Computational Statistics and Data Analysis 75 (2014) 203–216

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Finding the limit of diverging components in three-way Candecomp/Parafac—A demonstration of its practical merits

Alwin Stegeman *

Heijmans Institute for Psychological Research, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 20 June 2013
Received in revised form 11 February 2014
Accepted 11 February 2014
Available online 18 February 2014

Keywords:
Candecomp
Parafac
Diverging components
Degeneracy
Uniqueness

ABSTRACT

Three-way Candecomp/Parafac (CP) is a three-way generalization of principal component analysis (PCA) for matrices. Contrary to PCA, a CP decomposition is rotationally unique under mild conditions. However, a CP analysis may be hampered by the non-existence of a best-fitting CP decomposition with $R \geq 2$ components. In this case, fitting CP to a three-way data array results in diverging CP components. Recently, it has been shown that this can be solved by fitting a decomposition with several interaction terms, using initial values obtained from the diverging CP decomposition. The new decomposition is called CP_{limit}, since it is the limit of the diverging CP decomposition. The practical merits of this procedure are demonstrated for a well-known three-way dataset of TV-ratings. CP_{limit} finds main components with the same interpretation as Tucker models or when imposing orthogonality in CP. However, CP_{limit} has higher joint fit of the main components than Tucker models, contains only one small interaction term, and does not impose the unnatural constraint of orthogonality. The uniqueness properties of the CP_{limit} decomposition are discussed in detail.

1. Introduction

1.1. Three-way data and the CP decomposition

Three-way data are data that can be arranged in a three-dimensional array or three-way array. Such data is found in many different contexts. For example: scores on various anxiety scales of a number of individuals in various situations; scores on various competences of a number of workers by several different assessors; scores on food quality indicators of a number of food products by several different judges; and a number of consecutive fMRI measurements for different areas of the brain for different individuals. The three sets of entities associated with three-way datasets are called the three *modes* of the array.

We denote a three-way array as $\mathcal{Z}$ and its entries as z_{ijk}, where the subscripts correspond to row i, column j, and frontal slice k. An $I \times J \times K$ array $\mathcal{Z}$ has frontal slices $\mathbf{Z}_k$ of size $I \times J$. Entry z_{ijk} is entry (i, j) of matrix $\mathbf{Z}_k$.

In this paper, we consider the three-way Candecomp/Parafac (CP) decomposition of a three-way array $\mathcal{Z}$. The CP model is:

$$\mathcal{Z} = \sum_{r=1}^{R} g_r \, (\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r) + \mathcal{E}, \tag{1}$$

* Tel.: +31 50 363 6193; fax: +31 50 363 6304.
E-mail address: a.w.stegeman@rug.nl.
URL: http://www.gmw.rug.nl/~stegeman.

http://dx.doi.org/10.1016/j.csda.2014.02.010
0167-9473/

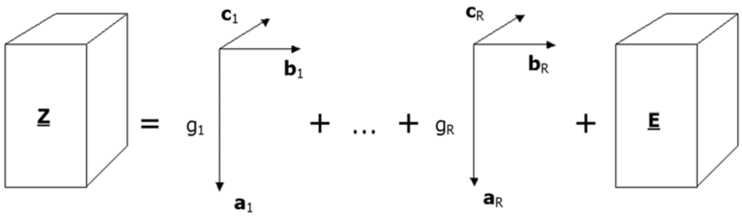

Fig. 1. Graphical representation of the CP model.

where $\circ$ denotes the outer vector product. For column vectors $\mathbf{a}$ and $\mathbf{b}$, the matrix $\mathbf{a}\circ\mathbf{b} = \mathbf{ab}^T$ has (i, j) entry $a_i b_j$. For column vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$, the three-way array $\mathbf{a}\circ\mathbf{b}\circ\mathbf{c}$ has (i, j, k) entry $a_i b_j c_k$. The rank of $\mathcal{Z}$ is defined as the minimal R for which $\mathcal{Z}$ satisfies (1) with all zero residual array $\mathcal{E}$. A three-way array has rank 1 if it is of the form $\mathbf{a}\circ\mathbf{b}\circ\mathbf{c}$ for nonzero vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$. A graphical representation of the CP model is depicted in Fig. 1.

The vectors $\mathbf{a}_r$, $\mathbf{b}_r$, and $\mathbf{c}_r$ in (1) are assumed to have a fixed length, i.e., $\mathbf{a}_r^T\mathbf{a}_r = l_a$, $\mathbf{b}_r^T\mathbf{b}_r = l_b$, and $\mathbf{c}_r^T\mathbf{c}_r = l_c$ for fixed positive numbers l_a, l_b, l_c. The weights g_r in (1) are assumed to be positive. We write a CP decomposition as $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{g})$, where $\mathbf{A} = [\mathbf{a}_1 \mid \dots \mid \mathbf{a}_R]$, $\mathbf{B} = [\mathbf{b}_1 \mid \dots \mid \mathbf{b}_R]$, $\mathbf{C} = [\mathbf{c}_1 \mid \dots \mid \mathbf{c}_R]$, and $\mathbf{g}^T = (g_1 \;\dots\; g_R)^T$. For a fixed number R of components, a best-fitting CP decomposition is found by an iterative CP algorithm that tries to minimize the sum-of-squares of the residual array $\mathcal{E}$. For an overview and comparison of CP algorithms see e.g. Tomasi and Bro (2006), and Comon et al. (2009). Fitting a CP decomposition to $\mathcal{Z}$ is equivalent to finding a best rank-R approximation to $\mathcal{Z}$.

The CP model was proposed independently by Carroll and Chang (1970) and Harshman (1970) as a method for exploratory component analysis of multi-way arrays. However, its origins in mathematics date back to Hitchcock (1927a,b). We consider the real-valued CP model, i.e., we assume the data array $\mathcal{Z}$ and the component matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ to be real-valued. The real-valued CP model is used in psychology and chemistry; see Harshman and Lundy (1994), Siciliano and Mooijaart (1997), Kiers and Van Mechelen (2001), Kroonenberg (2008), and Smilde et al. (2004). Complex-valued applications of CP occur in signal processing and telecommunications research; see e.g. Sidiropoulos et al. (2000a,b), and De Lathauwer and Castaing (2007). Here, the decompositions are mostly used to separate signal sources from an observed mixture of signals. A general overview of applications of CP and related decompositions can be found in Kolda and Bader (2009) and Acar and Yener (2009).

A matrix form of the CP model is:

$$\mathbf{Z}_k = \mathbf{A}\,\mathbf{C}_k\,\mathbf{G}\,\mathbf{B}^T + \mathbf{E}_k, \quad k = 1, \dots, K, \tag{2}$$

where $\mathbf{C}_k$ is the $R \times R$ diagonal matrix with row k of $\mathbf{C}$ as diagonal, $\mathbf{G}$ is the diagonal matrix with $\mathbf{g}$ as diagonal, and $\mathbf{E}_k$ is frontal slice k of array $\mathcal{E}$. For $K = 1$, the model has the same form as principal component analysis (PCA).

For an individual entry of $\mathcal{Z}$, the CP model is written as

$$z_{ijk} = \sum_{r=1}^{R} g_r\, a_{ir} b_{jr} c_{kr} + e_{ijk}. \tag{3}$$

For later use, we introduce the following more general component model for three-way arrays due to Tucker (1966):

$$\mathcal{Z} = \sum_{r=1}^{R}\sum_{p=1}^{P}\sum_{q=1}^{Q} g_{rpq}\,(\mathbf{a}_r \circ \mathbf{b}_p \circ \mathbf{c}_q) + \mathcal{E}. \tag{4}$$

The Tucker model (4) reduces to the CP model (1) if the weights g_{rpq} are zero for $(r, p, q) \neq (r, r, r)$. We will refer to terms with $(r, p, q) \neq (r, r, r)$ as *interaction terms*. For an individual entry z_{ijk}, the Tucker model is written as

$$z_{ijk} = \sum_{r=1}^{R}\sum_{p=1}^{P}\sum_{q=1}^{Q} g_{rpq}\, a_{ir} b_{jp} c_{kq} + e_{ijk}. \tag{5}$$

The most attractive feature of CP is that, for fixed residuals, the decomposition is unique up to permutation and scaling under mild conditions. That is, the only alternative CP decompositions yielding the same fitted array are obtained by permuting the R terms in (1) or by rescaling and counterscaling the vectors within each term $(\mathbf{a}_r \circ \mathbf{b} \circ \mathbf{c}_r)$. Sufficient uniqueness conditions for three-way CP can be found in Domanov and De Lathauwer (2013a,b) and the references therein. This uniqueness property does not hold for the Tucker model (4), nor for PCA.

1.2. Diverging CP components

A potential problem in the application of CP is that a best-fitting CP model does not exist for every data array $\mathcal{Z}$ (De Silva and Lim, 2008). In that case, trying to fit CP yields several (groups of) diverging components. In a group of diverging components, the corresponding columns in $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ become nearly linearly dependent and the corresponding weights

g_r become arbitrarily large as the CP algorithm keeps running (Krijnen et al., 2008). However, the sum of the diverging components remains relatively small and contributes to a better fit of the CP model. Early conjectures of non-existence of a best-fitting CP model as the cause of diverging components appeared in Kruskal et al. (1989). In practice, the presence of diverging component also results in very slow convergence of the CP algorithm.

In almost all observed cases of diverging components, the columns of **A**, **B**, **C** in the same group of diverging components become nearly equal up to sign (Stegeman and De Lathauwer, 2011). Hence, the contributions of some diverging components are nearly cancelled out by the contributions of other diverging components. For example, for a group of two diverging components s and t we may have $\mathbf{a}_s \approx \mathbf{a}_t$, $\mathbf{b}_s \approx -\mathbf{b}_t$, $\mathbf{c}_s \approx \mathbf{c}_t$, and $g_s \approx g_t$, where g_s and g_t are considerably large in magnitude. For examples of groups of more than two diverging components, or more groups of diverging components, see Stegeman (2006, 2012).

Unfortunately, there are only few general results indicating for which data array $\mathcal{Z}$ there is and is not a best-fitting CP model with R components. We know that the case $R = 1$ always has a best-fitting CP model, and that $2 \times 2 \times 2$ arrays of rank 3 have no best-fitting CP model with $R = 2$ (De Silva and Lim, 2008). Stegeman (2008, 2013b) proves the existence or non-existence of best-fitting CP models for generic $I \times J \times 2$ arrays, for all values of I, J, R. Stegeman (2006) describes how diverging components occur for $I \times I \times 2$ arrays in case no best-fitting CP model exists. Stegeman (2007) extends this approach to several $I \times J \times 3$ arrays.

