Constrained Candecomp/Parafac via the Lasso

Paolo Giordani
Dipartimento di Scienze Statistiche
Sapienza Università di Roma
E-mail: paolo.giordani@uniroma1.it

Roberto Rocci
Dipartimento di Studi Economico-Finanziari e Metodi Quantitativi
Università "Tor Vergata"
E-mail: roberto.rocci@uniroma2.it

Abstract: The Candecomp/Parafac (CP) model is a well-known tool for summarizing a three-way array by extracting a limited number of components. Unfortunately, in some cases, the model suffers from the so-called degeneracy, that is solution with diverging and uninterpretable components. To avoid degeneracy orthogonality constraints are usually applied to one of the component matrices. This solves the problem only from a theoretical point of view because the existence of orthogonal components underlying the data is not guaranteed. For this purpose, we consider some variants of the CP model where the orthogonality constraints are relaxed either by constraining only a pair, or a subset, of components or by stimulating the CP solution to be possibly orthogonal. We clarify theoretically that only the latter approach, based on the *Least absolute shrinkage and selection operator* (Lasso), is effective for solving the degeneracy problem. The results of two applications to real life data show its usefulness in practice.

Keywords: Candecomp/Parafac, Degeneracy, Lasso, Orthogonality constraints.

1. Introduction

In three-way analysis the available data usually refer to a set of I observation units on which a number J of quantitative variables are collected during K occasions. Such a data set can be stored in the array $\underline{\mathbf{X}}$ of order ($I \times J \times K$) with generic element x_{ijk} expressing the score of the i-th observation unit ($i = 1, ..., I$) with respect to the j-th variable ($j = 1, ..., J$) at occasion k ($k = 1, ..., K$). The Candecomp/Parafac (CP) model, independently proposed by Carroll & Chang (1970) and Harshman (1970), is a well-known tool for summarizing $\underline{\mathbf{X}}$ by seeking a limited number S of components for the observation unit, variable and occasion modes. The CP model can be written as

$$\mathbf{X}_{\mathrm{A}} = \mathbf{A}(\mathbf{C} \bullet \mathbf{B})' + \mathbf{E}_{\mathrm{A}}, \tag{1}$$

where $\mathbf{X}_{\mathrm{A}}$ is the matrix of order ($I \times JK$) obtained by juxtaposing the frontal slabs of $\underline{\mathbf{X}}$ next to each other. Similarly, $\mathbf{E}_{\mathrm{A}}$ is the error matrix of order ($I \times JK$). Furthermore, $\mathbf{A} \equiv [a_{is}, i = 1, ..., I; s = 1, ..., S]$ of order ($I \times S$) , $\mathbf{B} \equiv [\mathrm{b}_{js}, j = 1, ..., J; s = 1, ..., S]$ of order ($J \times S$) and $\mathbf{C} \equiv [c_{ks}, k = 1,...,K; s = 1, ..., S]$ of order ($K \times S$) are the component matrices for the observation unit, variable and occasion modes, respectively, and • denotes the Khatri-Rao product, i.e. the columwise Kronecker product. The CP model satisfies several desirable properties. Among them, it is worth mentioning that the solution is unique up to scaling and permuting columns of the three component matrices under mild conditions (Kruskal, 1977; Jiang & Sidiropoulos, 2004; Stegeman et al., 2006; Stegeman, 2009a). The model is strongly related to the concept of tensorial rank of a three-way array $\underline{\mathbf{Y}}$, that is the minimum number of components necessary to decompose $\underline{\mathbf{Y}}$ in CP form. It follows that the solution with S components provides the best approximation of $\underline{\mathbf{X}}$ of tensorial rank S. The best approximation is usually computed by using the least-squares criterion. In fact, the optimal component matrices are obtained in such a way to minimize the loss

$$\| \mathbf{X}_A - \mathbf{A}(\mathbf{C}\bullet\mathbf{B})' \|^2, \tag{2}$$

where $\|.\|^2$ denotes the Frobenius norm. In the literature there exist several alternating least squares (ALS) algorithms for the minimization of (2). See, for an overview and a comparison, Tomasi & Bro (2006). Unfortunately, in some situations, CP suffers from the so-called degeneracy, firstly mentioned by Harshman & Lundy (1984). The CP degeneracy has been deeply analyzed in the literature by several authors (one can refer to, e.g., ten Berge et al., 1988; Kruskal et al., 1989; Mitchell & Burdick, 1994; Rayens & Mitchell, 1997; Paatero, 2000; Stegeman, 2006, 2007, 2008, 2009b; De Silva & Lim, 2008, Krijnen et al., 2008, Stegeman & De Lathauver, 2009; Rocci & Giordani, 2010). In the presence of degeneracy, the computation time of a CP algorithm dramatically increases because the values of the loss function decrease very slowly. At the same time the parameters diverge very quickly and some columns in the component matrices tend to be collinear. Therefore, although the loss function value converges to a stable value after an abnormal time, the parameters values are divergent. When a CP solution is degenerate, the highly collinear components usually are uninterpretable. It is recognized that this phenomenon occurs because the CP loss may not have a global minimum. In other terms, the best S-rank approximation of $\underline{\mathbf{X}}$ may not exist. This can be explained observing that, when $S > 1$, there exist arrays of rank $T > S$ that are the limit of sequences of S-rank arrays. When S is the smallest integer for which this holds, then such arrays are of rank equal to T, but of border rank equal to S (Bini, 1980). If so, these arrays can be approximated arbitrary well by arrays of lower rank. See, for instance, Stegeman (2006, 2007), De Silva & Lim (2008), Rocci & Giordani (2010). To avoid degeneracy, it is recommended to minimize (2) subject to some constraints. The most frequently used is to require the columnwise orthogonality of one of the component matrices. As noted by Harshman & De Sarbo (1984) and Harshman & Lundy (1984) the orthogonality constraints allow us to avoid a dimension that splits into two highly correlated versions of itself and, as a consequence, this hopefully allows us to detect additional weaker, but real, dimensions previously discarded. Nevertheless, many authors consider the orthogonality constraints as a good solution only from a technical point of view because they do not accept to interpret the presence of degeneracy as an indication of the presence of orthogonal factors underlying the data.

