KM 6(1982)
pag. 65-94

TUCKALS3: A program for three-mode principal component analysis

Pieter M. Kroonenberg, vakgroep Wijsgerige en Empirische Pedagogiek,
Rijksuniversiteit Leiden, Schuttersveld 9^{V}, Leiden

VOORWOORD

Het voorliggende artikel is een verkorte versie van de programmabeschrijving van een computerprogramma voor drieweg principale componentenanalyse, TUCKALS3 genaamd. De presentatie van de theorie is hier met opzet kort gehouden, omdat deze theorie elders uitvoerig is beschreven. De bespreking van het voorbeeld is wel speciaal voor dit artikel geschreven en ik heb gepoogd zo veel mogelijk aspecten van het programma naar voren te brengen via het voorbeeld. De gedetailleerde invoerbeschijving is niet opgenomen aangezien die toch alleen van belang is voor directe gebruikers, maar wel is een samenvatting van de invoer als appendix toegevoegd om een indruk te geven van de mogelijkheden van het programma.
Het programma zelf, samen met het parallelle programma voor een beperktere vorm van drieweg principale componentenanalyse - TUCKALS2 - zijn vanaf ongeveer 1 januari 1982 samen met de programmabeschrijvingen en een voorbeeld data set bij de vakgroep Datatheorie (Breestraat 70, Leiden) te verkrijgen. Op het ogenblik draait het programma nog uitsluitend op de Leidse IBM/Amdahl configuratie en is nog niet uitgetest op andere machines. Met name conversie naar CDC hangt af van de vraag naar het programma van CDC-gebruikers.
Voor geinteresseerden is op verzoek bij mij een eerste versie van een geannoteerde bibliografie over drieweg factor- en componentenanalyse te verkrijgen.

SUMMARY

After a relatively non-technical account of three-mode component analysis of three-way data, several features of a computer program to perform such an analysis, TUCKALS3, are described. A detailed analysis of data on the similarities between Dutch political parties is presented to illustrate how three-mode principal component analysis may be used to unravel complex relationships.

INTRODUCTION

The three-mode principal component model - here referred to as Tucker3 model - was first formulated within the context of the social sciences by Tucker (1963), and in subsequent articles Tucker (1964, 1966) extended especially the mathematical description and the programming aspects of it. In the multidimensional scaling context references to his model occur frequently (Carroll & Chang (1972), Takane, Young & De Leeuw (1977), Jennrich (1972), etc.), as the Tucker3 model is the general model comprising various other individual differences models. A discussion of the relation between multidimensional scaling and three-mode factor analysis can be found in Tucker (1972), Carroll & Wish (1974), Takane, Young & De Leeuw (1977), Carroll & Arabie (1980). The method used here has been fully described by Kroonenberg & De Leeuw (1980).

INFORMAL DESCRIPTION

When a researcher has collected information of a number of subjects on a large number of variables, he or she often wants to know if linear combinations of these variables can be found, which explain the larger part of the variation present in the data. Such linear combinations are called components, latent variables or factors. Another way of asking the same question is: "Can the scores on the variables be explained by linear combinations of a set of underlying or latent variables?" If a small number of components is sufficient to explain the larger part of the variation, one might look upon these components as latent variables, which capture the essential information present in the data. As an example one could imagine that the scores on a battery of tests (= measured variables) are determined by linear combinations of such latent variables as the arithmetic and verbal content of the tests. The latent variables - arithmetic content and verbal content - can be found by a standard principle component analysis.

Suppose now in the same example that the researcher has administered the battery of tests a number of times under various conditions of stress and time limitations. The data are now classified by three different types of quantities: subjects, tests, and conditions. Again the researcher is interested in (1) the components of the variables which

explain the larger part of the variation in the data. Moreover, he or she is interested (2) if different groups or general characteristics can be defined for subjects as well. Said differently, he or she wants to know if it is possible to conceive of the individual subjects as linear combinations of 'idealized subjects'. In the example we could suppose that the subjects are linear combinations of an exclusively verbally gifted person and an exclusively mathematically gifted person. Such persons are clearly ideal types. Finally a similar question could arise with respect to conditions: (3) can the conditions be characterized by a set of 'idealized' or 'prototype' conditions?

Each of the three questions can be answered by performing principal component analyses for each type of quantities or 'mode'. In fact, the same variation present in the data is analyzed in three different ways. In other words the components extracted are in some way related. The question is of course: how. We will, in order to avoid confusion, call the components of the variables latent variables, the components of the subjects idealized subjects, and the components of the conditions prototype conditions.

With respect to the relations between the components of the three modes one could ask questions like: "Do idealized subject 1 and idealized subject 2 react differently to latent variable 2 in prototype condition 1?" Or one could ask: "Is the relation between the idealized subjects and the latent variables different under the various prototype conditions?" By performing three separate component analyses such questions are not immediately answerable, as one does not know how to relate the various components. The three-mode principal component model, however, specifies explicitly how the relations between the components can be determined. The three-mode matrix, which embodies the relations between the various components is called the core matrix, as it is supposed to contain essential relations or characteristics of the data. In a sense the component matrices for the three modes and the core matrix are all that is necessary to give a representation of the data within the framework of three-mode principal component analysis. However, various amounts of auxiliary information are necessary for the proper interpretation of the outcome of the analysis. TUCKALS3 provides such information, and the value of it will be discussed in connection with the output example.

Three-mode principal component analysis has been used in a wide variety of research problems. Below some typical applications in the social sciences are given.

1. A classical example of three-fold classified data can be found in the work of Osgood and associates (e.g. Osgood, Suci & Tannenbaum, 1957). In the development and application of so-called semantic differential scaling subjects have to judge various concepts using bipolar scales. Such data have been traditionally analysed averaged over subjects, but the advent of three-mode principal component analysis and similar techniques has seen an interest in analyzing the subject mode as well. The aim was to detect if individual differences existed with regard to the semantic organisation of the relation between scales and concepts. An example of such a study can be found in Snyder & Wiggins (1970).
2. Endler, Hunt & Rosenstein (1962) collected data for the development of a "Stimulus-Response Inventory of Anxiousness". The inventory consisted of eleven different (precarious) situations, for each of which the respondent had to answer how he would react. He had to give the intensity of his reaction for 14 different categories. Situations were, for example 'going for the first time on a date', 'going to an important application for a job'. Answer categories were, for instance: 'heart is beating quicker', 'perspiring', 'enjoing the challenge'. Using three-mode principal component analysis Levin (1964) not only found situation and reaction types, but also different kinds of subjects, who reacted in different ways to the various situation types.
3. Jones & Young (1972) collected data about the social structure of a small, closed and naturally formed group (staff and students of an institute of psychology). Staff, students and other personnell of the institute judged the similarity between the members of the scientific staff and doctoral students, where each similarity could be scored on a seven-point scale. Although such data can be analysed fruitfully with various other models, also three mode principal component analysis is appropriate for them. Especially if one also wants to evaluate individual differences. In this study the interest was primarily concentrated on the dimensions of interpersonal perception (as emerged in the ratee or stimulus space), and the use of subjects of these dimensions (as shown in the core matrix). The subjects (as raters) were divided in groups based on professional interest (in psychology). The three groups that could be found placed differential weights on the dimensions of the stimulus space (status, political persuasion and professional interests).