Diverging components cannot be interpreted and may thus be a serious problem in the practical use of CP. In simulation studies with randomly sampled data $\mathcal{Z}$ (from a continuous distribution), diverging components occur very often, with percentages of up to 50, 60, or even 100 (Stegeman, 2006, 2007, 2008, 2012, 2013a).

Diverging components can be avoided by imposing constraints in CP such that a best-fitting CP model exists. Examples are: orthogonality constraints on the columns of (one or more of) **A**, **B**, **C** (Harshman and Lundy, 1984; Krijnen et al., 2008), nonnegativity constraints on **A**, **B**, **C** when the data array is nonnegative (Lim and Comon, 2009), or constraining the magnitude of the inner products between pairs of columns of **A**, **B**, **C** (Lim and Comon, 2010). However, these constraints are not suitable for all applications of CP.

A different approach to deal with diverging components is as follows. Let $\mathcal{Y}^{(n)}$ denote the array formed by the CP decomposition $(\mathbf{A}^{(n)}, \mathbf{B}^{(n)}, \mathbf{C}^{(n)}, \mathbf{g}^{(n)})$ after the n-th iteration of a CP algorithm. For data $\mathcal{Z}$ of rank larger than R, the array $\mathcal{Y}^{(n)}$ will converge to the boundary of the set of $I \times J \times K$ arrays with rank at most R (i.e., a perfect fitting CP decomposition with R components). Indeed, if a CP algorithm is designed to minimize the sum-of-squares difference between $\mathcal{Y}^{(n)}$ and $\mathcal{Z}$, then $\mathcal{Y}^{(n)}$ will move from within the set of rank-R arrays to a boundary point $\mathcal{X}$ of that set. We call $\mathcal{X}$ an *optimal boundary point* if it has minimal sum-of-squares difference with $\mathcal{Z}$, for 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 there is no best-fitting CP model for $\mathcal{Z}$. 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 components. The approach we consider in this paper tries to find the limit $\mathcal{X}$ and its nondiverging decomposition, which has some interaction terms as in (4). We call this decomposition of $\mathcal{X}$ the $\mathrm{CP}_{\mathrm{limit}}$ decomposition.

Note that diverging components occur whenever $\mathcal{Y}^{(n)}$ converges to a limit $\mathcal{X}$ with rank larger than R. In the sequel, we assume that $\mathcal{X}$ is an optimal boundary point with rank larger than R, and that no optimal boundary points exist with rank less than or equal to R. If the latter is not true, then a best-fitting CP model does exist and the diverging components can be avoided by choosing suitable initial values for the CP algorithm (Paatero, 2000; Stegeman, 2009).

1.3. Outline of the paper

In this paper, we discuss recently proposed algorithms to find the optimal boundary point $\mathcal{X}$ and its nondiverging $\mathrm{CP}_{\mathrm{limit}}$ decomposition. Algorithms to find $\mathcal{X}$ directly, whether it has rank R or larger, exist only for $R = 2$ (Rocci and Giordani, 2010), and for $I \times J \times 2$ arrays (Stegeman and De Lathauwer, 2009; Stegeman, 2010). These algorithms are fast and diverging components 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, Stegeman (2012, 2013a) proposes the following approach. Suppose trying to find a best-fitting CP model for $\mathcal{Z}$ results in diverging components and one is convinced that no best-fitting CP model exists. Then the form of the $\mathrm{CP}_{\mathrm{limit}}$ decomposition of the limit $\mathcal{X}$ can be determined from the number of groups of diverging components in the CP sequence $\mathcal{Y}^{(n)}$, and the numbers of diverging components in each group. That is, in each case, the form of the $\mathrm{CP}_{\mathrm{limit}}$ decomposition is dictated by the mathematical results of Stegeman (2012, 2013a). The nondiverging $\mathrm{CP}_{\mathrm{limit}}$ decomposition of $\mathcal{X}$ can be found by fitting this form of decomposition to $\mathcal{Z}$, using initial values obtained from the CP sequence $\mathcal{Y}^{(n)}$. A more detailed discussion of this approach follows in Section 2.

In Section 3, we apply this method to a previously analyzed dataset of TV-ratings, for which one group of two diverging components occurs when trying to fit CP with $R = 3$ (Lundy et al., 1989; Harshman, 2004). The method yields a $\mathrm{CP}_{\mathrm{limit}}$ decomposition of $\mathcal{X}$ that contains three main components. Their interpretation is the same as the components found by CP with an orthogonality constraint, various Tucker decompositions, and the analysis of Lundy et al. (1989). We explain in detail why the $\mathrm{CP}_{\mathrm{limit}}$ solution is to be preferred to the solutions of the other models. This is the first application of the approach of Stegeman (2012, 2013a) to a real-life dataset.

In Section 4, we study in detail the uniqueness properties of the $\mathrm{CP}_{\mathrm{limit}}$ decomposition of $\mathcal{X}$, which is not a CP decomposition. We show that the $\mathrm{CP}_{\mathrm{limit}}$ decomposition allows some more transformational freedom than CP, which can be fixed by using standard rotation criteria. For the TV-ratings data, this does not affect the interpretation of the solution obtained in Section 3. Finally, in Section 5, we provide a discussion of our findings.

A CP_{limit} decomposition contains only few more terms than the corresponding CP decomposition. As we explain in Section 2, when only groups of two diverging components are present, the CP_{limit} decomposition is the smallest nondiverging decomposition with R components in each mode. Therefore, CP_{limit} it is easier to interpret than a Tucker decomposition. Also, the uniqueness properties of CP_{limit} are more attractive than the complete non-uniqueness of Tucker. The CP_{limit} approach is a good alternative when imposing constraints in CP (to avoid diverging components) is not appropriate.

2. The optimal boundary point and its decomposition

Here, we discuss the approach of Stegeman (2012, 2013a) in more detail. First, however, we introduce some notation. A three-way array may be multiplied by a matrix in one of its modes. The multiplication of $I \times J \times K$ array $\mathcal{Y}$ by matrices $\mathbf{S}$ ($I_2 \times I$), $\mathbf{T}$ ($J_2 \times J$), and $\mathbf{U}$ ($K_2 \times K$), is denoted as $\mathcal{Y}_2 = (\mathbf{S}, \mathbf{T}, \mathbf{U}) \cdot \mathcal{Y}$. The result of the multiplication is an $I_2 \times J_2 \times K_2$ array $\mathcal{Y}_2$ with entries

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

where s_{ir}, t_{jp}, and u_{kq} are entries of $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{U}$, respectively. Note that multiplication of a three-way array by vectors is defined in the same way. In that case, some or all of I_2, J_2, K_2 are equal to 1.

Using this notation, the Tucker model (4) can be written as $(\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathcal{G}$, where the $R \times P \times Q$ array $\mathcal{G}$ has entries g_{rpq}. Array $\mathcal{G}$ is known as the *core array*. Analogously, the CP decomposition $\sum_{r=1}^{R} g_r\, (\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r)$ can be written as $(\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathcal{D}_R$, where $\mathcal{D}_R$ is the $R \times R \times R$ array with entries $d_{rrr} = g_r$ and zeros elsewhere. Hence, $\mathcal{D}_R$ is a three-way generalization of a diagonal matrix.

Recall that we are interested in a nondiverging CP_{limit} decomposition of the limit $\mathcal{X}$ of a rank-R CP sequence $\mathcal{Y}^{(n)}$ featuring diverging components. Here, the limit $\mathcal{X}$ has rank larger than R. For $R = 2$, it has been shown by De Silva and Lim (2008) that $\mathcal{X}$ has rank 3, and a decomposition exists of the form $\mathcal{X} = (\mathbf{S}, \mathbf{T}, \mathbf{U}) \cdot \mathcal{G}$, with

$$\mathcal{G} = \left[\begin{array}{cc|cc} g_{111} & 0 & 0 & g_{122} \\ 0 & g_{221} & 0 & 0 \end{array} \right], \tag{7}$$

where the 2×2 frontal slices of $\mathcal{G}$ are given side by side. The CP_{limit} decomposition $\mathcal{X} = (\mathbf{S}, \mathbf{T}, \mathbf{U}) \cdot \mathcal{G}$ can be written as

$$\mathcal{X} = g_{111}\, (\mathbf{s}_1 \circ \mathbf{t}_1 \circ \mathbf{u}_1) + g_{221}\, (\mathbf{s}_2 \circ \mathbf{t}_2 \circ \mathbf{u}_1) + g_{122}\, (\mathbf{s}_1 \circ \mathbf{t}_2 \circ \mathbf{u}_2). \tag{8}$$

Hence, instead of the two terms in CP we now have three terms, two of which are interaction terms. For $R = 2$, the limit $\mathcal{X}$ and its CP_{limit} decomposition can be found directly (i.e., without first running a CP algorithm); see Rocci and Giordani (2010).

For $R = 3$ and a group of three diverging components (with corresponding columns of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ nearly identical up to sign), the form of the CP_{limit} decomposition can be found in Stegeman (2012). For $R = 4$ and a group of four diverging components (with corresponding columns of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ nearly identical up to sign), see Stegeman (2013a). Recall that cases with diverging components that are not identical up to sign have been constructed but are exceptional (Stegeman and De Lathauwer, 2011). Also, cases with groups of more than four diverging components have been found in simulation studies, but are very rare in practice.