In this paper, we are going to introduce some variants of the CP model relaxing the orthogonality constraints on one component matrix taking into account that they may be too restrictive for the data array under observation. In particular, two closely related approaches will be analyzed. In the first one, we impose the orthogonality constraints only on a pair or a subset of factors. This is done by decomposing the component matrix in the QR form, say $\mathbf{A} = \mathbf{QR}$ where $\mathbf{Q}$ is columnwise orthonormal and $\mathbf{R}$ upper triangular, and imposing that one or more elements of $\mathbf{R}$ are equal to zero. These constraints may fail in avoiding degeneracy and then a second approach is proposed, which could integrate the first one, where the CP solution is stimulated to be possibly orthogonal. The (unconstrained) CP model and the one with orthogonality constraints are seen as two opposite competitors within which one chooses the best model. Instead of a discrete choice, we propose to choose the best model in a *continuum* ranging from the standard CP model and its orthogonality constrained version. From a technical point of view, this is obtained by imposing an upper bound, say λ, on the sum of the off-diagonal elements of $\mathbf{R}$ taken in absolute value. A λ equal to zero corresponds to the model with the orthogonality constraints while a λ that tends to infinity implies the unconstrained CP model. In both the approaches, the concept of best model is in terms of predictive power measured by means of cross-validation techniques. Although such proposals are

introduced as a remedy for the CP degeneracy, we cannot exclude their application in order to get the best model regardless whether degeneracy occurs or not.
The paper is organized as follows. In the next section we describe the CP model with some pairs of components constrained to be orthogonal and we clarify how it may fail to handle degeneracy. Then, in Section 3 we introduce a constrained version of the CP model such that the solution is orthogonal as much as possible. This is done using the *Least absolute shrinkage and selection operator* (Lasso) proposed by Tibshirani (1996). Section 4 contains the results of some applications on two real data sets, while some concluding remarks are given in Section 5.

2. A constrained version of the Candecomp/Parafac model with pair-wise orthogonal components

We already recalled that in the literature it is implicitly recommended to limit as much as possible the use of orthogonality constraints in the CP model. This is so because a constrained CP model may be too restrictive leading to 'non-real' underlying components. In this respect, it can be fruitful to limit the emphasis of the orthogonality constraints for one mode by requiring that only a pair or a subset of the CP components are orthogonal.

2.1 The model

In order to develop a constrained version of the CP model with pair-wise orthogonal components, the QR-factorization can be used. Without loss of generality, we focus on **A**. The QR-factorization of **A** is **A** = **QR** where **Q** is a columnwise orthonormal matrix of order ($I \times S$) and **R** is an upper triangular matrix of order ($S \times S$). The orthogonality of **A** can be imposed by requiring that **R** is a diagonal matrix of order S, i.e. the off-diagonal elements of **R** are equal to 0. For reasons that will be clarified in the next section, we fix the diagonal elements of **R** to 1. This can always be done by rescaling the columns of **A**, it does not affect the fit of the CP solution provided that such a rescaling is compensated in **B** and/or **C**. Suppose now that we are interested in imposing that components s and s' are orthogonal. Since **Q** is columnwise orthonormal, the orthogonality of s and s' holds if and only if $\mathbf{r}_s'\mathbf{r}_{s'} = 0$, where $\mathbf{r}_s$ and $\mathbf{r}_{s'}$ are the s-th and s'-th columns of **R**, respectively. Taking into account that **R** is upper triangular and contains 1's in its main diagonal (diag(**R**) = **1**), a sufficient condition for getting $\mathbf{r}_s'\mathbf{r}_{s'} = 0$ is that some off-diagonal elements of **R** contained in $\mathbf{r}_s$ and/or $\mathbf{r}_{s'}$ are zero. The following example with $S = 3$ helps to clarify this point. Since

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & r_{13} \\ 0 & 1 & r_{23} \\ 0 & 0 & 1 \end{bmatrix},$$

if $r_{12} = 0$ then components 1 and 2 are orthogonal, if $r_{12} = r_{13} = 0$ then component 1 is orthogonal to components 2 and 3, if $r_{12} = r_{13} = r_{23} = 0$ then all the components are orthogonal.
The pair-wise orthogonal CP (Pw-O-CP) model can then be written as

$$\mathbf{X}_A = \mathbf{QR}(\mathbf{C}\bullet\mathbf{B})' + \mathbf{E}_A, \tag{3}$$

where $\mathbf{Q}'\mathbf{Q} = \mathbf{I}$ and **R** is an upper triangular matrix of order S with the main diagonal elements equal to one and some upper-off-diagonal elements fixed in advance equal to zero in order to guarantee that pair(s) of components are orthogonal. Note that the Pw-O-CP model offers a discrete set of choices among which the best model should be chosen depending on the number of orthogonal pairs of components ranging from zero (CP model) to $S(S - 1)/2$ (CP model with orthogonality constraints).

2.2 An ALS algorithm

The optimal solution of the Pw-O-CP model can be found by solving the following problem.

$$\min_{\mathbf{Q},\mathbf{R},\mathbf{B},\mathbf{C}} \| \mathbf{X}_A - \mathbf{QR}(\mathbf{C}\bullet\mathbf{B})' \|^2,$$
$$\text{s.t.} \quad \mathbf{Q}'\mathbf{Q} = \mathbf{I}, \tag{4}$$
$$\mathbf{R} \text{ upper triangular with diag}(\mathbf{R}) = \mathbf{1} \text{ and fixed-in-advance zero elements.}$$

The constrained minimization of (4) can be performed by implementing a suitable alternating least squares (ALS) algorithm, which represents a modification of the standard ALS algorithms for determining the optimal (unconstrained) CP solution.

*Updates of **B** and **C***

The basic ALS algorithm for the minimization of (2) consists of iteratively solving three multivariate regression problems with respect to **A**, **B** and **C**. The steps for updating **B** and **C** can also be used in the minimization of (4). In order to update **B** and **C**, it can be convenient to rewrite the CP model in (1) in terms of $\mathbf{X}_B$ and $\mathbf{X}_C$ containing, respectively, the horizontal and lateral slabs of $\underline{\mathbf{X}}$ next to each other. Equivalent formulations of the CP model are then given by

$$\mathbf{X}_B = \mathbf{B}(\mathbf{A}\bullet\mathbf{C})' + \mathbf{E}_B, \tag{5}$$
$$\mathbf{X}_C = \mathbf{C}(\mathbf{B}\bullet\mathbf{A})' + \mathbf{E}_C. \tag{6}$$

From (5) it is easy to see that the optimal value of **B**, given **A** and **C**, is

$$\mathbf{B} = \mathbf{X}_B(\mathbf{A}\bullet\mathbf{C})[(\mathbf{A}\bullet\mathbf{C})'(\mathbf{A}\bullet\mathbf{C})]^{-1}. \tag{7}$$

Similarly, given **A** and **B,** the update of **C** can be derived from (6) as

$$\mathbf{C} = \mathbf{X}_C(\mathbf{B}\bullet\mathbf{A})[(\mathbf{B}\bullet\mathbf{A})'(\mathbf{B}\bullet\mathbf{A})]^{-1}. \tag{8}$$

