(1988)

In A. Di Ciaccio & G. Bove (Ed.)

Multiway '88. Software guide.

Roma.: Università di Roma. "La Sapienza"

CHAPTER 9

p. 93-103

p. 105-114

TUCKALS2

Three-Mode Principal Component Analysis with extended core matrix

Kroonenberg P.M.

0. Introduction.

TUCKALS2 is a program to perform a three-mode principal component analysis, in which components are computed over only two of the three modes, and in which the third mode retains its original order. The technique was developed by Tucker (1972) building on his earlier work (Tucker, 1966). Improved estimation procedures were devised by Kroonenberg & De Leeuw (1980). The technique has been fully described and illustrated by Kroonenberg (1983a), and an annotated bibliography is Kroonenberg (1983b).

Three-mode principal component analysis is a technique to deal with data which can be classified by three kinds of entities (called *modes*), say subjects, variables, and occasions. These terms should be considered generic , rather than specific ones. Three-mode data can be arranged into a three-dimensional block or array **X**. The three modes will be called A, B, and C, respectively (see Figure 1). The orders of **X** are I, J, and K (upper case), and i, j, and k (lower case) are the indices for the elements of the respective modes.

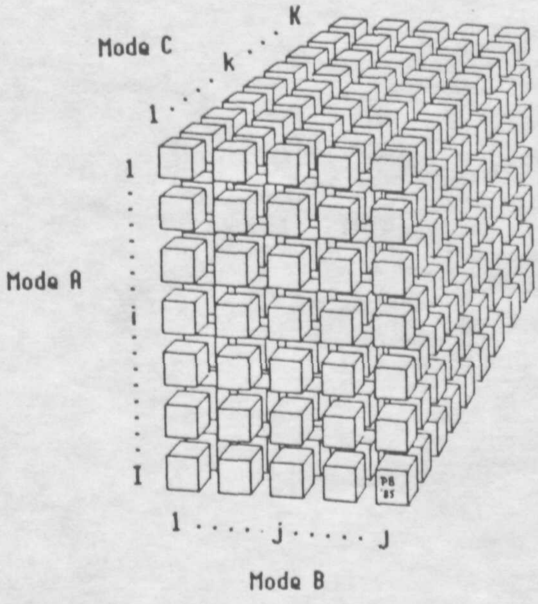

Figure 1 Data matrix **X**

A three-mode matrix can be seen as composed of two-mode submatrices alled *slices*, and of one-mode submatrices (or vectors), called *fibers*. These vo-way submatrices will be referred to as *frontal* slices, *horizontal* slices, and *iteral* slices (Figure 2). The fibers will be called *rows*, *columns*, and *tubes* Figure 3). Throughout this text $\mathbf{X}_k$ will denote the k-th of K frontal slice of $\mathbf{X}$.

