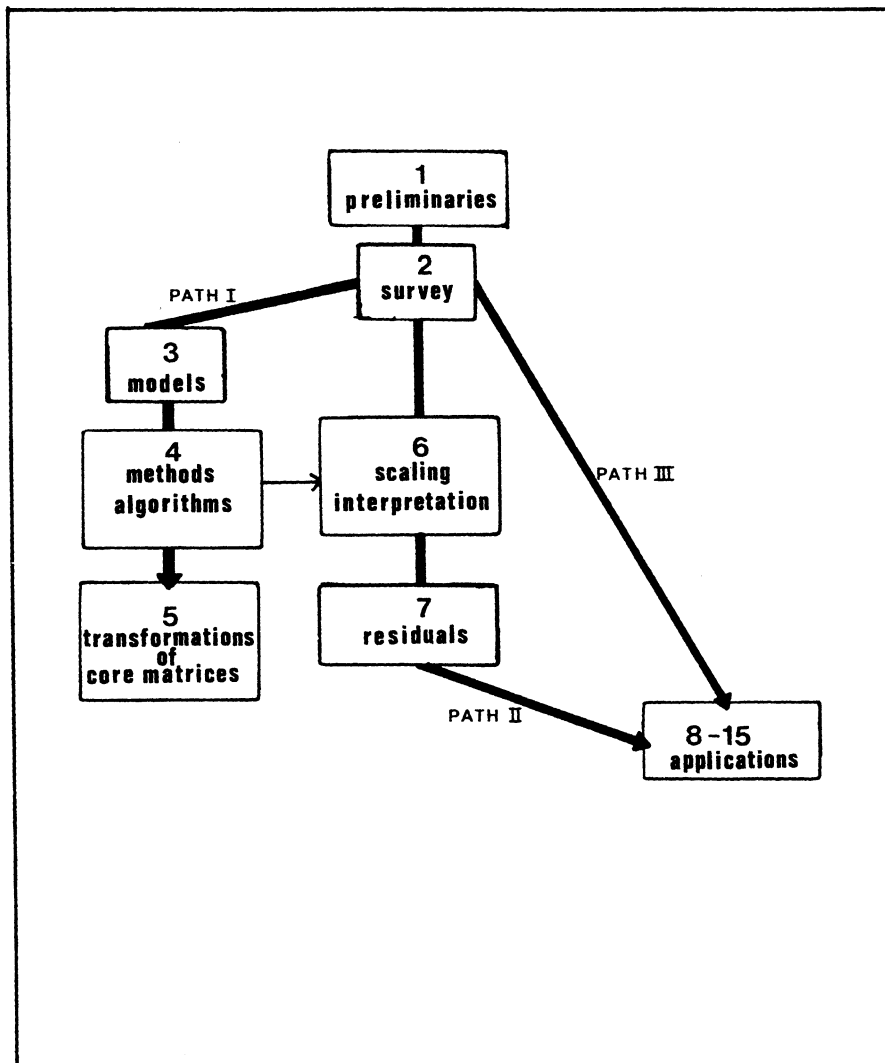


PRELIMINARIES



1.1 INTRODUCTION

Investigating the relationships between variables is a favourite research activity of social scientists. They often want to explore the structure of a large body of data. To understand this organization the data have to be condensed in one way or another, and the raw data have to be combined to form summary measures which are more easily comprehended.

Among the most popular methods to achieve such condensation and summarization are principal component analysis and multidimensional scaling. Different variants exist in both cases, and their appropriateness varies with the research design. 'Standard' principal component analysis is applied when observations are available for a number of variables, and it is desired to condense these variables to a smaller amount of independent 'latent' variables or components. Similarly, 'standard' multidimensional scaling is applied when similarity measures are available for a number of variables, in this context usually called stimuli, and insight is desired into their structural organisation.

In many research designs, observations on variables have been made under a number of conditions, or at various points in time, or similarity measures have been produced by a number of persons, etc. In such cases, where the data can be classified by three kinds of quantities, or modes, e.g. subjects, variables, and conditions, the standard variants no longer suffice.

The extra mode in the design requires an extension of those standard techniques. It is, of course, possible to mould the data into the standard two-mode (or two-way) format by rearranging them

into two-mode matrices, but this entails losing a part of the information which could be very important for the understanding of the organization of the data as a whole.

Since the introduction of three-mode principal component analysis by Tucker in 1964, and of individual differences scaling by Bloxom (1968b), Carroll & Chang (1970), and Horan (1969), considerable progress has been made in finding ways to confront the summarization and condensation of three-mode data. This has mainly been done by adapting the standard techniques to make them fit the problems created by the extra mode. That this introduces many complications will become clear in the sequel.