4. Van de Geer (1974) presented an example of time-series data with a relatively large number of variables and only a few time-points. More in particular, data were available of 188 hospitals on 27 variables measured in 11 consecutive years. The aim of this study was to determine if there were various kind of hospitals which showed differential growth rates or growth patterns. Typical results of this study were that over the years large hospitals stay large relative to the initially smaller ones, and that all hospitals grew roughly in the same manner. There were, however, a small number of hospitals which showed a different growth pattern, especially in a special group of variables.

FORMAL PROBLEM DESCRIPTION

A lxmxn three-mode matrix $Z \in R^{lxmxn}$ is defined as the collection of elements:

$$\{z_{ijk} \mid i=1,\ldots,l;\ j=1,\ldots,m;\ k=1,\ldots,n\}$$

These elements can be placed in a three-dimensional block with the index i running along the vertical axis, the index j along the horizontal axis, and the index k along the 'depth' axis. We will use the word mode to indicate a collection of indices by which the data can be classified. For instance, in semantic differential studies one collects scores of a number of persons on a set of bipolar scales for a collection of attributes. These data can be classified by persons, scales, and attributes, each of these therefore determines a mode of the data.

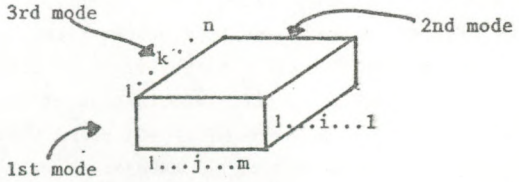


Figure 1 A three-mode matrix

A three-mode matrix can also be seen as a collection 'normal' (= two-mode) matrices. This can be done in three different ways, as is shown below in Figure 2.

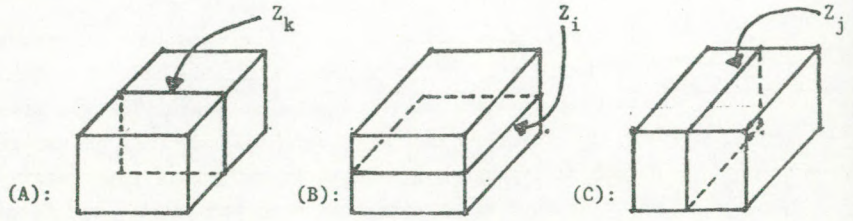


Figure 2 Three different ways to view a three-mode matrix as a collection two-mode matrices.

The Tucker3 model is the factorization of the three-mode matrix matrix $Z = \{ z_{ijk} \}$, $Z \in R^{lxmxn}$ such that

$$z_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} g_{ip} h_{jq} e_{kr} c_{pqr} \qquad \begin{matrix} \text{for } i=1,\ldots,l; \\ j=1,\ldots,m; \\ k=1,\ldots,n, \end{matrix} \qquad (1)$$

where the coefficients g_{ip}, h_{jq}, and e_{kr} are the elements of the columnwise orthonormal matrices $G \in K^{lxs}$, $H \in K^{mxt}$, $E \in K^{nxu}$ respectively, and the c_{pqr} are the elements of the so-called three-mode core matrix $C \in R^{sxtu}$.

One can interprete G as the matrix which contains the scores of the variables of the first mode on their components. (In the factor analytic literature these scores are usually called "loadings". Note, however, that the length of the components are here scaled to be equal to 1, and not equal to the eigenvalues). The interpretation of H and E is analoguous. The number of components of the first, second and third mode are s, t, and u respectively. In the original matrix Z every element of the matrix represents a specific combination of categories of the original variables. In a similar manner each element of the core matrix represents a unique combination of categories of the components. Thus the core matrix describes the basic relations that exist between the various collections of variables as expressed through their components.

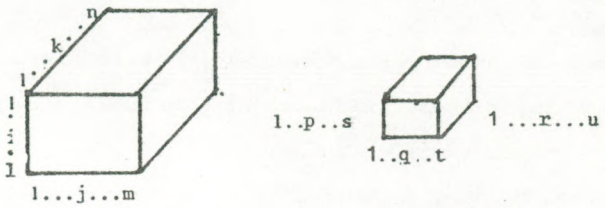


Figure 3 Data array versus core matrix

The matrix formulation of the model is

$$Z = GC(H' \otimes E') \qquad (1)$$

where $Z \in R^{kxnm}$ and $C \in R^{sxtu}$ are now ordinary two-mode matrices by making use of so-called combination modes (Tucker, 1966, 281), and $\otimes$ denotes the Kronecker product (Tucker, 1966, 283ff). We will not introduce special notation to distinguish between the two-mode and three-mode versions of Z and C, as the appropriate one is indicated by the real space whereof it is an element.

If we would compute all the principal components, thus s=l, t=m, and u=n, then we would decompose any data matrix exactly in its components. However, in practical applications one is just interested in the two, three or four first principal components. This precludes in general finding an exact factorisation of Z in G, H, E, and C. One, therefore, has to settle for an approximation, i.e. one has to find G, H, E, and C such that the difference between the model and the data is minimal according to some loss function, or in slightly different terms, we have to look for a best approximate factorization of the matrix Z into G, H, E, and C according to the Tucker3 model.

In our case we define the loss function to be the mean squared one, and propose to search for those G, H, E, and C such that

$$f(G,H,E,C) = \|Z - GC(H' \otimes E')\|^2 \qquad (2)$$

is minimal. Where $\|\ .\ \|$ denotes the Euclidean norm. The minimization has to be carried out under the restrictions of the model:

G, H, E are columnwise orthonormal matrices (3)

In a more semantic notation (2) can be written as

$$SSQ(Residual) = SSQ(Data - 'Reconstructed\ Data') \qquad (4)$$

where SSQ means Sum of SQuares. By 'reconstructed data' is meant the estimated data values on the basis of model (1). Via some algebraic manipulations it can be shown that (4) can be transformed into

$$SSQ(Residual) = SSQ(Data) - SSQ(Reconstructed\ Data), \text{ or}$$

$$SSQ(Residual) = SSQ(Data) - SSQ(Fit). \qquad (5)$$

As the SSQ(Data) or SSQ(Total) is fixed, minimizing the SSQ(Residual) is equal to maximizing the SSQ(Fit). This maximization problem can be written as

$$\max_{G,H,E} SSQ(Fit) = \max_{G,H,E} Tr\ G'Z(HH' \otimes EE')Z'G. \qquad (6)$$

The maximization can be carried out via a so-called Alternating Least Squares algorithm, which maximizes the objective funtion over G, H, and E in turn holding the other two parameter matrices fixed. This procedure is repeated until convergence. Details about the algorithm are given in the User's Guide to TUCKALS3, and the convergence properties are discussed in Kroonenberg & De Leeuw (1980).