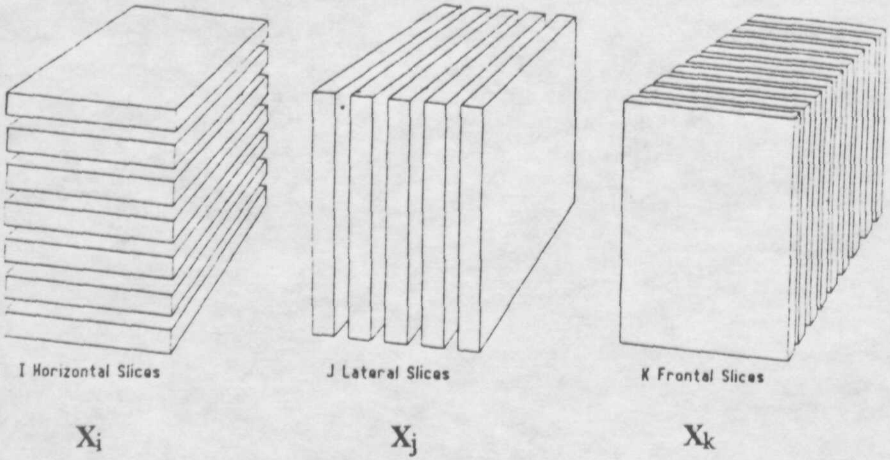

Figure 2 Slices, the two-way submatrices of **X**

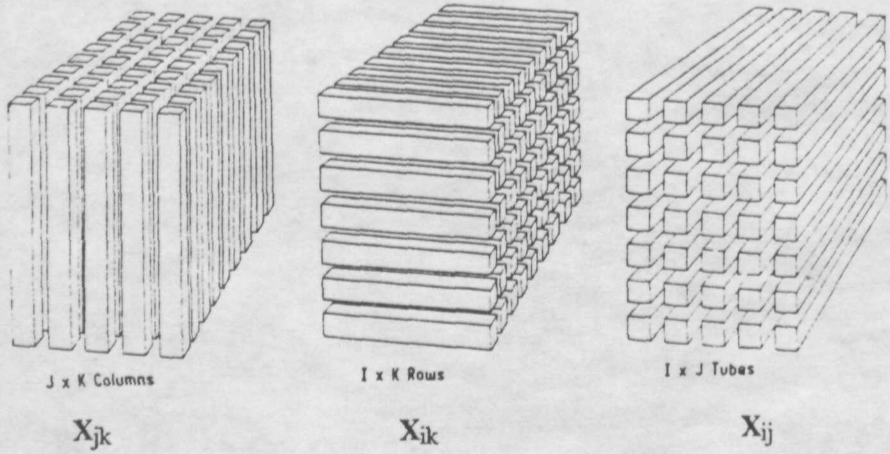

Figure 3 Fibers, the one-way submatrices of **X**

The matrices of component loadings are named after the modes they refer to, but as usual, the names of vectors and matrices are printed in bold face. Thus **A** is the *component matrix* for Mode A and so on. The *core matrix* is denoted by **H**. The terminology presented here is largely based on Harshman and Lundy (1984a,b). The only difference lies in the choice of A and B. Harshman and Lundy call Mode B what is called Mode A here, and vice versa.

1. Characteristics of input data.

TUCKALS2 is a three-mode program which is primarily geared towards metric three-way three-mode data, which are fully crossed with respect to all modes. There are no special provisions for conditionality, nor for missing data. The program may be used for three-way two-mode data, such as multiple covariance matrices or (double-centred) (dis)similarity matrices. In the latter case it is implicitly assumed that the dissimilarities are equal to squared distances rather than ordinary distances. If this is unacceptable, corrections should be made before the analysis proper. There are no specific provisions in the program for nonmetric data, such as optimal scaling or similar procedures for handling ordinal or nominal data.

2. Data manipulation.

Several centrings can be performed in the program, primarily on frontal slices of the three-way matrix, such as centring rows, columns or frontal slices, and standardization of frontal slices, but the program is not specifically geared towards comprehensive data manipulation. In practice, the centring options suffice for most data sets, especially as by transposing the data matrix all desired centrings can be performed. Centring on three modes at the same time is seldom necessary. However, for full data manipulation, the program NDISMIS3 by Brouwer (1985) or Harshman and Lundy's PARAFAC (q.v.) can be used; they contain most of the desired options for centring and standardization. The latter program includes an (iterative) standardization procedure for simultaneously standardizing two or three modes.

3. Mathematical models

The program handles the Tucker2 model, in which orthonormal components are computed for two of the three modes. The weights for combinations of components of the first two modes for each of the elements of the third mode are computed as well. They form together the core matrix **H** which has orders equal to the number of components of two of the modes times the size of the third mode, i.e. PxQxK.

The model is formally described as

$$x_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} a_{ip} b_{jq} h_{pqk} + e_{ijk} \, .$$

where $i=1,..,I$, $j=1,..,J$, and $k=1,..,K$; P and Q are the number of components for the first two modes, and $\mathbf{A} = (a_{ip})$ and $\mathbf{B} = (b_{jq})$ are the component matrices of the first and second mode respectively. $\mathbf{H} = (h_{pqk})$ is the PxQxK core matrix, and $\mathbf{E} = (e_{ijk})$ the three-mode matrix with errors of approximation. A matrix formulation of the model is

$$\mathbf{X}_k = \mathbf{A}\mathbf{H}_k\mathbf{B}' + \mathbf{E}_k, \quad k=1,\ldots,K$$

in which the $\mathbf{H}_k$ are the (unrestricted) *individual characteric matrices* .