*Update of **Q***

From (4) it is

$$\min_{\mathbf{Q}} \|\mathbf{X}_A - \mathbf{QR}(\mathbf{C}\bullet\mathbf{B})' \|^2,$$
$$\text{s.t.} \quad \mathbf{Q}'\mathbf{Q}=\mathbf{I}. \tag{9}$$

The problem in (9) can be recognized as a matrix regression problem subject to orthonormality constraints. Its solution can be obtained as follows. By exploiting the norm of (9) and taking into account the constraints on **Q,** it is

$$\|\mathbf{X}_A - \mathbf{QR}(\mathbf{C}\bullet\mathbf{B})' \|^2 = \|\mathbf{X}_A\|^2 + \text{tr}\{(\mathbf{C}\bullet\mathbf{B})\mathbf{R}'\mathbf{R}(\mathbf{C}\bullet\mathbf{B})'\} - 2\text{tr}\{(\mathbf{C}\bullet\mathbf{B})\mathbf{R}'\mathbf{Q}'\mathbf{X}_A\}. \tag{10}$$

The constrained minimization of (10) is then equivalent to

$$\max_{\mathbf{Q}} \text{tr}\{\mathbf{Q}'\mathbf{X}_A(\mathbf{C}\bullet\mathbf{B})\mathbf{R}'\},$$
$$\text{s.t.} \quad \mathbf{Q}'\mathbf{Q}=\mathbf{I}, \tag{11}$$

the solution of which can be found by performing the SVD decomposition of $\mathbf{X}_A(\mathbf{C}\bullet\mathbf{B})\mathbf{R}'$ (Cliff, 1966). Indicating by **U** and **V** the matrices holding the left and right hand singular vectors of $\mathbf{X}_A(\mathbf{C}\bullet\mathbf{B})\mathbf{R}'$, respectively, it is

$$\mathbf{Q} = \mathbf{U}\mathbf{V}'. \tag{12}$$

Update of ***R***
In vec notation the problem to be solved is

$$\min_{\mathbf{R}} \| \mathrm{vec}(\mathbf{X}_A) - [(\mathbf{C}\bullet\mathbf{B})\otimes\mathbf{Q}]\mathrm{vec}(\mathbf{R}) \|^2,$$
$$\text{s.t. } \mathbf{R} \text{ upper triangular with } \mathrm{diag}(\mathbf{R}) = \mathbf{1} \text{ and fixed-in-advance zero elements,} \tag{13}$$

where vec(**Y**) is the vector obtained by stacking the columns of **Y** below one another and ⊗ denotes the Kronecker right product. The loss function can be rewritten as

$$\| \mathrm{vec}(\mathbf{X}_A) - [(\mathbf{C}\bullet\mathbf{B})\otimes\mathbf{Q}]\mathrm{vec}(\mathbf{R}) \|^2 = \| \mathrm{vec}(\mathbf{X}_A) - [(\mathbf{C}\bullet\mathbf{B})\otimes\mathbf{Q}][\mathrm{vec}(\mathbf{I})+\mathbf{K}\mathbf{r}] \|^2, \tag{14}$$

where **r** is the vector containing the nonzero elements of **R** above the diagonal and **K** is the matrix mapping **r** into the nonzero upper off-diagonal elements of vec(**R**). For instance, if

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & r_{13} \\ 0 & 1 & r_{23} \\ 0 & 0 & 1 \end{bmatrix},$$

i.e. we want to constrain only the first two components to be orthogonal, then

$$\mathbf{K} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \ \mathbf{r} = \begin{bmatrix} r_{13} \\ r_{23} \end{bmatrix} \text{ and } \mathrm{vec}(\mathbf{R}) = \mathrm{vec}(\mathbf{I}) + \mathbf{K}\mathbf{r}.$$

The loss (14) can also be written as

$$\|\{\mathrm{vec}(\mathbf{X}_A) - [(\mathbf{C}\bullet\mathbf{B})\otimes\mathbf{Q}]\mathrm{vec}(\mathbf{I})\} - \{[(\mathbf{C}\bullet\mathbf{B})\otimes\mathbf{Q}]\mathbf{K}\}\mathbf{r}]\|^2 = \|\mathbf{y} - \mathbf{Z}\mathbf{r}\|^2, \tag{15}$$

where $\mathbf{y} = \{\mathrm{vec}(\mathbf{X}_A) - [(\mathbf{C}\bullet\mathbf{B})\otimes\mathbf{Q}]\mathrm{vec}(\mathbf{I})\}$ and $\mathbf{Z} = [(\mathbf{C}\bullet\mathbf{B})\otimes\mathbf{Q}]\mathbf{K}$, i.e. the matrix containing the columns of $(\mathbf{C}\bullet\mathbf{B})\otimes\mathbf{Q}$ corresponding to the free elements of **R**. From (15) it is clear that the update of **r** boils down to a standard regression problem and

$$\mathbf{r} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}. \tag{16}$$

2.3 Pw-O-CP and degeneracy
In principle the application of Pw-O-CP may represent a remedy for the CP degeneracy. Nonetheless, in our analysis we saw that it often fails because it tends to find solutions suffering

from degeneracy. Specifically, suppose to apply Pw-O-CP with S = 3 to data for which CP degeneracy occurs. For instance, we require two pairs of components to be orthogonal ($r_{12} = r_{13} = 0$). We frequently found that the obtained solution is characterized by the standard features of degeneracy with the unconstrained pair of components highly collinear. To understand this point, let us consider the following array

$$\mathbf{X}_A = \left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \varepsilon \end{array}\right].$$

It is easy to show that the CP model has a perfect fit on this array with

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1/\varepsilon \\ 0 & 0 & 1 \end{bmatrix},\ \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/\varepsilon & 1 \end{bmatrix} \text{ and } \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & \varepsilon \end{bmatrix},$$

where the matrix **A** satisfies the required constraints. When $\varepsilon \to 0$ the solution tends to degeneracy. Therefore, although this in not always the case (it is easy to generate data arrays suffering from CP degeneracy for which a Pw-O-CP solution is not degenerate), in practice, Pw-O-CP does not represent a powerful remedy to degeneracy.