A point which needs to be mentioned is the initialization of G_o, H_o, and E_o, the starting matrices of the algorithm. It seems appropriate to search for initializations which would increase the chances of convergence to a global maximum rather than a local one. The approach taken here is to choose G_o, H_o, and E_o such that they will solve the maximization if it had an exact solution. It can be shown that the eigenvectors associated with the largest eigenvalues of $U = ZZ'$ ($Z \in R^{lxmn}$), $V = ZZ'$ ($Z \in R^{mxnl}$), and $W = ZZ'$ ($Z \in R^{nxlm}$) will solve the maximiztion problem exactly if there exists such a solution. These eigenvectors are, therefore, used to initialize the algorithm. This initial solution is, in fact, the Method I solution of Tucker (1966).

JOINT PLOTS

After the components have been constructed the core matrix will provide the information about the relations of the components. It is, also very instructive to investigate the common space of two modes (say, subjects and variables). The components can be scaled in such a manner that they can be plotted together in one figure. Such joint plots are provided in the program (see Input Summary, parameter ICPLOT).

The joint plots of every pair of component matrices for each component of the third mode, say E, are constructed in such a way that g_i (i=1,..,s) and h_j (j=1,..,t) - i.e. the columns of G and H - are close to each other. Closeness is measured as the sum of all sxt squared distances.

The plots are constructed as follows: For each component r of E, G and H are scaled by dividing the core plane C_r between them (by using a singular value decomposition), and weighting the so scaled G and H by the relative number of elements in the modes to make the distances comparable:

$$GC_rH' = GU\Lambda V'H' = (\frac{m}{I})^{\frac{1}{4}}(GU\Lambda^{\frac{1}{2}})\ (\frac{I}{m})^{\frac{1}{4}}(HV\Lambda^{\frac{1}{2}})' = G^*H^{*\prime}$$

with

$$G^* = (\frac{m}{I})^{\frac{1}{4}}GU\Lambda^{\frac{1}{2}} \quad \text{and} \quad H^* = (\frac{I}{m})^{\frac{1}{4}}HV\Lambda^{\frac{1}{2}}.$$

When C_r is not square only the first min(s,t) components can be used.

For further remarks on this procedure one could consult Appendix 2 of Kroonenberg & De Leeuw (1977). (Note, however, that in the second section I and m should be interchanged.)

ANALYSIS OF RESIDUALS

For a proper assessment of the role of the elements of the various modes, it is necessary to have some insight in the structure of the residuals. A residual in this context is the difference between the data and the estimated data 'reconstructed' from the estimated model parameters. Large residuals or residual sums of squares (SSQ(Residual)) indicate that a particular element does not fit very well in the structure determined by the other quantities. However, an extremely large residual sum of squares, often combined with a very large total sum of squares (SSQ(Total)) is often indicative for some clerical error in the data.
The size of a SSQ(Residual) depends on the SSQ(Total). Therefore, the relative residual sum of squares (=SSQ(Residual)/SSQ(Total)) should be used for the assessment of the role of an element of a mode. Also the so-called fit/residual ratio (see below) provides valuable information to that end. It is good to realize that in general elements with large SSQ(Total)'s will be fitted better than those with small SSQ(Total)'s due to the least squares procedures used. Sometimes it is wise to rerun an analysis without an element which has an incomparable larger total sum of squares than the other elements in the analysis.

The residuals can be analysed at two levels:

a. From formula (5) we see that

$$SSQ(Total) = SSQ(Fit) + SSQ(Residual) \qquad (7)$$

For each element e of each mode it can be shown that

$$SSQ(Total_e) = SSQ(Fit_e) + SSQ(Residual_e) \qquad (8)$$

Using (8) we can examine

$$SSQ(Fit_e)/SSQ(Residual_e) \qquad (9)$$

or the fit/residual ratios of the elements of each mode. High fit/residual ratios indicate good agreement with the model, while low fit/residual ratios mean bad agreement with the model.

The program provides for each mode a plot of the SSQ(Residual) versus the SSQ(Fit) from which the relative performance of the points can be gauged, i.e. both relative to each other, and relative to the overall fit/residual ratio. Probably the easiest way to examine the plot is by connecting the point (0,0) with the point which has as its coordinates the average SSQ(Fit) and the average SSQ(Residual). Points above the line have a relative worse fit than average, and points below the line have a relatively better fit than average. As sums of squares are plotted, rather than proportions of total of sums of squares, the sizes of the total sums of squares are reflected in the plots. They can be found by adding the x- and y- values in the plot. This means that points with equal total sums of squares lie on lines which make an angle of -45 degrees with the positive X-axis. Note that due to (possibly) unequal scaling this line need not be under an angle of -45° at first sight.

b. A more detailed analysis of the residuals is possible by writing the lxmxn block of squared residuals to an external unit (tape, disk or card). The unit number is transferred to the program by the parameter IOURES (see Input specifications). These squared residuals could be inspected individually, or, by averaging over one of the modes, the 'interaction' of two modes may be investigated.

SCALING OF INPUT DATA

In standard two-mode principal component analysis the input data are often transformed to standard scores without much thought to the consequences. In three-mode analysis the question of scaling must be approached with more care, as there are many ways to standardize or centre the data. The two basic rules are:

(1) those means should be eliminated (i.e. set equal to zero), which cannot be interpreted or which are incomparable within a mode;

(2) those variances should be eliminated (i.e. set equal to one), which are based on arbitrary units of measurement or which are incomparable within a mode. If all quantities are measured in the same (possibly arbitrary) units it is not necessary to eliminate the variances.

Table 1 Similarities between parties

PvdA group

compared parties

standards	CPN	PSP	PvdA	KVP	ARP	VVD	CHU	SGP	GPV	BP
CPN	9	7	7	4	4	2	4	2	2	3
PSP	8	9	7	4	4	3	4	2	2	2
PvdA	6	7	9	6	6	3	(5)	2	1	1
KVP	1	2	6	9	6	6	6	2	3	3
ARP	1	2	5	5	9	(4)	7	4	(5)	2
VVD	0	2	3	6	5	9	7	4	4	(5)
CHU	0	1	4	5	7	6	9	(4)	5	3
SGP	1	2	3	3	(6)	4	5	9	8	6
GPV	1	2	2	3	(6)	4	6	8	9	6
BP	1	2	2	3	4	(6)	5	6	6	9

KVP group

compared parties

standards	CPN	PSP	PvdA	KVP	ARP	VVD	CHU	SGP	GPV	BP
CPN	9	8	7	4	3	2	3	3	2	(5)
PSP	7	9	6	3	3	2	4	4	3	(5)
PvdA	5	7	9	7	5	2	(5)	2	2	2
KVP	1	3	6	9	7	(4)	7	3	3	1
ARP	1	1	5	6	9	(3)	7	(5)	(5)	2
VVD	0	1	3	6	5	9	7	3	4	(5)
CHU	1	2	3	6	8	5	9	5	5	2
SGP	1	3	2	3	(6)	3	6	9	8	(4)
GPV	1	3	3	4	(6)	2	6	8	9	(4)
BP	4	4	2	3	3	(5)	4	5	6	9

Very common procedures are

(a) centring or standardizing the variables over all subject-condition combinations;

(b) centring or standardizing the variables over all subjects for each condition separately;

(c) double-centring, i.e. centring per condition over both variables and subjects.