When instead of direct fitting of the original data, indirect fitting is used for cross-product or covariance matrices, mostly **A** and **B** will become identical or sign permuted versions of each other, and the core matrix **H** will in general be symmetric with possibly sign inversions. The Tucker2 model is then identical to the IDIOSCAL model of Carroll & Chang (1970,1972). When three-mode data fitted directly, and the $\mathbf{H}_k$ are restricted to be diagonal, the model is an orthonormal version of PARAFAC (q.v.), and when the component matrices **A** and **B** are no longer required to be orthogonal then the model is equal to the basic PARAFAC model. Finally when in the above case the input frontal slices are symmetric generally the component matrices will be symmetric as well, and the model is equal to the INDSCAL model (q.v.).

The model is essentially non-stochastic and data-analytic. It suffers from rotational indeterminacy of the components, but this indeterminacy allows for a nonrestricted, easy to fit, model. The indeterminacy implies that after a solution is obtained the orthonormal solution may be transformed in several ways without loss of fit, if the appropriate inverse transformations are applied as well. The present program has no transformational capabilities for the component matrices, but transformed solutions can be reintroduced into the program to evaluate especially the core matrix after transformation, but also the effect of the transformations on the redistribution of variability over the components can be assessed. Incorporated in the program is , however, an orthonormal transformation procedure to diagonalize the core matrix as much as possible. In a new experimental version of the program, also a no-singular transformation procedure operating on the core matrix is included, which gives a PARAFAC solution if such a solution exists. If not it provides either an approximation to the PARAFAC solution, or it degenerates in a similar manner as PARAFAC does (for details see Harshman and Lundy, 1984a, and Brouwer and Kroonenberg, 1985).

4. Optimization algorithm

The estimation of the Tucker2 model is achieved via an alternating least squares algorithm which minimizes the loss function

$$\sum_{k=1}^{K} ||\mathbf{X}_k - \mathbf{A}\mathbf{H}_k\mathbf{B}'||^2$$

The minimization problem can be reduced by solving first for $\mathbf{H}$ as $\mathbf{H}^*_k=\mathbf{A}'\mathbf{X}_k\mathbf{B}$, and substituting $\mathbf{H}^*$ into the loss function to obtain

$$\sum_{k=1}^{K} ||\mathbf{X}_k - \mathbf{A}\mathbf{A}'\mathbf{X}_k\mathbf{B}\mathbf{B}'||^2$$

This last loss function can be solved via cyclically estimating **A** for fixed **B** , followed by **B** for fixed **A**, and then **A** for fixed **B** again, etc. Each subproblem is an eigenvalue-eigenvector problem of a dimension equal to the number of components for the mode in question, and it can be handled efficiently by using a Jacobi procedure embedded in Bauer-Rutishauser's simultaneous iteration method.

To start iterations, the solutions obtained via Tucker's Method I are used, which will already provide the solution if an exact solution exists. As in virtually all problems of this kind, only convergence to a local minimum is assured, however, the specific initial configuration has shown to steer the algorithm in the proper direction. The general impression is that local minima do not form a serious problem.

5. Results

The **primary output** of the program consists of the following parts

1. The information on the overall fit of the model, and several partitionings of this fit by the elements (i.e. variables, subjects, occasions) of each mode, as well as by the component combinations via the extended core matrix;

2. Components scaled in several ways;

3. Core matrix scaled in several ways;

Optional **supplementary information** includes

4. Input data;

5. (Optionally) removed means and scale factors, and scaled data;

6. Initial configurations;

7. Iteration history;

8. Residuals, fitted data , squared residuals;

9. Analysis of variance of squared residuals;

10. Joint plot of the first two modes, based on the average core slice;

11. Distances (inner products) of points in the joint plot;

12. Component scores for all first-third mode combinations on the components of the second mode;

13. Many plots can be produced to visually inspect the solutions;

14. Coordinates of components, joint plot, and component scores, core matrix, (squared) residuals, fitted data, and fits per element can be written to external units;

14. External configurations can be read in, to restart analyses, to evaluate results from other studies, to evaluate component spaces after transformation, to construct core matrices for PARAFAC components (as in PFCORE , q.v.).