A way to avoid degeneracy can be to impose an upper bound on the off-diagonal elements of **R**. This is clear in the previous example but it can also be shown in general. To this end, we recall that degeneracy occurs when during the iterations the objective loss function seems to converge while the component matrices diverge, i.e. some elements tend to plus or minus infinity, and two or more columns of them tends to be collinear. In this situation the spectral condition number of matrix **A**, defined as the ratio between the maximum and the minimum singular values of **A**, i.e. $\sigma_{max}(\mathbf{A})/\sigma_{min}(\mathbf{A})$, tends to infinity. The connection between condition numbers and CP degeneracy has been already analyzed in the literature, see for instance Krijnen et al. (2008). A bound on the condition number of **A** can be imposed if we consider a CP model where **A** = **QR** and the matrix **R** has the diagonal elements equal to 1 and the off-diagonal elements in absolute value less than or equal to a positive constant λ. In fact, in this case it is possible to show that the condition number of **A** is bounded by a finite monotone increasing function of λ. The proof is based on the following inequality due to Guggenheimer et al. (1995), for a given square matrix **Y** of order S,

$$\frac{\sigma_{\max}(\mathbf{Y})}{\sigma_{\min}(\mathbf{Y})} \le \frac{2}{\det(\mathbf{Y})}\left(\frac{\mathrm{tr}\mathbf{Y}'\mathbf{Y}}{S}\right)^{S/2}. \qquad (17)$$

In our case we have

$$\frac{\sigma_{\max}(\mathbf{A})}{\sigma_{\min}(\mathbf{A})} = \frac{\sigma_{\max}(\mathbf{R})}{\sigma_{\min}(\mathbf{R})} \le 2\left(\frac{\mathrm{tr}\mathbf{R}'\mathbf{R}}{S}\right)^{S/2} \le 2\left[1+\lambda^2(S-1)/2\right]^{S/2}. \qquad (18)$$

It follows that if in absolute value the upper-off-diagonal elements of **R** are bounded by a positive constant λ, then the columns cannot be collinear or nearly collinear. It is important to note that by imposing the constraint diag(**R**) = **1**, the degree of collinearity among the columns of **A** is bounded by the magnitude of the off-diagonal elements of **A**. For the sake of completeness, it is interesting to note that the upper bound in (18) is not sharp because for λ = 0 the bound is 2 while the condition number is equal to 1. A sharper bound can be found in Merikoski et al. (1997).

On the basis of the above considerations, in the next section, we propose to apply a bound to the elements of **R** that can be used with or without the Pw-O-CP approach.

3. A constrained version of the Candecomp/Parafac model via the Lasso

In the previous section, we saw that in Pw-O-CP the orthogonality constraints on **A** are relaxed allowing for orthogonal pairs of components. These were fixed in advance by the researcher constraining to zero some elements of **R**. This proposal does not remarkably restrict the model because the orientations of the axes are not strongly constrained, but, unfortunately, does not necessarily solve the degeneracy problem because the off-diagonal elements of **R** can take arbitrarily large values in absolute value. As explained in the previous section, an improvement able to cope with the degeneracy problem can be done by bounding the off-diagonal elements of **R**. A way to do so is given by the use of the *Least absolute shrinkage and selection operator* (Lasso) technique (Tibshirani, 1996). Lasso is a procedure widely used in regression analysis aiming at shrinking some regression coefficients toward zero and setting others to zero. This is done by imposing that the sum of the absolute values of the regression coefficients is lower than a (positive) threshold λ fixed in advance. Let **y** be the vector of the response variable and **X** be the matrix of the *p* explanatory variables, the Lasso problem can be written as

$$\min_{\boldsymbol{\beta}} \| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \|^2, \quad \text{s.t.} \sum_{j=1}^{p} |\beta_j| \le \lambda, \tag{19}$$

where $\boldsymbol{\beta}$ is the vector of the regression coefficients. It can be shown that the Lasso estimates of the regression coefficients can be found by solving a constrained quadratic programming problem. Differently from other existing shrinkage procedures, a nice property of Lasso is its geometry that tends to produce some zero coefficients (see, for more details, Tibshirani, 1996). In our case this may correspond to find orthogonal components.

The features of the Lasso can be used in the CP domain on one of the three component matrices, without loss of generality, once more we concentrate our attention on **A**. We already saw that **A** = **QR** with diag(**R**) = **1**, and that suitable zero constraints on the elements of **R** allow us to obtain pairs of orthogonal factors. We also apply the Lasso constraints

$$\sum_{s=1}^{S-1} \sum_{s'=s+1}^{S} |r_{ss'}| \le \lambda . \tag{20}$$

Using (20) it follows that the loss function to be minimized is

$$\begin{aligned} &\min_{\mathbf{Q},\mathbf{R},\mathbf{B},\mathbf{C}} \| \mathbf{X}_{\mathrm{A}} - \mathbf{QR}(\mathbf{C} \bullet \mathbf{B})' \|^2, \\ &\text{s.t.} \sum_{s=1}^{S-1} \sum_{s'=s+1}^{S} |r_{ss'}| \le \lambda, \\ &\quad \mathbf{Q}'\mathbf{Q} = \mathbf{I}. \end{aligned} \tag{21}$$

We refer to the problem in (21) as Candecomp/Parafac with Lasso constraints (in brief, CP-Lasso). When λ = 0 it is **R** = **I** and, therefore, **A** is orthonormal. For λ > 0, CP-Lasso solutions often have some off-diagonal elements of **R** equal to zero that may give orthogonal pairs of components. In general, the Lasso limits the magnitude of the upper-diagonal elements of **R** allowing us to obtain a non-degenerate solution with pair-wise orthogonal components because abnormal off-diagonal elements of **R** (higher than λ in absolute value) are unfeasible. Note that it does not necessarily hold that zero constrained elements of **R** lead to orthogonality for at least a pair of components. For instance, when S = 3, if r_{23} = 0, then no pair-wise orthogonal components are determined. By solving (21) for different values of λ we have a *continuum* of solutions that ranges from the orthogonal one when λ = 0 to the standard CP one when $\lambda \to \infty$.

3.1 The ALS algorithm

The optimal parameter matrices can be found by implementing an ALS algorithm for the minimization of (21). Since the Lasso constraints affect only the elements of **R**, the updates of **Q**, **B** and **C** can be done in the same way as described in Section 2.2. To update **R** we can operate as follows.

