Simultaneous component analysis by means of Tucker3

Alwin Stegeman

Group Science, Engineering and Technology, KU Leuven – Kulak,
E. Sabbelaan 53, B-8500 Kortrijk, Belgium,

and

Department of Electrical Engineering (ESAT), KU Leuven,
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium,

alwin.stegeman@kuleuven.be,
http://www.alwinstegeman.nl

The author would like to thank Kim Shifren (Department of Psychology, Towson University) for permission to use the PANAS dataset, and Henk Kiers (Department of Psychometrics and Statistics, University of Groningen) for suggestions on how to rotate a Tucker3 solution to another one. Research supported by: (1) Research Council KU Leuven: C1 project c16/15/059-nD; (2) the Belgian Federal Science Policy Office: IUAP P7 (DYSCO II, Dynamical systems, control and optimization, 2012-2017)

Simultaneous component analysis by means of Tucker3

Abstract

A new model for simultaneous component analysis (SCA) is introduced that contains the existing SCA models with common loading matrix as special cases. The new SCA-T3 model is a multi-set generalization of the Tucker3 model for component analysis of three-way data. For each mode (observational units, variables, sets) a different number of components can be chosen and the obtained solution can be rotated without loss of fit to facilitate interpretation. SCA-T3 can be fitted on centered multi-set data and also on the corresponding covariance matrices. For this purpose, alternating least squares algorithms are derived. SCA-T3 is evaluated in a simulation study and its practical merits are demonstrated for several benchmark datasets.

Keywords: simultaneous components analysis, multi-set data, Tucker, Parafac, rotation

1 Introduction

Simultaneous component analysis (SCA) aims to summarize observed (centered) scores of variables in samples of several subpopulations into a small number of components for each sample. Such data are also known as multi-set data, where each set consists of observations of the same variables in a sample from one subpopulation. When no constraints are imposed in the SCA problem, the best summary in terms of explained variance is given by principal component analysis (PCA) for each sample separately. To facilitate the comparison of the components found for each subpopulation, several constraints have been proposed for SCA. Imposing the component weights matrices to be congruent (i.e., columnwise proportional) guarantees equal definitions of the components as linear combinations of the observed variables for each subpopulation. This method is referred to as SCA-W. Alternatively, one may impose that the structure matrices are congruent (SCA-S) to establish equal interpretation of the components across subpopulations. Yet another possibility is to impose pattern congruence (SCA-P) to obtain proportional regression weights for optimally reconstructing the

variables from the components. Relations between SCA-W, SCA-S, and SCA-P are discussed in Kiers and Ten Berge (1994). The authors show that SCA-W always has largest explained variance, followed by SCA-P, and then SCA-S. A disadvantage of SCA-W is that it discriminates poorly between subpopulations with clearly different correlation structures. Conversely, there can be a large gap in explained variance between doing separate PCAs and SCA-S when correlation structures are very similar in subpopulations. Hence, not one method of SCA seems to be preferred in all cases.

Working within the framework of SCA-P, Timmerman and Kiers (2003) consider three other variants of SCA. The SCA-PF2 model is based on the multi-set Parafac2 model (Harshman, 1972) and is equal to SCA-P with identical component covariances across subpopulations. The SCA-IND model is equal to SCA-PF2 with uncorrelated components in all subpopulations. The SCA-ECP model is a constrained version of SCA-PF2 with identical component variances across subpopulations. Timmerman and Kiers (2003) apply these four SCA models to a 20-item mood test administered to 12 individuals diagnosed with Parkinson's disease at a number of consecutive days that differs per subject. Here, the multi-set data is organized as days-by-items for each patient, and each sample or set now corresponds to one patient.

In this paper, we introduce a new SCA model that is a multi-set generalization of the Tucker3 model for component analysis of three-way data (Tucker, 1966; Kroonenberg, 2008). We name our model SCA-T3. The relation between SCA-T3 and Tucker3 is analogous to the relation between Parafac2 and the Parafac model for three-way data (Harshman, 1970; Carroll and Chang, 1970): the component score matrix is replaced by component score matrices for each subpopulation. Contrary to Parafac and Parafac2, a Tucker3 solution is not rotationally unique and rotations can be applied to obtain an interpretable solution. The same holds for SCA-T3. Also, different numbers of components can be chosen for the observational units, variables, and subpopulations. For Parafac and Parafac2 the number of components is the same for all three modes of the data. SCA-T3 contains SCA-P, SCA-PF2, SCA-IND, and SCA-ECP as special cases. Hence, the SCA-T3 model is more general and more versatile than the existing SCA models within the SCA-P framework. Its added value will be shown in applications to several datasets in the literature.

This paper is organized as follows. In section 2 we give a formal presentation of the existing SCA models and our SCA-T3 model. This is preceded by a formal introduction of

Parafac and Tucker3 and related concepts. In section 3 we derive an alternating least squares (ALS) algorithm to fit SCA-T3. We also propose a novel method to fit the SCA models on observed covariance matrices. We discuss how to compute various measures of explained variance for the SCA models, and how an SCA-T3 solution may be rotated to obtain a convenient interpretation. In section 4 the SCA-T3 model and algorithms are evaluated in a Monte Carlo simulation study. In section 5 we fit SCA-T3 to several datasets in the literature and compare the solutions to those of existing SCA models. Finally, in section 6 we present a discussion of our findings.

We use the following notation: $\mathcal{Y}$, $\mathbf{Y}$, $\mathbf{y}$, y are used for a three-way array, a matrix, a column vector, and a scalar, respectively. All arrays, matrices, vectors, and scalars are real-valued. Matrix transpose and inverse are denoted as $\mathbf{Y}^T$ and $\mathbf{Y}^{-1}$, respectively. An zero matrix of size $p \times q$ is denoted by $\mathbf{O}_{p,q}$. An zero column vector is denoted by $\mathbf{0}$. A $p \times p$ matrix $\mathbf{Y}$ is called orthonormal if $\mathbf{Y}^T\mathbf{Y} = \mathbf{Y}\mathbf{Y}^T = \mathbf{I}_p$. A $p \times q$ matrix has orthogonal columns if $\mathbf{Y}^T\mathbf{Y}$ is diagonal. The diagonal matrix containing entries of vector $\mathbf{y}$ as its diagonal is denoted as $\text{diag}(\mathbf{y})$. The matrix Frobenius norm $\|\mathbf{Y}\|$ is defined as $\text{trace}(\mathbf{Y}^T\mathbf{Y})^{1/2}$.

2 Parafac, Tucker3, and the SCA models

In section 2.1 we briefly discuss the Parafac and Tucker3 models, and related concepts concerning three-way arrays. In section 2.2 we give formal definitions of the four SCA models considered in Timmerman and Kiers (2003), and of our new SCA-T3 model. We show that SCA-T3 contains the four existing SCA models as special cases.

2.1 Parafac and Tucker3 and related concepts

Let a sample of N observed scores on J variables for K conditions or occassions be collected in the $N \times J \times K$ three-way array $\mathcal{X}$ with $N \times J$ slices $\mathbf{X}_k$ for $k = 1, \ldots, K$. For component analysis of $\mathcal{X}$ commonly two models are used which both are related to principal component analysis (PCA) or the singular value decomposition (SVD) for matrices. The first model we discuss is known as Parafac or Candecomp (Harshman, 1970; Carroll and Chang, 1970) and models $\mathbf{X}_k$ (after centering and possibly normalizing $\mathcal{X}$) as $\mathbf{A}\,\mathbf{C}_k\,\mathbf{B}^T$, where $\mathbf{A}$ is an $N \times R$ matrix of component scores, $\mathbf{C}_k$ is an $R \times R$ diagonal matrix, and $\mathbf{B}$ is a $J \times R$ loading matrix. Hence, centered observations $\mathbf{X}_k$ are explained by scores on R common components

and a common loading matrix, where the strength or variance of the components may be different for different k. Parafac is fitted by minimizing $\sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{A}\,\mathbf{C}_k\,\mathbf{B}^T\|^2$, which shows its link with PCA. Indeed, for $K = 1$ the PCA solution is obtained.

Let $K \times R$ matrix $\mathbf{C}$ contain the diagonal of $\mathbf{C}_k$ as its kth row. Parafac can be written as

$$\mathcal{X} \approx \sum_{r=1}^{R} (\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r)\,, \tag{1}$$

with $\mathbf{a}_r$, $\mathbf{b}_r$, and $\mathbf{c}_r$ denoting the rth columns of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, respectively, and $\circ$ denoting the outer vector product. That is, the $N \times J \times K$ outer product array $(\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r)$ has entry (i, j, k) equal to $a_{ir} b_{jr} c_{kr}$. The three-way form of Parafac (1) implies that $\mathcal{X}$ is approximated in least squares sense by a sum of R three-way arrays that have rank 1 (i.e., outer vector product form). The rank of a matrix $\mathbf{X}$ is defined as the smallest number of rank-1 matrices whose sum equals $\mathbf{X}$. Analogously, the rank of a three-way array $\mathcal{X}$ is defined as the smallest number of rank-1 arrays whose sum equals $\mathcal{X}$. In this sense, fitting Parafac to $\mathcal{X}$ implies computing a best rank-R approximation of $\mathcal{X}$. For matrices, the solution is given by the truncated SVD (Eckart and Young, 1936), while for three-way arrays we need to fit Parafac. Various algorithms to fit Parafac are available (e.g., Tomasi and Bro, 2006), of which the ALS algorithm is most often used. In Parafac ALS each of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ is updated while keeping the other two fixed. This boils down to OLS regression for each update. For example, the update of $\mathbf{B}$ is obtained from OLS regression in

$$\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{bmatrix} \approx \begin{bmatrix} \mathbf{A}\,\mathbf{C}_1 \\ \vdots \\ \mathbf{A}\,\mathbf{C}_K \end{bmatrix} \mathbf{B}^T = (\mathbf{C} \odot \mathbf{A})\,\mathbf{B}^T\,, \tag{2}$$

where $(\mathbf{C} \odot \mathbf{A}) = [\mathbf{c}_1 \otimes \mathbf{a}_1 \ \ldots \ \mathbf{c}_R \otimes \mathbf{a}_R]$ denotes the column-wise Khatri-Rao product, with $\otimes$ denoting the Kronecker product.

Despite its analogous relation to PCA, there are two fundamental differences between Parafac and PCA. First, unlike PCA a Parafac solution $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is rotationally unique under mild conditions (e.g., Domanov and De Lathauwer, 2013). And second, unlike PCA a best fitting Parafac model may not exist. That is, a three-way array $\mathcal{X}$ may not have a best rank-R approximation. In such cases several rank-1 terms become nearly linearly dependent and large in magnitude while fitting Parafac (Stegeman, 2006; De Silva and

Lim, 2008; Stegeman, 2014). Needless to say, this phenomenon should be avoided when an interpretable solution is needed.

The second model for component analysis of a three-way array is Tucker3 (Tucker, 1966), which is written analogous to (1) as

$$\mathcal{X} \approx \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} \left(\mathbf{a}_p \circ \mathbf{b}_q \circ \mathbf{c}_r \right). \tag{3}$$

We write a Tucker3 solution as $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathcal{G})$, where $\mathbf{A}$ has size $N \times P$, $\mathbf{B}$ has size $J \times Q$, $\mathbf{C}$ has size $K \times R$, and $\mathcal{G}$ is the $P \times Q \times R$ array (also known as core array) with entries g_{pqr}. Hence, $\mathcal{X}$ is approximated by rank-1 terms formed from all combinations of the columns of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, and each rank-1 term has a weight g_{pqr}. The Tucker3 model for slice $\mathbf{X}_k$ is written as $\mathbf{A} \left(\sum_{r=1}^{R} c_{kr} \, \mathbf{G}_r \right) \mathbf{B}^T$, where $\mathbf{G}_r$ is the rth $P \times Q$ slice of core array $\mathcal{G}$. By comparing (1) and (3) it can be seen that Parafac is a special case of Tucker3 with $P = Q = R$ and $g_{pqr} = 0$ unless $(p, q, r) = (r, r, r)$. Tucker3 is fitted by minimizing $\sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{A} \left(\sum_{r=1}^{R} c_{kr} \, \mathbf{G}_r \right) \mathbf{B}^T\|^2$ and an ALS algorithm has been derived in Kroonenberg and De Leeuw (1980). Other algorithms can be found in De Lathauwer, De Moor and Vandewalle (2000b), Savas and Lim (2010), and Ishteva, Absil, Van Huffel and De Lathauwer (2011).

More insight in the Tucker3 model can be obtained by writing it as $\mathcal{X} \approx (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathcal{G}$, where the latter array has (i, j, k) entry equal to $\sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr}$. Analogous to matrix multiplication, we have $(\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathcal{G} = (\mathbf{AS}, \mathbf{BT}, \mathbf{CU}) \cdot ((\mathbf{S}^{-1}, \mathbf{T}^{-1}, \mathbf{U}^{-1}) \cdot \mathcal{G})$ for nonsingular matrices $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{U}$. This shows that a Tucker3 solution is not rotationally unique. In fact, we may assume without loss of generality that $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are columnwise orthonormal. The rotational nonuniqueness may be exploited by trying to find orthogonal or oblique rotations that rotate the core $\mathcal{G}$ (and possible also $\mathbf{B}$ and/or $\mathbf{C}$) to simple structure (i.e., with only few large nonzero entries). This facilitates interpretation of the Tucker3 solution, analogous to rotating principal components. Rotation methods for this purpose have been developed by Kiers (1998ab) and will be discussed in relation to SCA-T3 in section 3.

For later use, we mention that Tucker3 can be written as $[\mathbf{X}_1 \ \ldots \ \mathbf{X}_K] \approx \mathbf{A} \, [\mathbf{G}_1 \ \ldots \ \mathbf{G}_R] \, (\mathbf{C}^T \otimes \mathbf{B}^T)$. By vectorizing the matrices (i.e., stacking the columns on top of each other) on both sides we obtain

$$\mathrm{Vec}([\mathbf{X}_1 \ \ldots \ \mathbf{X}_K]) \approx (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A}) \, \mathrm{Vec}([\mathbf{G}_1 \ \ldots \ \mathbf{G}_R]), \tag{4}$$

which can be used to obtain the update of $\mathcal{G}$ for fixed $\mathbf{A}, \mathbf{B}, \mathbf{C}$ in the Tucker3 ALS algorithm. When $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are columnwise orthonormal, the OLS regression in (4) features orthonormal predictors, and the explained variance due to each predictor adds up to the total explained variance. This will be elaborated in relation to SCA-T3 in section 3. Note that an optimal solution for Tucker3 is guaranteed to exist, since it is found by varying columnwise orthonormal $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ which implies a compact feasible set (see also De Lathauwer, De Moor and Vandewalle, 2000b). The updated core $\mathcal{G}$ can be written in terms of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, as implied by (4).