6. Technical information.

The program was originally written in portable FORTRAN-IV, but was adapted to FORTRAN77. It is designed for main frames, and it runs satisfactorily on machines like the IBM8083, CDC, Fujitsu, and under UNIX on Perkin Elmer and MicroVAX, and other machines.

The program has an option for dynamic array allocation and accordingly its size depends on the variable array size. The program itself is approximately 300K, and the variable array size depends on I, J, K., P, and Q. If no dynamic array allocation is used, the standard array space is 120K, which can easily be enlarged by changing only a few statements. A problem of 160 by 12 by 8 with two components for each mode runs in 434K memory, and a 12 by 12 by 11 with five components for each mode in 326K memory.

The input is based on fixed column entry, and the program has an editing facility for checking the input parameters without execution. The echo of the input parameters is at the same time a complete input description.

It is contemplated to extend the program to provide options for producing output in accordance with other standard programs for three-mode analysis, such as PARAFAC and STATIS. Further possible developments consist of porting the program to microcomputers by rewriting it into C, including some transformational procedures on both the components and the core matrix, including the Weesie and Van Houwelingen algorithm to allow for missing data, and possibly extending the program to handle four modes.

The program is available from the author (P.M. Kroonenberg, Department of Education, University of Leiden, P.O. Box 9507, 2300 RA Leiden, The Netherlands), and the costs are US$150 . Further details can be obtained from the above address.

7. Documentation

User's guide

Kroonenberg, P.M. & Brouwer, P. (1985). **User's guide to TUCKALS2 (version 4.0)** (WEP Reeks WR 85-12-RP). Leiden: Department of Education, Unversity of Leiden.

Technical references

Brouwer, P. (1985) **Gebruikers handleiding NDIMIS3 versie 2.0. Een programma voor het voorbewerken van 3-weg data** [User's guide to NDIMIS3 version 2.0. A program to preprocess 3-way data.]. Leiden: D.I.O.S., Faculty of Social Sciences, University of Leiden.

Brouwer, P. , & Kroonenberg, P.M. (1985). Comparison and evalutation of PARAFAC and TUCKALS for three-mode analysis. Paper presented at the Fourth European Meeting of the Psychometric Society. Cambridge, UK, July.

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**, 283-219.

Carroll, J.D., & Chang, J.J. (1972). **A generalization of INDSCAL allowing IDIOsyncratic reference systems as well as an analytic approximation to INDSCAL.** Paper presented at the Spring Meeting of the Pschometric Society, Princeton, New Jersey, March.

Harshman, R.A. , & Kroonenberg, P.M. (submitted). Overlooked solutions to Cattell's parallell proportional profiles problem: A perspective on three-mode analysis.

Harshman, R.A., & Lundy, M.E. (1984a). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H.G. Law, C.W. Snyder Jr., J.A. Hattie, and R.P. McDonald (Eds.), **Research methods for multimode data analysis** (pp.122-215). New York: Preager.

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

Kroonenberg, P.M. (1983). **Three-mode principal component analysis. Theory and applications.** Leiden: DSWO Press.

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.

Ten Berge, J.M.F., De Leeuw, J., & Kroonenberg, P.M. (1987). Some additional results on principal component analysis of three-mode data by means of alternating least squares algorithms. **Psychometrika, 52**, 183-191.

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

Tucker, L.R. (1972). Relations between multidimensional scaling and three-mode factor analysis. **Psychometrika, 37**, 3-27.

Weesie, H.M., & Van Houwelingen, J.C. (1983). **GEPCAM User's Manual.** Utrecht: Institute of Mathematical Statistics, University of Utrecht.

Applications

Kroonenberg, P.M. (1983). **Three-mode principal component analysis. Theory and applications.** Leiden: DSWO Press.

Kroonenberg, P.M. (1983) . Annotated bibliography of three-mode factor analysis. **Britisch Journal of Mathematical and Statistical Psychology, 36**, 81-113.