21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

Nonexistence of best low-rank approximations for real-valued three-way arrays and what to do about it*

Alwin Stegeman[1]

***Abstract*— Nonexistence of a best rank-R approximation of a three-way array hampers the practical use of the Canonical Polyadic Decomposition (CPD) for exploratory data analysis. We present theoretical results on nonexistence of a best rank-R approximation and propose a method to overcome this problem by augmenting the CPD with one or more interaction terms.**

***Keywords*: tensor decomposition, low-rank approximation, Canonical Polyadic Decomposition, diverging components.**

***AMS subject classifications*: 15A69, 49M27, 62H25.**

I. TENSOR RANK AND CPD

The rank of an order-3 tensor or three-way array $\mathcal{Y} \in \mathbb{R}^{I \times J \times K}$ is defined as the smallest number of rank-1 arrays whose sum equals $\mathcal{Y}$. A three-way array has rank 1 if it is the outer vector product of three nonzero vectors. The outer vector product $\mathbf{Y} = \mathbf{a} \circ \mathbf{b} = \mathbf{a}\,\mathbf{b}^T$ is a rank-1 matrix (or order-2 tensor) with entries $y_{ij} = a_i\, b_j$. The outer vector product $\mathcal{Y} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$ is rank-1 tensor with entries $y_{ijk} = a_i\, b_j\, c_k$. We have

$$\text{rank}(\mathcal{Y}) = \min\{\, R : \ \mathcal{Y} = \sum_{r=1}^{R} (\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r) \,\} . \quad (1)$$

The problem of finding a best rank-R approximation to $\mathcal{Z} \in \mathbb{R}^{I \times J \times K}$ can be denoted as

$$\begin{aligned} &\text{Minimize} \qquad \|\mathcal{Z} - \mathcal{Y}\|_F \\ &\text{subject to} \quad \mathcal{Y} \in S_R(I,J,K)\,, \end{aligned} \quad (2)$$

where $S_R(I,J,K)$ denotes the rank-R set

$$S_R(I,J,K) = \{\mathcal{Y} \in \mathbb{R}^{I \times J \times K} : \ \text{rank}(\mathcal{Y}) \le R\}\,, \quad (3)$$

and $\|\cdot\|_F$ denotes the Frobenius norm (i.e., the square root of the sum-of-squares). Problem (2) is equivalent to finding a best-fitting CPD to $\mathcal{Z}$, where we write the CPD as

$$\sum_{r=1}^{R} g_r\,(\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r)\,, \quad (4)$$

with $g_r \ge 0$ and the vectors $\mathbf{a}_r, \mathbf{b}_r, \mathbf{c}_r$ having unit norm [9] [10] [6] [1]. This model has applications in chemometrics, signal processing, the behavioral sciences, algebraic complexity theory, and data mining in general [12]. An attractive feature of the CPD is its rotational uniqueness property under mild conditions, as opposed to two-way techniques as principal component analysis or factor analysis [4], [5]. For later use, we define the component matrices $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_R]$, $\mathbf{B} = [\mathbf{b}_1 \ldots \mathbf{b}_R]$, $\mathbf{C} = [\mathbf{c}_1 \ldots, \mathbf{c}_R]$, and the vector of weights $\mathbf{g} = (g_1 \ldots g_R)^T$. Also, we define the following more general decomposition due to [30] which is also known as multilinear or higher-order singular value decomposition (HOSVD) [2]:

$$\sum_{r=1}^{R}\sum_{p=1}^{P}\sum_{q=1}^{Q} g_{rpq}\,(\mathbf{a}_r \circ \mathbf{b}_p \circ \mathbf{c}_q)\,. \quad (5)$$

The HOSVD does not have the rotational uniqueness property of the CPD. Hence, to obtain an interpretable HOSVD solution, a rotation method should be applied.