1.2 SOME EXAMPLES

To get a general idea of the kind of research problems three-mode principal component analysis can handle, we first give some examples of typical applications.

Semantic differential data. A classical example of three-way classified data can be found in the work of Osgood and associates (e.g. Osgood, Suci & Tannenbaum, 1957). In the development and application of semantic differential scaling, subjects have to judge various concepts using bi-polar scales of adjectives. Such data used to be analysed averaged over subjects, but the advent of three-mode principal component analysis and similar techniques has made it possible to analyse the subject mode as well in order to detect individual differences with regard to the semantic organization of the relations between scales and concepts. An example of such a study can be found in Snyder & Wiggins (1970). In Chapter 9 we present an example of this type of data.

Similarity data. Three-way similarity data consisting of stimuli \times stimuli \times subjects are generally analysed with individual differences scaling programs, such as INDSCAL (Carroll & Chang, 1970) and ALSCAL (Takane, Young, & De Leeuw, 1977). However, when the data are asymmetric and/or a more general model is required,

three-mode principal component analysis can provide useful insight. See Chapter 3 for details on individual differences scaling and its relation with three-mode component analysis, and Chapter 11 for an empirical comparison of the techniques on the same data.

Asymmetric similarity data. Van der Kloot & Van den Boogaard (1978) collected data from 60 subjects who rated 31 stimulus persons on 11 personality trait scales. In the original report the data, which can be considered asymmetric similarity data, were first averaged over subjects, and subsequently analysed by canonical discriminant analysis using the stimulus persons as groups. Van der Kloot & Kroonenberg (1982) used three-mode principal component analysis on the original data to assess the individual differences and the extent to which the subjects shared the common stimulus and scale configurations. A summary of the results can be found in Chapter 10. The example in Chapter 2 on similarities between Dutch political parties also falls into this class of applications.

Multivariate longitudinal data. In the social sciences, multivariate longitudinal data pose problems for many standard techniques. (See Visser (1982) for a detailed review of techniques useful for such data in psychology). There are often too few observations and/or too many points in time for the 'structural approach' to the analysis of covariance matrices (Jöreskog & Sörbom, 1977), or too few points in time and/or too many variables for multivariate analysis of time series by some kind of ARIMA model (see e.g. Glass, Wilson & Gottman, 1975, or Cook & Campbell, 1979, Ch. 6). In such situations three-mode principal component analysis can be very useful, especially for exploratory purposes.

Lammers (1974) presented an example of longitudinal data with a relatively large number of variables (22), and only a limited number of points in time (11 years). The aim of the study was to determine whether some of the 188 hospitals measured showed different growth patterns or growth rates compared to the other hospitals. A re-analysis of these data is presented in Chapter 13.

In Chapter 14 we present a three-mode analysis of typical learning data collected by Bus (1982). In this case there were only

six observational units (children), who had scores on five tests, but measures were available for 37 more or less consecutive weeks.

Three-way contingency tables. One of the ways to study interactions in large two-way contingency tables is by correspondence analysis. In its most common form, this technique is an analysis of the dependencies of the column and row categories of a contingency table by means of a so-called 'singular value decomposition' (see section 2.2) of the standardized residuals. A similar procedure can be defined for three-way tables using three-mode principal component analysis instead of the singular value decomposition. This approach is outlined and illustrated in Chapter 15 with data from three different elections for the election wards or precincts of Leiden.

1.3 ORGANIZATION OF THIS BOOK

The aim of this book is to treat three-mode principal component analysis with all its possibilities and limitations. We will pay attention to both theoretical and practical aspects of the technique, and therefore the level of the exposition will vary in mathematical sophistication. The theory is mainly dealt with in Chapters 3, 4, and 5, and the applications in Chapters 8 through 15. Chapter 2 provides a quick run-through of the entire book, and Chapters 6 and 7 are intermediary in the sense that they treat the theory necessary for a detailed understanding of the technical aspects of the applications and their interpretation.

The reader only interested in the technical aspects of three-mode principal component analysis, (or *three-mode analysis* for short), and its relation to other models and methods of analysis, should follow PATH I in Fig. 1.1, reading only Chapters 1 (preliminaries), 2 (survey), 3 (models), 4 (methods and algorithms), and 5 (transformations of core matrices). In Chapter 6 (scaling and interpretation) and 7 (residuals) some further technical information on matters that precede and follow a three-mode analysis can be found.