When not all components are diverging, or multiple groups of diverging components occur, Stegeman (2012, 2013a) proposes the following CP_{limit} decomposition of $\mathcal{X}$. Each group of d_j diverging components converges to its own limit $\mathcal{X}_j$ with its 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 CP_{limit} decomposition of $\mathcal{X}_j$ is known, as explained above. Each nondiverging component stays nondiverging in the limit, and corresponds to a $d_j = 1$. For example, suppose $R = 3$ and we have one group of two diverging components. The limit $\mathcal{X}$ then has 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 components, and $\mathcal{X}_2 = g_3\, (\mathbf{s}_3 \circ \mathbf{t}_3 \circ \mathbf{u}_3)$ being the limit of the nondiverging component. Hence, the CP_{limit} decomposition of $\mathcal{X}$ is of the form

$$\mathcal{X} = g_{111}\, (\mathbf{s}_1 \circ \mathbf{t}_1 \circ \mathbf{u}_1) + g_{221}\, (\mathbf{s}_2 \circ \mathbf{t}_2 \circ \mathbf{u}_1) + g_{122}\, (\mathbf{s}_1 \circ \mathbf{t}_2 \circ \mathbf{u}_2) + g_3\, (\mathbf{s}_3 \circ \mathbf{t}_3 \circ \mathbf{u}_3). \tag{9}$$

The assumption of Stegeman (2012, 2013a) that groups of diverging components each converge to their respective limits, is confirmed by simulation studies. The limit $\mathcal{X}$ and its CP_{limit} decomposition $\mathcal{X} = \sum_{j=1}^{m} \mathcal{X}_j = \sum_{j=1}^{m} (\mathbf{S}_j, \mathbf{T}_j, \mathbf{U}_j) \cdot \mathcal{G}_j$ may be obtained by fitting the appropriate decomposition form to the data $\mathcal{Z}$. For this, the alternating least squares algorithm of Kiers and Smilde (1998) for fitting a constrained Tucker model can be used. In this algorithm, each of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and the core $\mathcal{G}$ is estimated by solving a regression, while keeping the others fixed. The regression to estimate the core $\mathcal{G}$ features only the unconstrained core entries. Initial values for this algorithm can be obtained from the CP decomposition of $\mathcal{Y}^{(n)}$, which features diverging components. Matlab codes are available online for finding the correct form of CP_{limit}, obtaining initial values, and fitting it to a dataset that yields diverging CP components; see http://www.gmw.rug.nl/~stegeman. For more details, we refer to Stegeman (2012, 2013a).

In our application in Section 3 we have $R = 3$ and two diverging components and we fit the CP_{limit} decomposition (9) to the data $\mathcal{Z}$. The CP_{limit} decomposition has four terms and is the smallest nondiverging Tucker decomposition with a constrained $3 \times 3 \times 3$ core. Indeed, the simplest $3 \times 3 \times 3$ core has three nonzeros and corresponds to CP with $R = 3$ (which cannot be fit due to diverging components). Concerning the part of the two diverging components, we can reason as follows. Imposing a Tucker model with $2 \times 2 \times 2$ core of rank 2 is not feasible. A generic rank-2 core can be transformed to a simple form with two nonzeros and corresponds to CP with $R = 2$ (De Silva and Lim, 2008). The complete model is then again CP with $R = 3$ and yields diverging components. For a $2 \times 2 \times 2$ core of rank 3, we have two possibilities. First, the $2 \times 2 \times 2$ rank-3 core lies on the boundary between rank-2 and rank-3 arrays, and can be transformed to simple form (7) (De Silva and Lim, 2008). We obtain the CP_{limit} decomposition (9). Alternatively, the $2 \times 2 \times 2$ rank-3 core is generic and can be transformed to simple form with four nonzeros (De Silva and Lim, 2008). This yields a decomposition with one more term than CP_{limit}. Hence, for the part of the two diverging components, the first three terms of CP_{limit} in (9) represent the smallest nondiverging Tucker decomposition with a constrained $2 \times 2 \times 2$ core.

Since the CP_{limit} decomposition of $\mathcal{X}$ is not of CP form, one may wonder what its uniqueness properties are. This has been studied in Stegeman (2012) for limits of groups of two diverging components. In the example above, the order of $\mathcal{X}_1$ and $\mathcal{X}_2$ may of course be reversed. As shown in Stegeman (2012), there are also transformations possible within the CP_{limit} decomposition of $\mathcal{X}_1$. Namely, there exist 2×2 matrices $\mathbf{L}, \mathbf{M}, \mathbf{N}$ such that $(\mathbf{L}, \mathbf{M}, \mathbf{N}) \cdot \mathcal{H}_1 = \mathcal{G}_1$, with $\mathcal{G}_1$ and $\mathcal{H}_1$ of the form (7). However, the separation of the limit $\mathcal{X}$ into the limits $\mathcal{X}_1$ and $\mathcal{X}_2$ is unique. In Section 4, we will study in more detail the transformations $(\mathbf{L}, \mathbf{M}, \mathbf{N}) \cdot \mathcal{H}_1 = \mathcal{G}_1$.

3. Application to TV-ratings data

Here, we present a three-way component analysis of TV-ratings data. The data consists 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. The data are previously analyzed by Lundy et al. (1989) and also feature in Harshman (2004). Next, we describe the data, which is given as rating scales (mode 1) by TV shows (mode 2) by persons (mode 3). The rating scales and TV shows are given in Table 1. The possible scores on the rating scales are $-6, -5, \ldots, -1, 0, 1, \ldots, 5, 6$. After deleting subjects with missing data, 30 persons (mode 3) are kept.

The result of a CP analysis of this $16 \times 15 \times 30$ array $\mathcal{Z}$ will be R components for the TV shows in the columns of $\mathbf{B}$ ($15 \times R$), loadings of the rating scales on each component in the columns of $\mathbf{A}$ ($16 \times R$), and loadings of the persons on each component in the columns of $\mathbf{C}$ ($30 \times R$). Additionally, each component r has a weight g_r; see (1). The loading of rating scale i and person k on component r is then given by the product $g_r\, a_{ir} c_{kr}$. We scale the CP decomposition $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{g})$ such that $\mathbf{a}_r^T \mathbf{a}_r = 16$, $\mathbf{b}_r^T \mathbf{b}_r = 15$, and $\mathbf{c}_r^T \mathbf{c}_r = 4$. Hence, the mean squared component score in $\mathbf{b}_r$ equals 1, the mean squared rating scale loading in $\mathbf{a}_r$ also equals 1, and the sum of squared person loadings in $\mathbf{c}_r$ equals 4. Furthermore, whenever possible, we apply sign changes to the columns of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ such that $\mathbf{C}$ does not contain negative loadings. Unless specified otherwise, the person loadings in $\mathbf{C}$ are positive.

The result of a Tucker analysis of $\mathcal{Z}$ is given analogously by matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ (now with R, P, Q columns, respectively) and an $R \times P \times Q$ core array $\mathcal{G}$; see (4). We apply the same scaling to $\mathbf{A}, \mathbf{B}, \mathbf{C}$ as in CP.

Like PCA, the CP and Tucker decompositions are multiplicative in nature. Hence, the data needs to be centered and possibly scaled prior to analysis. This is called preprocessing. For a detailed discussion of (three-way) preprocessing we refer to Bro and Smilde (2003), and Kiers (2006). A three-way array can be centered across one, two, or three modes. For example, centering across the first mode $(z_{ijk} - z_{\bullet jk})$ removes offset terms depending only on j or k. A three-way array can be scaled within one mode only. For example, scaling within the third mode (z_{ijk}/σ_k) can be used to set the variance of each frontal slice to 1. In the CP decomposition of $\mathcal{Z}$, centering across the first mode centers the columns of $\mathbf{A}$, and scaling within the third mode scales the rows of $\mathbf{C}$.

The exact preprocessing used by Lundy et al. (1989) is not reported. However, for ratings data Harshman and De Sarbo (1984) recommend centering across rating scales and stimuli, and standardizing within rating scales and subjects. The latter is done iteratively and approximately. Accordingly, we use centering across rating scales and TV shows to remove all one-way main effects. We standardize within the persons mode to remove differences between extreme and moderate response styles. Our CP solutions have close resemblance to those obtained by Lundy et al. (1989), which indicates that our preprocessing closely resembles theirs.

3.1. Choosing an appropriate three-way component model

In Lundy et al. (1989) and Harshman (2004) only the CP analysis of the TV-ratings data is considered. Here, we take a step back and first discuss which Tucker or CP models might be suitable for this dataset. Various methods have been proposed for choosing the number(s) of components in Tucker and CP models. They compare fit percentages and numbers of free parameters in the models (e.g. Ceulemans and Kiers, 2006). The fit percentage for any three-way model is defined as

$$100 - 100\,\frac{\text{ssq}(\mathcal{E})}{\text{ssq}(\mathcal{Z})}, \tag{10}$$

Table 1
Rating scales and TV shows in the TV-ratings dataset.

Rating scales	TV shows
1. Thrilling ... Boring	1. Mash
2. Intelligent ... Idiotic	2. Charlie's Angels
3. Erotic ... Not Erotic	3. All in the Family
4. Sensitive ... Insensitive	4. 60 Minutes
5. Interesting ... Uninteresting	5. The Tonight Show
6. Fast ... Slow	6. Let us Make a Deal
7. Intellectually Stimulating ... Intellectually Dull	7. The Waltons
8. Violent ... Peaceful	8. Saturday Night Live
9. Caring ... Callous	9. News (any channel; national edition)
10. Satirical ... Not Satirical	10. Kojak
11. Informative ... Uninformative	11. Mork and Mindy
12. Touching ... "Leaves Me Cold"	12. Jacques Cousteau
13. Deep ... Shallow	13. Football
14. Tasteful ... Crude	14. Little House on the Prairie
15. Real ... Fantasy	15. Wild Kingdom
16. Funny ... Not Funny	

Table 2
Fit percentages for various Tucker and CP models fitted to the TV-ratings data.

$R \times P \times Q$	fit %	$R \times P \times Q$	fit %	$R \times P \times Q$	fit %
$1 \times 2 \times 2$	30.06	$2 \times 3 \times 3$	43.65	$3 \times 3 \times 4$	51.88
$2 \times 1 \times 2$	29.52	$3 \times 2 \times 3$	43.59	$3 \times 4 \times 4$	52.82
$2 \times 2 \times 1$	41.05	$3 \times 3 \times 1$	48.32	$4 \times 3 \times 4$	52.58
$2 \times 2 \times 2$	41.96	$3 \times 3 \times 2$	50.20	$4 \times 4 \times 1$	50.26
CP with $R = 2$	41.96	$3 \times 3 \times 3$	51.16	$4 \times 4 \times 2$	52.28
$3 \times 2 \times 2$	42.69	CP with $R = 3$	50.76	$4 \times 4 \times 3$	53.43
$2 \times 3 \times 2$	42.73	$4 \times 3 \times 3$	51.65	$4 \times 4 \times 4$	54.51
$2 \times 2 \times 3$	42.66	$3 \times 4 \times 3$	51.83	CP with $R = 4$	53.79

where $\mathcal{E}$ denotes the residual array, and ssq(·) denotes the sum-of-squares. In Table 2 the fit percentages for various Tucker and CP models are reported.