Finally, we discuss the link between Tucker3 and the matrix SVD. Let a mode-i fiber of $\mathcal{X}$ be defined as a vector obtained from $\mathcal{X}$ by varying index i and keeping the other two indices fixed. That is, a mode-1 fiber is a column of some $\mathbf{X}_k$, a mode-2 fiber is a row of some $\mathbf{X}_k$, and a mode-3 fiber is a vector $(x_{ij1}, \ldots, x_{ijK})^T$ for fixed row i and column j indices. When Tucker3 fits perfectly, the linear spaces of mode-1, mode-2, and mode-3 fibers are given by the column spaces of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, respectively. This is analogous to the matrix SVD, which describes the row and column spaces of a matrix. A crucial difference with matrices is, however, that a three-way array may have different ranks for the three fiber spaces. Moreover, these may be different from the rank of the three-way array as defined above. In contrast, a matrix has row rank equal to column rank equal to its rank. More details on the link between Tucker3 and the SVD can be found in De Lathauwer, De Moor and Vandewalle (2000a).

Numerous applications of Parafac, Tucker3, and related models exist. The reader is referred to the following books and overview papers: Smilde, Bro and Geladi (2004), Kroonenberg (2008), Kolda and Bader (2009), Acar and Yener (2009), Comon and De Lathauwer (2010), De Lathauwer (2010).

2.2 Formal definition of the SCA models

Let the observed scores on J variables for sample k be collected in the $N_k \times J$ matrix $\mathbf{X}_k$, $k = 1, \ldots, K$. We assume that the columns of $\mathbf{X}_k$ are centered, so that $\mathrm{Cov}(\mathbf{X}_k) = N_k^{-1}\mathbf{X}_k^T\mathbf{X}_k$. Table 1 contains the formal definitions of the SCA models for both the observed scores $\mathbf{X}_k$ and the observed covariance matrices.

[Insert Table 1 here.]

First, we discuss the four existing SCA models. Each model is of the form $\mathbf{X}_k \approx \mathbf{F}_k \mathbf{B}^T$, where $N_k \times R$ matrix $\mathbf{F}_k$ contains component scores for sample k and $J \times R$ matrix $\mathbf{B}$ is the common loading matrix. The SCA models are fitted by minimizing $\sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{F}_k \mathbf{B}^T\|^2$. The component scores $\mathbf{F}_k$ are linear combinations of the observed scores $\mathbf{X}_k$ and, hence, we have $\mathrm{Cov}(\mathbf{F}_k) = N_k^{-1}\mathbf{F}_k^T\mathbf{F}_k$. As model for $\mathrm{Cov}(\mathbf{X}_k)$ we obtain $\mathbf{B}\,\mathrm{Cov}(\mathbf{F}_k)\,\mathbf{B}^T$. The SCA models differ in their assumptions for $\mathbf{F}_k$.

In SCA-P these are all assumptions and the component covariance matrices $\mathrm{Cov}(\mathbf{F}_k) = \mathbf{\Phi}_k$ are unconstrained for each sample k. In SCA-PF2 we have $\mathbf{F}_k = \mathbf{A}_k \mathbf{C}_k$, with $N_k^{-1}\mathbf{A}_k^T\mathbf{A}_k = \mathbf{\Phi}$ not depending on k and $\mathbf{C}_k$ an $R \times R$ diagonal matrix. Without loss of generality we may assume that the columns of $\mathbf{A}_k$ have sum-of-squares equal to N_k (i.e., they have variance one). Hence, in SCA-PF2 $\mathrm{Cov}(\mathbf{F}_k) = \mathbf{C}_k \mathbf{\Phi} \mathbf{C}_k$ and the component correlations in $\mathbf{\Phi}$ are equal for all k, but their variances in $\mathbf{C}_k^2$ may differ for each k. In SCA-IND we have the same model as SCA-PF2 with the additional restriction that the components are uncorrelated in each sample: $\mathbf{\Phi} = \mathbf{I}_R$. SCA-ECP is the same model as SCA-PF2 with the additional restriction that $\mathbf{C}_k = \mathbf{I}_R$. Hence, in SCA-ECP the component variances and correlations are equal for all samples. In SCA-ECP we may impose $\mathbf{\Phi} = \mathbf{I}_R$ without loss of generality and, hence, the model is a constrained version of SCA-IND (Timmerman and Kiers, 2003).

The SCA-PF2 model for $\mathbf{X}_k$ is equal to Parafac2 (Harshman, 1972) and equal to Parafac when $\mathbf{A}_k = \mathbf{A}$; see section 2.1. Solutions of SCA-PF2 and SCA-IND are usually rotationally unique for $K \geq 4$, with permutation and scaling being the only ambiguities in the solution (Ten Berge and Kiers, 1996; Kiers, Ten Berge and Bro, 1999). Solutions of SCA-P and SCA-ECP can be rotated to obtain an interpretable loading matrix $\mathbf{B}$, analogous to rotating principal components. It is known that Parafac2 is not unique for $K = 2$ (Ten Berge and Kiers, 1996), which makes application of SCA-PF2 and SCA-IND problematic for two samples. Indeed, this type of nonuniqueness is not fixed by applying a rotation.

Next, we introduce the new SCA-T3 model. SCA-T3 is related to the Tucker3 model (section 2.1) in the same way Parafac2 is related to Parafac. Namely, SCA-T3 is obtained from Tucker3 by replacing $\mathbf{A}$ by $\mathbf{A}_k$ of size $N_k \times P$ and imposing the constraint $N_k^{-1}\mathbf{A}_k^T\mathbf{A}_k = \mathbf{\Phi}$. In SCA-T3 the component covariance matrix $\mathrm{Cov}(\mathbf{F}_k)$ has Tucker3 structure. The four existing SCA models are special cases of SCA-T3. For SCA-PF2 this follows from the fact that Parafac is a special case of Tucker3 (section 2.1). Indeed, when setting $P = Q = R$ and

$(\mathbf{G}_r)_{pq} = 0$ for $(p,q) \neq (r,r)$, $r = 1, \ldots, R$, in SCA-T3 we obtain SCA-PF2. Since SCA-IND and SCA-ECP are special cases of SCA-PF2, they are also special cases of SCA-T3. Also SCA-P is a special case of SCA-T3. Let $P = Q$ and $R = K$ and $\mathbf{C} = \mathbf{I}_K$. Then the SCA-T3 model for $\mathbf{X}_k$ reduces to $\mathbf{A}_k\mathbf{G}_k\mathbf{B}^T$, with $\mathbf{G}_k$ of size $P \times P$. Clearly, we can absorb $\mathbf{G}_k$ into $\mathbf{A}_k$ and obtain the SCA-P model. We conclude that SCA-T3 contains the four existing SCA models in Table 1 as special cases.

Note that in practical applications the fact that, formally, SCA-T3 includes SCA-P as a special case is of little significance. Indeed, one would be interested in comparing the fit of different SCA models with the same number of components in the variables mode (R for the existing SCA models, Q for SCA-T3) and setting $R = K$ in SCA-T3 is usually not of practical interest. In fact, formally, SCA-T3 with $P = Q = R < K$ is a special case of SCA-P.

The main differences between SCA-T3 and the four other SCA models are the following. In SCA-T3 we are allowed to choose different numbers of components in the three modes of the data. Indeed, matrices $\mathbf{A}_k$ contain component scores of P components, while loading matrix $\mathbf{B}$ contains loadings of J variables on Q components, and $\mathbf{C}$ contains weights for K samples on R components. The components in the three modes are linked via the entries of the core array $\mathcal{G}$. Analogous to Tucker3 the core array can be rotated without loss of fit to obtain a solution in which only few entries of $\mathcal{G}$ are large, which will be elaborated in section 3. This makes interpretation of the solution much easier, a feature shared with SCA-P and SCA-ECP. However, SCA-T3 offers more insight in differences between samples than SCA-P and SCA-ECP. Indeed, SCA-ECP yields the same covariance model for all samples, while in SCA-P correlated components may complicate interpretation (Timmerman and Kiers, 2003). Interpretation of an SCA-T3 solution will be illustrated in applications in section 5. A potential problem of SCA-PF2 is that so-called diverging solutions may occur while trying to fit the model to multi-set data. In that case no optimal solution exists and some columns of $\mathbf{A}_k$, $\mathbf{B}$, and $\mathbf{C}$ become nearly linearly dependent and their norm increases without bound. Such cases have been observed in simulation studies (Stegeman and Lam, 2016) and also occur for Parafac (section 2.1). In Tucker3 and SCA-T3 there always exists an optimal solution.

Note that from the Parafac and Parafac2 names one may expect SCA-T3 to be named Tucker2 instead of Tucker3. However, Tucker2 is an existing special case of Tucker3 in

which the number of components equals the size of the three-way array in one mode and the component matrix for that mode is set to identity (e.g., $R = K$ and $\mathbf{C} = \mathbf{I}_K$). Therefore, we have chosen to name our model SCA-T3 instead.

3 Fitting the SCA models

Algorithms to fit the four existing SCA models in Table 1 to observed (and centered) $\mathbf{X}_k$ are the following. SCA-P is solved via PCA on $[\mathbf{X}_1^T \ \ldots \ \mathbf{X}_K^T]^T$ (Kiers and Ten Berge, 1994). For SCA-PF2 an ALS algorithm is derived in Kiers et al. (1999). As shown in Timmerman and Kiers (2003) the ALS algorithm for SCA-PF2 can be adapted easily to SCA-IND and SCA-ECP. In section 3.1 we derive an ALS algorithm for SCA-T3 analogous to the algorithm for SCA-PF2. In section 3.2 we propose a new method to fit the SCA models to observed $\mathrm{Cov}(\mathbf{X}_k)$, $k = 1, \ldots, K$. In section 3.3 we discuss rotation methods for SCA-T3. In section 3.4 we discuss how to compute various measures of explained variance in SCA-T3 and the exisiting SCA models. Finally, in section 3.5 we discuss how a suitable SCA model may be selected for a particular dataset.

3.1 ALS algorithm for SCA-T3

To fit SCA-T3 we minimize $\sum_k \|\mathbf{X}_k - \mathbf{A}_k \left(\sum_{r=1}^R c_{kr}\, \mathbf{G}_r\right) \mathbf{B}^T\|^2$ over $\mathbf{A}_k$, $\mathbf{B}$, $\mathbf{C}$, and the core $\mathcal{G}$. Recall that $N_k^{-1}\mathbf{A}_k^T\mathbf{A}_k = \mathbf{\Phi}$ does not depend on k. We assume that $\mathrm{rank}(\mathbf{A}_k) = P$ for $k = 1, \ldots, K$. Then we can set $N_k^{-1}\mathbf{A}_k^T\mathbf{A}_k = \mathbf{I}_P$ without loss of generality. Indeed, for $\mathbf{\Phi} = \mathbf{V}^T\mathbf{V}$ we can replace $\mathbf{A}_k$ by $\overline{\mathbf{A}}_k = \mathbf{A}_k\mathbf{V}^{-1}$ and each $\mathbf{G}_r$ by $\overline{\mathbf{G}}_r = \mathbf{V}\,\mathbf{G}_r$ to obtain a solution with $N_k^{-1}\overline{\mathbf{A}}_k^T\overline{\mathbf{A}}_k = \mathbf{I}_P$. We may also set $\mathbf{B}^T\mathbf{B} = \mathbf{I}_Q$ without loss of generality, as will be explained below. Note that when $\mathrm{rank}(\mathbf{B}) < Q$ or $\mathrm{rank}(\mathbf{C}) < R$ the SCA-T3 model can be reduced to a model with less components without loss of fit (as for Tucker3).

Analogous to the ALS algorithm for SCA-PF2 in Kiers et al. (1999) we alternatingly minimize the objective function over $\mathbf{A}_k$, $k = 1, \ldots, K$, for fixed $\mathbf{B}$, $\mathbf{C}$, $\mathcal{G}$, and over $\mathbf{B}$, $\mathbf{C}$, and $\mathcal{G}$ for fixed $\mathbf{A}_k$, $k = 1, \ldots, K$. We need to reparameterize the model as $\widetilde{\mathbf{A}}_k \left(\sum_{r=1}^R \tilde{c}_{kr}\, \mathbf{G}_r\right) \mathbf{B}^T$, $k = 1, \ldots, K$, with $\widetilde{\mathbf{A}}_k^T\widetilde{\mathbf{A}}_k = \mathbf{I}_P$. After convergence we obtain $\mathbf{A}_k$ and $\mathbf{C}$ from $\widetilde{\mathbf{A}}_k$ and $\widetilde{\mathbf{C}}$, respectively, by rescaling. The two main steps of the algorithm are as follows.

Step 1. Minimizing the objective function over $\widetilde{\mathbf{A}}_k$ subject to $\widetilde{\mathbf{A}}_k^T\widetilde{\mathbf{A}}_k = \mathbf{I}_P$ is equivalent

to minimizing $\|\mathbf{X}_k\,\mathbf{B}\left(\sum_{r=1}^{R}\tilde{c}_{kr}\,\mathbf{G}_r^T\right)-\widetilde{\mathbf{A}}_k\|^2$ over $\widetilde{\mathbf{A}}_k$, since the only term depending on $\widetilde{\mathbf{A}}_k$ is the same in the expansion of both objective functions. The solution is found by computing the singular value decomposition $\mathbf{P}_k\mathbf{\Delta}_k\mathbf{Q}_k^T$ of $\mathbf{X}_k\,\mathbf{B}\left(\sum_{r=1}^{R}\tilde{c}_{kr}\,\mathbf{G}_r^T\right)$ ($N_k \times P$), and $\widetilde{\mathbf{A}}_k = \mathbf{P}_k\mathbf{Q}_k^T$ is the optimal solution, $k = 1, \ldots, K$ (Golub and Van Loan, 1996, section 12.4).

Step 2. The problem of minimizing the objective function over $\mathbf{B}$, $\widetilde{\mathbf{C}}$ and $\mathcal{G}$ reduces to minimizing $\sum_k \|\widetilde{\mathbf{A}}_k^T\mathbf{X}_k - \left(\sum_{r=1}^{R}\tilde{c}_{kr}\,\mathbf{G}_r\right)\mathbf{B}^T\|^2$, since the terms depending on $\mathbf{B}$, $\widetilde{\mathbf{C}}$ and $\mathcal{G}$ are identical in the expansion of both objective functions when $\widetilde{\mathbf{A}}_k^T\widetilde{\mathbf{A}}_k = \mathbf{I}_P$. The latter objective is equivalent to fitting the Tucker2 model with (P, Q, R) components to the $P \times J \times K$ array with $P \times J$ slices $\widetilde{\mathbf{A}}_k^T\mathbf{X}_k$, $k = 1, \ldots, K$. We apply one iteration of the Tucker2 ALS algorithm (Kroonenberg and De Leeuw, 1980) to update each of $\mathbf{B}$, $\widetilde{\mathbf{C}}$ and $\mathcal{G}$. Analogous to Tucker3 we take $\mathbf{B}$ and $\widetilde{\mathbf{C}}$ columnwise orthonormal without loss of generality (section 2.1).