*Update of **R***

For the sake of simplicity, we consider only the case where none of the off-diagonal elements are set to zero in advance. The general case can be easily derived by the present taking into account the update of **R** shown for the Pw-O-CP. The problem to be solved is

$$\min_{\mathbf{R}} \| \mathbf{X}_{\mathrm{A}} - \mathbf{QR}(\mathbf{C} \bullet \mathbf{B})' \|^2, \quad \text{s.t.} \sum_{s=1}^{S-1} \sum_{s'=s+1}^{S} |r_{ss'}| \leq \lambda. \tag{22}$$

The loss function in (22) can be rewritten as in (14) using **K** and **r** introduced in Section 2.2. Letting $\mathbf{y} = \{\mathrm{vec}(\mathbf{X}_{\mathrm{A}}) - [(\mathbf{C} \bullet \mathbf{B}) \otimes \mathbf{Q}]\mathrm{vec}(\mathbf{I})\}$ and $\mathbf{Z} = [(\mathbf{C} \bullet \mathbf{B})' \otimes \mathbf{Q}]\mathbf{K}$, it follows that (22) boils down to

$$\min_{\mathbf{r}} \| \mathbf{y} - \mathbf{Zr} \|^2, \quad \text{s.t.} \ \mathbf{1}'|\mathbf{r}| \leq \lambda, \tag{23}$$

where |**r**| is the vector containing the elements of **r** in absolute value. The minimization problem in (23) is equivalent to (19) and, thus, can be recognized as a standard Lasso problem. As explained by Tibshirani (1996), the constraints in (23) involving the absolute values can be exploited as a set of 2^H linear constraints where H denotes the length of **r**. Every constraint takes the form $\boldsymbol{\alpha}_l'\mathbf{r} \leq \lambda$, in which $\boldsymbol{\alpha}_l$'s (l = 1 , …, 2^H) are all possible vectors with elements ±l. Instead of 2^H constraints and H variables, the Lasso constraints can also be written as a set of $2H$+1 constraints with respect to $2H$ variables obtained replacing every element of **r**, say r_h, by $r_h^+ - r_h^-$ where r_h^+ and r_h^- are non-negative. In fact, we have

$$\mathbf{1}'|\mathbf{r}| \leq \lambda \Leftrightarrow \sum_h |r_h| \leq \lambda \Leftrightarrow \sum_h r_h^+ - \sum_h r_h^- \leq \lambda, r_h^+ \geq 0, r_h^- \geq 0, h = 1, \ldots, H. \tag{24}$$

In the following we will adopt the transform in (24) implementing one of the existing algorithms to solve LS problems subject to linear constraints. See, for instance, Gill et al. (1981) and Lawson & Hanson (1995). In our case $H = (S^2 - S)/2$ and then we should consider a problem with $(S^2 - S)$ variables and $(S^2 - S) + 1$ constraints. However, it is important to note that in CP-Lasso the number of constraints and variables can be decreased taking into account that, without loss of generality, we can always get non-negative elements on the secondary diagonal of **R** by absorbing the sign changes in **Q** and/or **B** and/or **C**. To clarify this point, let us consider the following example. Suppose that we have a CP-Lasso solution (**Q**,**R**,**B**,**C**) with

$$\mathbf{R} = \begin{bmatrix} 1 & -2 & x & x \\ 0 & 1 & 3 & x \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where x's denote real numbers. Starting from r_{12}=−2, if we post-multiply **R** by

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

we obtain a non-negative value for r_{12} (=2) but, at the same time, we get r_{22}=−1. To avoid this, it is sufficient to pre-multiply **R** by **T**. Summing up, this consists of replacing **R** by

$$\mathbf{TRT} = \begin{bmatrix} 1 & 2 & x & x \\ 0 & 1 & -3 & x \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and compensating such a scaling in, respectively, **Q** and **B** or **C** without affecting the fit of the CP solution. The new solution could be (**QT**,**TRT**,**BT**,**C**). **T** is the identity matrix with the second diagonal element equal to −1 because we need to change the sign of r_{12} which is located in the second column of **R**. Since the post-multiplication of **R** by **T** change the signs of all the elements of the second column of **R**, it is necessary to pre-multiply the resulting matrix **RT** by **T** in order to have the second diagonal elements equal to 1. In general, we can change the sign of $r_{s-1,s}$ by pre- and post-multiplying **R** by **T** which is the identity matrix with $t_{ss} = -1$. It is important to note that such operations do not affect the sub-matrix of **R** with vertices r_{11}, $r_{s-1,s-1}$. This implies that the change of sign of $r_{s-1,s}$ does not affect $r_{s'-1,s'}$ for $s' < s$. It follows that starting from r_{12}, we can change the sign sequentially to all the elements of the second diagonal of **R.** Such a property reduces the computational complexity of the Lasso constraints. In fact, by using the fact that $S-1$ elements of **r** (the number of elements of the second diagonal of **R**) are surely non-negative, we can discard (S-1) variables of the form r_h^-. It follows that we have $(S\text{-}1)^2$ variables with $(S\text{-}1)^2 + 1$ constraints.

The steps of the ALS algorithm can be summarized as follows.

Step 0: Randomly or rationally generate starting values for **Q**, **R** and **C**.
Step 1: Update **B** using (7), keeping fixed **Q**, **R** and **C**.
Step 2: Update **C** using (8), keeping fixed **Q**, **R** and **B**.
Step 3: Update **Q** using (12), keeping fixed **R**, **B** and **C**.
Step 4: Update **R** by the minimization of (23), keeping fixed **Q**, **B** and **C**.
Step 5: Check convergence.

In the next section, we will apply the Pw-O-CP and CP-Lasso to real data sets already examined in the literature in order to test their effectiveness.

4. Application

In this section we discuss how CP-Lasso works in practice by means of two applications.

4.1 TV data

The data we considered are the famous TV data (Lundy et al., 1989). They consist of ratings of 15 American television shows on 16 scales made by 40 psychology students at the University of Western Ontario in 1981. However, the analysis we made were based on the ratings given by a subset of 30 students in order to avoid the presence of missing items in the data array. Therefore, the data under investigation were stored in an array $\underline{\mathbf{X}}$ of order (16×15×30). Prior to fitting the models to the data, the data were preprocessed by centering across TV programs and scales and normalizing within the scales.

It is well-known that CP degeneracy occurs for S = 3. To avoid it, Lundy et al. (1989) considered CP with orthogonality constraints on the columns of the component matrix for the scales, i.e. **A**.

We applied CP-Lasso with three components to the TV data constraining the columns of **A** considering different values of λ ranging from 0 (i.e. orthogonal constraints) to 4.