Since the number(s) of components in Table 2 are relatively small, we do not use a formal method to choose an appropriate model. As can be seen, the fit percentage increases with a relatively large amount when the numbers of components in the rating scales and TV show modes increase simultaneously. This increase of fit occurs up to three components and is much less when going to four components. In contrast, the fit does not increase much when the number of components in the person mode is increased. Also, we observe that the fit of the CP model is rather close to the fit of the Tucker model with (R, R, R) components. Based on these considerations, we select the Tucker models with (3, 3, 1), (3, 3, 2), (3, 3, 3) components and the CP model with $R = 3$ components as suitable models for the TV-ratings data. The fit percentages of these models are around 50%, which is acceptable.

For later use, we define the congruence coefficient between two three-way components (i.e., rank-1 terms in a CP or Tucker decomposition). For any two components, we use the congruence coefficient to define their closeness. The congruence coefficient of vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ is defined as

$$cc_A(1, 2) = \frac{\mathbf{a}_1^T \mathbf{a}_2}{\sqrt{\mathrm{ssq}(\mathbf{a}_1)}\sqrt{\mathrm{ssq}(\mathbf{a}_2)}}, \tag{11}$$

which is between −1 and 1 (Tucker, 1951). We define the congruence coefficient between components $(\mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1)$ and $(\mathbf{a}_2 \circ \mathbf{b}_2 \circ \mathbf{c}_2)$ as

$$cc(1, 2) = cc_A(1, 2)\, cc_B(1, 2)\, cc_C(1, 2), \tag{12}$$

where $cc_B(1, 2)$ and $cc_C(1, 2)$ are defined analogous to (11). The two components are nearly identical up to sign when $cc(1, 2)$ is close to −1 or 1.

3.2. The CP solutions for $R = 2$ and $R = 3$

For ease of presentation we also present the CP solution for $R = 2$. We fit the CP model using the standard alternating least squares algorithm. We run the algorithm 11 times, 10 times with random initial values and one time with initial values based on the singular value decompositions of the three matrix unfoldings of the data. Of these 11 runs, we keep the CP solution with the highest fit percentage. The convergence criterion of the algorithm is set at `1e-9`.

For $R = 2$, each of the 11 runs yields nearly the same solution. The CP fit is 41.96%, and the two components are nearly orthogonal ($cc(1, 2) = 0.002$). The weights of the components are $g_1 = 1.46$ and $g_2 = 1.01$, and the components would

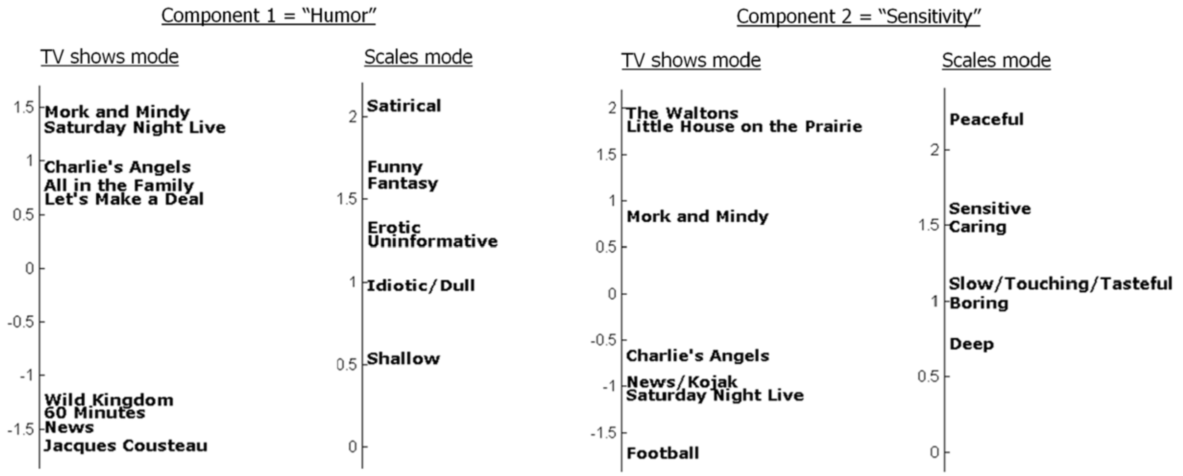

Fig. 2. TV show scores and rating scale loadings for the CP solution with $R = 2$.

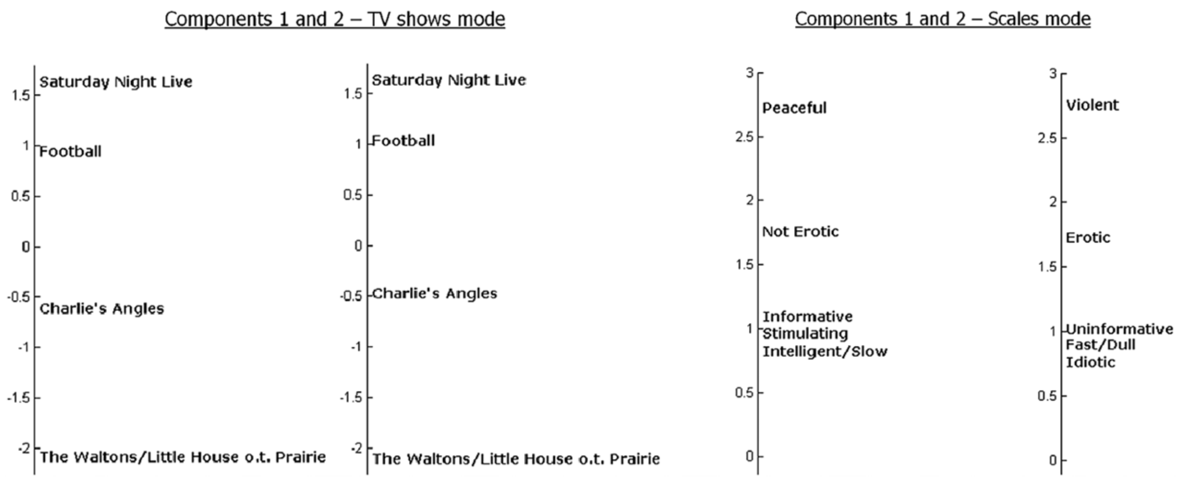

Fig. 3. TV show scores and rating scale loadings for the diverging components in the CP solution with $R = 3$.

explain 28.46% and 13.59% respectively if they were the only component. In Fig. 2 the TV show scores and the rating scale loadings are depicted in a one-way plot for the two components. Note that the rating scale plots are unipolar, with the absolute values of the loadings plotted and the labels indicating negative or positive loadings. As can be seen, the rating scales and TV shows of the first component correspond to "Humor", while those of the second component correspond to "Sensitivity". This is similar to the CP solution found by Lundy et al. (1989).

In general, when interpreting a CP solution for a three-way array also the third mode (persons in our case) is considered. However, since we have no additional information on the persons to aid interpretation of the CP components, we only consider the rating scale and TV show modes. All person loadings in the CP solution are positive. Hence, the components have the same interpretation for all persons.

Next, we repeat the procedure for $R = 3$. Each of the 11 runs yields nearly the same solution. The CP fit is 50.76%, and the congruence coefficients are $cc(1, 2) = -0.996$, $cc(1, 3) = -0.13$, and $cc(2, 3) = 0.12$. The weights of the components are $g_1 = 15.23$, $g_2 = 15.39$, and $g_3 = 1.52$. Hence, components 1 and 2 are nearly identical up to sign and have large weights. Also, the number of iterations of the CP algorithm is very large (around 8500). This is a case of two diverging components. The interpretation of these two components is not clear. In Fig. 3 it can be seen that their TV show scores are nearly identical while their rating scale loadings are nearly the opposite of each other. The third and nondiverging component has congruence coefficient 0.93 with the "Humor" component of the $R = 2$ solution. Hence, it also represents "Humor". The fit of the two diverging components added together equals 20.11% (if they were the only components). The fit of the third component equals 24.38% if it was the only component.

To avoid diverging components, we fit the CP model with $R = 3$ under the constraint of orthogonal TV show components. We denote this model as $\mathrm{CP}_{\mathrm{orth}}(B)$, to indicate that the columns of $\mathbf{B}$ are orthogonal. Lundy et al. (1989) also fit this model. Each of the 11 runs yields nearly the same solution. The fit percentage is 50.22, where the contribution of each component

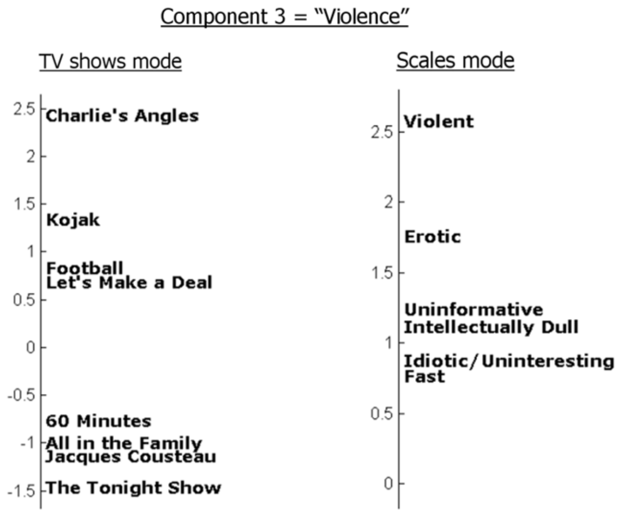

Fig. 4. TV show scores and rating scale loadings for the third component in the CP solution with $R = 3$ and orthogonal TV show components.

Table 3
Congruence coefficients between the components of the CP solutions with $R = 2$ and $R = 3$, and $R = 3$ with orthogonal TV shows components.