Subjects, variables and conditions indicate here the first, second, and third mode quantities, respectively.

The decision which centring or standardization is appropriate in any particular data set depends on the researcher's assessment of the origin of the variability of his or her data, in other words on which means and variances can be meaningful interpreted. For a more extensive discussion one could consult Kroonenberg (1981).

EXAMPLE: SIMILARITIES BETWEEN POLITICAL PARTIES

Data

To illustrate the working of the program we will use some very old data on similarities between Dutch political parties (De Gruijter, 1967). The choice of the data was mainly guided by the small size of the data set, and the reasonable interpretability. De Gruijter used a group of 82 members of political student organizations at the University of Leiden; three of the students were not used in the analysis. On the basis of their preference for a particular party the students were divided into six groups, namely into a PSP, PvdA, KVP, ARP, VVD, and a CHU group respectively. The ten parties which were then (1966) in Parliament - CPN,PSP, PvdA, KVP, ARP, VVD, CHU, SGP, GPV, Boerenpartij (BP) - were used as stimuli. De Gruijter confronted the students with all possible triads of parties, and asked them to indicate for each triad which two parties were most alike, and which two were least alike. For each preference group he computed the number of times (summed over all subjects in that group - n_g) that in all triads with stimulus parties i and j the similarity between i and j was considered to be greater than that between i (the standard) and the third stimulus. As each party was compared with all combinations of the other parties the sums for the standards are equal to $n_g\binom{9}{2}$. The data have thus the form of 6 matrices (preference groups) of 10 standards by 10 compared parties.

Table 2 General characteristics of the solutions

	type of solution	
	3x3x2	2x2x2
Standardized total sum of squares - SSQ (Total)	1.00	1.00
Approximation of SSQ (Fit) derived from separate PCA on mode 1	.94	.83
Approximation of SSQ (Fit) derived from separate PCA on mode 2	.94	.83
Approximation of SSQ (Fit) derived from separate PCA on mode 3	.97	.97
Fitted sum of squares from simultaneous estimation - SSQ (Fit)	.92	.82
Residual sum of squares: SSQ (Res) = SSQ (Total) - SSQ (Fit)	.08	.18
Improvement in fit compared to initial configuration	.001	.004

In the present analysis the data matrices have been divided by the number of persons in a preference group (cf. Table 4), as in a previous analysis (which is presented in the User's Guide to TUCKALS3), it was shown that some of the differences observed between groups could be attributed to different sizes of the groups. The unequal sizes obscured any real differences in variability between parties. In addition, the main diagonal elements of each data matrix, which were left blank in De Gruijter's analysis) were set to 9, indicating that a party is more similar to itself than to any other party. Note that the data matrices can be and are asymmetric as there is no necessity for a party to be chosen over another when compared with a standard as often as it is considered alike to the same party when itself is the standard. Note also that all row sums are now $\binom{10}{2} = 45$. The data matrices were double-centred before the analysis proper, as is customarily done with similarities. An option to do this and other types of centring is included in the program. Table 1 gives, as an example, the data,adjusted for group sample size,for the KVP and the PvdA group.

In the following we will report the results primarily to illustrate the program, rather than to shed light on the political scene in 1966 as seen through the eyes of 79 students. We will, therefore, primarily follow the flow of the output.

Fit

In principal the analysis which is reported here is the one with three components each for the first and second mode, and with two components for the third mode; from now on called the 3x3x2-solution. It is sometimes compared with another analysis with two components for each of the modes, or the 2x2x2-solution.

From Table 2 it can be seen that with three components for the party modes we exhaust the variability in the data, as 92% of the total sum of squares was accounted for by the model. Even the two component solution was already satisfactory. The 'approximate total fit' from the initial configuration for each of the modes (which make up the standard Tucker (1966) Method I solution) are upper bounds for the SSQ(Fit) of the simultaneous solution. Obviously the smallest of the three is the least upper bound, in this case the one based on the second mode (.94). The initial configurations are used as starting points for the main TUCKALS3 algorithm. The improvement in fit indicates how much the iterative process improves the simultaneous solution over the starting one. In this case this improvement is negligible, in other words we could have settled for the old method as far as fit is concerned, but we could not have known that beforehand.

Table 3 Party spaces (based on mode 1 - parties used as standards)

	3x3x2 - solution			2x2x2 - solution	
	1	2	3	1	2
CPN	.48	-.25	-.04	.48	-.25
PSP	.48	-.17	.04	.48	-.18
PvdA	.43	.18	.15	.43	.18
KVP	.01	.50	-.09	.01	.50
ARP	-.15	.30	.40	-.15	.30
VVD	-.20	.22	-.62	-.20	.21
CHU	-.22	.30	-.00	-.22	.29
GPV	-.33	-.33	.29	-.33	-.33
SGP	-.33	-.28	.34	-.33	-.27
BP	-.17	-.46	-.47	-.17	-.46
component weight	.61	.21	.11	.61	.21