II. NONEXISTENCE OF A BEST RANK-R APPROXIMATION

Unfortunately, a best rank-R approximation of $\mathcal{Z}$ need not exist for $R \ge 2$ because the set $S_R(I,J,K)$ is not closed [3]. In such a case, trying to compute a best rank-R approximation yields a rank-R sequence converging to a boundary point $\mathcal{X}$ of $S_R(I,J,K)$ with rank$(\mathcal{X}) > R$. As a result, while running the iterative CPD algorithm, the decrease of the objective function becomes very slow, and some (groups of) columns of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ become nearly linearly dependent, while the corresponding weights g_r increase without bound [14] [13]. This phenomenon is known as "diverging components" or "degenerate solutions" or "diverging rank-1 terms". Needless to say, diverging rank-1 terms should be avoided if an interpretation of the rank-1 terms is needed. In simulation studies with random $\mathcal{Z}$, diverging rank-1 terms occur very often [20] [22] [21] [25].

Nonexistence of a best rank-R approximation can be avoided by imposing constraints on the rank-1 terms in $(\mathbf{A},\mathbf{B},\mathbf{C})$. Imposing orthogonality constraints on (one of) the component matrices guarantees existence of a best rank-R approximation [7] [13], and the same is true for nonnegative $\mathcal{Z}$ under the restriction of nonnegative $\mathbf{A},\mathbf{B},\mathbf{C}$ [15]. Also, [16] show that constraining the magnitude of the inner products between pairs of columns of $\mathbf{A},\mathbf{B},\mathbf{C}$ guarantees existence of a best rank-R approximation. However, imposing constraints will not be suitable for all CPD applications. As an alternative to deal with diverging rank-1 terms, methods have been developed to obtain the limit point $\mathcal{X}$ of the diverging rank-R sequence and a sparse decomposition of $\mathcal{X}$ [29] [19] [25] [26].

There are not many theoretical results on the (non)existence of a best rank-R approximation for specific

*This work is supported by the Dutch Organisation for Scientific Research (NWO), VIDI grant 452-08-001.

[1]Alwin Stegeman is with the Heijmans Institute for Psychological Research, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands, a.w.stegeman@rug.nl, http://www.gmw.rug.nl/∼stegeman.

three-way arrays or sizes. It has been proven that $2 \times 2 \times 2$ arrays of rank 3 do not have a best rank-2 approximation [3]. Proofs of (non)existence of a best rank-R approximation for generic $I \times J \times 2$ arrays can be found in [22] [27]; see Table I. Although diverging rank-1 terms may also occur due to a bad choice of starting point for the iterative algorithm [18] [23], if trying many random starting points does not help, then this is strong evidence for nonexistence of a best rank-R approximation.

TABLE I
EXISTENCE OF A BEST RANK-R APPROXIMATION FOR GENERIC $I \times J \times 2$ ARRAYS $\mathcal{Z}$, WITH $I \geq J \geq 2$ AND $R \geq 2$ [22] [27].

size of $\mathcal{Z}$	rank($\mathcal{Z}$)	R	Best rank-R ?
$I = J$	$I+1$	$R \geq I+1$	always
$I = J$	$I+1$	$R = I$	zero volume
$I = J$	$I+1$	$R < I$	positive volume
$I = J$	I	$R \geq I$	always
$I = J$	I	$R < I$	positive volume
$I > J$	$\min(I, 2J)$	$R \geq \min(I, 2J)$	always
$I > J$	$\min(I, 2J)$	$\min(I, 2J) > R > J$	almost everywhere
$I > J$	$\min(I, 2J)$	$R = J$	positive volume
$I > J$	$\min(I, 2J)$	$R < J$	positive volume