	$R = 2$		$R = 3$			$R = 3$ orth.		
	"H"	"S"	C.1	C.2	"H"	"H"	"S"	"V"
"Humor"	1	0.00	−0.15	0.15	0.93	0.96	0.00	0.11
"Sensitivity"		1	−0.41	0.46	0.01	0.00	0.94	0.08
Component 1			1	−0.99	−0.13	−0.20	−0.30	0.30
Component 2				1	0.12	0.19	0.36	−0.24
"Humor"					1	0.95	0.03	−0.03
"Humor"						1	0	0
"Sensitivity"							1	0
"Violence"								1

is: 27.19%, 13.04%, and 9.99%. The weights of the components are $g_1 = 1.43$, $g_2 = 0.99$, and $g_3 = 0.87$. The interpretation of the first component is "Humor" and it has congruence coefficient 0.96 with the "Humor" component of the $R = 2$ solution. The interpretation of the second component is "Sensitivity" and it has congruence coefficient 0.94 with the "Sensitivity" component of the $R = 2$ solution. The interpretation of the third component is "Violence", as can be seen in Fig. 4. This is similar to the CP solution found by Lundy et al. (1989). The orthogonality constraint is rather arbitrary and unintuitive, especially since "Sensitivity" and "Violence" do not seem uncorrelated. A similar CP solution is obtained for $\text{CP}_{\text{orth}}(A)$, i.e., when the columns of the rating scale loadings are orthogonal. The fit of $\text{CP}_{\text{orth}}(A)$ is 50.02%.

Table 3 shows comparisons between the components of the CP solutions with $R = 2$ and $R = 3$, and $\text{CP}_{\text{orth}}(B)$. As can be seen, the two diverging components in the $R = 3$ solution relate to "Sensitivity" and "Violence". In Section 3.4, we find the CP_{limit} decomposition of the limit of the CP solution with $R = 3$ and diverging components. As we will show, CP_{limit} also contains the "Humor", "Sensitivity", and "Violence" components, but does not impose the constraint of orthogonality.

3.3. The Tucker solutions with $(3, 3, Q)$ components

Here, we present the Tucker solutions with (3, 3, 1), (3, 3, 2), and (3, 3, 3) components. We fit the Tucker model using the alternating least squares algorithm (e.g. Kroonenberg and De Leeuw, 1980)). As for CP, we take the best of 11 runs and use convergence criterion `1e-9`. For rotating an obtained Tucker solution $(\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathcal{G}$, we try two methods. First, we use the joint orthomax procedure of Kiers (1998a) such that a balance is found between simplicity in the core array and the rating scales and TV show matrices. That is, orthonormal $\mathbf{S}_A, \mathbf{S}_B, \mathbf{S}_C$ are found such that in the rotated Tucker solution $(\mathbf{AS}_A, \mathbf{BS}_B, \mathbf{CS}_C) \cdot ((\mathbf{S}_A^T, \mathbf{S}_B^T, \mathbf{S}_C^T) \cdot \mathcal{G})$ the matrices $\mathbf{AS}_A$ and $\mathbf{BS}_B$ have simple structure according to the varimax criterion, and also the rotated core $(\mathbf{S}_A^T, \mathbf{S}_B^T, \mathbf{S}_C^T) \cdot \mathcal{G}$ has simple structure. The 'standard weights' of Kiers (1998a) are used to balance these multiple objectives. Second, we use the Simplimax procedure of Kiers (1998b), in which oblique rotation matrices $\mathbf{S}_A, \mathbf{S}_B, \mathbf{S}_C$ are found such that the rotated core $(\mathbf{S}_A^T, \mathbf{S}_B^T, \mathbf{S}_C^T) \cdot \mathcal{G}$ has simple structure. As objective, the rotated core should approximate as close as possible a core with three nonzero entries. Note that the Tucker solutions are such that $\mathbf{A}, \mathbf{B}, \mathbf{C}$ have orthogonal columns. Hence, the joint orthomax procedure of Kiers (1998a) results in orthogonal rotated components, while the Simplimax procedure of Kiers (1998b) results in oblique rotated components.

Table 4
Congruence coefficients between the three large components of the rotated Tucker solutions, the CP_{limit} and rotated CP_{limit} solutions, and the $CP_{orth}(B)$ solution with $R = 3$ and orthogonal TV show components. The rotations for Tucker are joint orthomax (JOM) and Simplimax (SIM). The rotations for CP_{limit} are Prodmin for the rating scales mode, and either Prodmin (rot.1) or Quartimin (rot.2) for the TV shows mode.

	Rotation	(3, 3, 1)		(3, 3, 2)		(3, 3, 3)		$CP_{orth}(B)$		CP_{limit}		CP_{limit} rotated		
		Congr.	Fit %	Congr.	Fit %	Congr.	Fit %	Congr.	Fit %	Congr.	Fit %	Rotation	Congr.	Fit %
"Humor"	JOM	0.97	27.22	0.97	28.00	0.97	28.29	1	27.19	0.95	24.37	rot.1	0.95	24.37
	SIM	0.91	27.54	0.93	24.82	0.94	26.38					rot.2	0.95	24.37
"Sensitivity"	JOM	0.83	10.38	0.84	10.43	0.84	10.30	1	13.04	0.81	10.75	rot.1	0.83	11.05
	SIM	0.82	11.82	0.69	13.64	0.74	11.66					rot.2	0.83	11.05
"Violence"	JOM	0.84	9.01	0.80	8.68	0.79	8.70	1	9.99	0.86	7.62	rot.1	0.83	8.59
	SIM	0.69	8.96	0.62	4.13	0.54	6.65					rot.2	0.88	7.07
Large terms	JOM	–	46.61	–	47.11	–	47.29	–	50.22	–	49.32	rot.1	–	48.08
	SIM	–	48.32	–	48.24	–	48.20					rot.2	–	49.09
total		–	48.32	–	50.20	–	51.16	–	50.22	–	50.76		–	50.76

In all three Tucker solutions, there are only three large terms after rotation. The three large terms are of the form

$$g_1\,(\mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1) + g_2\,(\mathbf{a}_2 \circ \mathbf{b}_2 \circ \mathbf{c}_1) + g_3\,(\mathbf{a}_3 \circ \mathbf{b}_3 \circ \mathbf{c}_1). \tag{13}$$

Hence, they share the same person loadings vector $\mathbf{c}_1$ (which contains only positive entries). Each of the terms is large in the sense that the weights g_1, g_2, g_3 are large in magnitude, and in the sense that the explained variance due to each term is large. For orthogonal components, the explained variances of all terms in the decomposition add up to the total explained variance (i.e., the fit percentage). In Table 4 the fit percentages of the large terms and the remaining terms are given for all rotated Tucker solutions. Also, the congruence coefficients between the large terms and the components of the $CP_{orth}(B)$ solution are given. As can be seen, when using the joint orthomax rotation procedure we find about the same "Humor", "Sensitivity", and "Violence" components as in the $CP_{orth}(B)$ solution. When using the Simplimax rotation procedure, however, the "Sensitivity" and "Violence" components are (much) less clearly present. There is no clear alternative interpretation of the "Violence" component for the three Simplimax rotated Tucker solutions. For all rotated Tucker solutions, the joint fit of the three large terms is smaller than the fit of the $CP_{orth}(B)$ solution. When using the joint orthomax rotation procedure, the Tucker components are orthogonal, which is not intuitive for "Sensitivity" and "Violence". When using Simplimax, oblique components are obtained, but (at least) one large component has no clear interpretation.

Apart from the three large terms, the rotated Tucker solutions contain a lot of small terms that are hardly interesting for interpretation. Also, the person loading vectors other than $\mathbf{c}_1$ in (13) contain both positive and negative entries. This implies that the interpretation of most small terms may differ for different persons.

3.4. *The* CP_{limit} *decomposition for* $R = 3$

Here, we apply the method of Stegeman (2012, 2013a) to the CP solution with $R = 3$ and featuring two diverging components. As explained in Section 2, we fit a CP_{limit} decomposition of the form (9) to the data $\mathcal{Z}$. The first three terms in (9) are the limit of the two diverging components and the fourth term is the limit of the nondiverging "Humor" component. Initial values for the CP_{limit} decomposition are obtained from the CP solution with $R = 3$ (for details see Stegeman, 2012). We fit the CP_{limit} decomposition to $\mathcal{Z}$ using the alternating least squares algorithm of Kiers and Smilde (1998) with convergence criterion 1e−9. The vectors $\mathbf{s}_r, \mathbf{t}_r, \mathbf{u}_r$ are scaled analogous to the columns of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ in the CP solutions.

The number of iterations needed is only 79. The fit percentage of the CP_{limit} decomposition is 50.7571. For the CP solution with $R = 3$ this is 50.7569%, indicating that the limit point $\mathcal{X}$ is indeed a little closer to $\mathcal{Z}$ than the solution array $\mathcal{Y}^{(n)}$ of the CP solution. For the four terms in (9) the fit percentages are: 7.62, 10.75, 1.55, and 24.37 (when each of them was the only term); see Table 4. The weights are $g_{111} = 0.99$, $g_{221} = 0.95$, $g_{122} = -0.33$, and $g_3 = 1.52$. The last term is very close to the nondiverging "Humor" component of the CP solution. For the interpretation of the first three terms, we compare them to the $CP_{orth}(B)$ solution (see Table 4). We find that the first term in (9) has congruence coefficient 0.86 with the "Violence" component. Hence, $\mathbf{s}_1$ and $\mathbf{t}_1$ relate to "Violent" rating scales and TV shows. The second term has congruence coefficient 0.81 with the "Sensitivity" component. Hence, $\mathbf{s}_2$ and $\mathbf{t}_2$ relate to "Sensitive" rating scales and TV shows. Therefore, the small interaction term $(\mathbf{s}_1 \circ \mathbf{t}_2 \circ \mathbf{u}_2)$ relates "Violent" rating scales to "Sensitive" TV shows. Its person loadings vector $\mathbf{u}_2$ contains both positive and negative entries. Hence, its interpretation differs for different persons (if it is of interest at all).

The congruence coefficients between the four terms in (9) are all smaller than 0.15 in magnitude. Hence, all nasty properties of the CP solution with diverging components have vanished. Instead, we have a nondiverging CP_{limit} decomposition of the limit point that is easy to obtain. It contains the "Humor", "Sensitivity", and "Violence" components that are also in the orthogonally rotated Tucker solutions and the $CP_{orth}(B)$ solution. However, now the components are oblique, which is more intuitive. Also, we have only one small term and not many as in the Tucker solutions. Moreover, the fit of the three large terms together is 49.32% (see Table 4), which is larger than in the Tucker solutions and only slightly smaller than the fit of the CP solution with orthogonal TV show components.