The ALS procedure above guarantees monotonic convergence of the SCA-T3 objective function. We stop the ALS algorithm when the relative decrease of the objective function drops below 10^{-7}. After convergence we set $\mathbf{A}_k = N_k^{1/2}\widetilde{\mathbf{A}}_k$ and $c_{kr} = N_k^{-1/2}\tilde{c}_{kr}$ to obtain a valid SCA-T3 solution. Note that $\mathbf{C}$ will not be columnwise orthogonal unless the N_k are equal for all k.

We can choose between using random or rational starting values. As random starting values we obtain $\widetilde{\mathbf{A}}_k$, $\mathbf{B}$, and $\widetilde{\mathbf{C}}$ as orthonormal bases of the column spaces of matrices with entries randomly sampled from the standard normal distribution. The starting core $\mathcal{G}$ is then found via the update in the Tucker2 ALS algorithm. As rational starting values we obtain $\widetilde{\mathbf{A}}_k$ via PCA on $\mathbf{X}_k$. After the $P \times J \times K$ array with slices $\widetilde{\mathbf{A}}_k^T\mathbf{X}_k$ is constructed, starting $\mathbf{B}$ and $\widetilde{\mathbf{C}}$ are likewise obtained via PCA on corresponding matrix unfoldings of the array. Finally, the starting core $\mathcal{G}$ is again computed via the update in the Tucker2 ALS algorithm. Since the objective function is not convex the ALS algorithm may terminate in a local minimum. A common remedy is to run the algorithm multiple times with different starting values. In the simulation study in section 4 we determine the number of runs needed for SCA-T3 models of various sizes.

3.2 Fitting SCA models to observed covariance matrices

When only observed covariance matrices are available, can the SCA models in Table 1 be fitted in their covariance form? Below we propose a new heuristic method for this purpose. First we discuss SCA-P. Let $\mathbf{X}_{\text{all}} = [\mathbf{X}_1^T \ \ldots \ \mathbf{X}_K^T]^T$ be rescaled such that its columns have equal sum-of-squares $\sum_k N_k$ (as suggested by Timmerman and Kiers, 2003). We assume the same scaling for $\mathbf{F}_{\text{all}} = [\mathbf{F}_1^T \ \ldots \ \mathbf{F}_K^T]^T$. Recall that SCA-P minimizes $\sum_k \|\mathbf{X}_k - \mathbf{F}_k \mathbf{B}^T\|^2$ without any constraints on $N_k^{-1}\mathbf{F}_k^T\mathbf{F}_k = \mathbf{\Phi}_k$. Let $\mathbf{U}\,\mathbf{S}\,\mathbf{V}^T$ be the SVD of $(\sum_k N_k)^{-1/2}\,\mathbf{X}_{\text{all}}$, with $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_J$ and $\mathbf{S} = \text{diag}(s_1, s_2, \ldots, s_J)$ containing the singular values in decreasing order. Let $\mathbf{U}_R = [\mathbf{U}_{1,R}^T \ \ldots \ \mathbf{U}_{K,R}^T]^T$ consist of the first R columns of $\mathbf{U}$, with block $\mathbf{U}_{k,R}$ having N_k rows. Let $\mathbf{V}_R$ consist of the first R columns of $\mathbf{V}$, and let $\mathbf{S}_R = \text{diag}(s_1, \ldots, s_R)$. The unrotated SCA-P solution is given by $\mathbf{F}_k = (\sum_k N_k)^{1/2}\,\mathbf{U}_{k,R}$ and $\mathbf{B} = \mathbf{V}_R\,\mathbf{S}_R$ (Kiers and Ten Berge, 1994). This implies that

$$\left(\sum_{k=1}^{K} \frac{N_k}{\sum_k N_k}\,\text{Cov}(\mathbf{X}_k)\right) = (\sum_k N_k)^{-1}\,\mathbf{X}_{\text{all}}^T\mathbf{X}_{\text{all}} = \mathbf{V}\,\mathbf{S}^2\,\mathbf{V}^T\,, \tag{5}$$

has best rank-R approximation $\mathbf{V}_R\,\mathbf{S}_R^2\,\mathbf{V}_R^T = \mathbf{B}\,\mathbf{B}^T$ (Eckart and Young, 1936). Hence, when only observed covariance matrices are available the optimal unrotated SCA-P loading matrix $\mathbf{B}$ can be obtained from a PCA on the (rescaled) weighted sum of covariance matrices. Given this $\mathbf{B}$, the OLS estimates of the component covariance matrices can be obtained via Penrose regression as $\mathbf{\Phi}_k = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\text{Cov}(\mathbf{X}_k)\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}$ (Penrose, 1956). Note that the weighted sum in (5) turns into a regular sum when SCA-P is fitted to $\widetilde{\mathbf{X}}_k = N_k^{-1/2}\mathbf{X}_k$ and the normalization by $(\sum_k N_k)$ is ignored in the above. For the other SCA models PCA is not used to obtain a solution and it is not clear how the minimizer $\mathbf{B}$ of $\sum_k \|\mathbf{X}_k - \mathbf{F}_k\,\mathbf{B}^T\|^2$ should be obtained when only $\text{Cov}(\mathbf{X}_k)$ are available.

Next, we consider minimizing $\sum_k \|\text{Cov}(\mathbf{X}_k) - \mathbf{B}\,\text{Cov}(\mathbf{F}_k)\,\mathbf{B}^T\|^2$, where the model for $\text{Cov}(\mathbf{F}_k)$ depends on the SCA model; see Table 1. Kiers (1993) has proposed an algorithm to fit SCA-PF2 to observed covariance matrices, but it is rather complicated and converges very slowly (as stated in Kiers et al., 1999). Also, the algorithm of Kiers (1993) cannot be adapted to SCA-IND. We propose a heuristic method to fit the SCA models to observed covariance matrices, by using the ALS algorithms of the SCA models for observed scores. Our approach consists of two steps. First, the eigenvalue decompositions of the observed covariance matrices $\text{Cov}(\mathbf{X}_k) = \mathbf{V}_k\mathbf{D}_k\mathbf{V}_k^T$ are computed. Second, we compute $\mathbf{Y}_k = \mathbf{D}_k^{1/2}\mathbf{V}_k^T$

so that $\text{Cov}(\mathbf{X}_k) = \mathbf{Y}_k^T\mathbf{Y}_k$ and fit the SCA model to $\mathbf{Y}_k$ $(J \times J)$, $k = 1, \ldots, K$. For example, for SCA-PF2 we fit $\mathbf{Y}_k \approx \mathbf{A}_k\mathbf{C}_k\mathbf{B}^T$ with $\mathbf{A}_k^T\mathbf{A}_k = \mathbf{\Phi}$. Hence, we do not use the normalization N_k^{-1} here. The ALS algorithm for SCA-PF2 is easily adapted to this purpose. It follows that $\text{Cov}(\mathbf{X}_k) = \mathbf{Y}_k^T\mathbf{Y}_k \approx \mathbf{B}\,\mathbf{C}_k\,\mathbf{\Phi}\,\mathbf{C}_k\,\mathbf{B}^T$, with equality for perfect fit. However, this procedure does not necessarily minimize $\sum_k \|\text{Cov}(\mathbf{X}_k) - \mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T\|^2$. Yet it provides an elegant and relatively simple and fast algorithm to obtain a very good approximation nonetheless. In Lam (2015) a simulation study with noisefree data is conducted to compare the approach above to Kiers (1993) for SCA-PF2, and fitting the model as above yields more accurate results. Variants of the approach above to fit SCA models to observed covariance matrices are used in Stegeman and Lam (2014) for three-mode factor analysis and in Stegeman and Lam (2016) for multi-set factor analysis. Apart from Kiers (1993) no special purpose algorithms are available to fit the SCA models to observed covariance matrices. In the simulation study in section 4 it will be shown that fitting SCA-T3 as above yields accurate results for noisy multi-set data. Moreover, for noisefree data following a unique SCA-T3 model (with imposed zeros in $\mathcal{G}$; see section 3.3) the correct solution is retrieved.

3.3 Obtaining an interpretable solution for SCA-T3

In step 2 of the ALS algorithm for SCA-T3 (section 3.1) a Tucker2 solution $(\mathbf{I}_P, \mathbf{B}, \widetilde{\mathbf{C}}) \cdot \mathcal{G}$ is obtained. After convergence this Tucker2 solution may be rotated analogous to Tucker3 solutions to obtain an interpretable solution (section 2.1). In particular, orthonormal matrices $\mathbf{S}$ $(P \times P)$, $\mathbf{T}$ $(Q \times Q)$, and $\mathbf{U}$ $(R \times R)$ satisfy $(\mathbf{S}, \mathbf{BT}, \widetilde{\mathbf{C}}\mathbf{U}) \cdot ((\mathbf{S}^T, \mathbf{T}^T, \mathbf{U}^T) \cdot \mathcal{G})$. After rotating we set $\mathbf{A}_k = N_k^{1/2}\widetilde{\mathbf{A}}_k\mathbf{S}$ so that $N_k^{-1}\mathbf{A}_k^T\mathbf{A}_k = \mathbf{I}_P$ still holds. Also, $(\mathbf{BT})^T(\mathbf{BT}) = \mathbf{I}_Q$, and $(\widetilde{\mathbf{C}}\mathbf{U})^T(\widetilde{\mathbf{C}}\mathbf{U}) = \mathbf{I}_R$. The rotational freedom may be used to find rotation matrices such that the rotated core $(\mathbf{S}^T, \mathbf{T}^T, \mathbf{U}^T) \cdot \mathcal{G}$ and the rotated loading matrix $\mathbf{BT}$ (or perhaps also the rotated $\widetilde{\mathbf{C}}\mathbf{U}$) have simple structure, analogous to rotation methods for PCA. For this purpose, the joint orthomax rotation method for Tucker3 of Kiers (1998b) may be used, in which the weighted objective of a simple core and simple component matrices in the three modes is obtained in the orthomax sense. A component matrix may also be given weight zero, indicating that simple structure is not an objective for this matrix. For SCA-T3 the weight of the first component matrix is zero, since this is the orthogonal rotation of the components $\mathbf{A}_k$ and does not need to be simple. Application of joint orthomax involves specifying the modes to rotate and the weights for each mode. For practical applications it

may be convenient to choose the natural weights proposed in Kiers (1998b) that take into account the sizes of the core and the component matrices.

A different rotation method is Simplimax (Kiers, 1998a) that applies oblique rotations with the aim of obtaining a simple core only. Simplimax can be used in two ways: either the required number m of nonzero core entries is specified in advance or the pattern of nonzero core entries is specified. The Simplimax algorithm then finds oblique rotation matrices $\mathbf{S}, \mathbf{T}, \mathbf{U}$ such that the rotated core has smallest sum-of-squares of the smallest $PQR - m$ entries, or of the entries required to be zero. Hence, this rotation method only yields a simple core and may result in oblique rotated components $\widetilde{\mathbf{A}}_k\mathbf{S}$.

A third method to obtain an interpretable SCA-T3 solution is to impose zero restrictions in the core $\mathcal{G}$ while fitting the SCA-T3 model. Such Tucker3 models arise naturally in several applications in chemometrics (Smilde et al., 2004). ALS algorithms to fit such constrained Tucker3 models can be found in Rocci (1992) and Kiers and Smilde (1998). However, when imposing zero restrictions in the core one should make sure the model is unique and non-trivial (Ten Berge and Smilde, 2002). Indeed, if the model is not unique then alternative component matrices and core exists resulting in the same fitted model array, thus hampering unambiguous interpretation of the model. Nontriviality means that the pattern of zeros in the core should not be possible to obtain by rotating a general core of the same size. In that case the model would not truly be a model but rather an artifact. See Ten Berge and Smilde (2002) for a discussion.

In the applications in section 5 we only use the joint orthomax rotation method of Kiers (1998b), and use natural weights to balance between a simple loading matrix $\mathbf{B}$, a simple core array $\mathcal{G}$, and a simple weight matrix $\widetilde{\mathbf{C}}$ by rotating orthogonally in all three modes.

After orthogonal rotations have been applied in SCA-T3, we rescale the solution such that it matches the scaling applied in Timmerman and Kiers (2003) for SCA-PF2. The SCA-PF2 model can be written in matrix form as

$$\begin{bmatrix} \mathbf{A}_1\,\mathbf{C}_1 \\ \vdots \\ \mathbf{A}_K\,\mathbf{C}_K \end{bmatrix} \mathbf{B}^T\,. \tag{6}$$

In Timmerman and Kiers (2003) the matrix on the left-hand side has column sum-of-squares equal to $\sum_k N_k$, which defines the scaling of the columns of $\mathbf{C}$ when each $\mathbf{A}_k$ has column sum-of-squares equal to N_k. The scaling coefficients are then absorbed in $\mathbf{B}$. Analogously,

the SCA-T3 model can be written as

$$\begin{bmatrix} \mathbf{A}_1 \left(\sum_{r=1}^{R} c_{1r}\, \mathbf{G}_r \right) \\ \vdots \\ \mathbf{A}_K \left(\sum_{r=1}^{R} c_{Kr}\, \mathbf{G}_r \right) \end{bmatrix} \mathbf{B}^T \,. \tag{7}$$

We keep $N_k^{-1}\mathbf{A}_k^T\mathbf{A}_k = \mathbf{I}_P$ and require that the matrix on the left-hand side of (7) has column sum-of-squares equal to $\sum_k N_k$. This implies the scaling of the columns of $\mathbf{B}$. Additionally, we normalize the largest absolute entry in each $\mathbf{G}_r$ to one and apply the reverse normalization to column r of $\mathbf{C}$. This ensures that the entries in $\mathbf{B}$ and $\mathbf{C}$ are comparable in size to those in SCA-PF2 when the scaling of Timmerman and Kiers (2003) is used.

3.4 Computing explained variance measures for the SCA models

The SCA models are fitted to (centered) observed scores $\mathbf{X}_k$ by minimizing $\|\mathbf{X}_{\text{all}} - \mathbf{F}_{\text{all}}\,\mathbf{B}^T\|^2$, with $\mathbf{X}_{\text{all}} = [\mathbf{X}_1^T \ \ldots \ \mathbf{X}_K^T]^T$, $\mathbf{F}_{\text{all}} = [\mathbf{F}_1^T \ \ldots \ \mathbf{F}_K^T]^T$, and the model for $\mathbf{F}_k$ depending on the SCA model (Table 1). The update for $\mathbf{B}$ is obtained via OLS regression in all SCA algorithms. Since regression and residual are orthogonal, it follows that the overall fit percentage can be expressed as

$$100 - 100 \cdot \frac{\|\mathbf{X}_{\text{all}} - \mathbf{F}_{\text{all}}\,\mathbf{B}^T\|^2}{\|\mathbf{X}_{\text{all}}\|^2} = 100 \cdot \frac{\|\mathbf{F}_{\text{all}}\,\mathbf{B}^T\|^2}{\|\mathbf{X}_{\text{all}}\|^2} \,. \tag{8}$$

Although the fit is not maximized for each sample k separately, we can still compute the fit percentage per sample to identify samples for which the model does not fit well. The following lemma states that equality (8) holds analogously for each sample when SCA-P, SCA-PF2, SCA-IND, or SCA-T3 are fitted. Its proof can be found in the appendix.