III. FINDING THE LIMIT POINT AND ITS SPARSE DECOMPOSITION

The following approach to deal with diverging rank-1 terms is an alternative to imposing constraints in CPD. Let $\mathcal{Y}^{(n)}$ denote the array formed by the CPD $(\mathbf{A}^{(n)}, \mathbf{B}^{(n)}, \mathbf{C}^{(n)}, \mathbf{g}^{(n)})$ after the n-th iteration of a CPD algorithm. For data $\mathcal{Z}$ of rank larger than R, the array $\mathcal{Y}^{(n)}$ will converge to the boundary of the rank-R set $S_R(I, J, K)$. Indeed, if a CPD algorithm is designed to minimize $\|\mathcal{Z} - \mathcal{Y}^{(n)}\|_F$, then $\mathcal{Y}^{(n)}$ will move from within the rank-R set to a boundary point $\mathcal{X}$ of the rank-R set. We call $\mathcal{X}$ an *optimal boundary point* if $\|\mathcal{Z} - \mathcal{X}\|_F$ is minimal over all boundary points of the rank-R set. If there is no optimal boundary point $\mathcal{X}$ with rank less than or equal to R, then a best rank-R approximation to $\mathcal{Z}$ does not exist. In that case, the rank-R sequence $\mathcal{Y}^{(n)}$ converges to a limit $\mathcal{X}$ with rank larger than R and will feature diverging rank-1 terms [13]. Next, we consider the problem of finding the limit $\mathcal{X}$ and a nondiverging sparse decomposition of $\mathcal{X}$ that can be interpreted. We refer to this decomposition of $\mathcal{X}$ as the CP_{limit} decomposition.

Algorithms to find the limit $\mathcal{X}$ directly, whether it has rank R or larger, exist only for $R = 2$ [19], and for $I \times J \times 2$ arrays [29] [24]. These algorithms are fast and diverging rank-1 terms do not occur. For $R \geq 3$ and $I \times J \times K$ arrays with $\min(I, J, K) \geq 3$ such algorithms have not been found. As an alternative, [25] [26] proposes the following approach. Suppose trying to find a best rank-R approximation for $\mathcal{Z}$ results in diverging rank-1 terms and one is convinced that no best rank-R approximation exists. Then the form of the CP_{limit} decomposition of the limit $\mathcal{X}$ can be determined from the number of groups of diverging rank-1 terms in the CPD sequence $\mathcal{Y}^{(n)}$, and the numbers of diverging rank-1 terms in each group. That is, in each case, the form of the CP_{limit} decomposition is dictated by the mathematical results of [25] [26]. The nondiverging CP_{limit} decomposition of $\mathcal{X}$ can be found by fitting this form of decomposition to $\mathcal{Z}$, using initial values obtained from the CPD sequence $(\mathbf{A}^{(n)}, \mathbf{B}^{(n)}, \mathbf{C}^{(n)}, \mathbf{g}^{(n)})$.

IV. THE FORM OF THE CP_{limit} DECOMPOSITION

For the description of the form of the CP_{limit} decomposition, we need the following notation. A three-way array may be multiplied by a matrix in one of its modes. The multiplication of $\mathcal{Y} \in \mathbb{R}^{I \times J \times K}$ by matrices $\mathbf{S} \in \mathbb{R}^{I_2 \times I}$, $\mathbf{T} \in \mathbb{R}^{J_2 \times J}$, and $\mathbf{U} \in \mathbb{R}^{K_2 \times K}$, is denoted as $\mathcal{Y}_2 = (\mathbf{S}, \mathbf{T}, \mathbf{U}) \cdot \mathcal{Y}$. The result of the multiplication is an $\mathcal{Y}_2 \in \mathbb{R}^{I_2 \times J_2 \times K_2}$ with entries

$$y_{ijk}^{(2)} = \sum_{r=1}^{I} \sum_{p=1}^{J} \sum_{q=1}^{K} s_{ir}\, t_{jp}\, u_{kq}\, y_{rpq}\,, \qquad (6)$$

where s_{ir}, t_{jp}, and u_{kq} are entries of $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{U}$, respectively. The HOSVD (5) can be written as $(\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathcal{G}$, where $\mathcal{G} \in \mathbb{R}^{R \times P \times Q}$ has entries g_{rpq} and is known as the *core array*. Analogously, the CPD (4) can be written as $(\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathcal{D}_R$, where $\mathcal{D}_R \in \mathbb{R}^{R \times R \times R}$ has entries $d_{rrr} = g_r$ and zeros elsewhere. Hence, $\mathcal{D}_R$ is a three-way generalization of a diagonal matrix.