Contrary to CP, there is some rotational freedom in CP_{limit}. This will be discussed in detail in Section 4. The rotational freedom implies that, after fixing the scaling and permutation ambiguities, two obtained (unrotated) CP_{limit} solutions with identical fitted model arrays $\mathcal{Z} - \mathcal{E}$ may differ when different starting values are used. Here, we obtained starting values from the CP solution with $R = 3$ and diverging components. However, as we will see in Section 4, fixing the rotational freedom with standard rotation criteria does not change the conclusions about the CP_{limit} solutions (also see Table 4).

As can be seen in Table 2, the Tucker solutions, the $\text{CP}_{\text{orth}}(B)$ solution, and the CP_{limit} solution have fit percentages that are close together. Also, when comparing the fitted model arrays (i.e., $\mathcal{Z} - \mathcal{E}$), we find that fit percentages of one model array with respect to another are larger than 96% except for the (3, 3, 1) Tucker solution, which has fit percentages larger than 93%. Given the fact that these models fit nearly equally well, the choice of a suitable decomposition for the TV-ratings data boils down to the number of terms in the model, the joint fit of the three main components, and the reasonability of the assumptions made.

4. Uniqueness of the CP_{limit} decomposition

Here, we have a closer look at the uniqueness properties of the CP_{limit} decomposition (9). We consider a general CP_{limit} decomposition that is the limit of a CP sequence $\mathcal{Y}^{(n)}$ featuring one or more groups of two diverging components. As explained in Section 2, Stegeman (2012) shows that the decomposition of the limit point $\mathcal{X}$ is of the form $\mathcal{X} = \sum_{j=1}^{m} (\mathbf{S}_j, \mathbf{T}_j, \mathbf{U}_j) \cdot \mathcal{G}_j$, where $\mathcal{G}_j$ is either $2 \times 2 \times 2$ and equal to (7) (for the limit of a group of two diverging components) or $1 \times 1 \times 1$ and nonzero (for the limit of a nondiverging component). Let an alternative CP_{limit} decomposition be given by $\mathcal{X} = \sum_{j=1}^{m} (\bar{\mathbf{S}}_j, \bar{\mathbf{T}}_j, \bar{\mathbf{U}}_j) \cdot \mathcal{H}_j$, where the sizes of the $\mathcal{H}_j$ match those of the $\mathcal{G}_j$ up to a permutation of the summands. The CP_{limit} decomposition is called *essentially unique* if it follows that $\bar{\mathbf{S}}_j = \mathbf{S}_{\pi(j)}\mathbf{L}_{\pi(j)}$, $\bar{\mathbf{T}}_j = \mathbf{T}_{\pi(j)}\mathbf{M}_{\pi(j)}$, $\bar{\mathbf{U}}_j = \mathbf{U}_{\pi(j)}\mathbf{N}_{\pi(j)}$, and $\mathcal{H}_j = (\mathbf{L}_{\pi(j)}^{-1}, \mathbf{M}_{\pi(j)}^{-1}, \mathbf{N}_{\pi(j)}^{-1}) \cdot \mathcal{G}_{\pi(j)}$, for nonsingular matrices $\mathbf{L}_{\pi(j)}, \mathbf{M}_{\pi(j)}, \mathbf{N}_{\pi(j)}$, and a permutation π of $(1, \ldots, m)$. Hence, the only existing ambiguities are nonsingular transformations between the matrices $\mathbf{S}_j, \mathbf{T}_j, \mathbf{U}_j$ and the arrays $\mathcal{G}_j$, and a permutation of the summands. Decompositions of this type (without zero restrictions on $\mathcal{G}_j$) are known as decompositions in block terms, and were introduced by De Lathauwer (2008).

Stegeman (2012) has shown that, under the conditions stated above, the ambiguities in the CP_{limit} decomposition of $\mathcal{X}$ are those under essential uniqueness, and those of the form $(\mathbf{L}_j, \mathbf{M}_j, \mathbf{N}_j) \cdot \mathcal{H}_j = \mathcal{G}_j$, where $\mathcal{G}_j$ and $\mathcal{H}_j$ are of the form (7). This implies that the terms $(\mathbf{S}_j, \mathbf{T}_j, \mathbf{U}_j) \cdot \mathcal{G}_j$ remain separated in alternative decompositions. Hence, the limit points $\mathcal{X}_j$ of the groups of two diverging components and of the nondiverging components are unique. In this section, we consider the transformations $(\mathbf{L}, \mathbf{M}, \mathbf{N}) \cdot \mathcal{H} = \mathcal{G}$, with $\mathcal{G}$ and $\mathcal{H}$ of the form (7). For ease of presentation, we drop the subscript j and consider the decomposition $(\mathbf{S}, \mathbf{T}, \mathbf{U}) \cdot \mathcal{G}$ with $\mathcal{G}$ of the form (7). Our analysis below sharpens the uniqueness result of Stegeman (2012) and is relevant for the interpretation of the CP_{limit} decomposition of $\mathcal{X}$. Let

$$\mathbf{L} = \begin{bmatrix} l_1 & l_2 \\ l_3 & l_4 \end{bmatrix}, \qquad \mathbf{M} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix}, \qquad \mathbf{N} = \begin{bmatrix} n_1 & n_2 \\ n_3 & n_4 \end{bmatrix}. \tag{14}$$

In the decomposition $(\mathbf{S}, \mathbf{T}, \mathbf{U}) \cdot \mathcal{G} = (\mathbf{S}, \mathbf{T}, \mathbf{U}) \cdot ((\mathbf{L}, \mathbf{M}, \mathbf{N}) \cdot \mathcal{H}) = (\mathbf{SL}, \mathbf{TM}, \mathbf{UN}) \cdot \mathcal{H}$, we require $\mathbf{L}, \mathbf{M}, \mathbf{N}$ to have columns of length 1 to fix the scaling ambiguity. We have the following result.

Lemma 4.1. *Let* $(\mathbf{L}, \mathbf{M}, \mathbf{N}) \cdot \mathcal{H} = \mathcal{G}$, *with* $\mathcal{G}$ *and* $\mathcal{H}$ *of the form* (7) *and* $\mathbf{L}, \mathbf{M}, \mathbf{N}$ *nonsingular with columns of length* 1. *Then*

$$\mathbf{L} = \begin{bmatrix} 1 & l_2 \\ 0 & g_{221}/h_{221} \end{bmatrix}, \qquad \mathbf{M} = \begin{bmatrix} g_{111}/h_{111} & 0 \\ m_3 & 1 \end{bmatrix}, \qquad \mathbf{N} = \begin{bmatrix} 1 & n_2 \\ 0 & g_{122}/h_{122} \end{bmatrix}, \tag{15}$$

with

$$l_2^2 + (g_{221}/h_{221})^2 = (g_{111}/h_{111})^2 + m_3^2 = n_2^2 + (g_{122}/h_{122})^2 = 1, \tag{16}$$

$$h_{111}\, m_3 + h_{221}\, l_2 + h_{122}\, n_2 = 0. \tag{17}$$

Proof. The equations for the eight entries of $(\mathbf{L}, \mathbf{M}, \mathbf{N}) \cdot \mathcal{H} = \mathcal{G}$ are as follows:

$$n_1\,(h_{111}l_1m_1 + h_{221}l_2m_2) + n_2\,(h_{122}l_1m_2) = g_{111}, \tag{18}$$

$$n_1\,(h_{111}l_3m_1 + h_{221}l_4m_2) + n_2\,(h_{122}l_3m_2) = 0, \tag{19}$$

$$n_1\,(h_{111}l_1m_3 + h_{221}l_2m_4) + n_2\,(h_{122}l_1m_4) = 0, \tag{20}$$

$$n_1\,(h_{111}l_3m_3 + h_{221}l_4m_4) + n_2\,(h_{122}l_3m_4) = g_{221}, \tag{21}$$

$$n_3\,(h_{111}l_1m_1 + h_{221}l_2m_2) + n_4\,(h_{122}l_1m_2) = 0, \tag{22}$$

$$n_3\,(h_{111}l_3m_1 + h_{221}l_4m_2) + n_4\,(h_{122}l_3m_2) = 0, \tag{23}$$

$$n_3\,(h_{111}l_1m_3 + h_{221}l_2m_4) + n_4\,(h_{122}l_1m_4) = g_{122}, \tag{24}$$

$$n_3\,(h_{111}l_3m_3 + h_{221}l_4m_4) + n_4\,(h_{122}l_3m_4) = 0. \tag{25}$$

Since $n_3 = n_4 = 0$ is not allowed as a solution ($\mathbf{N}$ must be nonsingular), Eqs. (22), (23) and (25) in (n_3, n_4) must be proportional in their coefficients. Proportionality of (22) and (23) is equivalent to $(l_2 l_3 - l_1 l_4)\, m_2^2 = 0$. Hence, $-\det(\mathbf{L})\, m_2^2 = 0$, which implies $m_2 = 0$. Proportionality of (23) and (25) is equivalent to $(m_1 m_4 - m_2 m_3)\, l_3^2 = 0$. Hence, $\det(\mathbf{M})\, l_3^2 = 0$, which implies $l_3 = 0$.

Substituting $m_2 = 0$ in (22) yields $n_3\,(h_{111} l_1 m_1) = 0$. Since $l_3 = m_2 = 0$, setting either $l_1 = 0$ or $m_1 = 0$ violates the nonsingularity of $\mathbf{L}$ and $\mathbf{M}$, respectively. Hence, we obtain $n_3 = 0$. It can be verified that Eqs. (19), (22), (23) and (25) are now satisfied.

Since the columns of $\mathbf{L}$, $\mathbf{M}$, $\mathbf{N}$ have length 1, it follows that $l_1 = m_4 = n_1 = 1$. Eq. (18) then yields $m_1 = g_{111}/h_{111}$, Eq. (21) yields $l_4 = g_{221}/h_{221}$, and Eq. (24) yields $n_4 = g_{222}/h_{222}$. Eq. (16) follows from the fact that $\mathbf{L}$, $\mathbf{M}$, $\mathbf{N}$ have length-1 columns. Finally, Eq. (20) is equal to (17). This completes the proof. □

An example of $\mathbf{L}$, $\mathbf{M}$, $\mathbf{N}$ for which $(\mathbf{L}, \mathbf{M}, \mathbf{N}) \cdot \mathcal{H} = \mathcal{G}$ is

$$\mathbf{L} = \begin{bmatrix} 1 & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} \end{bmatrix}, \qquad \mathbf{M} = \begin{bmatrix} 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1 \end{bmatrix}, \qquad \mathbf{N} = \begin{bmatrix} 1 & 2/\sqrt{5} \\ 0 & 1/\sqrt{5} \end{bmatrix}, \tag{26}$$

with $g_{111} = g_{221} = g_{122} = 1$, $h_{111} = h_{221} = \sqrt{2}$, and $h_{122} = \sqrt{5}$.