Lemma 3.1 *Let $\mathbf{X}_k$ be centered observed scores, $k = 1, \ldots, K$. When* SCA-P *(via* PCA*),* SCA-PF2*,* SCA-IND*, or* SCA-T3 *(via* ALS*) are fitted to $\mathbf{X}_k$, $k = 1, \ldots, K$, the fit percentage for sample k can be computed as*

$$100 - 100 \cdot \frac{\|\mathbf{X}_k - \mathbf{F}_k\,\mathbf{B}^T\|^2}{\|\mathbf{X}_k\|^2} = 100 \cdot \frac{\|\mathbf{F}_k\,\mathbf{B}^T\|^2}{\|\mathbf{X}_k\|^2} \,. \tag{9}$$

□

The rotations that are possible in SCA-P, SCA-ECP, and SCA-T3 leave the fitted model matrices invariant. That is, rotating does not change the matrices $\mathbf{F}_{\text{all}}\,\mathbf{B}^T$ and $\mathbf{F}_k\,\mathbf{B}^T$, $k = 1, \ldots, K$. Hence, these rotations do not change the fit percentages in (8) and (9).

Under SCA-IND the columns of $\mathbf{F}_{\text{all}}$ are orthogonal, since $\mathbf{F}_{\text{all}}^T\mathbf{F}_{\text{all}} = \sum_k \mathbf{F}_k^T\mathbf{F}_k = \sum_k N_k\,\mathbf{C}_k^2$. The same is true for SCA-ECP with $\mathbf{C}_k = \mathbf{I}_R$. (Recall that in SCA-ECP we can set $\mathbf{\Phi} = \mathbf{I}_R$ without loss of generality.) Hence, for these models we can define the fit percentage due to component r as

$$100 - 100 \cdot \frac{\|\mathbf{X}_{\text{all}} - \mathbf{F}_{\text{all},r}\,\mathbf{b}_r^T\|^2}{\|\mathbf{X}_{\text{all}}\|^2} = 100 \cdot \frac{\|\mathbf{F}_{\text{all},r}\,\mathbf{b}_r^T\|^2}{\|\mathbf{X}_{\text{all}}\|^2}\,, \tag{10}$$

where $\mathbf{F}_{\text{all},r}$ denotes column r of $\mathbf{F}_{\text{all}}$. Moreover, the fit percentages (10) for components $r = 1, \ldots, R$ sum up to the overall fit percentage (8).

For SCA-T3 we can compute the fit percentage due to each term (p, q, r) in the model. This is done as follows. After step 2 of the ALS algorithm for SCA-T3 (section 3.1) the fitted model for $\mathbf{X}_k$ can be written as $\widetilde{\mathbf{A}}_k\,\mathbf{G}\,(\tilde{\mathbf{c}}_k^{(\text{row})} \otimes \mathbf{B}^T)$, with $\mathbf{G} = [\mathbf{G}_1\ \ldots\ \mathbf{G}_R]$ and $\tilde{\mathbf{c}}_k^{(\text{row})}$ the kth row of $\widetilde{\mathbf{C}}$ as a column vector. Analogous to (4) for Tucker3 (section 2.1), we obtain

$$\text{Vec}(\mathbf{X}_{\text{all}}) \approx \begin{bmatrix} (\tilde{\mathbf{c}}_1^{(\text{row})})^T \otimes \mathbf{B} \otimes \widetilde{\mathbf{A}}_1 \\ \vdots \\ (\tilde{\mathbf{c}}_K^{(\text{row})})^T \otimes \mathbf{B} \otimes \widetilde{\mathbf{A}}_K \end{bmatrix} \text{Vec}(\mathbf{G})\,. \tag{11}$$

The matrix on the right-hand side is columnwise orthonormal, since

$$\begin{aligned} \sum_{k=1}^{K}((\tilde{\mathbf{c}}_k^{(\text{row})})^T \otimes \mathbf{B} \otimes \widetilde{\mathbf{A}}_k)^T((\tilde{\mathbf{c}}_k^{(\text{row})})^T \otimes \mathbf{B} \otimes \widetilde{\mathbf{A}}_k) &= \sum_{k=1}^{K}(\tilde{\mathbf{c}}_k^{(\text{row})}(\tilde{\mathbf{c}}_k^{(\text{row})})^T \otimes \mathbf{B}^T\mathbf{B} \otimes \widetilde{\mathbf{A}}_k^T\widetilde{\mathbf{A}}_k) \\ &= \left(\sum_{k=1}^{K}\tilde{\mathbf{c}}_k^{(\text{row})}(\tilde{\mathbf{c}}_k^{(\text{row})})^T\right) \otimes \mathbf{I}_Q \otimes \mathbf{I}_P \\ &= \widetilde{\mathbf{C}}^T\widetilde{\mathbf{C}} \otimes \mathbf{I}_Q \otimes \mathbf{I}_P \\ &= \mathbf{I}_R \otimes \mathbf{I}_Q \otimes \mathbf{I}_P = \mathbf{I}_{PQR}\,. \end{aligned} \tag{12}$$

The update of $\mathcal{G}$ in step 2 of the ALS algorithm for SCA-T3 in section 3.1 is computed via OLS regression in (11). Indeed, this is identical to the update of $\mathcal{G}$ in the ALS algorithm for Tucker2 fitted to the $P \times J \times K$ array with slices $\widetilde{\mathbf{A}}_k^T\mathbf{X}_k$, $k = 1, \ldots, K$. After convergence of the ALS algorithm for SCA-T3 we set $\mathbf{A}_k = N_k^{1/2}\widetilde{\mathbf{A}}_k$ and $\mathbf{c}_k^{(\text{row})} = N_k^{-1/2}\,\tilde{\mathbf{c}}_k^{(\text{row})}$, $k = 1, \ldots, K$. For the blocks of the matrix in (11), we have $(\mathbf{c}_k^{(\text{row})})^T \otimes \mathbf{B} \otimes \mathbf{A}_k = (\tilde{\mathbf{c}}_k^{(\text{row})})^T \otimes \mathbf{B} \otimes \widetilde{\mathbf{A}}_k$. It follows that for SCA-T3 we can compute the fit percentage due to term (p, q, r) as

$$100 - 100 \cdot \frac{\|\mathbf{X}_{\text{all}} - \mathbf{A}_{\text{all},p}\,c_{kr}\,g_{pqr}\,\mathbf{b}_q^T\|^2}{\|\mathbf{X}_{\text{all}}\|^2} = 100 \cdot \frac{\|\mathbf{A}_{\text{all},p}\,c_{kr}\,g_{pqr}\,\mathbf{b}_q^T\|^2}{\|\mathbf{X}_{\text{all}}\|^2}\,, \tag{13}$$

where $\mathbf{A}_{\text{all},p}$ denotes column p of $\mathbf{A}_{\text{all}} = [\mathbf{A}_1^T \ \ldots \ \mathbf{A}_K^T]^T$. Moreover, summing the fit percentages (13) for all p, q, and r yields the overall fit percentage (8). Orthogonal rotation of the Tucker2 solution in step 2 of the SCA-T3 algorithm (section 3.3) does not change the columnwise orthonormality of the matrix in (11). Hence, the fit percentage due to term (p, q, r) can be computed via (13) also after orthogonal rotation. The fit percentages indicate which terms should be used to interpret the SCA-T3 solution. This provides a more reliable measure for interpretation than considering 'large' core entries only.

When the SCA models are fitted on observed covariance matrices and we use our heuristic method explained in section 3.2, we compute the fit percentages as follows. We have $\text{Cov}(\mathbf{X}_k) = \mathbf{Y}_k^T\mathbf{Y}_k$ and fit an SCA model as $\mathbf{Y}_k \approx \mathbf{F}_k\,\mathbf{B}^T$, where the form of $\mathbf{F}_k$ depends on the SCA model (Table 1). The overall fit percentage (8) can be computed as $100 \cdot (\sum_k \text{trace}(\mathbf{B}\,\mathbf{F}_k^T\mathbf{F}_k\,\mathbf{B}^T)/(\sum_k \text{trace}(\text{Cov}(\mathbf{X}_k)))$. Analogously, the fit percentage (9) for sample k equals $100 \cdot (\text{trace}(\mathbf{B}\,\mathbf{F}_k^T\mathbf{F}_k\,\mathbf{B}^T)/(\text{trace}(\text{Cov}(\mathbf{X}_k)))$. For SCA-T3 also the fit percentage due to each term (p, q, r) can be computed analogous to (13).

3.5 SCA model selection

Since SCA-T3 adds a fifth and more general model to the already existing four SCA models in the SCA-P framework, one may wonder how to select an appropriate SCA model and number(s) of components for a given dataset at hand. For the four existing SCA models an excellent discussion of model selection is included in Timmerman and Kiers (2003). Next, we summarize the guidelines in this discussion and discuss model selection for the SCA-T3 model. In general, it is desirable to choose the most constrained model with a relatively small number of components that fits the data well and has a clear substantive interpretation. Apart from the fit percentage versus number of components trade-off, issues of model stability are also important to take into account in model selection. Model stability can be assessed via cross-validation and split-half analysis.

Cross-validation aims to assess the predictive validity of the estimated model. For the SCA models Timmerman and Kiers (2003) advise to use the Expectation-Maximization cross-validation (EM-CV) method of Louwerse, Smilde, and Kiers (1999). In EM-CV the complete dataset $\mathbf{X}_k$, $k = 1, \ldots, K$, is split up into M about equal-sized parts, by randomly asigning each observation to one of the M parts. Then the SCA model is fitted to the dataset while treating one of the M parts as missing data in the EM framework. In the ALS

algorithm the missing values are replaced with model estimates after each iteration. After convergence the predictive error sum-of-squares (PRESS) is computed as the sum-of-squares of the true missing data minus the estimated missing data. This procedure is repeated for all M parts separately, and the final PRESS value is the sum of the M individual PRESS values. As such, PRESS is used as an indicator of the predictive accuracy of the model when a part of the data is left out. For more details on PRESS we refer to Timmerman and Kiers (2003) and Louwerse et al. (1999). The EM-CV procedure described above can also be used for SCA-T3. Since only the fitted model is needed in its entirety, no rotations need to be applied.

In a split-half analysis the complete dataset is randomly split up into two about equal-sized parts (same procedure as in EM-CV), after which the SCA model is fitted to both halves separately. The obtained two sets of estimates can then be compared to assess model stability. Timmerman and Kiers (2003) propose to use loading matrix $\mathbf{B}$ for this purpose, and define the split-half stability (SHS) coefficient as the mean absolute difference between the two estimates of $\mathbf{B}$ (after appropriately permutating and sign changing their columns). The final SHS is then obtained as the average of the SHS coefficients obtained from 50 random splits of the data. For SCA models that do not yield a unique loading matrix estimate (up to permutation and scaling) such as SCA-P, SCA-ECP, and SCA-T3, we need to decide how to compare the two estimates of $\mathbf{B}$. For SCA-P and SCA-ECP Timmerman and Kiers (2003) propose to use Varimax rotation of the two estimated loading matrices separately. For SCA-T3 we can apply joint orthomax rotation of Kiers (1998b) to obtain a simple core and loading matrix in both estimated models separately. Alternatively, we can rotate one SCA-T3 solution to the other one by minimizing the criterion (17) as will be explained in section 4. Indeed, when only the subspace of the loading matrix is identified it makes sense to use this measure for closeness of the two subspaces. Naturally, the same reasoning applies to SCA-P and SCA-ECP.

For each SCA model and number(s) of components the PRESS and SHS values and the fit percentage can be reported. Models with too large PRESS or SHS values are then discarded, using a chosen threshold value that is specific for the dataset at hand. We are then left with various SCA models of various sizes and their fit percentages. An appropriate SCA model may now be chosen by considering the fit percentage and the interpretability of the solution. For SCA-T3 we need to choose (R, P, Q) instead of just one number of components. For

Tucker3 several methods have been proposed for chosing (R, P, Q), an intuitive one being the convex hull method of Ceulemans and Kiers (2006). Here, a graph is plotted of number of free parameters in the Tucker3 model versus fit percentage. Only the models on the convex hull of this graph are considered as suitable, since they offer relatively the most fit with relatively the smallest number of model parameters. For SCA-T3 this method may also be used, after unstable models have been discarded. Note, however, that the convex hull method does not consider interpretability of the solution. This may be a subjective reason to favor a model that is slightly less stable than desirable or does not lie on the convex hull in the plot of parameters versus fit.

In an application in section 5 it will be illustrated how the above criteria for model selection can be balanced in practice.

4 Simulation study

We conduct a Monte Carlo simulation study to evaluate the SCA-T3 model and ALS algorithm with respect to the occurrence of local minima and the recovery of underlying true loading matrix $\mathbf{B}$, weights matrix $\mathbf{C}$, and core array $\mathcal{G}$. In section 4.1 we generate noisefree SCA-T3 datasets and determine the number of runs with different (random) starting values that are needed to find the global minimum. We consider both fitting SCA-T3 to observed scores (using the ALS algorithm of section 3.1) and to observed covariance matrices (using the heuristic algorithm of section 3.2). In section 4.2 we generate noisy SCA-T3 datasets and focus on recovery of the underlying true parameters.

4.1 Occurrence of local minima

We set $J = 6$, $K = 5$, $N_1 = 50$, $N_2 = 100$, $N_3 = 50$, $N_4 = 150$, and $N_5 = 250$. True matrices $\widetilde{\mathbf{A}}_k$, $\mathbf{B}$ and $\widetilde{\mathbf{C}}$ are generated as orthonormal bases of matrices with entries randomly sampled from the standard normal distribution. For these values of J, K, and N_k, we consider two true core arrays: $(P, Q, R) = (3, 3, 2)$ and

$$[\mathbf{G}_1 \,|\, \mathbf{G}_2] = \left[\begin{array}{ccc|ccc} 1.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.8 \\ 0 & 0 & -1 & 0 & 0.6 & 0 \end{array}\right], \tag{14}$$

and $(P, Q, R) = (3, 3, 1)$ and

$$\mathbf{G} = \begin{bmatrix} 1.5 & 0 & 0 \\ 0 & 0.8 & 0 \\ 0 & 0 & 0.6 \end{bmatrix}. \tag{15}$$

We also consider core (15) with $K = 2$ and J and N_k as above, since we use this model in the simulation study in section 4.2. Additionally, we consider the SCA-T3 model size of the application in section 5.1: $J = 20$, $K = 12$, N_k equal to 70, 53, 70, 69, 68, 70, 68, 69, 69, 70, 70, and 71, $(P, Q, R) = (4, 4, 3)$, and core $\mathcal{G}$ equal to

$$[\mathbf{G}_1 \,|\, \mathbf{G}_2 \,|\, \mathbf{G}_3] = \left[\begin{array}{cccc|cccc|cccc} 0 & 0.6 & 0 & 0.4 & 0 & 0 & 0.3 & 0.5 & 0.3 & 0 & 0 & 0 \\ 0.9 & 0 & 0 & 0 & 0 & 0.5 & 0.8 & -0.3 & 0 & 0 & 0.3 & 0 \\ 0.3 & 0 & 0 & 1.0 & -0.8 & 0 & 1.0 & 0 & 0 & 1.0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 & 0 & 0.3 & 0 & 0 & -0.4 & 0 & 0 & 0.3 \end{array}\right]. \tag{16}$$

After true $\widetilde{\mathbf{A}}_k$, $\mathbf{B}$, $\widetilde{\mathbf{C}}$, and $\mathcal{G}$ have been generated, the scores data is constructed as $\mathbf{X}_k = \widetilde{\mathbf{A}}_k \left(\sum_{r=1}^{R} \tilde{c}_{kr}\, \mathbf{G}_r\right) \mathbf{B}^T$, $k = 1, \ldots, K$. For each model size we generate 100 noisefree datasets. For each dataset we fit the SCA-T3 ALS model of section 3.1 with convergence criterion 10^{-7}, one rationally started run and a varying number of randomly started runs. For each dataset we increase the number of runs by 10 when the run with highest fit percentage (8) has fit less than 99.99 percent. In Table 2 the numbers of datasets with best run less than 99.99, 99.9, and 99.0 percent fit are reported, where the number of runs is the minimum for which at most 5 datasets have their best run with fit less than 99.99 percent. We also report the frequencies for half that number of runs and when only the rationally started run is used.