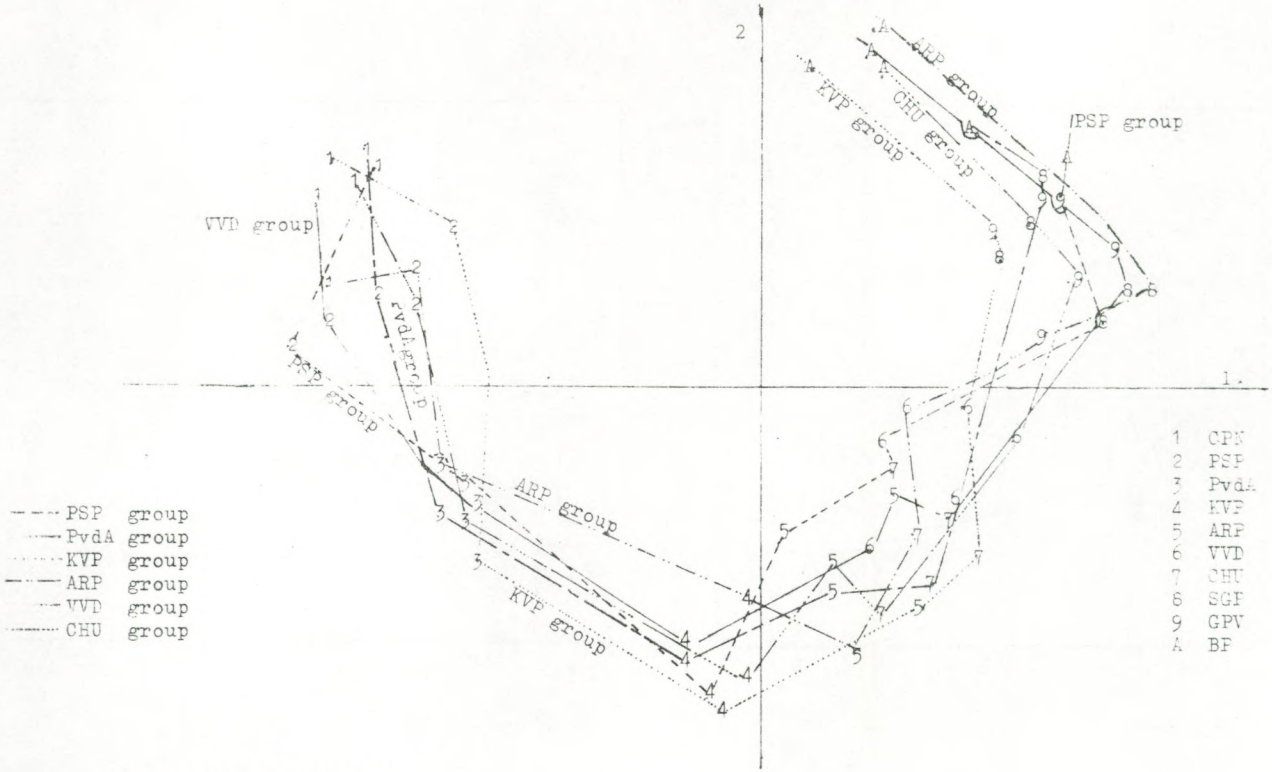
Configurations for the three modes

De Gruijter had to symmetricize the matrices in his analysis, due to the inability of earlier multidimensional scaling programs to handle asymmetric data. He analysed each preference group separately, rather than simultaneously as was done here. His results are displayed in Figure 4. The advantage of the present approach is that one space can be found for all groups together, and in addition the fit of this common configuration for each group can be assessed. De Gruijter extracted only two dimensions, and concluded that a 'horseshoe' could be found for each of the preference groups separately (see Figure 4), and not surprisingly in the present analysis the first two components of the common space exhibit a horseshoe as well (see Table 3 and Figure 5). Of course we get two common spaces, one for the first mode (standards), and one for the second mode (compared parties), but the two are hardly different from one another as can be seen from their respective solutions. In other words the asymmetry present in the data is very small; a conclusion also reached by De Gruijter using different means.

Horseshoes always pose interpretational problems. Often but not always both the projections on the axes, and the position along the horseshoe are candidates for interpretation. Kruskal & Wish (1978, p.88,89), Levelt, Van de Geer, and Plomp (1966, p. 173,174), and Gifi (1980, p.73) discuss horseshoes, their origins, and their interpretations. In the present example the interpetation of the position along the horseshoe is very clear cut, viz. from left-wing (CPN) to right-wing (BP).

The party space is open to a more complex interpretation than just the horseshoe. In particular, the first dimension also shows a left-right distinction, so that we are now faced with the problem that we have not a priori scaled the parties on this dimension, and that we do not know which of the GPV, SGP, and BP the students considered the most right-wing party. This information would have made it possible to choose between interpreting the horseshoe or the axis. The second axis separates the big and ideologically or politically flexible parties with governmental experience from the small and dogmatic parties which have never borne governmental responsibility. Which of the three mentioned characteristics the students really used or used more often is not possible to assess without additional independent infor-

Figure 4 Two dimensional configuration according to De Gruijter (1967)



Source: Wolters(1975)

Fig. 5 Party space of mode 1 (standards) - 3x3x2-solution

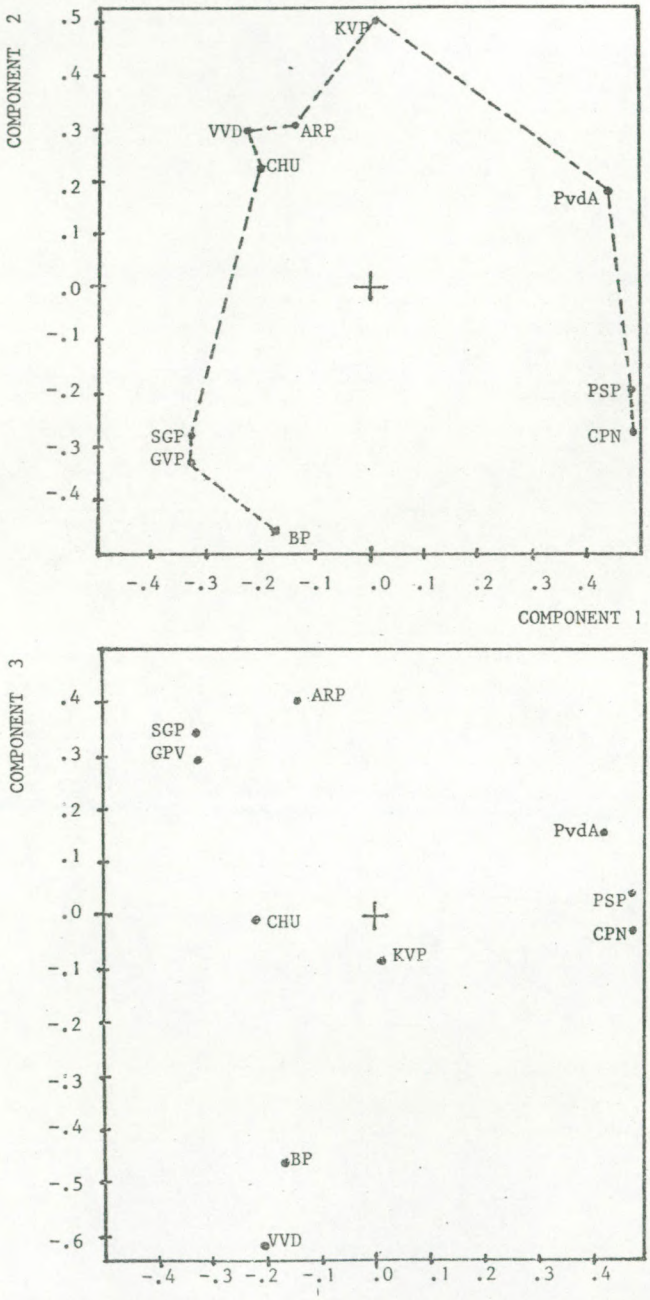


Table 4 Party preference space

party preference	3x3x2-solution 1	3x3x2-solution 2	2x2x2-solution 1	2x2x2-solution 2	number of students in group N_g
PSP	.42	.68	.42	.52	8
PvdA	.42	.34	.42	.22	15
KVP	.40	-.59	.40	-.79	11
ARP	.40	-.19	.40	-.09	10
VVD	.42	-.10	.42	.21	9
CHU	.38	-.18	.38	-.12	9
component weight	.91	.01	.81	.007	