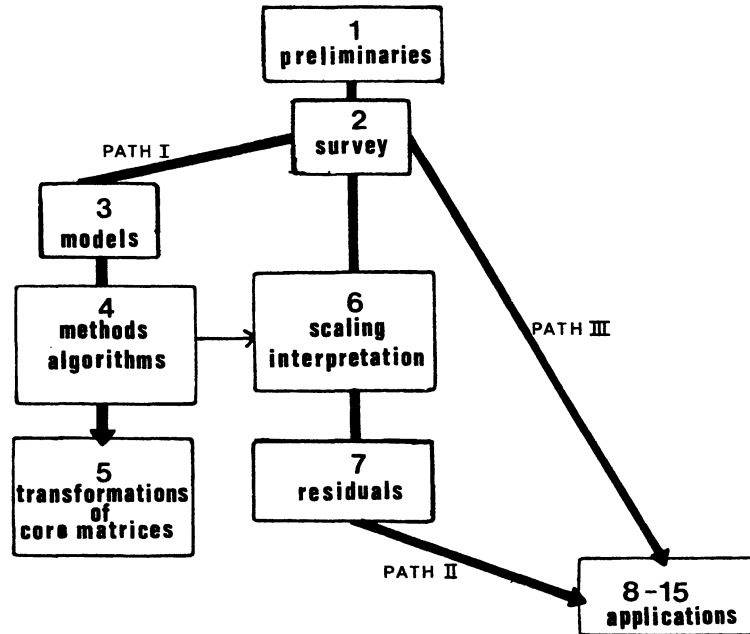


Fig. 1.1 Organization of this book

A reader only interested in the scope of the methods, and whose interest does not go beyond a basic notion of what three-mode analysis is about, should follow PATH III and read Chapters 1, 2, and the applications. However, in order to take full advantage of three-mode principal component analysis in practical situations it is best to read Chapters 6 and 7 as well, i.e. following PATH II, as in Chapter 6 the scaling of the input is discussed, as well as some interpretational aspects of the output which are helpful in understanding the peculiarities of the data at hand, and in Chapter 7 methods are described to assess the quality of the solutions obtained.

In order to facilitate the reading of single chapters, a three-mode glossary of the major terms used in this book has been included as next section.

For people looking for specific applications within their field of interest a large number of pertinent papers have been included in the references. To improve the usefulness of their inclusion these papers have been classified according to subject matter and data type. A list of papers referring to computer programs has been included as well.

1.4 THREE-MODE GLOSSARY

Basic terms

- | | |
|-------------------------|---|
| combination-mode (ij) | - Cartesian product of two (elementary) modes i and j; "i outer loop, j inner loop"; see Tucker (1966a, p. 281) |
| combination-mode matrix | - two-mode matrix with one (elementary) mode (usually columns) and one combination-mode (usually rows). |
| core matrix | - three-mode matrix, which contains the relations between the components of the various modes; its size is usually $s \times t \times u$, where s, t, and u are the number of components for the first, second, and third mode respectively. |
| element (of a mode) | - generic term for a variable (subject, condition, etc.) in a mode. |
| extended core matrix | - three-mode core matrix, of which one of the dimensions is equal to the number of elements in that mode; its size is usually $s \times t \times n$, where n is the number of elements in the third mode. |
| frontal plane | - $s \times t$ -slice of an (extended) core matrix, or $l \times m$ -slice of a three-mode data matrix. |

- mode (or elementary mode); way - collection of indices by which the data can be classified; way and mode are here used as synonyms; for a different usage of the word 'mode' in the same context see Carroll & Arabie (1980).
- reduced mode - mode of which principal components have been computed.
- three-mode matrix (-array) - collection of numbers which can be classified in three (different) ways. i.e. using three indices; the numbers can thus be arranged in a three-dimensional block.

Methods

- Covariance structure approach - In this method the subject mode is treated as a random variable, and the analysis is performed on the combination-mode covariance matrix of the other two modes. Solutions can be obtained by maximum likelihood estimation, or generalized least squares procedures. An a priori structure for the component matrix and the core matrix can be specified.
- Alternating Least Squares (ALS) - An iterative method to solve large and complex models by breaking up the total number of parameters in a number of groups, each of which can be estimated conditional on the fixed values of the parameters in the other groups.
- Partial Least Squares (PLS) - See Alternating Least Squares

- Tucker's (1966a) Method I - Standard principal component analysis on each of the three combination-mode matrices, and subsequent combination of the three solutions to form the core matrix.
- Tucker's (1966a) Method II - Standard principal component analysis on two combination-mode matrices, combined with a clever juggling to compute an approximate core matrix and the third principal component matrix without resorting to solving the eigenvalue - eigenvector problem for the largest mode. Appropriate for data sets with one very large mode, usually individuals.
- Tucker's (1966a) Method III- Method to analyse multitrait-multi-methodlike covariance and correlation matrices. Forerunner of the covariance structure approach.