Next to fitting SCA-T3 to observed scores $\mathbf{X}_k$, we also fit SCA-T3 to noisefree $\text{Cov}(\mathbf{X}_k)$ using the heuristic algorithm of section 3.2. The noisefree $\text{Cov}(\mathbf{X}_k)$ is given by $\mathbf{B} \left(\sum_{r=1}^{R} \tilde{c}_{kr}\, \mathbf{G}_r^T\right) \left(\sum_{r=1}^{R} \tilde{c}_{kr}\, \mathbf{G}_r\right) \mathbf{B}^T$. We consider the same true models as above.

[Insert Table 2 here.]

The results are as follows (see Table 2). For core (15) only the rationally started run is needed to obtain a fit of 100 percent. For core (14) 20 or 30 runs are sufficient to obtain a best run with fit larger than 99.99 percent for almost all datasets. For the larger model with core (16), however, a much larger number of runs is needed. This is important to keep in

mind when fitting SCA-T3. The rationally started run is of not much use for cores (14) and (16), since for the vast majority of datasets it yields fit less than 99.99 percent.

4.2 Recovery of model parameters

We consider the same true models as in section 4.1 and the scores data is now generated as $\mathbf{X}_k = \mathbf{X}_k^{(\text{model})} + \sigma_k \mathbf{E}_k$, with $\mathbf{X}_k^{(\text{model})} = \widetilde{\mathbf{A}}_k \left(\sum_{r=1}^{R} \tilde{c}_{kr} \mathbf{G}_r \right) \mathbf{B}^T$ and the entries of $\mathbf{E}_k$ randomly sampled from the standard normal distribution. The noise strength σ_k is defined as $\sigma_k = s_k \, \|\mathbf{X}_k^{(\text{model})}\| / (J \, N_k)^{1/2}$ such that the expected sum-of-squares of the noise term equals $s_k^2 \, \|\mathbf{X}_k^{(\text{model})}\|^2$ (analogous to Kiers et al., 1999, who set $s_k = 0.5$). We consider two noise levels, $s_k = 0.5$ and $s_k = 0.7$, that are equal for all samples k. Additionally, we consider unequal s_k by taking the first K values from the sequence 0.4, 0.6, 0.5, 0.7, 0.4, 0.4, 0.6, 0.5, 0.7, 0.4, 0.6, 0.6. For each true model we generate 100 noise instances and fit SCA-T3 to the resulting $\mathbf{X}_k$, $k = 1, \ldots, K$, using the ALS algorithm of section 3.1. For each dataset, the ALS algorithm is run $m - 1$ times with random starting values and one time with rationally chosen starting values, with the number of runs m taken from Table 2 with the following exceptions: for core (15) we use $m = 10$, and for core (16) we use $m = 50$. The estimates of the run with the best fit are kept.

After estimates $\widehat{\mathbf{B}}$, $\widehat{\mathbf{C}}$, and $\widehat{\mathcal{G}}$ have been obtained, we apply orthogonal rotations $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{U}$ such that

$$\|\widehat{\mathbf{B}} \, \mathbf{T} - \mathbf{B}\|^2 + \|\widehat{\mathbf{C}} \, \mathbf{U} - \widetilde{\mathbf{C}}\|^2 + \sum_{h=1}^{R} \|\mathbf{S}^T \left(\sum_r u_{rh} \, \widehat{\mathbf{G}}_r \right) \mathbf{T} - \mathbf{G}_h\|^2 \,, \tag{17}$$

is minimal. Hence, the rotations minimize the sum of the norms of the differences between the rotated estimates and the true $\mathbf{B}$, $\widetilde{\mathbf{C}}$, and $\mathcal{G}$, respectively. This is achieved by iterating over $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{U}$, where each update boils down to a Varimax rotation when the core is unfolded as a matrix in the appropriate way. This method was also suggested in Kiers (2004). As starting rotations for $\mathbf{T}$ and $\mathbf{U}$ we use the Varimax rotations of $\widehat{\mathbf{B}}$ to $\mathbf{B}$ and of $\widehat{\mathbf{C}}$ to $\widetilde{\mathbf{C}}$, respectively. The starting $\mathbf{S}$ is then taken equal to its update for fixed $\mathbf{T}$ and $\mathbf{U}$. In the simulations below the minimization of (17) does not take more than 10 iterations in all replications.

Note that analogous to SCA-PF2 the model SCA-T3 features a sign indeterminacy of each row of $\widetilde{\mathbf{C}}$ separately. Indeed, we can multiply row k of $\widetilde{\mathbf{C}}$ by -1 and $\widetilde{\mathbf{A}}_k$ by -1 and

obtain the same fitted model. This sign indeterminacy is not taken into account in the rotations above. Therefore, we consider all possible sign changes of the rows of $\widehat{\mathbf{C}}$, compute the minimum of (17), and keep the sign changed $\widehat{\mathbf{C}}$ with smallest value of (17). A discussion of this sign indeterminacy in SCA-PF2 can be found in Helwig (2013).

In the above it is implicitly assumed that for the SCA-T3 model without noise we can recover the true $\widetilde{\mathbf{A}}_k$, $\mathbf{B}$, $\widetilde{\mathbf{C}}$, and $\mathcal{G}$ after rotating when the fit percentage is 100 percent. This has been verified numerically for the chosen models above. This is not always true, however. For example, with core (14) and $K = 2$ (and $N_1 = 50$, $N_2 = 100$) we obtain different $\widetilde{\mathbf{A}}_k$, $\widetilde{\mathbf{C}}$, and $\mathcal{G}$ after rotating even though we have perfect fit. With $K = 2$ and core (15) we do not have this problem. Note that for $K = 2$ we cannot apply SCA-PF2 due to nonuniqueness, as mentioned in section 2.2.

The estimation accuracy is evaluated by means of the mean absolute deviation (MAD) and root mean square error (RMSE), which are defined for estimate $\widehat{\mathbf{B}}$ and true value $\mathbf{B}$ as

$$\text{MAD}(\mathbf{B}) = (J\,Q)^{-1} \sum_{j=1}^{J} \sum_{q=1}^{Q} |b_{jq} - \hat{b}_{jq}|\,, \qquad \text{RMSE}(\mathbf{B}) = \left((J\,Q)^{-1} \sum_{j=1}^{J} \sum_{q=1}^{Q} (b_{jq} - \hat{b}_{jq})^2 \right)^{1/2}. \tag{18}$$

The MAD and RMSE for $\widetilde{\mathbf{C}}$ and $\mathcal{G}$ are defined analogously, where $\mathcal{G}$ is first unfolded as a matrix.

In Table 3 the results are presented. As can be seen, the recovery is very good for all estimates and models when the noise strength is $s_k = 0.5$. As expected, for $s_k = 0.7$ the recovery is less accurate. For unequal s_k the recovery is mostly less accurate when compared to $s_k = 0.5$, but more accurate when compared to $s_k = 0.7$. When comparing the results for different true models, it can be seen that recovery of $\mathbf{B}$ is best for core (16), second best for core (14), and worst (but still very good) for core (15), with little difference between $K = 5$ and $K = 2$. Recovery of $\widetilde{\mathbf{C}}$ and $\mathcal{G}$ is best for core (15), second best for core (16), and worst (but still good for $\widetilde{\mathbf{C}}$) for core (14). In general, recovery of $\mathcal{G}$ seems to be most difficult. This is probably due to the zeros imposed in the true cores. It may be surprising that the largest model with core (16) has better recovery results than the model with core (14). However, the larger value of $K = 12$ may contribute to more stable estimation results.

Next, we generate multi-set data from true covariance matrices, and use the heuristic method described in section 3.2 to estimate $\mathbf{B}$, $\widetilde{\mathbf{C}}$, and $\mathcal{G}$. The data for sample k is generated as $\mathbf{X}_k = \mathbf{Z}_k\,(\mathbf{\Sigma}_k)^{1/2}$, where $\mathbf{Z}_k$ is an $N_k \times J$ matrix with entries randomly sampled

from the standard normal distribution, and $\mathbf{\Sigma}_k$ is the true covariance matrix under SCA-T3. The algorithm of section 3.2 is then applied to $\text{Cov}(\mathbf{X}_k)$, $k = 1, \ldots, K$. After estimation, rotations to the true $\mathbf{B}$, $\widetilde{\mathbf{C}}$, and $\mathcal{G}$ are applied as above. We use the same true models as in section 4.1 and the number of randomly started runs is taken as above. It was verified numerically that for $\text{Cov}(\mathbf{X}_k) = \mathbf{\Sigma}_k$ the true $\mathbf{B}$, $\widetilde{\mathbf{C}}$, and $\mathcal{G}$ are recovered when the fit is 100 percent. The results are given in Table 3. As can be seen, recovery is very good for all estimates and models. Compared to scores data with noise strength $s_k = 0.5$, recovery is better for $\mathbf{B}$ and $\mathcal{G}$ but slightly worse for $\widetilde{\mathbf{C}}$. We conclude that the heuristic algorithm of section 3.2 for fitting SCA-T3 to observed covariance matrices can accurately retrieve the underlying model parameters.

[Insert Table 3 here.]

5 Applications

Here we apply the five SCA models in Table 1 to benchmark datasets in the literature and show the added value of SCA-T3. In section 5.1 we apply SCA-T3 to the 20-item mood test administered to 12 individuals diagnosed with Parkinson's disease. This dataset was analyzed by Timmerman and Kiers (2003) using the four existing SCA models. In section 5.2 we consider a dataset of scores on 12 tests by high school students of high and low IQ and high and low socioeconomic status, which was also analyzed in McGaw and Jöreskog (1971). In section 5.1 the SCA models are fitted to the observed scores $\mathbf{X}_k$, while in section 5.2 the SCA models are fitted to the observed covariance matrices $\text{Cov}(\mathbf{X}_k)$.

5.1 PANAS and Parkinson's disease

In Shifren, Hooker, Wood and Nesselroade (1997) 12 individuals diagnosed with Parkinson's disease were administered the Positive and Negative Affect Scale (PANAS; Watson, Clark and Tellegen, 1988) on consecutive days to study their mood structure over time. The PANAS consists of 20 items, 10 measuring positive affect and 10 measuring negative affect, scored on a 5-point Likert scale. The 12 subjects scored the PANAS on consecutive days over a period varying from 53 to 71 days. The column centered scores of subject k are collected

in $\mathbf{X}_k$ ($N_k \times J$), where N_k is the number of days subject k scored the PANAS, and $J = 20$ is the number of items. Hence, we have $K = 12$ matrices $\mathbf{X}_k$. Values of N_k are 70, 53, 70, 69, 68, 70, 68, 69, 69, 70, 70, and 71. To study the interindividual and intraindividual differences in mood structure for the subjects with Parkinson's disease Timmerman and Kiers (2003) apply the four existing SCA models. The general component structure is then given by loading matrix $\mathbf{B}$, while $\mathbf{C}$ contains weights indicating interindividual differences, and $\mathbf{A}_k$ contains the intraindividual variation for subject k. Scaling of the columns of $\mathbf{X}_k$ is applied such that $[\mathbf{X}_1^T \ \ldots \ \mathbf{X}_K^T]^T$ has column sum-of-squares equal to $\sum_k N_k = 817$. For the four existing SCA models the PRESS and SHS values and fit percentages are computed for one through five components. To obtain the PRESS value the EM-CV method is used as described in section 3.5, where the total number of $817 \cdot 20 = 16340$ observations is randomly split into 150 parts (149 parts of size 109 and one part of size 99). Using the (subjective) criteria of PRESS ≤ 12000 and SHS ≤ 0.10, three stable models are obtained: SCA-IND with $R = 2$ (fit 42.8%) and SCA-P with $R = 2$ (fit 43.5%) and $R = 4$ (fit 58.4%). After orthogonal Procrustes rotation of the loading matrix of SCA-P with $R = 2$ to that of the SCA-IND model, the two solutions are found to be very similar with the largest difference being the moderate correlations for some subjects between the two components in the SCA-P solution (while these are uncorrelated in SCA-IND). The two components are interpreted as Introversion and Emotional Instability. After Varimax rotation, the solution for SCA-P with $R = 4$ yields a more detailed description of the data with higher fit that also features the Emotional Instability component. The Introversion component is split up into Arousal and Nervousness components. The fourth component is a combination of the items "ashamed", "excited" and "proud". Timmerman and Kiers (2003) conclude that the SCA-IND solution with $R = 2$ is preferred for a parsimonious description of the main features of the data, while the SCA-P solution with $R = 4$ yields a more detailed picture. Interpretation of the latter solution is more difficult, however, due to moderately correlated components for some subjects.

Next, we apply SCA-T3 to the dataset with numbers of components $P \in \{2, 3, 4\}$, $Q \in \{2, 3, 4\}$, and $R \in \{1, 2, 3, 4\}$. For each model we compute PRESS and SHS values and the fit percentage. For PRESS the same procedure is used as in Timmerman and Kiers (2003) where for each dataset with missing values SCA-T3 is run once with rationally chosen starting values. For SHS we split the dataset 50 times in random halves and compute

SCA-T3 solutions for both halves using one run with rationally chosen starting values. The two estimates of loading matrix $\mathbf{B}$ are then rotated in two ways (section 3.3): using joint orthomax rotation to simple cores and loading matrices in both solutions separately (yielding SHS1), and rotating one solution to the other by minimizing the criterion (17) (yielding SHS2). Final values of SHS1 and SHS2 are the average of the MAD between the two estimates of the loading matrix over the 50 replications. For all models we obtain SHS2 $<$ SHS1. For the fit percentage the SCA-T3 model is fitted using 49 runs with random starting values and one run with rationally chosen starting values, where the solution with the highest fit is kept. In Table 4 the SCA-T3 models are listed with PRESS $\leq$ 12000 and SHS2 $\leq$ 0.10.