mation. Finally, the third axis which defies my interpretational abilities indicates that BP and VVD are alike, and both unlike the ARP. That the third axis is not a fluke of the technique or some accidental effect, can be seen from the data itself. For each of the preference groups (for examples see Table 1) we see primarily a central central band of high similarities, which causes the horseshoe. Not fitting into this pattern are exactly the effects mentioned: the VVD and the BP are more alike than the horseshoe would predict, and the ARP is less alike especially than the VVD. The same can be said to a lesser extent for the too close agreement of the ARP and the SGP, GPV. A further point worth mentioning is that the bending back of the horseshoe to make the BP somewhat similar to the CPN (probably because their similar extremist and unflexible approach to politics) can be seen in the data by the slight increase of the similarities in the NE and SW corner of the data matrices. This effect is strongest for the KVP preference group, which is a party in the middle of the political spectrum, and least so for the PSP preference group near the end of the horseshoe. The circumplex structure is clearly not complete. Finally we want to point out that the method to solve the three-mode model makes that the solutions are not nested, i.e. the first two components of the 3x3x2-solution are not equal to the two components of the 2x2x2-solution. That the difference is very small in the present example is besides the point.

Table 5 Frontal planes of the core matrix

3x3x2 solution

Core plane C_1 belonging to the first component of the party preference space

	components of "compared parties"								
	$C_1 = (c_{pq1})$			$C_1^2 = (c_{pq1}^2)$			standardized $(c_{pq1}^2/SSQ(Total))$		
	1	2	3	1	2	3	1	2	3
components 1	19.	-.03	-.01	361			.60		
of the 2	.06	11.	-.2		121			.21	
standards 3	.01	.03	8.			64			.11

sum = first component weight=.91

Core plane C_2 belonging to the second component of the party preference space

	components of "compared parties"								
	$C_2 = (c_{pq2})$			$C_2^2 = (c_{pq2}^2)$			standardized $(c_{pq2}^2/SSQ(Total))$		
	1	2	3	1	2	3	1	2	3
components 1	.5	1.3	1.2	.3	1.7	1.5			
of the 2	1.1	-0.6	-0.4	1.2	.4	.2	>.003		
standards 3	.9	-0.1	-0.4	.8	.0	.2			

sum = second component weight=.01

Note: The program also prints the core matrix in two other ways, i.e. as horizontal and as lateral planes. The sums of the horizontal planes are equal to the component weights of the first mode standards. The sums of the lateral planes are equal to the component weights of the second mode ("compared" parties).

The party preference space is on the whole hardly interesting. As was to be expected from the similarity of the solutions from the separate analyses, the loadings on the first component are virtually equal. In addition, the second component is not worth bothering about as it accounts only for 1% of the total variation, and it is probably just reflecting random error. On the other hand, it should be remembered that only the first and second modes were centred, and not the third mode. This means that the first component of the third mode still reflects the average scoring level of the six groups. Technically the average frontal plane of the double-centred three-mode matrix is not zero, while the average lateral and average horizontal planes are zero. For the present data the second component of the third mode is too small under any circumstances, but in other data the second component might contain valuable information about differences between the elements of the third mode even though it is far smaller than the first component.

Core matrix

Above we noted that really only the first component of the space for the preference groups was interesting, therefore we will only discuss the interpretation of the first core plane.

The core matrix indicates how the various components of the three modes relate to one another. For instance, the element c_{111} (=19) of the core matrix indicates the strength of the relation between the first components of the three modes, and c_{221} (=11) the strength of the relation between the second components of the first and second mode with the first of the third mode. The interpretation of the elements of the core matrix is facilitated if one knows that the sum over all squared elements of the core matrix is equal to the SSQ(Fit). In other words the c^2_{pqr}'s indicate how much the combination of the p-th component of the first mode, the q-th component of the second mode, and the r-th component of the third mode contributes to the overall fit of the model, or how much of the total variation is accounted for by this particular combination of components. Thus as Table 5 shows, 60% of the SSQ(Total) is accounted for by the combination of the first components of the three modes, another 21% by c^2_{221}, and 11% by c^2_{331}. Together with the negligible contributions of the other elements of the first frontal plane these contributions sum to

Table 6 Sums of squares broken down by party preference

3x3x2-solution

Party preference	SSQ(Total) standardized	SSQ(Fit) standardized	SSQ(Fit) relative fit	SSQ(Res) standardized	SSQ(Res) relative fit
PSP	.18	.16	.93	.01	.07
PvdA	.17	.16	.95	.01	.05
KVP	.16	.15	.93	.01	.07
ARP	.16	.14	.91	.01	.09
VVD	.18	.16	.92	.01	.08
CHU	.15	.14	.91	.01	.09
overall	1.00	.92		.08	
	SSQ(Total)	= SSQ(Fit)	+	SSQ(Res)	

Notes: 1) 'standardized' means devided by the overall SSQ (Total)
2) relative fit = SSQ (Fit of a party)/SSQ (Total of a party)

91%, which is equal to the weight of the first component of the third mode, as it should be. The core matrix thus breaks the SSQ(Fit) up into small parts, through which the (possibly) complex relations between components can be analysed. It is in this way that we can interpret the core matrix as the generalization of the eigenvalues or singular values of the Eckart-Young decomposition. It constitutes a further partitioning of the 'explained' variation as is indicated by the eigenvalues of the standard principal components. The present example is in a way too simple to make full use of the interpretational possibilities of the core matrix, as all off-diagonal elements are virtually zero.