Next, we introduce the general form of the CP_{limit} decomposition of the limit $\mathcal{X}$. For $R = 2$ and two diverging rank-1 terms, there exists a decomposition $\mathcal{X} = (\mathbf{S}, \mathbf{T}, \mathbf{U}) \cdot \mathcal{G}$, with $\mathcal{G}$ given by [3]

$$\mathcal{G} = \left[\begin{array}{cc|cc} * & 0 & 0 & * \\ 0 & * & 0 & 0 \end{array}\right], \qquad (7)$$

where the 2×2 frontal slices of $\mathcal{G}$ are given side by side, and $*$ denotes a nonzero entry. Here, $\text{rank}(\mathcal{X}) = \text{rank}(\mathcal{G}) = 3$, and CP_{limit} has three rank-1 terms, which is one more than the CPD with $R = 2$.

For $R = 3$ and three diverging rank-1 terms, there exists a decomposition $\mathcal{X} = (\mathbf{S}, \mathbf{T}, \mathbf{U}) \cdot \mathcal{G}$, with $\mathcal{G}$ given by [25]

$$\left[\begin{array}{ccc|ccc|ccc} * & 0 & 0 & 0 & * & 0 & 0 & 0 & * \\ 0 & * & 0 & 0 & 0 & * & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]. \qquad (8)$$

We have $\text{rank}(\mathcal{X}) = \text{rank}(\mathcal{G}) = 5$.

For $R = 4$ and four diverging rank-1 terms, there exists a decomposition $\mathcal{X} = (\mathbf{S}, \mathbf{T}, \mathbf{U}) \cdot \mathcal{G}$, with $\mathcal{G}$ given by [26]

$$\left[\begin{array}{cccc|cccc|cccc|cccc} * & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * \\ 0 & * & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]. \qquad (9)$$

We have $\mathrm{rank}(\mathcal{X}) = \mathrm{rank}(\mathcal{G}) \geq 7$.

For groups of more than four diverging rank-1 terms, the decomposition form for the limit $\mathcal{X}$ may be obtained analogous to the proof for $R = 4$ in [26]. Considering the forms of $\mathcal{G}$ above, it may not be surprising that these can be seen as three-way generalizations of Jordan blocks [26].

When not all rank-1 terms are diverging, or multiple groups of diverging rank-1 terms occur, [25] [26] proposes the following $\mathrm{CP_{limit}}$ decomposition of $\mathcal{X}$. Each group of d_j diverging rank-1 terms converges to its own limit $\mathcal{X}_j$ with its respective $\mathrm{CP_{limit}}$ decomposition $\mathcal{X}_j = (\mathbf{S}_j, \mathbf{T}_j, \mathbf{U}_j) \cdot \mathcal{G}_j$, where $\mathcal{G}_j$ has size $d_j \times d_j \times d_j$. For $d_j \in \{2, 3, 4\}$ the form of the $\mathrm{CP_{limit}}$ decomposition of $\mathcal{X}_j$ is as described above. Each nondiverging rank-1 term stays nondiverging in the limit, and corresponds to a $d_j = 1$. The complete $\mathrm{CP_{limit}}$ decomposition of $\mathcal{X}$ is given by $\mathcal{X} = \sum_{j=1}^{m} \mathcal{X}_j = \sum_{j=1}^{m} (\mathbf{S}_j, \mathbf{T}_j, \mathbf{U}_j) \cdot \mathcal{G}_j$, and can be seen as a three-way generalization of the Jordan canonical form for matrices [26].

As an example, suppose $R = 3$ and we have one group of two diverging components. The limit $\mathcal{X}$ then has $\mathrm{CP_{limit}}$ decomposition $\mathcal{X} = \mathcal{X}_1 + \mathcal{X}_2$, where $\mathcal{X}_1 = (\mathbf{S}_1, \mathbf{T}_1, \mathbf{U}_1) \cdot \mathcal{G}_1$ with $\mathcal{G}_1$ as in (7) being the limit of the two diverging rank-1 terms, and $\mathcal{X}_2 = g_3 (\mathbf{s}_3 \circ \mathbf{t}_3 \circ \mathbf{u}_3)$ being the limit of the nondiverging rank-1 term. Hence, the $\mathrm{CP_{limit}}$ decomposition of $\mathcal{X}$ is of the form

$$\mathcal{X} = g_{111} (\mathbf{s}_1 \circ \mathbf{t}_1 \circ \mathbf{u}_1) + g_{221} (\mathbf{s}_2 \circ \mathbf{t}_2 \circ \mathbf{u}_1) + g_{122} (\mathbf{s}_1 \circ \mathbf{t}_2 \circ \mathbf{u}_2) + g_3 (\mathbf{s}_3 \circ \mathbf{t}_3 \circ \mathbf{u}_3) \,. \tag{10}$$

