psychological Measurement*, *16*, 295–305.

Krijnen, W.P., & Kiers, H.A.L. (1993). A confirmatory and an exploratory approach to simple structure in PARAFAC. In J.H.L. Oud & R.A.W. van Blokland-Vogelesang (Eds.), *Advances in longitudinal and multivariate analysis in the behavioral sciences* (pp. 165–177). Leiden: DSWO Press.

Kruskal, J.B. (1976). More factors than subjects, tests and treatments: an indeterminacy theorem for canonical decomposition & individual differences scaling. *Psychometrika*, *41*, 281–293.

Kruskal, J.B. (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.

Kruskal, J.B. (1989). Rank, decomposition, and uniqueness for 3-way and n-way arrays. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 7–18). Amsterdam: Elsevier.

Kruskal, J.B., Harshman, R.A., & Lundy, M.E. (1983). *Some relationships between Tucker's three-mode factor analysis and PARAFAC/CANDECOMP*. Paper presented at the annual meeting of the Psychometric Society, Los Angeles.

Kruskal, J.B., Harshman, R.A., & Lundy, M.E. (1985). *Several mathematical relationships between PARAFAC/CANDECOMP and three-mode factor analysis*. Paper presented at the annual meeting of the classification society, St. John's, Newfoundland.

Kruskal, J.B., Harshman, R.A., & Lundy, M.E. (1989). How 3-MFA data can cause degenerate PARAFAC solutions, among other relationships. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 115–122). Amsterdam: Elsevier.

Leurgans, S., & Ross, R.T. (1992). Multilinear models: applications in spectroscopy. *Statistical Science*, *7*, 289–319.

Lim, L.-H. (2005). Optimal solutions to non-negative Parafac/multilinear NMF always exist. Talk at the *Workshop on tensor decompositions and applications*, CIRM, Luminy, Marseille, France. Available online at http://www.etis.ensea.fr/~wtda.

Luenberger, D.G. (1969). *Optimization by vector space methods*. New York: Wiley.

Magnus, J.R., & Neudecker, H. (1979). The commutation matrix: some properties for applications. *The Annals of Statistics*, *7*, 381–394.

Meyer, J.P. (1980). Causal attribution for success and failure. A multivariate investigation of dimensionally, formation, and consequences. *Journal of Personality and Social Psychology*, *38*, 704–718.

Mitchell, B.C., & Burdick, D.S. (1994). Slowly converging PARAFAC sequences: swamps and two-factor degeneracies. *Journal of Chemometrics*, *8*, 155–168.

Ortega, J.M., & Rheinboldt, W.C. (1970). *Iterative solution of nonlinear equations in several variables*. San Diego: Academic Press.

Paatero, P. (1999). The multilinear engine—a table-driven least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *Journal of Computational and Graphical Statistics*, *8*, 854–888.

Paatero, P. (2000). Construction and analysis of degenerate PARAFAC models. *Journal of Chemometrics*, *14*, 285–299.

Rudin, W. (1976). *Principles of mathematical analysis* (3rd ed.). New York: McGraw-Hill.

Schur, J. (1911). Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *Journal für die Reine und Angewante Mathematiek*, *140*, 1–28.

Smilde, A. (1992). Three-way analysis: Problems and prospects. *Chemometrics and Intelligent Laboratory Systems*, *15*, 143–157.

Stegeman, A. (2006). Degeneracy in Candecomp/Parafac explained for $p \times p \times 2$ arrays of rank $p + 1$ or higher. *Psychometrika*, *71*, 483–501.

Stegeman, A. (2008a). Low-rank approximation of generic $p \times q \times 2$ arrays and diverging components in the Candecomp/Parafac model. *SIAM Journal on Matrix Analysis and Applications*, to appear.

Stegeman, A. (2008b). Degeneracy in Candecomp/Parafac and Indscal explained for several three-sliced arrays with a two-valued typical rank. *Psychometrika*, to appear.

Stegeman, A., & Sidiropoulos, N.D. (2007). On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition. *Linear Algebra and Its Applications*, *420*, 540–552.

Styan, G.P.H. (1973). Hadamard products and multivariate statistical analysis. *Linear Algebra and Its Applications*, *6*, 217–240.

Ten Berge, J.M.F., Kiers, H.A.L., & De Leeuw, J. (1988). Explicit CANDECOMP-PARAFAC solutions for a contrived $2 \times 2 \times 2$ array of rank three. *Psychometrika*, *53*, 579–584.

Ten Berge, J.M.F., Kiers, H.A.L., & Krijnen, W.P. (1993). The problem of negative saliences and asymmetry in IND-SCAL. *Journal of Classification*, *10*, 115–124.

Tomasi, G., & Bro, R. (2006). A comparison of algorithms for fitting the Parafac model. *Computational Statistics and Data Analysis*, *50*, 1700–1734.

Zijlstra, B.J.H., & Kiers, H.A.L. (2002). Degenerate solutions obtained from several variants of factor analysis. *Journal of Chemometrics*, *16*, 596–605.