Table 7 Sums of squares broken down by parties used as standards

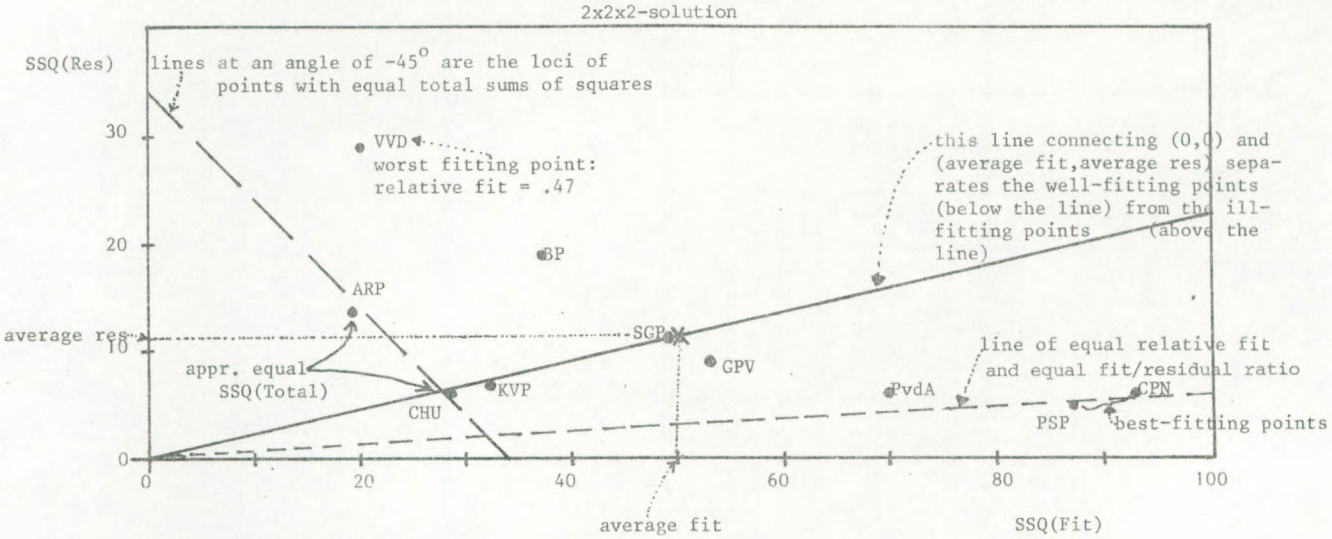
	3x3x2-solution			2x2x2-solution			
Party	SSQ (Total) standardized	SSQ(Fit) standardized	SSQ(Fit) relative fit	SSQ(Fit) standardized	SSQ(Fit) relative fit	+ fit residual ratio	Improvement in relative fit due to third component
CPN	.16	.16	.95	.15	.94	16.5	.01
PSP	.15	.15	.95	.15	.95	18.5	.00
PvdA	.13	.12	.94	.12	.92	11.5	.02
KVP	.06	.05	.84	.05	.82	4.6	.02
ARP	.06	.05	.88	.03	.59	1.4	.29
VVD	.08	.08	.93	.03	.41	0.7	.52
CHU	.06	.05	.83	.05	.83	4.8	.00
SGP	.10	.10	.95	.09	.85	5.8	.10
GPV	.10	.10	.95	.08	.82	4.5	.13
BP	.09	.09	.90	.06	.66	1.9	.24
overall	1.00	.92		.82			

Notes: 1. 'standardized' means devided by the overall SSQ (Total)
2. relative fit = SSQ (Fit of a party)/SSQ (Total of a party)
3. fit/residual ratio = SSQ (Fit of a party)/SSQ (Residual of a party)
\+ not yet printed in the program, but may well be in a next version

Assessing the quality of fit

In essence the analysis could stop with the above interpretations. All that the technique could offer towards breaking the complex relationships down into small intelligible pieces is contained in the information by the program so far. However, it is good to have some ancillary information available to assess if there are no irregularities in the data like outliers, unduly influential points, points

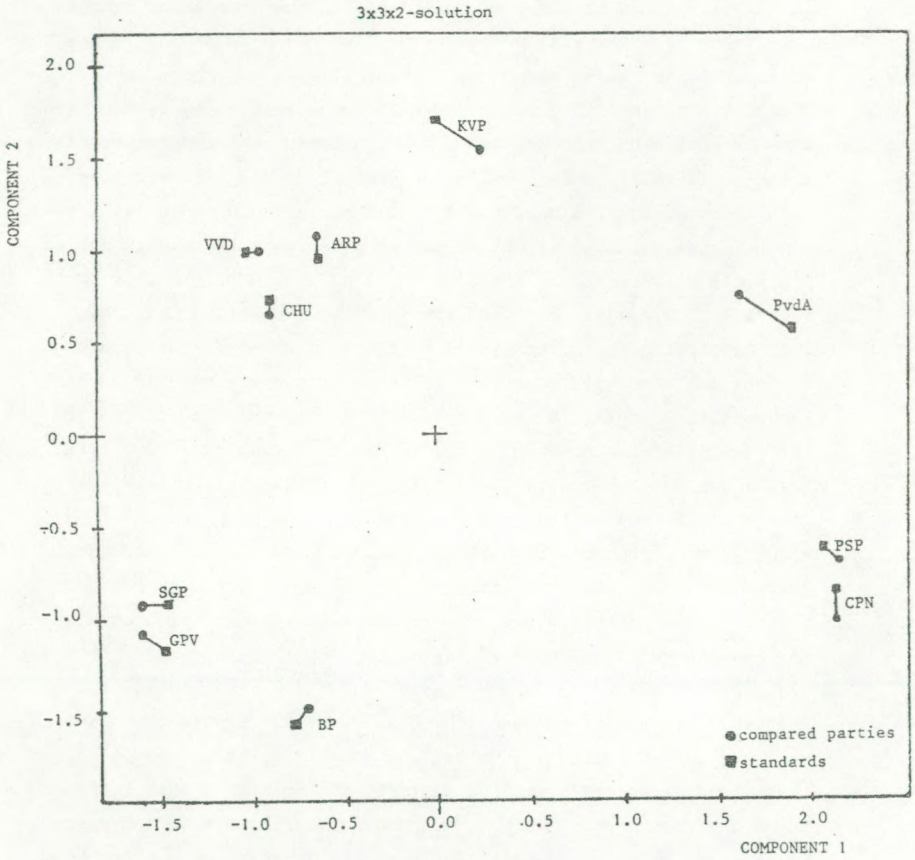
Fig. 6 Fitted sums of squares versus residual sums of squares for mode 1 (parties used as standards)



which are not sufficiently accounted for, etc. An attractive way to investigate such questions is to inspect the residual sums of squares in conjunction with the fitted sums of squares (see also "Analysis of Residuals"). Whereas the core matrix informs us about the contributions of the components and their interrelationships, so do the sums of squares broken down by the elements of the modes tell us about their contributions to the solutions.

Let us first turn to the sums of squares of the preference groups (see Table 6). Now that we have equalized the n_g's of the groups, it is possible to assess the overall variability between the groups. The ARP, CHU, and KVP groups tended to judge relatively less consistent or outspoken than the PSP and VVD groups, as a comparison of their SSQ(Total)'s shows. Although very often elements with larger total sums of squares tend to be fitted better than those with smaller sums of squares, this effect is hardly present in these data due to the small differences in the SSQ(Total)'s. The SSQ(Fit)'s or better relative fits are all high and rather alike indicating that the common solution is shared by all to the same extent. The sums of squares for the third mode thus tells us that everything is in order. Far more interesting is the comparison between the sums of squares of the stimuli (here shown for the first mode) in the 3x3x2 and 2x2x2-solutions (see Table 7). The total sums of squares show that parties in the middle look alike, and thus the students have not one mind which parties resemble each other the most. Most of the students agreed, however, about the similarities of the parties at the end points of the political spectrum to other parties, leading to higher and lower scores, and thus to larger SSQ(Total)'s (see also Table 1). The SSQ(Fit)'s for the 2x2x2-solution show that the relative fit of the VVD, ARP, and BP leaves much to be desired. Figure 6 summarizes most of the information of Table 7, and makes it easy to spot especially the ill-fitting and the well-fitting points. Table 7 also shows that not much is gained by adding a fourth axis to the party space. Not only is the overall fit of the 3x3x2-solution very good, but also all parties now fit more or less equally well. In other words the third component was all that was needed to accomodate remaining anomalies.