Models

- CANDECOMP - Carroll & Chang (1970). T3 with a three-way identity matrix as core matrix, or equivalent to INDSCAL with different reduced modes.
- IDIOSCAL - Carroll & Chang (1972). As T2, but the two reduced modes are equal, and thus the extended core matrix is symmetric in its frontal planes. Allows for both idiosyncratic rotations of axes in the common stimulus space, and individually different weighting of these axes. Component matrices are not necessarily orthogonal.

- INDSCAL - Carroll & Chang (1970). As IDIOSCAL, but with the additional restriction, that the frontal planes are diagonal, i.e. no idiosyncratic rotations are allowed. The model can also be interpreted as having three reduced modes of equal numbers of components, and a three-mode identity core matrix.
- PARAFAC - Harshman (1970, 1972a,b, 1976). Parallel profiles factor analysis. PARAFAC1 is equal to CANDECOMP. PARAFAC2 is similar to IDIOSCAL, but it specifies a common weighting of the axes of the stimulus space. However, idiosyncratic rotations of these axes are allowed.
- Three-mode Scaling - Tucker (1972a). As the Tucker3 model, but two of three reduced modes are equal. Core matrix has symmetric frontal planes.
- Tucker2 model (T2) - Israelsson (1969). Model specifies two unequal reduced modes with an unrestricted extended core matrix.
- Tucker3 model (T3) - Tucker (1966a). Three unequal reduced modes with an unrestricted core matrix.
- Tucker's common factor model - Tucker (1966a). As T3, but unique variances are specified for the combination-mode covariance matrix.

Terms with special definitions

- component - vector of loadings (e.g. g_p, h_q, e_r)
- component weight - eigenvalue, indicating the amount of variation explained by the component corresponding to the eigenvalue.

i-mode	- first mode
j-mode	- second mode
k-mode	- third mode
SS(Fit)	- sum of squares of the estimated data values, derived from the fitted model
SS(Res)	- residual sum of squares
SS(Tot)	- total sum of squares of the data
standardized component weight	- component weight divided by the total sum of squares of the data
standardized sum of squares (St.SS)	- sum of squares divided by the total sum of squares of the data.
variation	- general term to indicate the sum of squares, generally of data values; depending on their scaling variations may be sums of squares, average sums of squares, or variances.

1.5 NOTATION

Matrices

$R^{a \times b}$	set of real matrices with a rows and b columns
$[[X]], X \in R^{a \times b}$	$\sqrt{\sum_{i=1}^a \sum_{j=1}^b x_{ij}^2}$; Euclidean norm
$\text{tr } X, X \in R^{a \times a}$	$\sum_{i=1}^a x_{ii}$; trace of X
$C = X \otimes Y, X \in R^{a \times b}, Y \in R^{c \times d}$	$c_{mn} = x_{ij} y_{kl}$; $m=1, \dots, ac$; $n=1, \dots, bd$ Kronecker product
$\text{diag}(X)$	$d_{ij} = x_{ij}$ if $i=j$, and 0 otherwise

Data

$Z = \{z_{ijk}\}$	three-mode data matrix; $i=1,\dots,\ell$ (rows), $j=1,\dots,m$ (columns), $k=1,\dots,n$ (frontal planes)
$\hat{Z} = \{\hat{z}_{ijk}\}$	three-mode matrix with data values estimated from the fitted model

Model

<u>first mode</u>	<u>second mode</u>	<u>third mode</u>	
ℓ	m	n	number of elements
i	j	k	index of elements
s	t	u	number of components ($\ell \geq s$; $m \geq t$; $n \geq u$)
p	q	r	index of components
$G = \{g_{is}\}$	$H = \{h_{jq}\}$	$E = \{e_{kr}\}$	component matrix (orthonormal for Tucker models); for definition see <i>Theorem 4.1</i> .
$G = \{g_{is}\}$	$H = \{h_{jq}\}$	$E = \{e_{kr}\}$	component matrix for Tucker Methods for definition see <i>Theorem 4.2</i> .
g_p, g_p	h_q, h_q	e_r, e_r	components, vectors of loadings
λ_p	μ_q	ν_r	standardized component weights (eigenvalues of P (or P), Q (or Q), R (or R), respectively [for definition see <i>Theorem 4.1</i> (or <i>Theorem 4.2</i>)]
$C = \{c_{pqr}\}$			core matrix of Tucker3 model
$C = \{c_{pqr}\}$			core matrix of Tucker3 model for Tucker Methods
$\hat{C} = \{\hat{c}_{pqr}\}$			extended core matrix of Tucker2 model
$\hat{C} = \{\hat{c}_{pqr}\}$			extended core matrix for Tucker Methods