The assumption of [25] [26] that groups of diverging rank-1 terms each converge to their respective limits, is confirmed by simulation studies. The limit $\mathcal{X}$ and its $\mathrm{CP_{limit}}$ decomposition may be obtained by fitting the appropriate $\mathrm{CP_{limit}}$ form to the data $\mathcal{Z}$. For this, the alternating least squares algorithm [11] for fitting a constrained HOSVD (5) can be used. Matlab codes are available online for finding the correct form of $\mathrm{CP_{limit}}$, obtaining initial values, and fitting it to a dataset that yields diverging rank-1 terms; see `http://www.gmw.rug.nl/~stegeman`.

V. Application to TV-ratings data

We apply the approach above to a well-known three-way dataset for which diverging rank-1 terms occur. The data consist of ratings of 15 American TV shows on 16 rating scales, made by 40 subjects in 1981. The subjects were introductory psychology students at the University of Western Ontario, Canada, who were familiar with the shows. After deleting subjects with missing data, we keep 30 persons. The data are previously analyzed by [17] and also feature in [8].

Appropriate component models for the TV-ratings data are the CPD with $R = 3$ or the HOSVD with $R = P = 3$ and $Q \in \{1, 2, 3\}$ [28]. Fitting the CPD with $R = 2$ yields a "Humor" component and a "Sensitivity" component. Fitting the CPD with $R = 3$ yields a similar "Humor" component and two diverging rank-1 terms. Imposing orthogonality among the TV show components yields "Humor", "Sensitivity", and "Violence" components. Three similar components are obtained by fitting the HOSVD models, which also impose orthogonality. However, orthogonality between "Sensitivity" and "Violence" does not seem intuitive. Also, fitting the HOSVD involves applying a rotation to an interpretable solution. This results in many small rank-1 terms next to the three main components.

Next, we fit the $\mathrm{CP_{limit}}$ decomposition for $R = 3$ and two diverging rank-1 terms, which is of the form (10) and does not impose orthogonality. The same nondiverging "Humor" component is obtained (term four in (10)), as well as "Sensitivity" and "Violence" components (terms one and two in (10)). The third term in (10) is small and represents an interaction between "Sensitivity" and "Violence". Hence, the $\mathrm{CP_{limit}}$ decomposition is interpretable and does not impose the unintuitive constraint of orthogonality [28]. Also, contrary to the HOSVD models, $\mathrm{CP_{limit}}$ contains only one small rank-1 term. All appropriate models have fit percentages around 50, so the choice for a model boils down to the form of the model itself. It is clear that $\mathrm{CP_{limit}}$ has the most desirable properties for the TV-ratings data.

The uniqueness properties of the $\mathrm{CP_{limit}}$ decomposition in (10) are not the same as for the CPD. Although the nondiverging rank-1 term is unique, there is some transformational freedom within the first three terms of (10) [25] [28]. However, this transformational freedom can be fixed by applying standard rotation criteria and this does change the interpretation of $\mathrm{CP_{limit}}$ for the TV-ratings data [28].

References

[1] 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.

[2] De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, **21**, 1253–1278.

[3] De Silva, V., & Lim, L.-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, **30**, 1084–1127.

[4] Domanov, I., & De Lathauwer, L. (2013a). On the uniqueness of the canonical polyadic decomposition of third-order tensors - Part I: basic results and uniqueness of one factor matrix. *SIAM Journal on Matrix Analysis and Applications*, **34**, 855–875.

[5] Domanov, I., & De Lathauwer, L. (2013b). On the uniqueness of the canonical polyadic decomposition of third-order tensors - Part II: uniqueness of the overall decomposition. *SIAM Journal on Matrix Analysis and Applications*, **34**, 876–903.