[Insert Table 4 here.]

It would be unfair to use SHS2 $\leq$ 0.10 as stability criterion for the SCA-T3 models while Timmerman and Kiers (2003) use 0.10 as theshold for two loading matrices that are Varimax rotated individually. Hence, we also consider computing the SHS via Procrustes rotation of one loading matrix to the other for SCA-ECP and SCA-P. The SCA-ECP models all have PRESS $>$ 12000 and only one SCA-P model (with three components) has PRESS $\leq$ 12000 but SHS $>$ 0.10. Using Procrustes rotation to compute SHS for SCA-P with three components yields SHS=0.09, which makes this an additional stable model using this definition of SHS. Varimax rotation of its loading matrix results in components very similar to the Emotional Stability, Arousal, and Nervousness components found in the Varimax rotated SCA-P solution with four components.

As can be seen in Table 4 the SCA-T3 models up to $(P, Q, R) = (2, 2, 4)$ have nearly the same fit percentage and all have $Q = 2$ components for the 20 items of the PANAS. Based on the criteria of parsimony and interpretability the SCA-IND solution with two components is preferred to these SCA-T3 models, since its fit is only slightly worse. The other two models in Table 4 have $Q = 4$ and higher fit percentage. Of these two models we prefer the SCA-T3 model that has $R = 3$ components for the 12 subjects instead of $R = 4$. Below, we present the solution of SCA-T3 with $(P, Q, R) = (4, 4, 3)$ after joint orthomax rotation to a simple core, loading matrix $\mathbf{B}$, and weights matrix $\widetilde{\mathbf{C}}$. Rotating to a simple core and $\mathbf{B}$ yields very similar results. By taking the best run out of $m - 1$ runs with random starting values (for large m) and one run with rationally chosen starting values, we found two distinct solutions

with nearly the same fit of 57.1 percent. The two solutions differ mostly in two columns of $\mathbf{B}$ and the core array $\mathcal{G}$. Below we present the solution with the clearest structure in $\mathbf{B}$.

In Table 5 the loading matrix $\mathbf{B}$ can be found. We label the columns of $\mathbf{B}$ as B1, B2, B3, and B4, respectively. Column B1 is similar to the Introversion component in Timmerman and Kiers (2003), except that its loadings are a bit larger in magnitude. Column B2 is similar to the fourth component in the SCA-P solution above. Column B3 can be interpreted as Perseverance, since it combines the items "afraid" and "scared" with "determined" and "strong". Column B4 is similar to the Emotional Instability component of the SCA-P solution above, except that items "jittery", "hostile", and "strong" are not loading on it. Columns B3 and B4 are mixed in the other solution we found, resulting in less clear structure in $\mathbf{B}$.

The subject weights matrix $\mathbf{C}$ is given in Table 6, together with fit percentage (9) for each subject and the sum of the variances of each $\mathbf{X}_k$. As can be seen, subjects with high variability in their scores also tend to have a high fit percentage. The interindividual differences are quite large in terms of variability in the scores and weights in matrix $\mathbf{C}$. Below, we also consider some of the time courses in $\mathbf{A}_k$. The core entries and associated explained variances (13) can be found in Table 7. After rotating, only four core entries have fit larger than 4 percent. Hence, although the SCA-T3 model contains 48 combinations of columns of $\mathbf{A}_k$, $\mathbf{B}$, and $\mathbf{C}$, joint orthomax rotation yields a relatively simple solution in terms of interpretability. The combination of the Introversion component B1 and weights C3 with time course $\mathrm{A3}_k$ has explained variance 20.9 percent, which is the dominant term in the solution. Introversion component B1 combined with weights C1 and time course $\mathrm{A1}_k$ has the second largest explained variance.

[Insert Tables 5,6,7 here.]

To plot the time course associated with Introversion component B1 for subject k, we compute $\mathbf{A}_k \left(\sum_{r=1}^{R} c_{kr} \left(\mathbf{G}_r \right)_{(:,1)} \right)$, with $\left(\mathbf{G}_r \right)_{(:,1)}$ denoting column 1 of $\mathbf{G}_r$. In Figure 1 these time courses are depicted for subjects 5 and 10, which have large and small amounts of variability in their scores, respectively. As can be seen, the time courses mimic the variability present in the scores for these two subjects. Similar time courses, including the change point around 15 days for subject 5, are found in the SCA-IND and SCA-P solutions in Timmerman and

Kiers (2003).

[Insert Figure 1 here.]

To summarize, for the PANAS dataset SCA-T3 yields stable models with $Q = 2$ or $Q = 4$ components for the 20 items. This is analogous to the findings of Timmerman and Kiers (2003). After rotating the SCA-T3 solution with $(P, Q, R) = (4, 4, 3)$ only a few terms with large explained variance are obtained, which simplifies interpretation. Compared to the SCA-P solution with four components, the SCA-T3 solution offers the following advantages. Unlike SCA-P the components in $\mathbf{A}_k$ are uncorrelated in SCA-T3, which eases interpretation. Indeed, Timmerman and Kiers (2003) state that the SCA-P solution features component covariance matrices $\mathbf{\Phi}_k$, $k = 1, \ldots, 12$, with small to large off-diagonal entries for different k which makes the solution difficult to interpret. Next, unlike the existing SCA models the number of components can be different for different modes. Hence, when four components are desired for the 20 items, the number of components in the subjects mode may be chosen smaller. Another feature of SCA-T3 that makes interpretation easier. Time courses corresponding to specific components in $\mathbf{B}$ can be plotted as for the existing SCA models. Hence, for the PANAS dataset SCA-T3 offers a more versatile model with stable solutions that are easy to interpret. As such, it is an interesting addition to the existing SCA models.

5.2 Ability tests in low and high IQ and socioeconomic status groups

In McGaw and Jöreskog (1971) a large dataset of test scores by high school children is analyzed with multi-set factor analysis using maximum likelihood. The sample of 11743 children is split up into $K = 4$ groups with low and high IQ and low and high socioeconomic status (SES). Group sizes are $N_1 = 4491$ (low IQ, low SES), $N_2 = 1336$ (low IQ, high SES), $N_3 = 939$ (high IQ, low SES), and $N_4 = 4977$ (high IQ, high SES). The dataset consists of rescaled correlation matrices $\mathbf{S}_k$, $k = 1, 2, 3, 4$, for $J = 12$ tests administered to all children (given in McGaw and Jöreskog, 1971). Applying maximum likelihood factor analysis to the pooled correlation matrix yields a nice structure in the loading matrix for four factors after

Varimax rotation: each factor represents a distinct set of three tests. In a subsequent multi-set factor analysis for the four groups the small entries in this Varimax rotated loading matrix are fixed at zero in the common loading matrix. Below, we fit the five SCA models in Table 1 to $\mathbf{S}_k$, $k = 1, 2, 3, 4$, and wish to obtain a loading matrix $\mathbf{B}$ with the nice structure above without having to resort to setting loadings equal to zero. Note that we cannot compute PRESS and SHS values, since the observed scores are not available.

For the four existing SCA models we use $R = 4$ components and fit them to $\mathbf{S}_k$, $k = 1, 2, 3, 4$, using the heuristic method described in section 3.2. The iterative ALS algorithms for SCA-PF2, SCA-IND, and SCA-ECP are run 49 times with random starting values and once with rationally chosen starting values. The solution of the run with the highest fit is kept. In SCA-PF2 the solution is scaled such that $[(\mathbf{A}_1\mathbf{C}_1)^T \ \ldots \ (\mathbf{A}_4\mathbf{C}_4)^T]^T$ has column sum-of-squares equal to $K = 4$ and $\mathbf{A}_k$ has column sum-of-squares one (see section 3.2). The scaling coefficients are then absorbed in $\mathbf{B}$. This is done analogously for the other SCA models. We call the obtained loading matrix $\mathbf{B}$ separated when each test has absolute loading larger than 0.4 on exactly one factor and each factor has loadings larger than 0.4 on a distinct set of three tests. Otherwise we call the loading matrix mixed. For SCA-P (fit 63.5%) we obtain a separated $\mathbf{B}$ after Varimax rotation. The component covariance matrices $\mathbf{\Phi}_k$ are roughly diagonal with largest off-diagonal entry equal to 0.14. For SCA-PF2 (fit 63.5%) and SCA-IND (fit 63.4%) we obtain mixed loading matrices. For SCA-ECP (fit 62.1%) we obtain a separated $\mathbf{B}$ after Varimax rotation, and have set $\mathbf{\Phi} = \mathbf{I}_4$. Hence, both SCA-P and SCA-ECP yield a separated $\mathbf{B}$ after Varimax rotation. In the SCA-P solution the differences between the four groups boil down to differences in component variances. In SCA-ECP the model is the same for all four groups, which makes it inappropriate to analyze group differences. For SCA-P we also use PCA on the weighted average of $\mathbf{S}_k$; see (5). This also yields a separated $\mathbf{B}$ after Varimax rotation. Estimating $\mathbf{\Phi}_k$ via Penrose regression (section 3.2) yields largest off-diagonal entry equal to 0.18. This SCA-P solution is very similar to the one obtained via the heuristic method. Next, we fit SCA-T3 and show that a solution with separated $\mathbf{B}$ can be obtained that contains more information on group differences than the SCA-P solutions.

We fit SCA-T3 with $(P, 4, R)$ components for $P \in \{1, \ldots, 5\}$ and $Q \in \{1, 2\}$. We use joint orthomax rotation to a simple core and loading matrix $\mathbf{B}$ for $Q = 1$ and also to a simple weight matrix $\widetilde{\mathbf{C}}$ for $Q = 2$. For $P \leq 3$ we obtain a mixed $\mathbf{B}$ for both $Q = 1$ and $Q = 2$.

For $P \geq 4$ we obtain a seperated $\mathbf{B}$ for both $Q = 1$ and $Q = 2$. The fit is 62.6% for the $(4, 4, 1)$ and $(5, 4, 1)$ solutions, and 63.4% for the $(4, 4, 2)$ and $(5, 4, 2)$ solutions. Hence, using $P = 5$ hardly adds explained variance with respect to $P = 4$. Below, we present the $(4, 4, 2)$ solution since it contains more information on group differences than the (4,4,1) solution.

The loading matrix $\mathbf{B}$ is Table 8 is clearly separated: each component represents a distinct group of three tests. The group weight matrix $\mathbf{C}$ in Table 9 shows that the main difference is between high and low IQ. Column C1 corresponds to low IQ and column C2 corresponds to high IQ. The fit percentage per group does not differ much. The core entries and associated explained variances in Table 10 show that the solution is dominated by three terms with explained variance larger than 10 percent, and three terms with explained variance around 6 percent. For the high IQ group components B1 and B3 are dominant, while for the low IQ group all four components play a role with B2 and B3 showing more variability. In general the low IQ group shows more variability in the solution, and also in the rescaled correlation matrices $\mathbf{S}_k$ (see Table 9).

[Insert Tables 8,9,10 here.]

Finally, we make a detailed comparison of the SCA-T3 solution to the SCA-P solutions. We pick the SCA-P solution that was obtained using the heuristic method in section 3.2, since it does not put weights on the $\mathbf{S}_k$ based on sample size N_k (and neither do the other SCA methods used above). The SCA-P solution obtained from PCA on the weighted average of $\mathbf{S}_k$ is very similar, however. The loading matrix $\mathbf{B}$ in the SCA-P solution is separated and very similar to $\mathbf{B}$ in the SCA-T3 solution in Table 8. In Table 11 the variance of the components in the SCA-P solution (diagonals of $\mathbf{\Phi}_k$) are given for the corresponding columns of $\mathbf{B}$ in the SCA-T3 solution. Also given are the explained variances of each component in each sample, if it were the only component. These are computed as $100 \cdot \text{trace}(\mathbf{b}_r \, (\mathbf{\Phi}_k)_{rr} \, \mathbf{b}_r^T)/\text{trace}(\mathbf{S}_k)$; see the end of section 3.4. It is important to note that these explained variances should be interpreted with care, since they only add up to the explained variance per group when the components are orthogonal (which is not true for the SCA-P solution). The component variances should also be interpreted with care, since they cannot be related unambiguously to explained variances. Hence, a first observation in the comparison between the SCA-P and SCA-T3 solutions is that interpretation of the explained variances due to each term is much

easier in the SCA-T3 solution; see Table 10 in which the explained variances add up to the total explained variance. Ignoring these ambiguities and comparing the SCA-P and SCA-T3 solutions, we observe the following. SCA-T3 summarizes the four groups into two group components, corresponding to low and high IQ (Table 9), and indicates the strength of the test components in $\mathbf{B}$ for the two group components in the core array $\mathcal{G}$ (Table 10). In the SCA-P solution this information should be gathered from one table of explained variances (Table 11), which is less clear for the practitioner. The SCA-P and SCA-T3 solutions both yield B2 and B3 as strongest components for the low IQ groups, and B1 and B3 as strongest for the high IQ groups. However, the differences in explained variance are more pronounced in the SCA-T3 solution. Also, there is considerable variation among the two low IQ groups in the SCA-P solution, which is summarized more clearly in the group components of the SCA-T3 solution. We conclude that the SCA-T3 solution offers a clearer picture of group differences than the SCA-P solution.

[Insert Table 11 here.]

6 Discussion

We have introduced a new model for simultaneous component analysis based on the Tucker3 model for component analysis of three-way arrays. The new SCA-T3 model is more versatile than existing SCA models, since for each mode a different number of components can be chosen. The solution can be rotated to obtain simple structure in the loading matrix, core array, and the weight matrix for the samples. An optimal solution always exists for SCA-T3 and, hence, problems of diverging solutions do not occur. We have derived an ALS algorithm for fitting SCA-T3 to observed scores, and a heuristic method to fit SCA-T3 to observed covariance matrices. In the simulation study the algorithms are successful in recovering underlying components for generated scores data with added noise and for data generated from population covariance matrices. The application to the PANAS data for subjects with Parkinson's disease shows that SCA-T3 offers stable models that are easy to interpret after rotation of the solution. Moreover, compared to the SCA-P solution with four components (recommended in Timmerman and Kiers, 2003) the SCA-T3 solution is easier to interpret due

to the components $\mathbf{A}_k$ being uncorrelated. In the SCA-P solution the component covariance matrices $\mathbf{\Phi}_k$ feature small and large off-diagonal entries for different k. The application to the observed covariance matrices of ability tests for high school children shows that rotating the solution to simple structure is a key feature in obtaining an appropriate loading matrix. The SCA-PF2 and SCA-IND do not yield a simple loading structure for this dataset. The SCA-ECP solution does, but differences between groups cannot be taken into account in its covariance form. A comparison of the SCA-P and SCA-T3 solutions yields that although the SCA-P solution features roughly diagonal $\mathbf{\Phi}_k$, determining the strength of item components in each group k is problematic when considering either component variance or explained variance due to each component. In the SCA-T3 solution the explained variance due to each term adds up to total explained variance. Moreover, the strength of item components for each group is broken down into two levels: the group components and the explained variance of item components for the group components. This offers a clearer picture of group differences than the SCA-P solution. In sum, we conclude that SCA-T3 is a valuable addition to the existing SCA models.