The choice of the best value of λ, and indeed of the optimal model, was carried out in terms of predictive power considering cross-validation techniques (in the three-way context see, for instance, Louwerse et al., 1999). Specifically, we operated by removing from the array one frontal slice (i.e. one student) and then applying the model to the resulting array. The so-obtained estimated model was then used to estimate the values of the frontal slice not involved in the estimation. In formulas, the removed slab $\mathbf{X}_{..k}$ is estimated by

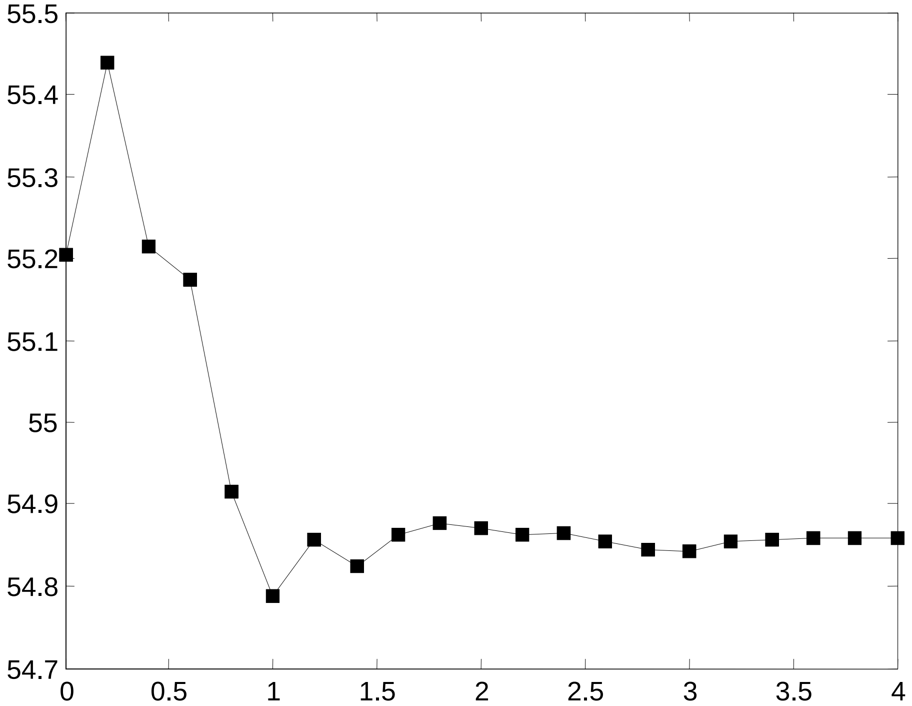
$$\mathrm{vec}(\hat{\mathbf{X}}_{..k}^{(-k)\prime}) = (\mathbf{B}^{(-k)}\bullet\mathbf{A}^{(-k)})[(\mathbf{B}^{(-k)}\bullet\mathbf{A}^{(-k)})'(\mathbf{B}^{(-k)}\bullet\mathbf{A}^{(-k)})]^{-1}(\mathbf{B}^{(-k)}\bullet\mathbf{A}^{(-k)})'\mathrm{vec}(\mathbf{X}'_{..k}), \tag{25}$$

where $\mathbf{A}^{(-k)}$ and $\mathbf{B}^{(-k)}$ are estimated from the data after removing the frontal slab $\mathbf{X}_{..k}$. Repeating this process in such a way that every frontal slice of the arrays was left out once, a measure of the predictive power of the model was then found as

$$CV_{CP} = \frac{\sum_{k=1}^{K}\left\|\mathbf{X}_{..k} - \hat{\mathbf{X}}_{..k}^{(-k)}\right\|^2}{\left\|\mathbf{X}_{\mathrm{A}}\right\|^2}100. \tag{26}$$

The index in (26) is the sum of squares of the differences between all the observed and fitted frontal slices divided by the sum of squares of the elements of $\underline{\mathbf{X}}$ expressed as a percentage. The motivation for such a cross-validation procedure is two-fold. On one side it allows us to reduce the computational burden in comparison with other cross-validation techniques for three-way models (for instance those based on eliminating single data points). On the other side it appears to be consistent with the parallel proportional profiles principle underlying the CP model in the sense that a CP model fits well the data if removing one frontal slice, the axis orientations do not strongly differ and, therefore, the parallel proportional profiles principle holds. An indication about the optimal λ is the minimum of the CV_{CP} index computed over an interval of λ values. We assumed the algorithm converged if the percentage of the residual sum of squares decreased less than 10^{-9} in two subsequent iterations. For every condition (λ value and frontal slice removed) the algorithm was run using five starting points. Two rational starts for **A**, **B** and **C** were considered using the solutions resulting from the standard CP and the CP with orthogonal constraints), respectively, with convergence criterion equal to 10^{-6}, and the remaining three starting points were independently generated from N(0,1). The obtained CV_{CP} values are reported in Figure 1.

Figure 1 Values of CV_{CP} computed on the TV data for different values of λ in the interval [0,4] with step = 0.2.



The minimum can be found when λ=1. The corresponding CV_{CP} is 54.79%. The QR decomposition of the optimal **A** obtained by CP-Lasso with λ = 1.0 was such that r_{12} = 0.12, r_{13} = 0, and r_{23} = 0.88, hence components 1 and 3 are orthogonal. The solution was easily interpretable as one can see inspecting the normalized component matrices for the TV programs and the scales (we do not report the component matrix for the students, which contains all positive scores).

Table 1. Component matrix for the scales (scores higher than 0.30 in absolute value are in bold)

Scales	Comp. 1	Comp. 2	Comp. 3
Thrilling - Boring	**0.30**	-0.11	0.10
Intelligent - Idiotic	0.17	**0.34**	-0.27
Erotic - Not Erotic	-0.01	**-0.32**	**0.37**
Sensitive – Insensitive	**-0.44**	0.04	-0.03
Interesting – Uninteresting	0.27	0.11	**-0.30**
Fast – Slow	0.25	-0.15	0.26
Intellectually Stimulating – Intellectually Dull	0.16	**0.34**	**-0.36**
Violent – Peaceful	0.25	-0.13	**0.54**
Caring – Callous	**-0.44**	0.01	0.09
Satirical – Non Satirical	-0.04	**-0.37**	-0.04
Informative – Uninformative	0.18	**0.36**	**-0.33**
Touching – Leave Me Cold	**-0.35**	-0.04	-0.02
Deep – Shallow	-0.15	0.24	-0.23
Tasteful – Crude	-0.26	0.17	-0.01
Real – Fantasy	0.14	**0.35**	-0.11
Funny – Not Funny	-0.11	**-0.35**	-0.09

Note: Negative scores refer to the left side of the bipolar scale

Table 2. Component matrix for the TV programs (the highest scores per row higher than 0.25 in absolute value are in bold)