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

[7] Harshman, R.A., & Lundy, M.E. (1984). Data preprocessing and the extended Parafac model, pp. 216–284 in: *Research Methods for Multimode Data Analysis*, H.G. Law, C.W. Snyder Jr., J.A. Hattie and R.P. McDonald (Eds.), Praeger, New York.

[8] Harshman, R.A. (2004). The problem and nature of degenerate solutions or decompositions of 3-way arrays. Talk at the *Tensor Decompositions Workshop*, Palo Alto, USA.

[9] Hitchcock, F.L. (1927). The expression of a tensor or a polyadic as a sum of products, *Journal of Mathematics and Physics*, **6**, 164–189.

[10] Hitchcock, F.L. (1927). Multiple invariants and generalized rank of a p-way matrix or tensor, *Journal of Mathematics and Physics*, **7**, 39–70.

[11] Kiers, H.A.L., & Smilde, A.K. (1998). Constrained three-mode factor analysis as a tool for parameter estimation with second-order instrumental data. *Journal of Chemometrics*, **12**, 125–147.

[12] Kolda, T.G., & Bader, B.W. (2009). Tensor decompositions and applications. *SIAM Review*, **51**, 455–500.

[13] Krijnen, W.P., Dijkstra, T.K., & Stegeman, A. (2008). On the non-existence of optimal solutions and the occurrence of "degeneracy" in the Candecomp/Parafac model. *Psychometrika*, **73**, 431–439.

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

[15] Lim, L.-H., & Comon, P. (2009). Nonnegative approximations of nonnegative tensors. *Journal of Chemometrics*, **23**, 432–441.

[16] Lim, L.-H., & Comon, P. (2010). Multiarray signal processing: tensor decomposition meets compressed sensing. *Comptes-Rendus de l'Académie des Sciences, Mécanique*, **338**, 311–320.

[17] Lundy, M.E., Harshman, R.A., & Kruskal, J.B. (1989). A two-stage procedure incorporating good features of both trilinear and quadrilinear models, pp. 123–130 in: *Multiway Data Analysis*, R. Coppi and S. Bolasco (Eds.), North-Holland.

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

[19] Rocci, R., & Giordani, P. (2010). A weak degeneracy revealing decomposition for the Candecomp/Parafac model. *Journal of Chemometrics*, **24**, 57–66.

[20] 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.

[21] Stegeman, A. (2007). Degeneracy in Candecomp/Parafac explained for several three-sliced arrays with a two-valued typical rank. *Psychometrika*, **72**, 601–619.

[22] Stegeman, A. (2008). 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*, **30**, 988–1007.

[23] Stegeman, A. (2009). Using the Simultaneous Generalized Schur Decomposition as a Candecomp/Parafac algorithm for ill-conditioned data. *Journal of Chemometrics*, **23**, 385–392.

[24] Stegeman, A. (2010). The Generalized Schur Decomposition and the rank-R set of real $I \times J \times 2$ arrays. Technical Report, available online as `arXiv:1011.3432`

[25] Stegeman, A. (2012). Candecomp/Parafac - from diverging components to a decomposition in block terms. *SIAM Journal on Matrix Analysis and Applications*, **33**, 291–316.

[26] Stegeman, A. (2013). A three-way Jordan canonical form as limit of low-rank tensor approximations. *SIAM Journal on Matrix Analysis and Applications*, **34**, 624–650.

[27] Stegeman, A. (2013). On the (non)existence of best low-rank approximations of generic $I \times J \times 2$ arrays. Technical Report, available online as `arXiv:1309.5727`

[28] Stegeman, A. (2014). Finding the limit of diverging components in three-way Candecomp/Parafac - a demonstration of its practical merits. *Computational Statistics & Data Analysis*, **75**, 203–216.

[29] Stegeman, A., & De Lathauwer, L. (2009). A method to avoid diverging components in the Candecomp/Parafac model for generic $I \times J \times 2$ arrays. *SIAM Journal on Matrix Analysis and Applications*, **30**, 1614–1638.

[30] Tucker, L.R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, **31**, 279–311.