The SCA-ECP and SCA-P models also have clusterwise variants in which each $\mathbf{X}_k$ is assigned to a cluster c with its own loading matrix $\mathbf{B}_c$. In clusterwise SCA-ECP the model is $\mathbf{X}_k \approx \mathbf{F}_k \mathbf{B}_c^T$ with $N_k^{-1}\mathbf{F}_k^T\mathbf{F}_k = \mathbf{\Phi}$, when k is assigned to cluster c (De Roover, Ceulemans, Timmerman, Vansteelandt, Stouten and Onghena, 2012; De Roover, Ceulemans, Timmerman, Nezlek and Onghena, 2013). Clusterwise SCA-P is the same model without restrictions on the component covariance matrices $N_k^{-1}\mathbf{F}_k^T\mathbf{F}_k$ (De Roover, Ceulemans, Timmerman and Onghena, 2013). Formally, these clusterwise SCA models are special cases of SCA-T3. For example, when there are two clusters we set $\mathbf{B} = [\mathbf{B}_1\ \mathbf{B}_2]$, $R = 2$, $\mathbf{G}_1 = [\mathbf{G}_{11}\ \mathbf{O}]$, $\mathbf{G}_2 = [\mathbf{O}\ \mathbf{G}_{22}]$, and $c_{kr} = 0$ if k is not in cluster r. This yields the model $\mathbf{X}_k \approx \mathbf{A}_k\,(c_{kc}\,\mathbf{G}_{cc})\,\mathbf{B}_c^T$ when k is in cluster c. In applications SCA-T3 could perhaps be used to check whether clusterwise SCA would be an appropriate model. For example, in the SCA-T3 solution for the ability tests dataset we observed a clear difference in component structure between the low and high IQ groups, resulting in group weights matrix $\mathbf{C}$ with roughly the form as above. When the low IQ groups would have large explained variances only for B1 and B2 and the high IQ groups only for B3 and B4, this would be an indication that a clusterwise SCA model may be useful. However, more research is needed to study the usefulness of SCA-T3 in such cases.

The SCA-PF2 and SCA-IND models have been extended to multi-set factor models by adding a diagonal matrix of unique variances to the corresponding SCA model for observed covariance matrices (Stegeman and Lam, 2016). The unique variances are estimated via minimum rank factor analysis (Ten Berge and Kiers, 1991), after which they are subtracted from the observed covariance matrices and the SCA model is fitted to the result. Obviously, this extension is also possible for SCA-T3 and would yield a multi-set factor model in which the model for the common parts can be rotated to obtain an interpretable solution. As such, these multi-set factor models are an alternative to the maximum likelihood approach introduced by Jöreskog (1971).

Appendix: proof of Lemma 3.1

We start by proving equality (9) for SCA-P fitted via PCA as described at the beginning of section 3.2. From the latter it can be verified that $\mathbf{F}_k = \mathbf{X}_k \mathbf{B} (\mathbf{B}^T\mathbf{B})^{-1}$. This implies that

$$\begin{aligned} \|\mathbf{X}_k - \mathbf{F}_k \mathbf{B}^T\|^2 &= \text{trace}(\mathbf{X}_k^T\mathbf{X}_k) - \text{trace}(\mathbf{B} (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X}_k^T\mathbf{X}_k) \\ &= \text{trace}(\mathbf{X}_k^T\mathbf{X}_k) - \text{trace}(\mathbf{B} (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X}_k^T\mathbf{X}_k\mathbf{B} (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T) \\ &= \|\mathbf{X}_k\|^2 - \|\mathbf{F}_k \mathbf{B}^T\|^2 , \end{aligned}$$

which is equivalent to (9).

Next, we consider SCA-PF2. It suffices to show that $\text{trace}(\mathbf{X}_k^T\mathbf{F}_k \mathbf{B}^T) = \text{trace}(\mathbf{B} \mathbf{F}_k^T\mathbf{F}_k \mathbf{B}^T)$. In SCA-PF2 we can set $\mathbf{F}_k = \widetilde{\mathbf{A}}_k \mathbf{H} \widetilde{\mathbf{C}}_k$, with $\widetilde{\mathbf{A}}_k^T\widetilde{\mathbf{A}}_k = \mathbf{I}_R$, $\mathbf{H}^T\mathbf{H} = \widetilde{\mathbf{\Phi}}$, and $\widetilde{\mathbf{C}}_k$ diagonal $R \times R$ (Kiers et al., 1999; Timmerman and Kiers, 2003). In step 2 of the ALS algorithm for SCA-PF2 the objective function is minimized over $\mathbf{H}$, $\widetilde{\mathbf{C}}_k$, and $\mathbf{B}$ for fixed $\widetilde{\mathbf{A}}_k$. Analogous to section 3.1, this boils down to fitting Parafac to the $R \times J \times K$ array with slices $\widetilde{\mathbf{A}}_k^T\mathbf{X}_k$, $k = 1, \ldots, K$. One iteration of the Parafac ALS algorithm is used to approximate $\widetilde{\mathbf{A}}_k^T\mathbf{X}_k \approx \mathbf{H} \widetilde{\mathbf{C}}_k \mathbf{B}^T$. Vectorizing on both sides yields $\text{Vec}(\widetilde{\mathbf{A}}_k^T\mathbf{X}_k) \approx (\mathbf{B} \odot \mathbf{H}) \tilde{\mathbf{c}}_k^{(\text{row})}$, with $\tilde{\mathbf{c}}_k^{(\text{row})}$ denoting the kth row of $\widetilde{\mathbf{C}}$ as a column vector. The OLS regression update of $\tilde{\mathbf{c}}_k^{(\text{row})}$ for fixed $\mathbf{H}$ and $\mathbf{B}$ implies that

$$\text{trace}(\text{Vec}(\widetilde{\mathbf{A}}_k^T\mathbf{X}_k)^T(\mathbf{B} \odot \mathbf{H}) \tilde{\mathbf{c}}_k^{(\text{row})}) = \text{trace}((\tilde{\mathbf{c}}_k^{(\text{row})})^T(\mathbf{B} \odot \mathbf{H})^T(\mathbf{B} \odot \mathbf{H}) \tilde{\mathbf{c}}_k^{(\text{row})}) ,$$

which is equivalent to $\text{trace}(\mathbf{X}_k^T\widetilde{\mathbf{A}}_k \mathbf{H} \widetilde{\mathbf{C}}_k \mathbf{B}^T) = \|\widetilde{\mathbf{A}}_k \mathbf{H} \widetilde{\mathbf{C}}_k \mathbf{B}^T\|^2$. This is the desired result.

The proof for SCA-IND follows from the above by setting $\widetilde{\mathbf{\Phi}} = \mathbf{I}_R$ (Timmerman and Kiers, 2003).

Finally, we consider SCA-T3 fitted by the ALS algorithm in section 3.1. The proof is analogous to the proof for SCA-PF2. We write $\mathbf{F}_k = \widetilde{\mathbf{A}}_k \left(\sum_r \tilde{c}_{kr} \mathbf{G}_r\right)$, with $\widetilde{\mathbf{A}}_k^T \widetilde{\mathbf{A}}_k = \mathbf{I}_P$. In step 2 of the ALS agorithm in section 3.1 we fit Tucker2 to the $P \times J \times K$ array with slices $\widetilde{\mathbf{A}}_k^T \mathbf{X}_k$, $k = 1, \ldots, K$. One iteration of the Tucker2 ALS algorithm is used to approximate $\widetilde{\mathbf{A}}_k^T \mathbf{X}_k \approx \left(\sum_r \tilde{c}_{kr} \mathbf{G}_r\right) \mathbf{B}^T$. Vectorizing on both sides yields $\text{Vec}(\widetilde{\mathbf{A}}_k^T \mathbf{X}_k) \approx (\mathbf{B} \otimes \mathbf{I}_P) \left[\text{Vec}(\mathbf{G}_1) \ \ldots \ \text{Vec}(\mathbf{G}_R)\right] \tilde{\mathbf{c}}_k^{(\text{row})}$. As above, the OLS regression update of $\tilde{\mathbf{c}}_k^{(\text{row})}$ for fixed $\mathbf{B}$ and $\mathcal{G}$ yields the result. This completes the proof.

Generally, the equality in (9) follows from an OLS regression update specifically for sample k. For SCA-ECP there is no such update and the equality does not hold. Indeed, fitting SCA-ECP to the PANAS dataset in section 5.1 yields inequality in (9).

References

Acar, E., & Yener, B. (2009). Unsupervised multiway data analysis: a literature survey. *IEEE Transactions on Knowledge and Data Engineering*, *21*, 1–15.

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 Journal of Mathematical and Statistical Psychology*, *59*, 133–150.

Comon, P., & De Lathauwer, L. (2010). Algebraic identification of under-determined mixtures, pp.325–366 in: *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, P. Comon and C. Jutten (Eds.), Academic Press.

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

De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000b). On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications*, *21*, 1324–1342.

De Lathauwer, L. (2010). Algebraic methods after prewhitening, pp.155–178 in: *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, P. Comon and C. Jutten (Eds.), Academic Press.

De Roover, K., Ceulemans, E., Timmerman, M.E., Vansteelandt, K., Stouten, J., & Onghena, P. (2012). Clusterwise simultaneous component analysis for analyzing structural differences in multivariate multiblock data. *Psychological Methods*, *17*, 100–119.

De Roover, K., Ceulemans, E., Timmerman, M.E., Nezlek, J.B., & Onghena, P. (2013). Modeling differences in the dimensionality of multiblock data by means of clusterwise simultaneous component analysis. *Psychometrika*, *78*, 648–668.

De Roover, K., Ceulemans, E., Timmerman, M.E., & Onghena, P. (2013). A clusterwise simultaneous component method for capturing within-cluster differences in component variances and correlations. *British Journal of Mathematical and Statistical Psychology*, *66*, 81–102.

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.

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

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, *1*, 211–218.

Golub, G.H., & Van Loan, C.F. (1996). *Matrix Computations, third edition.* Baltimore: John Hopkins University Press.

Green, B.F. (1952). The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika*, *17*, 429–440.

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

Harshman, R.A. (1972). Parafac2: Mathematical and technical notes. *UCLA Working Papers in Phonetics*, *22*, 30–44.

Helwig, N.E. (2013). The special sign indeterminacy of the direct-fitting Parafac2 model: Some implications, cautions, and recommendations for simultaneous component analysis. *Psychometrika*, *78*, 725–739.

Ishteva, M., Absil, P.-A., Van Huffel, S., & De Lathauwer, L. (2011). Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme. *SIAM Journal on Matrix Analysis and Applications*, *32*, 115–135.

Jöreskog, K.G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409–426.

Kiers, H.A.L. (1993). An alternating least squares algorithm for PARAFAC2 and three-way Dedicom. *Computational Statistics and Data Analysis*, *16*, 103–118.

Kiers, H.A.L., & Ten Berge, J.M.F. (1994). Hierarchical relations between methods for simultaneous component analysis and a technique for rotation to a simple simultaneous structure. *British Journal of Mathematical and Statistical Psychology*, *47*, 109–126.

Kiers, H.A.L. (1998a). Three-way SIMPLIMAX for oblique rotation of the three-mode factor analysis core to simple structure. *Computational Statistics and Data Analysis*, *28*, 307–324.

Kiers, H.A.L. (1998b). Joint orthomax rotation of the core and component matrices resulting from three-mode principal components analysis. *Journal of Classification*, *15*, 245–263.

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

Kiers, H.A.L., Ten Berge, J.M.F., & Bro, R. (1999). Parafac2–Part I. A direct fitting algorithm for the Parafac2 model. *Journal of Chemometrics*, *13*, 275–294.

Kiers, H.A.L. (2004). Bootstrap confidence intervals for three-way methods. *Journal of Chemometrics*, *18*, 22–36.

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

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.

Kroonenberg, P.M. (2008). *Applied Multiway Data Analysis*, Wiley Series in Probability and Statistics.

Lam, T.T.T. (2015). *Some new methods for three-mode factor analysis and multi-set factor analysis*. PhD Thesis. University of Groningen, The Netherlands.

Louwerse, D.J., Smilde, A.K., & Kiers, H.A.L. (1999). Cross-validation of multiway component models. *Journal of Chemometrics*, *13*, 491–510.

McGaw, B., & Jöreskog, K.G. (1971). Factorial invariance of ability measures in groups differing in intelligence and socioeconomic status. *British Journal of Mathematical and Statistical Psychology*, *24*, 154–168.

Penrose, R. (1956). On the best approximate solutions of linear matrix equations. *Mathematical Proceedings of the Cambridge Philosophical Society*, *52*, 17–19.

Rocci, R. (1992). Three-mode factor analysis with binary core and orthonormality constraints. *Journal of the Italian Statistical Society*, *1*, 413–422.

Savas, B., & Lim, L.-H. (2010). Quasi-Newton methods on Grassmannians and multilinear approximation of tensors. *SIAM Journal on Scientific Computing*, *32*, 3352–3393.

Shifren, K., Hooker, K, Wood, P., & Nesselroade, J.R. (1997). Structure and variation of mood in individuals with Parkinson's disease: a dynamic fator analysis. *Psychology and aging*, *12*, 328–339.

Smilde, A., Bro, R., & Geladi, P. (2004). *Multi-way Analysis: Applications in the Chemical*

Sciences. Chichester: Wiley.

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

Stegeman, A., & Lam, T.T.T. (2014). Three-mode factor analysis by means of Candecomp/Parafac. *Psychometrika*, *79*, 426–443.

Stegeman, A., & Lam, T.T.T. (2016). Multi-set factor analysis by means of Parafac2. *British Journal of Mathematical and Statistical Psychology*, *69*, 1–19.

Ten Berge, J.M.F., & Kiers, H.A.L. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika*, *56*, 309–315.

Ten Berge, J.M.F., & Kiers, H.A.L. (1996). Some uniqueness results for Parafac2. *Psychometrika*, *61*, 123–132.

Ten Berge, J.M.F., & Smilde, A.K. (2002). Non-triviality and identification of a constrained Tucker3 analysis. *Journal of Chemometrics*, *16*, 609–612.

Timmerman, M., & Kiers, H.A.L. (2003). Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences . *Psychometrika*, *68*, 105–121.

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

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

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.

Table 1: Formal definitions of five SCA models.