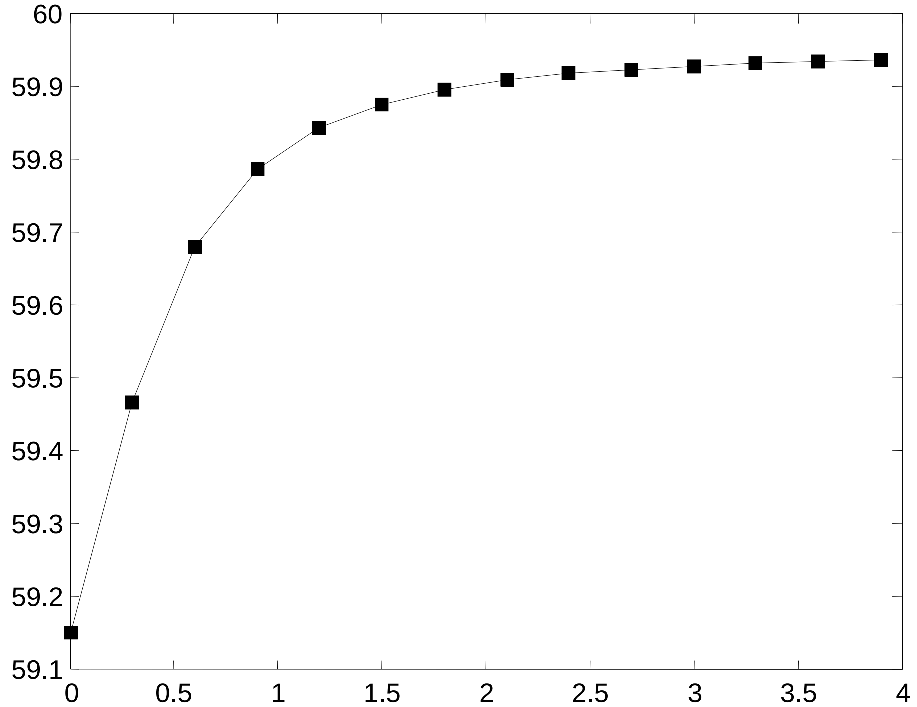
TV programs	Comp. 1	Comp. 2	Comp. 3
Mash	0.05	0.17	0.18
Charlie's angels	0.04	0.07	**-0.52**
All in the family	0.07	0.27	**0.33**
60 minutes	-0.15	**-0.27**	0.11
The tonight show	-0.18	0.26	**0.40**
Let's make a deal	-0.04	0.13	-0.17
The Waltons	**0.56**	-0.23	-0.15
Saturday night live	-0.30	**0.43**	0.29
News	-0.26	**-0.34**	-0.07
Kojak	-0.10	0.03	**-0.31**
Mork and Mindy	0.13	**0.38**	0.22
Jacques Cousteau	-0.05	**-0.33**	0.13
Football	**-0.36**	-0.08	-0.28
Little house on the prairie	**0.54**	-0.18	-0.18
Wild kingdom	0.06	**-0.29**	0.01

The first two components resemble the first two components found in Lundy et al. (1989). In particular, Component 1 is mainly related to Sensitive, Caring, Touching and Boring and can be interpreted as ‘Sensitivity’ as Component 2 from Lundy et al. (1989). The TV programs Little house on the prairie and The Waltons take the highest positive scores and Football the highest negative one. Component 2 strongly depends on scales related to ‘Humor’, as it is the case for Component 1 in Lundy et al. (1989). As expected, Saturday night life, Mork and Mindy and, with negative sign, News, Jacques Cousteau, Wild kingdom and 60 minutes take the highest scores in absolute value. Although Violence – Peaceful is the most important scale for Component 3 resulting from CP-Lasso (Peaceful with score = 0.54) and from Lundy et al. (1989), the two components sensibly differ. In fact, we found a relevant importance of Not Erotic, Intellectually Stimulating, Informative and Interesting (only the latter scale plays a relevant role in Lundy et al., 1989) and the TV programs well characterized by this component are The tonight show, All in the family and, with negative sign, Charlie’s angels and Kojak. This component seems to capture ‘thoughtful’ shows.
The application of Pw-O-CP with a number of orthogonal pairs of components ranging from 1 to 3 always led to degeneracy and, hence, the details are not reported here.

4.2 Sensory bread data

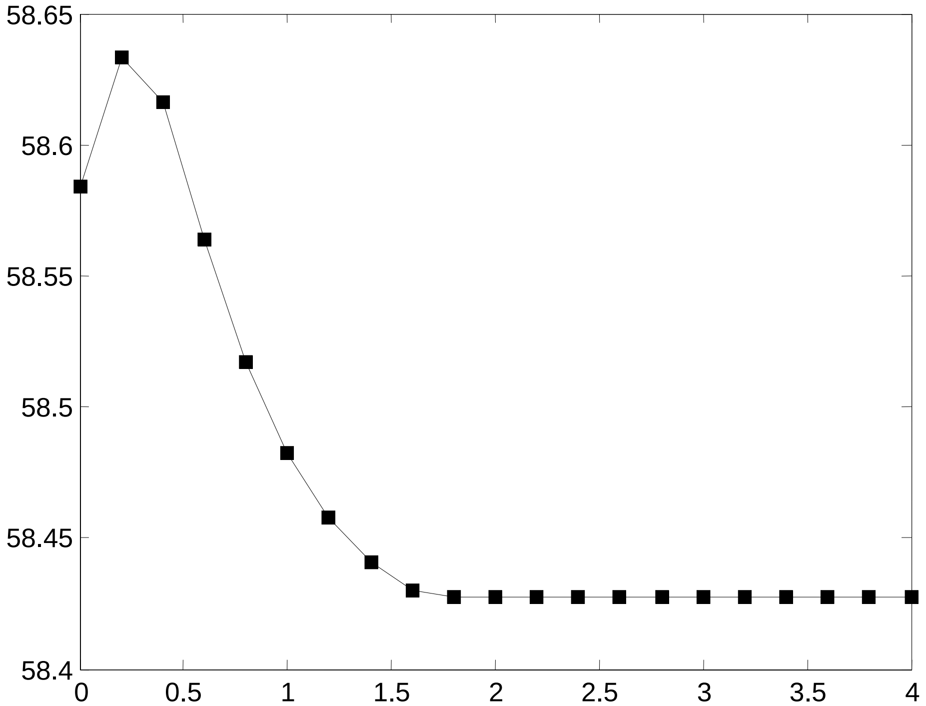
The Sensory bread data (Bro, 1998) refer to the assessments made by $K = 8$ judges with respect to $J = 11$ attributes on five different breads baked in replicates, hence $I = 10$. In the literature, the three-way array has been centered across the breads and then analyzed by CP using $S = 2$ components (see, for instance, Bro & Kiers, 2003). The solution is meaningful and does not show degeneracy. However, we used this data set to test the behavior of our procedure in case of non degeneracy. The same preprocessing and number of components were chosen as in Bro (1998) before applying CP-Lasso with λ ranging from 0 to 4 (step = 0.2).
We got the CV_{CP} values displayed in Figure 2, from which we easily concluded that the best model in terms of predictive power is CP-Lasso with $\lambda = 0$, that is CP with orthogonality constraints. The analysis of the component matrices of the two models show that negligible differences can be found and the interpretation of the extracted components does not vary and is similar to the one reported by Bro (1998).