Next, we show that for given $\mathcal{G}$, fixing two of l_2, m_3, n_2 will fix all variables satisfying (16) and (17). Without loss of generality, we assume l_2 and m_3 are fixed. Then h_{221} and h_{111} follow from (16). Next, n_2 and h_{122} must be found such that $n_2^2 + (g_{122}/h_{122})^2 = 1$ and (17) hold. This is done as follows. Eq. (17) implies $h_{122} n_2 = -(h_{111} m_3 + h_{221} l_2)$. The value of h_{122} (up to sign) then follows from $(h_{122} n_2)^2 + g_{122}^2 = h_{122}^2$. Finally, n_2 follows (up to sign) from the values of $h_{122} n_2$ and h_{122}. Hence, for any $l_2, m_3 \in (-1, 1)$ the transformation (15) is fixed (up to sign).

Next, we consider the implications of such transformations for the interpretation of the CP$_{\text{limit}}$ decomposition of $\mathcal{X}$ in Section 3.4. Let $\widetilde{\mathbf{S}} = [\mathbf{s}_1\ \ \mathbf{s}_2]$, $\widetilde{\mathbf{T}} = [\mathbf{t}_1\ \ \mathbf{t}_2]$, and $\widetilde{\mathbf{U}} = [\mathbf{u}_1\ \ \mathbf{u}_2]$. Suppose we have $\mathbf{L}$, $\mathbf{M}$, $\mathbf{N}$ as in (15) and $\mathcal{H}$ that satisfy $(\mathbf{L}, \mathbf{M}, \mathbf{N}) \cdot \mathcal{H} = \mathcal{G}$, for $\mathcal{G}$ as in (7). Hence, the alternative CP$_{\text{limit}}$ decomposition to (9) is $(\widetilde{\mathbf{S}}\mathbf{L}, \widetilde{\mathbf{T}}\mathbf{M}, \widetilde{\mathbf{U}}\mathbf{N}) \cdot \mathcal{H} + g_3\, (\mathbf{s}_3 \circ \mathbf{t}_3 \circ \mathbf{u}_3)$. Recall that $\mathbf{s}_1$ and $\mathbf{t}_1$ relate to "Violent" rating scales and TV shows, and $\mathbf{s}_2$ and $\mathbf{t}_2$ relate to "Sensitive" rating scales and TV shows. Since $l_3 = 0$ it follows that the first column of $\widetilde{\mathbf{S}}\mathbf{L}$ still relates to "Violent" rating scales, while the second column of $\widetilde{\mathbf{S}}\mathbf{L}$ is a mixture of "Violent" and "Sensitive" rating scales. Since $m_2 = 0$ it follows that the second column of $\widetilde{\mathbf{T}}\mathbf{M}$ still relates to "Sensitive" TV shows, while the first column of $\widetilde{\mathbf{T}}\mathbf{M}$ is a mixture of "Violent" and "Sensitive" TV shows. Hence, also in the alternative CP$_{\text{limit}}$ decomposition the interpretation involves "Humor", "Violence" and "Sensitivity". However, the last two concepts may be mixed together.

Suitable linear combinations of "Sensitive" and "Violent" rating scales and TV shows (fixing l_2 and m_3) may be determined as follows. For the rating scales loading matrix $\mathbf{S} = [\mathbf{s}_1\ \mathbf{s}_2\ \mathbf{s}_3]$, we have fixed $\mathbf{s}_1$ and $\mathbf{s}_3$, while the second column may be a linear combination $\alpha_1\, \mathbf{s}_1 + \alpha_2\, \mathbf{s}_2$, with $\alpha_1^2 + \alpha_2^2 = 1$. As a suitable criterion to determine α_1 and α_2 we propose to consider row complexity of the resulting $\mathbf{S}_{\text{rot}} = [\mathbf{s}_1\ \alpha_1\mathbf{s}_1 + \alpha_2\mathbf{s}_2\ \mathbf{s}_3]$. We suggest to minimize one of the following two criteria for row complexity of an $p \times r$ matrix $\boldsymbol{\Lambda}$:

$$\text{Prodmin} : f(\boldsymbol{\Lambda}) = \sum_{i=1}^{p} \prod_{j=1}^{r} \lambda_{ij}^2, \tag{27}$$

$$\text{Quartimin} : f(\boldsymbol{\Lambda}) = \sum_{i=1}^{p} \sum_{j=1}^{r} \sum_{k \neq j}^{r} \lambda_{ij}^2 \lambda_{ik}^2. \tag{28}$$

Criterion (27) was introduced by Thurstone (1935) and we refer to it as Prodmin. It is small when each row contains at least one entry close to zero. Criterion (28) is called Quartimin and was introduced by Carroll (1953). It is small when each row has at most one large (squared) entry. For more background on complexity criteria, we refer to Browne (2001).

The bipolar rating scales for the TV shows imply that components may have large loadings for the same rating scale yet still differ in interpretation. Indeed, the signs of the loadings may be opposite. For example, this is true for the "Sensitive" and "Violent" components and rating scale 8 on "Violent …Peaceful". Hence, for the row complexity of the transformed rating scale loadings in $\mathbf{S}_{\text{rot}}$, we use the Prodmin criterion (27). For the transformed TV show components matrix $\mathbf{T}_{\text{rot}} = [\beta_1\mathbf{t}_1 + \beta_2\mathbf{t}_2\ \mathbf{t}_2\ \mathbf{t}_3]$, with $\beta_1^2 + \beta_2^2 = 1$, we try both the Prodmin and the Quartimin row complexity criteria.

The results are as follows. For the rating scales, we obtain $\alpha_1 = -0.19$ and $\alpha_2 = 0.98$. Hence, the linear combination is rather close to the original column $\mathbf{s}_2$. For the TV show components and the Prodmin criterion, we obtain $\beta_1 = 0.99$ and $\beta_2 = -0.11$. The corresponding $\mathcal{H}$ has $h_{111} = 0.99$, $h_{221} = 1.01$, and $h_{122} = -0.45$. For the Quartimin criterion, β_1 is very close to 1 and $\beta_2 = 0.05$. We have $h_{111} = 0.99$, $h_{221} = 1.01$, $h_{122} = -0.36$. Hence, also here the linear combinations are close to the original column $\mathbf{t}_1$. The three large terms of the rotated CP$_{\text{limit}}$ decompositions still have the same interpretation; see Table 4. The fit of the large terms together is 48.08% when using Prodmin for the TV shows mode, and 49.09% when using Quartimin. This is only slightly less than for the original CP$_{\text{limit}}$ decomposition.

Note that after computing $\mathbf{L}$, $\mathbf{M}$, $\mathbf{N}$ and $\mathcal{H}$, we rescaled the rotated CP$_{\text{limit}}$ such that column sum-of-squares are 16 for $\mathbf{S}_{\text{rot}}$, 15 for $\mathbf{T}_{\text{rot}}$, and 4 for the rotated $\mathbf{U}_{\text{rot}}$. This is the same scaling as in the CP models.

Table 5
Best-fitting core array for the CP solution with $R = 3$ and orthogonal TV show components. Interaction weights larger than 0.2 in magnitude are in boldfont.

	Person loadings 1		
	"H" shows	"S" shows	"V" shows
"Humor" scales	0.96	−0.06	−0.08
"Sensitive" scales	0.02	−0.04	−0.12
"Violent" scales	0.10	**0.30**	0.04
	Person loadings 2		
	"H" shows	"S" shows	"V" shows
"Humor" scales	−0.05	**0.22**	−0.09
"Sensitive" scales	0.02	1.03	0.02
"Violent" scales	0.14	**−0.23**	0.03
	Person loadings 3		
	"H" shows	"S" shows	"V" shows
"Humor" scales	0.13	−0.16	**0.21**
"Sensitive" scales	−0.06	0.03	0.16
"Violent" scales	**−0.34**	**−0.21**	0.90

The original CP_{limit} was obtained with initial values from CP, and has smaller $g_{122} = -0.33$ than the h_{122} after rotation. Also, the joint explained variance of the three large terms is larger. An explanation for this may be the following. For the TV-ratings dataset, the number of person mode components does not matter much in terms of fit (see Table 2). On the other hand, the fit increases significantly when the numbers of components in the rating scales and TV shows modes are increased simultaneously. Hence, for initial values obtained from CP, terms one, two, and four of the CP_{limit} decomposition (9) mimic CP and their joint fit will be close to the CP fit. They contain no interaction term between rating scales and TV shows. The third term does contain such an interaction and, hence, it will be small.

5. Discussion

In this paper, we have demonstrated a novel method of Stegeman (2012, 2013a) to overcome the problem of non-existence of a best-fitting CP model for a three-way array $\mathcal{Z}$. The CP sequence $\mathcal{Y}^{(n)}$ featuring diverging components is used to determine the form of the CP_{limit} decomposition of its limit point $\mathcal{X}$, where CP_{limit} contains interaction terms. Next, this CP_{limit} decomposition is fitted to $\mathcal{Z}$ using initial values obtained from the CP decomposition of $\mathcal{Y}^{(n)}$. For the TV-ratings data, we showed that the CP_{limit} decomposition of $\mathcal{X}$ finds main components with the same interpretation as orthogonally rotated Tucker models or CP with an orthogonality constraint. However, CP_{limit} has higher joint fit of the main components than Tucker models, contains only one small interaction term, and does not impose the unnatural constraint of orthogonality. When there are only groups of two diverging components, as with the TV-ratings data, the CP_{limit} decomposition is the smallest nondiverging Tucker decomposition with a constrained $R \times R \times R$ core; see Section 2.

We also studied the uniqueness properties of the CP_{limit} decomposition of $\mathcal{X}$. Although there is some rotational freedom for the "Violence" and "Sensitivity" components, fixing this by standard rotation criteria did not change the interpretation of the main components, and their joint fit decreased only slightly. Moreover, any rotated CP_{limit} decomposition still features "Violence" and "Sensitivity" components or mixtures of them.