SCA model	model for $\mathbf{X}_k$	model for $\mathrm{Cov}(\mathbf{X}_k)$	$\mathrm{Cov}(\mathbf{A}_k)$
SCA-P	$\mathbf{A}_k\,\mathbf{B}^T$	$\mathbf{B}\,\mathbf{\Phi}_k\,\mathbf{B}^T$	$\mathbf{\Phi}_k$
SCA-PF2	$\mathbf{A}_k\,\mathbf{C}_k\,\mathbf{B}^T$	$\mathbf{B}\,\mathbf{C}_k\,\mathbf{\Phi}\,\mathbf{C}_k\,\mathbf{B}^T$	$\mathbf{\Phi}$
SCA-IND	$\mathbf{A}_k\,\mathbf{C}_k\,\mathbf{B}^T$	$\mathbf{B}\,\mathbf{C}_k^2\,\mathbf{B}^T$	$\mathbf{I}_R$
SCA-ECP	$\mathbf{A}_k\,\mathbf{B}^T$	$\mathbf{B}\,\mathbf{\Phi}\,\mathbf{B}^T$	$\mathbf{\Phi}$
SCA-T3	$\mathbf{A}_k\left(\sum_{r=1}^R c_{kr}\,\mathbf{G}_r\right)\mathbf{B}^T$	$\mathbf{B}\left(\sum_{r=1}^R c_{kr}\,\mathbf{G}_r^T\right)\mathbf{\Phi}\left(\sum_{r=1}^R c_{kr}\,\mathbf{G}_r\right)\mathbf{B}^T$	$\mathbf{\Phi}$

Table 2: Number of runs and number of datasets with fit percentage less than 99.99, 99.9, and 99.0 percent for 100 generated SCA-T3 noisefree datasets. The SCA-T3 model is either fitted on the observed scores or on the observed covariance matrices.

K	core	data	m	best of m runs	best of $m/2$ runs	rational start
$K=5$	(14)	scores	20	2, 1, 0	11, 10, 0	71, 60, 10
$K=5$	(15)	scores	1	0, 0, 0	-	0, 0, 0
$K=2$	(15)	scores	1	0, 0, 0	-	0, 0, 0
$K=12$	(16)	scores	350	5, 0, 0	19, 3, 0	96, 87, 8
$K=5$	(14)	cov	30	1, 0, 0	14, 6, 1	77, 55, 9
$K=5$	(15)	cov	1	0, 0, 0	-	0, 0, 0
$K=2$	(15)	cov	1	0, 0, 0	-	0, 0, 0
$K=12$	(16)	cov	340	4, 0, 0	18, 1, 0	94, 85, 6

Table 3: Means and standard deviations of MAD and RMSE values for SCA-T3 estimates of $\mathbf{B}$, $\widetilde{\mathbf{C}}$, and $\mathcal{G}$ when generating 100 noisy datasets. The SCA-T3 model is either fitted on the observed scores or on the observed covariance matrices. For observed scores the noise strength s_k is given.

K	core	data	noise	MAD($\mathbf{B}$)	RMSE($\mathbf{B}$)	MAD($\widetilde{\mathbf{C}}$)	RMSE($\widetilde{\mathbf{C}}$)	MAD($\mathcal{G}$)	RMSE($\mathcal{G}$)
$K = 5$	(14)	scores	$s_k = 0.5$	0.02	0.02	0.02	0.02	0.05	0.09
				(0.00)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
$K = 5$	(14)	scores	$s_k = 0.7$	0.02	0.03	0.05	0.08	0.10	0.15
				(0.01)	(0.01)	(0.03)	(0.04)	(0.03)	(0.03)
$K = 5$	(14)	scores	unequal	0.01	0.02	0.05	0.06	0.05	0.08
				(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$K = 5$	(15)	scores	$s_k = 0.5$	0.02	0.03	0.00	0.01	0.03	0.05
				(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
$K = 5$	(15)	scores	$s_k = 0.7$	0.03	0.04	0.00	0.00	0.06	0.09
				(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
$K = 5$	(15)	scores	unequal	0.03	0.04	0.00	0.00	0.03	0.04
				(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
$K = 2$	(15)	scores	$s_k = 0.5$	0.02	0.03	0.01	0.01	0.03	0.05
				(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$K = 2$	(15)	scores	$s_k = 0.7$	0.04	0.05	0.00	0.00	0.06	0.09
				(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
$K = 2$	(15)	scores	unequal	0.03	0.04	0.01	0.01	0.04	0.06
				(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$K = 12$	(16)	scores	$s_k = 0.5$	0.01	0.01	0.01	0.02	0.05	0.06
				(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)
$K = 12$	(16)	scores	$s_k = 0.7$	0.01	0.02	0.02	0.02	0.07	0.09
				(0.00)	(0.00)	(0.00)	(0.01)	(0.02)	(0.03)
$K = 12$	(16)	scores	unequal	0.01	0.01	0.01	0.01	0.04	0.05
				(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
$K = 5$	(14)	cov	-	0.01	0.01	0.02	0.02	0.03	0.04
				(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
$K = 5$	(15)	cov	-	0.00	0.01	0.02	0.02	0.02	0.03
				(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
$K = 2$	(15)	cov	-	0.01	0.01	0.03	0.03	0.03	0.04
				(0.00)	(0.00)	(0.02)	(0.02)	(0.01)	(0.01)
$K = 12$	(16)	cov	-	0.00	0.01	0.02	0.02	0.03	0.04
				(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)

Table 4: PRESS and SHS values and fit of SCA-T3 models with PRESS $\leq$ 12000 and SHS2 $\leq$ 0.10 applied to the PANAS data for subjects with Parkinson's disease.

(P, Q, R)	PRESS	SHS1	SHS2	Fit %
(2,2,2)	11814	0.11	0.09	43.2
(3,2,2)	11686	0.11	0.09	43.4
(4,2,2)	11671	0.12	0.09	43.4
(3,2,3)	11676	0.09	0.08	43.5
(4,2,3)	11664	0.09	0.08	43.5
(3,2,4)	11670	0.09	0.08	43.5
(4,2,4)	11699	0.09	0.08	43.5
(2,2,3)	11666	0.10	0.08	43.5
(2,2,4)	11670	0.09	0.08	43.5
(4,4,3)	11243	0.13	0.09	57.1
(4,4,4)	11016	0.15	0.10	58.2

Table 5: Loading matrix **B** of the rotated solution for SCA-T3 with $(P, Q, R) = (4, 4, 3)$ fitted to the PANAS data for subjects with Parkinson's disease.

item	B1	B2	B3	B4
jittery	**0.76**	-0.14	0.13	0.17
distressed	0.23	0.18	0.07	**0.66**
upset	0.26	0.06	-0.03	**0.68**
afraid	**0.68**	0.07	**0.54**	0.02
scared	**0.69**	0.07	**0.57**	-0.02
hostile	**0.62**	-0.16	-0.03	0.22
irritable	0.18	0.04	-0.13	**0.68**
guilty	0.08	-0.34	-0.34	**0.55**
ashamed	**0.56**	**-0.48**	0.12	0.11
nervous	**0.72**	-0.11	0.21	0.21
inspired	**-0.66**	-0.22	-0.08	0.36
excited	-0.05	**-0.70**	0.02	-0.18
determined	-0.28	0.00	**0.56**	0.03
interested	**-0.75**	-0.02	0.21	0.29
enthusiastic	**-0.71**	-0.13	0.16	0.23
attentive	**-0.68**	-0.14	0.26	0.26
proud	-0.05	**-0.53**	-0.14	-0.24
strong	-0.13	-0.26	**0.58**	-0.13
active	**-0.64**	0.09	0.39	0.08
alert	**-0.73**	-0.04	0.21	0.20

Note: absolute values larger than 0.40 are in boldfont.

Table 6: Subjects weight matrix $\mathbf{C}$ of the rotated solution for SCA-T3 with $(P, Q, R) = (4, 4, 3)$ fitted to the PANAS data for subjects with Parkinson's disease. Also given are the fit percentages (9) and the sum of the variances of the centered and scaled data $\mathbf{X}_k$.

subject	C1	C2	C3	Fit %	trace($N_k^{-1}\mathbf{X}_k^T\mathbf{X}_k$)
1	0.10	0.49	0.93	67.1	13.1
2	-0.46	1.91	0.30	66.8	38.7
3	1.09	0.20	-0.41	48.4	14.3
4	1.64	-0.47	0.88	61.0	34.5
5	-0.39	-0.55	2.39	85.7	42.3
6	0.46	0.69	0.68	57.9	15.1
7	1.04	0.63	0.21	45.0	19.1
8	0.84	0.01	-0.13	32.7	11.3
9	1.44	-0.35	-0.50	42.1	32.1
10	0.62	0.20	0.02	24.7	8.9
11	0.53	0.22	0.21	30.9	7.3
12	0.65	0.36	0.19	38.9	8.9

Table 7: Core entries and explained variances (in round brackets) of the rotated solution for SCA-T3 with $(P, Q, R) = (4, 4, 3)$ fitted to the PANAS data for subjects with Parkinson's disease.

	B1	B2	B3	B4	
$A1_k$	**0.51 (6.1)**	0.06 (0.0)	-0.03 (0.0)	0.06 (0.0)	C1
$A2_k$	0.03 (0.0)	-0.00 (0.0)	0.11 (0.1)	**0.81 (6.0)**	
$A3_k$	0.06 (0.1)	**1.00 (5.5)**	-0.03 (0.0)	-0.16 (0.2)	
$A4_k$	-0.02 (0.0)	0.07 (0.0)	0.62 (2.8)	-0.01 (0.0)	
$A1_k$	-0.08 (0.1)	-0.21 (0.1)	-0.11 (0.0)	-0.09 (0.0)	C2
$A2_k$	-0.39 (1.8)	0.04 (0.0)	1.00 (3.8)	-0.12 (0.1)	
$A3_k$	0.15 (0.3)	0.09 (0.0)	-0.76 (2.2)	0.82 (3.1)	
$A4_k$	-0.11 (0.2)	0.43 (0.5)	0.08 (0.0)	-0.05 (0.0)	
$A1_k$	0.01 (0.0)	-0.30 (0.4)	0.16 (0.2)	-0.45 (1.7)	C3
$A2_k$	-0.03 (0.0)	-0.20 (0.2)	0.15 (0.1)	0.09 (0.1)	
$A3_k$	**1.00 (20.9)**	-0.00 (0.0)	-0.02 (0.0)	0.05 (0.0)	
$A4_k$	-0.00 (0.0)	0.10 (0.0)	-0.02 (0.0)	-0.18 (0.3)	

Note: entries with explained variance larger than 4 percent are in boldfont.

Table 8: Loading matrix **B** of the rotated solution for SCA-T3 with $(P, Q, R) = (4, 4, 2)$ fitted to ability tests for high school children.

test	B1	B2	B3	B4
Vocabulary	-0.12	-0.05	**0.98**	-0.21
Information I	-0.03	-0.07	**1.01**	-0.13
Information II	-0.07	-0.02	**0.93**	-0.12
Spelling	-0.21	0.04	0.24	**0.63**
Punctuation	0.10	-0.06	0.23	**0.69**
English Usage	0.08	-0.08	0.25	**0.61**
Mechanical Reasoning	**0.72**	-0.09	0.28	-0.07
Visualization I	**0.67**	0.21	-0.00	0.08
Visualization II	**0.80**	-0.02	0.05	0.02
Table Reading	-0.11	**0.77**	0.12	-0.04
Clerical Checking	-0.11	**0.67**	0.07	0.22
Object Inspection	0.14	**0.77**	0.07	-0.13

Note: absolute values larger than 0.40 are in boldfont.

Table 9: Group weight matrix **C** of the rotated solution for SCA-T3 with $(P, Q, R) = (4, 4, 2)$ fitted to ability tests for high school children. Also given are the fit percentages (9) and the sum of the variances for each group.

group	C1	C2	Fit %	trace($\mathbf{S}_k$)
low IQ, low SES	1.16	0.25	61.7	13.4
low IQ, high SES	1.30	-0.06	61.0	16.9
high IQ, low SES	-0.03	0.97	67.9	9.9
high IQ, high SES	-0.19	0.93	65.3	9.9

Table 10: Core entries and explained variances (in round brackets) of the rotated solution for SCA-T3 with $(P, Q, R) = (4, 4, 2)$ fitted to ability tests for high school children.

	B1	B2	B3	B4	
$A1_k$	0.13 (0.2)	0.10 (0.1)	**0.83 (13.1)**	0.16 (0.2)	C1
$A2_k$	0.00 (0.0)	-0.02 (0.0)	-0.06 (0.1)	**-0.82 (5.7)**	
$A3_k$	0.09 (0.1)	**1.00 (10.5)**	-0.04 (0.0)	0.04 (0.0)	
$A4_k$	**0.80 (6.8)**	0.05 (0.0)	0.07 (0.1)	-0.00 (0.0)	
$A1_k$	-0.17 (0.2)	-0.54 (1.8)	-0.15 (0.3)	0.38 (0.8)	C2
$A2_k$	**1.00 (6.4)**	0.11 (0.1)	0.04 (0.0)	-0.02 (0.0)	
$A3_k$	0.11 (0.1)	-0.01 (0.0)	**0.98 (11.2)**	0.43 (1.0)	
$A4_k$	0.00 (0.0)	0.42 (1.1)	0.05 (0.0)	0.82 (3.5)	

Note: entries with explained variance larger than 4 percent are in boldfont.

Table 11: Component variances and explained variance (in round brackets) due to each component in each group, if it were the only component, for the SCA-P solution fitted to ability tests for high school children.

group	B1	B2	B3	B4
low IQ, low SES	0.95 (12.9)	1.34 (17.3)	0.83 (16.7)	1.05 (13.2)
low IQ, high SES	1.12 (12.2)	1.76 (18.0)	1.32 (21.1)	1.19 (11.9)
high IQ, low SES	0.98 (18.2)	0.44 (7.7)	0.97 (26.5)	0.86 (14.7)
high IQ, high SES	0.94 (17.4)	0.46 (8.0)	0.88 (24.1)	0.90 (15.3)

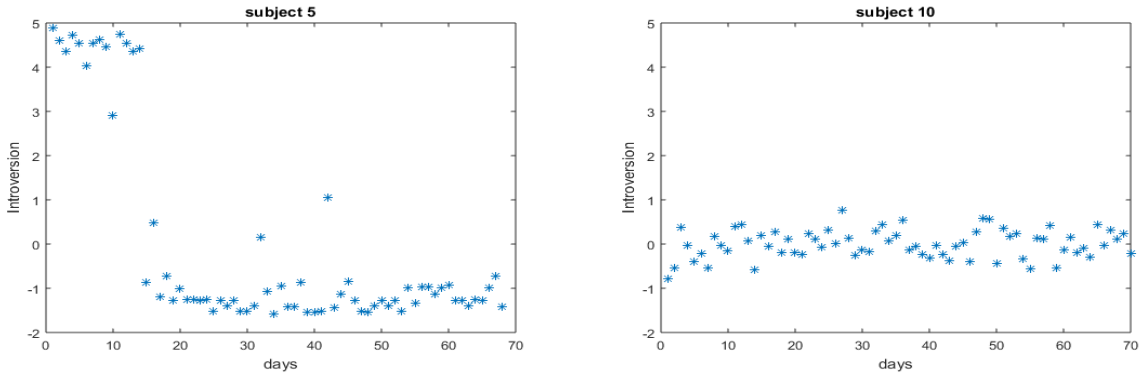

Figure 1: Time courses corresponding to the positive-negative affect component for subjects 5 and 10 for SCA-T3 with $(P, Q, R) = (4, 4, 3)$ fitted to the PANAS data for subjects with Parkinson's disease.