Figure 2 Values of CV_{CP} computed on the Sensory bread data for different values of λ in the interval [0,4] with step = 0.2.



One may wonder if there exists a reason for choosing the orthogonality constrained solution. In Rocci & Giordani (2010) we saw that removing judge n.6 from the analysis leads to degeneracy when running CP with two components. We can state that the data are affected by a sort of 'latent' degeneracy and this can motivate the choice of λ = 0. When we run CP-Lasso on the reduced data array, we found the *CV* values given in Figure 3.

Figure 3 Values of CV_{CP} computed on the Sensory bread data without judge n.6 for different values of λ in the interval [0,4] with step = 0.2.



The minimum is obtained for λ = 2. By inspecting Figure 3, we see that this value is not significant as the curve becomes flat on approximately the same value of CV_{CP}. In fact, the solution for λ = 2 has a degree of uninterpretability similar to the degenerate solution. In this case we can conclude that a global minimum for CV_{CP} does not exist and an indication for the optimal λ can be given by a local minimum. This can be found at λ = 0 and the corresponding solution is almost equal to the one found before.

5. Final remarks

This paper dealt with a variant of the CP model that has been proposed in order to avoid degeneracy. This paper dealt with a variant of the CP model that has been proposed in order to avoid degeneracy. Rather than constraining the CP solution to be fully orthogonal, we suggested to constraint the components to be possibly orthogonal and to limit the level of collinearity between the non orthogonal components. This was done by considering the QR decomposition of one of the component matrix, without loss of generality **A** = **QR**, and by constraining the sum of the absolute values of the off-diagonal elements of the upper triangular matrix **R** to be lower than a pre-specified threshold λ. In doing so, we took inspiration from the well-known shrinkage procedure called Lasso. The resulting variant of CP has been therefore labeled CP-Lasso. It allows us to relax the orthogonality constraints avoiding degeneracy.

A relevant choice to be made concerned the threshold λ. In the applications we proposed to determine the optimal value of λ according to the predictive power of the model. Thus, the optimal value of λ was the one for which the model had the best prediction in terms of a suitable cross-validation measure for three-way models. In the future, it can be interesting to deeply investigate this point by studying how to optimally choice the threshold λ both theoretically and in practice by means of simulation experiments and real life examples.

References

Bini, D. (1980). Border rank of a $p \times q \times 2$ tensor and the optimal approximation of a pair of bilinear forms. In Automata, Languages and Programming, (Lecture Notes in Computer Science, 85), de Bakker JW, van Leeuwen J (eds). Springer: New York, pp. 98-108.

Bro R. (1998). *Multi-way analysis in the food industry. Models Algorithms and Applications*. University of Amsterdam, 1998.

Bro R., & Kiers, H.A.L. (2003). A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics, 17*, 274-286.

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.

Cliff, N. (1966). Orthogonal Rotation to Congruence. *Psychometrika, 31*, 33-42.

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.

Gill, P.E., Murray, W., & Wright, M.H. (1981). *Practical Optimization*. Academic Press, London.

Guggenheimer, H.W., Edelman, A.S., & Johnson, C.R. (1995). A simple estimate of the condition number of a linear system. *The College Mathematics Journal, 26*, 2-5.

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

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

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

Jiang, T., & Sidiropoulos, N.D. (2004). Kruskal's permutation lemma and the identification of

Candecomp/Parafac and bilinear models with constant modulus constraints. *IEEE Transactions on Signal Processing, 52*, 2625-2636.
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.
Kruskal, J.B. (1977). Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Linear Algebra and its Applications, 18*, 95-138.
Kruskal, J.B., Harshman, R.A., & Lundy, M.E. (1989). How 3-MFA data can cause degenerate PARAFAC solutions, among other relationships. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 115-122). Elsevier, Amsterdam.
Lawson, C.L., & Hanson, R.J. (1995). *Solving Least Squares Problems*, (Classics in Applied Mathematics, Vol. 15). SIAM, Philadelphia.
Louwerse, D.J., Smilde, A.K., & Kiers, H.A.L. (1999). Cross-validation of multiway component models. *Journal of Chemometrics, 13*, 491-510.
Lundy, M.E., Harshman, R.A., & Kruskal, J.B. (1989). A two stage procedure incorporating good Features of both Trilinear and Quadrilinear Models. In *Multiway Data Analysis*, Coppi R, Bolasco S (eds). Elsevier, Amsterdam, 123-130.
Merikoski, J.K., Urpala, U., Virtanen, A., Tam, T.-Y., & Uhlig, F. (1997). A best upper bound for the 2-norm condition number of a matrix. *Linear Algebra and its Applications, 254*, 355-365.
Mitchell, B.C., & Burdick, D.S. (1994). Slowly converging Parafac sequences: swamps and two-factor degeneracies. *Journal of Chemometrics, 8*, 155-168.
Paatero, P. (2000). Construction and analysis of degenerate Parafac models. *Journal of Chemometrics, 14*, 285-299.
Rayens, W.S., & Mitchell, B.C. (1997). Two-factor degeneracies and a stabilization of Parafac. *Chemometrics and Intelligent Laboratory Systems, 38*, 173-181.
Rocci, R., & Giordani, P. (2010). A weak degeneracy decomposition for the CANDECOMP/PARAFAC model. *Journal of Chemometrics, 24*, 57-66.
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 and Indscal 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×q×2 arrays and diverging components in the Candecomp/Parafac model. *SIAM Journal on Matrix Analysis and Application,; 30*, 988-1007.
Stegeman, A. (2009a). On uniqueness conditions for Candecomp/Parafac and Indscal with full column rank in one mode. *Linear Algebra and its Applications, 431*, 211-227.
Stegeman, A. (2009b). Using the Simultaneous Generalized Schur Decomposition as a Candecomp/Parafac algorithm for ill-conditioned data. *Journal of Chemometrics, 23*, 385-392.
Stegeman, A., Ten Berge, J.M.F., & De Lathauwer, L. (2006). Sufficient conditions for uniqueness in Candecomp/Parafac and Indscal with random component matrices. *Psychometrika, 71*, 219-229.
Stegeman, A., & De Lathauver, 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.
ten Berge, J.M.F., Kiers, H.A.L., & De Leeuw, J. (1988). Explicit Candecomp/Parafac solutions for a contrived 2×2×2 array of rank three. *Psychometrika, 53*, 579-584.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B, 58*, 267-288.
Tomasi, G., & Bro, G. (2006). A comparison of algorithms for fitting the PARAFAC model. *Computational Statistics & Data Analysis, 50*, 1700-1734.