For the TV-ratings data, CP_{limit} provides a summary that is easier to interpret, more efficient, and more intuitive compared to Tucker and CP with orthogonal TV show components. Fitting CP_{limit} is easy and fast, by using an alternating least squares algorithm. Apart from imposing constraints in CP, this is the first solution for diverging CP components that is generally applicable. Further research is necessary to assess the practical usefulness of CP_{limit} for other three-way datasets featuring diverging CP components.

Lundy et al. (1989) are also interested in the size of possible interaction terms between the CP components. For the solution with $R = 3$ and orthogonal TV show components $(\mathbf{A}, \mathbf{B}, \mathbf{C})$, they compute the best fitting $3 \times 3 \times 3$ core array $\mathcal{G}$ such that the fit of $\mathcal{Z} \approx (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathcal{G}$ is maximized. This so-called PFCORE procedure uses the interaction sizes in $\mathcal{G}$ to interpret the solution. Lundy et al. (1989) obtain large interaction weights in $\mathcal{G}$ for "Humor" rating scales and "Sensitive" TV shows (negative for person loadings $\mathbf{c}_1$ and positive for $\mathbf{c}_3$) and "Violent" TV shows (positive for $\mathbf{c}_1$ and negative for $\mathbf{c}_3$). They conclude that these interaction terms show differences in the sense of humor among the raters. Raters with high $\mathbf{c}_1$ loading find "Violent" TV shows funny and "Sensitive" TV shows not, while the reverse is true for raters with high $\mathbf{c}_3$ loading.

However, our best-fitting core array $\mathcal{G}$ is rather different from Lundy et al. (1989); see Table 5. The interaction weights are not as large (Lundy et al. find two weights of 0.67 and 0.49 magnitude) and the larger weights are not in the same place. Hence, the PFCORE method does not appear to be robust. Moreover, this approach does not acknowledge the fact that the CP sequence featuring diverging components is converging to a CP_{limit} decomposition with *specific* interaction terms as proven in Stegeman (2012, 2013a).

In general, our analysis and application are in line with Harshman (2004) who states that diverging CP components occur when "CP is trying to model Tucker variation". However, the results of Stegeman (2012, 2013a) enable us to obtain the *exact* form of the CP_{limit} decomposition, that may be seen as a Tucker decomposition with a lot of weights g_{rpq} set to zero.

Matlab codes are available online for finding the correct form of CP_{limit}, obtaining initial values, and fitting it to a dataset that yields diverging CP components; see http://www.gmw.rug.nl/~stegeman.

Acknowledgments

The author would like to thank Henk Kiers, University of Groningen, for discussions on the subject of this paper and for his suggestions on how to fix the rotational freedom in CP_{limit}. This research is supported by the Dutch Organisation for Scientific Research (NWO), VIDI grant 452-08-001.

References

Acar, E., Yener, B., 2009. Unsupervised multiway data analysis: a literature survey. IEEE Trans. Knowl. Data Eng. 21, 1–15.
Bro, R., Smilde, A.K., 2003. Centering and scaling in component analysis. J. Chemometr. 17, 16–33.
Browne, M.W., 2001. An overview of analytic rotation in exploratory factor analysis. Multivariate Behav. Res. 36, 111–150.
Carroll, J.B., 1953. An analytic solution for approximating simple structure in factor analysis. Psychometrika 18, 23–28.
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.
Ceulemans, E., Kiers, H.A.L., 2006. Selecting among three-mode principal component models of different types and complexities: a numerical convex hull based method. British J. Math. Statist. Psych. 59, 133–150.
Comon, P., Luciani, X., de Almeida, A.L.F., 2009. Tensor decompositions, alternating least squares and other tales. J. Chemometr. 23, 393–405.
De Lathauwer, L., 2008. Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness. SIAM J. Matrix Anal. Appl. 30, 1033–1066.
De Lathauwer, L., Castaing, J., 2007. Tensor-based techniques for the blind separation of DS-CDMA signals. Signal Process. 87, 322–336.
De Silva, V., Lim, L.-H., 2008. Tensor rank and the ill-posedness of the best low-rank approximation problem. SIAM J. Matrix Anal. Appl. 30, 1084–1127.
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 J. Matrix Anal. Appl. 34, 855–875.
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 J. Matrix Anal. Appl. 34, 876–903.
Harshman, R.A., 1970. Foundations of the Parafac procedure: models and conditions for an explanatory multimodal factor analysis. In: UCLA Working Papers in Phonetics, Vol. 16. pp. 1–84.
Harshman, R.A., 2004. The problem and nature of degenerate solutions or decompositions of 3-way arrays. In: Talk at the Tensor Decompositions Workshop, Palo Alto, USA.
Harshman, R.A., De Sarbo, W.S., 1984. An application of Parafac to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques. In: Law, H.G., Snyder, C.W., Hattie, J.A., McDonald, R.P. (Eds.), Research Methods for Multimode Data Analysis. Praeger, New York, pp. 602–642.
Harshman, R.A., Lundy, M.E., 1984. Data preprocessing and the extended Parafac model. In: Law, H.G., Snyder, C.W., Hattie, J.A., McDonald, R.P. (Eds.), Research Methods for Multimode Data Analysis. Praeger, New York, pp. 216–284.
Harshman, R.A., Lundy, M.E., 1994. PARAFAC: parallel factor analysis. Comput. Stat. Data Anal. 18, 39–72.
Hitchcock, F.L., 1927a. The expression of a tensor or a polyadic as a sum of products. J. Math. Phys. 6, 164–189.
Hitchcock, F.L., 1927b. Multiple invariants and generalized rank of a p-way matrix or tensor. J. Math. Phys. 7, 39–70.
Kiers, H.A.L., 1998a. Joint orthomax rotation of the core and component matrices resulting from three-mode principal components analysis. J. Classification 15, 245–263.
Kiers, H.A.L., 1998b. Three-way Simplimax for oblique rotation of the three-mode factor analysis core to simple structure. Comput. Stat. Data Anal. 28, 307–324.
Kiers, H.A.L., 2006. Properties of and algorithms for fitting three-way component models with offset terms. Psychometrika 71, 231–256.
Kiers, H.A.L., Smilde, A.K., 1998. Constrained three-mode factor analysis as a tool for parameter estimation with second-order instrumental data. J. Chemometr. 12, 125–147.
Kiers, H.A.L., Van Mechelen, I., 2001. Three-way component analysis: principles and illustrative application. Psychological Methods 6, 84–110.
Kolda, T.G., Bader, B.W., 2009. Tensor decompositions and applications. SIAM Rev. 51, 455–500.
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.
Kroonenberg, P.M., 2008. Applied Multiway Data Analysis. In: Wiley Series in Probability and Statistics, Wiley, Hoboken, NJ.
Kroonenberg, P.M., De Leeuw, J., 1980. Principal component analysis of three-mode data by means of alternating least squares algorithms. Psychometrika 45, 69–97.
Kruskal, J.B., Harshman, R.A., Lundy, M.E., 1989. How 3-MFA data can cause degenerate Parafac solutions, among other relationships. In: Coppi, R., Bolasco, S. (Eds.), Multiway Data Analysis. North-Holland, pp. 115–121.
Lim, L.-H., Comon, P., 2009. Nonnegative approximations of nonnegative tensors. J. Chemometr. 23, 432–441.
Lim, L.-H., Comon, P., 2010. Multiarray signal processing: tensor decomposition meets compressed sensing. C. R. Acad. Sci. Mecanique 338, 311–320.
Lundy, M.E., Harshman, R.A., Kruskal, J.B., 1989. A two-stage procedure incorporating good features of both trilinear and quadrilinear models. In: Coppi, R., Bolasco, S. (Eds.), Multiway Data Analysis. North-Holland, pp. 123–130.
Paatero, P., 2000. Construction and analysis of degenerate Parafac models. J. Chemometr. 14, 285–299.
Rocci, R., Giordani, P., 2010. A weak degeneracy revealing decomposition for the Candecomp/Parafac model. J. Chemometr. 24, 57–66.
Siciliano, R., Mooijaart, A., 1997. Three-factor association models for three-way contingency tables. Comput. Stat. Data Anal. 24, 337–356.
Sidiropoulos, N.D., Giannakis, G., Bro, R., 2000a. Blind Parafac receivers for DS-CDMA systems. IEEE Trans. Signal Process. 48, 810–823.
Sidiropoulos, N.D., Bro, R., Giannakis, G., 2000b. Parallel factor analysis in sensor array processing. IEEE Trans. Signal Process. 48, 2377–2388.
Smilde, A., Bro, R., Geladi, P., 2004. Multi-way Analysis: Applications in the Chemical Sciences. Wiley, Chichester.
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., 2007. Degeneracy in Candecomp/Parafac explained for several three-sliced arrays with a two-valued typical rank. Psychometrika 72, 601–619.
Stegeman, A., 2008. Low-rank approximation of generic $p \times q \times 2$ arrays and diverging components in the Candecomp/Parafac model. SIAM J. Matrix Anal. Appl. 30, 988–1007.
Stegeman, A., 2009. Using the simultaneous generalized Schur decomposition as a Candecomp/Parafac algorithm for ill-conditioned data. J. Chemometr. 23, 385–392.
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.
Stegeman, A., 2012. Candecomp/Parafac: from diverging components to a decomposition in block terms. SIAM J. Matrix Anal. Appl. 33, 291–316.
Stegeman, A., 2013a. A three-way Jordan canonical form as limit of low-rank tensor approximations. SIAM J. Matrix Anal. Appl. 34, 624–650.
Stegeman, A., 2013b. 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.

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 J. Matrix Anal. Appl. 30, 1614–1638.
Stegeman, A., De Lathauwer, L., 2011. Are Diverging CP Components Always Nearly Propotional? Technical Report. Available online as arXiv:1110.1988.
Thurstone, L.L., 1935. The Vectors of Mind. University of Chicago Press, Chicago.
Tomasi, G., Bro, R., 2006. A Comparison of algorithms for fitting the Parafac model. Comput. Stat. Data Anal. 50, 1700–1734.
Tucker, L.R., 1951. A Method for Synthesis of Factor Analysis Studies. Personnel Research Section Report No. 984. Department of the Army, Washington, DC.
Tucker, L.R., 1966. Some mathematical notes on three-mode factor analysis. Psychometrika 31, 279–311.