Simultaneous component analysis by means of Tucker3
Short user guide to the Matlab codes

November 29, 2016

In the manuscript a model for simultaneous component analysis (SCA) based on Tucker3 is proposed. It can be used for component analysis in multi-set data, i.e., observations of the same variables in several different samples possibly of unequal size. 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. The SCA-T3 model is a new member of the family of SCA models in Timmerman and Kiers (2003). SCA-T3 approximates $\mathbf{X}_k$ by $\mathbf{A}_k \left(\sum_{r=1}^{R} c_{kr} \, \mathbf{G}_r \right) \mathbf{B}^T$ by minimizing $\sum_k \|\mathbf{X}_k - \mathbf{A}_k \left(\sum_{r=1}^{R} c_{kr} \, \mathbf{G}_r \right) \mathbf{B}^T\|^2$. For this purpose an alternating least squares (ALS) algorithm is derived in the manuscript. The notation is as follows: $\mathbf{A}_k$ $(N_k \times P)$ contains the centered component scores such that $N_k^{-1} \mathbf{A}_k^T \mathbf{A}_k = \mathbf{I}_P$, $\mathbf{B}$ $(J \times Q)$ is the common loading matrix, $\mathbf{C}$ $(K \times R)$ contains weights c_{kr} for the samples, and $P \times Q \times R$ core array $\mathcal{G}$ with $P \times Q$ slices $\mathbf{G}_r$ contains the weights of each triplet of components. Although SCA-T3 seems complicated at first, rotations may be used to obtain simple structure in the core $\mathcal{G}$, loading matrix $\mathbf{B}$, and weights matrix $\mathbf{C}$. This greatly facilitates interpretation. In the Matlab code the orthogonal rotation procedure of Kiers (1998) is used for this purpose. Next, we show how to use the Matlab code for fitting SCA-T3 to multi-set data.

The centered observed scores $\mathbf{X}_k$ are stacked into the matrix $\mathbf{X} = [\mathbf{X}_1^T \; \ldots \; \mathbf{X}_K^T]^T$. The sample sizes are stacked into the vector $\mathbf{N} = (N_1 \; \ldots \; N_K)^T$. After specifying the number of variables J, the number of samples K, and the numbers of components (P, Q, R), the following Matlab command fits SCA-T3 to the dataset:

$$[\texttt{A}, \texttt{B}, \texttt{C}, \texttt{G}, \texttt{EV}, \texttt{fp}, \texttt{fptot}] = \texttt{SCA\_T3}(\texttt{X}, \texttt{N}, \texttt{J}, \texttt{K}, \texttt{P}, \texttt{Q}, \texttt{R}, \texttt{conv}, \texttt{runs}, \texttt{rotation}, \texttt{direct})\,.$$

The ALS algorithm for SCA-T3 is stopped when the relative decrease of the objective function drops below `conv` (e.g., `conv=1e-7`). The ALS algorithm is started multiple times with the number of runs defined by `runs=[run1 run2]`. The solution for the run with the smallest value of the objective function is kept. The number of runs with random starting values is equal to `run1` and one run with rationally chosen starting values is added when `run2=1`.

The options for rotation to simple structure are specified by the vector `rotation=[opt1 opt2 opt3 opt4]`, with `opt1=0` for no rotation, `opt1=1` for rotation to simple core $\mathcal{G}$ and loading matrix $\mathbf{B}$, `opt1=2` for rotation to simple $\mathcal{G}$, $\mathbf{B}$, and $\mathbf{C}$, and `opt1=3` for rotation to simple $\mathcal{G}$ only. The parameters `opt2, opt3, opt4` indicate whether or not to rotate in the first, second, and third mode (yes=1, no=0). Finally, for fitting SCA-T3 to observed scores one must set `direct=1`.

The output of `SCA_T3.m` is as follows. Estimated component scores $\mathbf{A}_k$ are stacked in the matrix $\mathbf{A} = [\mathbf{A}_1^T \ \ldots \ \mathbf{A}_K^T]^T$. The estimated loading matrix and weights matrix are available as $\mathbf{B}$ and $\mathbf{C}$, respectively. The estimated core array is given as matrix $\mathbf{G} = [\mathbf{G}_1 \ \ldots \ \mathbf{G}_R]$. The explained variance due to each term corresponding to a core entry can be found in the matrix `EV`. The vector `fp` contains the fit percentage for each sample k, and `fptot` is the total fit percentage. The explained variances in `EV` add up to `fptot`. For more details and considerations of fitting SCA-T3 we refer to the manuscript.

SCA-T3 can also be fitted to observed covariance matrices $\mathrm{Cov}(\mathbf{X}_k)$ when the observed scores $\mathbf{X}_k$ themselves are not available. SCA-T3 in covariance form reads as $\mathrm{Cov}(\mathbf{X}_k) \approx \mathbf{B}\left(\sum_{r=1}^R c_{kr}\,\mathbf{G}_r^T\right)\left(\sum_{r=1}^R c_{kr}\,\mathbf{G}_r\right)\mathbf{B}^T$. In the manuscript the following heuristic procedure is used to fit this model. First, eigendecompositions are used to obtain $\mathbf{Y}_k$ such that $\mathrm{Cov}(\mathbf{X}_k) = \mathbf{Y}_k^T\mathbf{Y}_k$, $k = 1,\ldots,K$. Second, SCA-T3 is fitted as $\mathbf{Y}_k \approx \widetilde{\mathbf{A}}_k\left(\sum_{r=1}^R c_{kr}\,\mathbf{G}_r\right)\mathbf{B}^T$ with $\widetilde{\mathbf{A}}_k^T\widetilde{\mathbf{A}}_k = \mathbf{I}_P$. The following Matlab command fits SCA-T3 to observed covariance matrices:

$$[\mathtt{B}, \mathtt{C}, \mathtt{G}, \mathtt{EV}, \mathtt{fp}, \mathtt{fptot}] = \mathtt{SCA\_T3\_cov}(\mathtt{Sigma}, \mathtt{J}, \mathtt{K}, \mathtt{P}, \mathtt{Q}, \mathtt{R}, \mathtt{conv}, \mathtt{runs}, \mathtt{rotation})\,.$$

Here, the observed covariance matrices are stacked in the matrix `Sigma`$= [\mathrm{Cov}(\mathbf{X}_1) \ \ldots \ \mathrm{Cov}(\mathbf{X}_K)]$. The other input parameters are the same as above. Again, for details we refer to the manuscript.

The Matlab code also includes the files `varimcoco.m` and `orthmax2.m` which are programmed by Henk Kiers, Department of Psychometrics and Statistics, University of Groningen, and perform the orthogonal rotation procedure. Also included is the file `Data_IQSES.m` that contains rescaled correlation matrices of the ability test dataset analyzed in the manuscript and previously analyzed by McGaw and Jöreskog (1971).

References

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

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.

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.