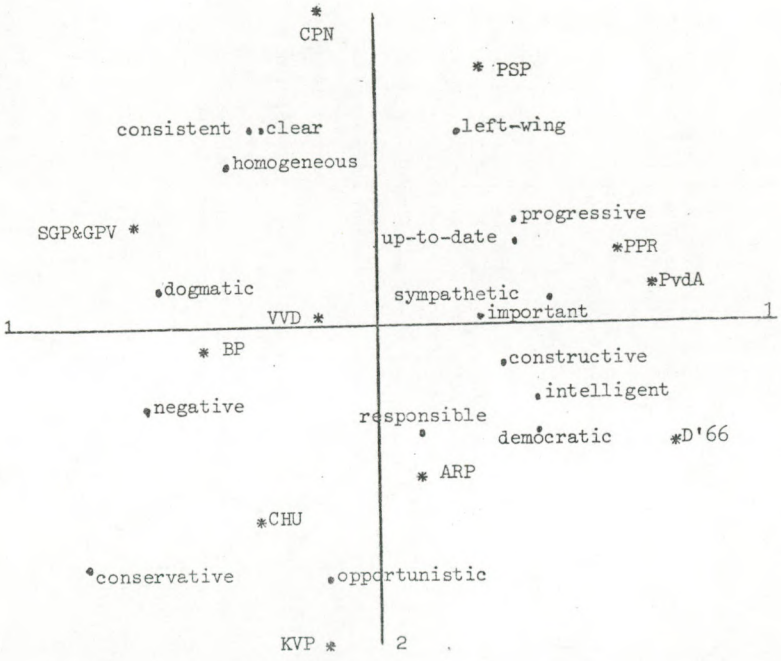
Fig. 7 Joint plot of first mode (standards) and second mode (compared parties)



Joint plots

Figure 7 shows a joint plot (see "Joint Plots") of the first and second mode. In this case all it tells us is that the original data matrices are virtually symmetric, which is by now no longer a surprise. For many other data sets, however, in which the first and second component are different types of variables, these joint plots are a major aid in interpretation. To illustrate this we show such a joint plot from another study of political parties in the same period (see Figure 8). Here staff and students from a psychology department indicated whether parties possessed certain attributes or not. For a more detailed account of these data and their three-mode analysis see Kroonenberg & De Leeuw (1977). It is left to the reader to decide where the judges' sympathy lay.

Fig. 8 Joint plot of attribute and political party space



INPUT SUMMARY

CARD 1 KEY WORD (COL.1-4) AND TITLE (COL.5-80)

'TITL'	title card
'SIZE'	compute necessary core space
'INFO'	produce the program information and this input summary
'. '	signals end of input (obligatory)

CARD 2 DATA SPECIFICATIONS:

COLUMN		DESCRIPTION
1- 5	L	number of elements in first mode (e.g. subjects)
6-10	M	number of elements in second mode (e.g. variables)
11-15	N	number of elements in third mode (e.g. conditions)
16-20	S	number of components in first mode (S should be smaller or equal to L)
21-25	T	number of components in second mode (T should be smaller or equal to M)
26-30	U	number of components in third mode (U should be smaller or equal to N)

CARD 3 INPUT AND OUTPUT SPECIFICATIONS:

1- 5	INP	unit number for data input (0=DEFAULT=STANDARD INPUT UNIT CARDS)
6-10	IPLOT	plot component scores; (0=NO=DEFAULT; 1=YES)
11-15	IPR	level of print-out; (0=STANDARD PRINT-OUT=DEFAULT; 1=ITERATION TABLE EXTRA)
16-20	IOUTG	unit number for output component scores of first mode other than printer (0=NO EXTRA OUTPUT=DEFAULT)
21-25	IOUTH	unit number for output component scores of second mode other than printer (0=NO EXTRA OUTPUT=DEFAULT)
26-30	IOUTE	unit number for output component scores of third mode other than printer (0=NO EXTRA OUTPUT=DEFAULT)
31-35	IOUTCP	unit number for joint plots other than printer (0=NO EXTRA OUTPUT=DEFAULT)
41-45	IDATPR	printing of input data (0=NO=DEFAULT;1=YES)
51-55	ISYM	restricted output due to symmetric frontal planes (0=NO=DEFAULT;1=YES)
56-60	INCONF	printing of initial configuration (0=NO=DEFAULT;1=YES)
61-65	IOURES	unit number for output individual squared residuals (0=NO OUTPUT=DEFAULT)
66-70	ICPLOT	printing of joint plots (0=NO=DEFAULT, 1=MODES 1&2,2=MODES 2&3,3=MODES 3&1,4=ALL MODES)
71-75	IOUFRS	unit number for output fitted and residual sums of squares per mode (0=NO OUTPUT OTHER THAN PRINT=DEFAULT)

CARD 4 ANALYSIS SPECIFICATIONS:

1- 5	IDBCEN	centring per frontal plane (0=NO,1=DOUBLE-CENTRING,2=CENTRE ROWS,3=CENTRE COLUMNS,4=CENTRE PLANE)
6-10	NIT	maximum number of iterations (DEFAULT=50)
11-15	MSTORE	internal storage mode of data (0=IN CORE=DEFAULT;1=ON DISK)
26-35	EPS1	convergence criterion for optimal factorisation (DEFAULT=.0001)
36-45	EPS2	convergence criterion for component matrices (DEFAULT=.0001)

CARD 5 LABELS FOR THE THREE MODES:

1-12	MLABI	label first mode
13-24	MLABJ	label second mode
25-36	MLABK	label third mode

CARD 6,7,8 VARIABLE INPUT FORMAT (THREE CARDS)

Here after DATA CARDS if the data are to be read from the same input unit as the input parameters otherwise the next KEY-AND-TITLE CARD should follow. If the present job is the last one the next KEY-AND-TITLE CARD should indicate end of input ('. ')

REFERENCES

Carroll, J.D. & Arabie, P., Multidimensional scaling. Annual Review of Psychology, 1980, 31, 607-649.

Carroll, J.D. & Chang, J.J. IDIOSCAL: A generalization of INDSCAL allowing IDIOsyncratic reference systems as well as an analytic approximation to INDSCAL. Paper presented at the Spring Meeting of the Psychometric Society, Princeton, N.J., March, 1972.

Carroll, J.D. & Wish, M. Models and methods for three-way multidimensional scaling. In D.H. Krantz et al. (Eds.) Contemporary Developments in Mathematical Psychology, Vol. II. San Francisco: W.H. Freeman & Cy., 1974.

De Gruijter, D. Cognitive structure of Dutch political parties (Report E019/67). Leiden, The Netherlands: Psychologisch Instituut, Rijksuniversiteit Leiden, 1967.

Endler, N.S., Hunt, J. Mc. V. & Rosenstein, A.J. An S-R inventory of anxiousness. Psychological Monographs, 1962, 76 (17, Whole Nr. 536).

Gifi, A. Niet-lineaire multivariate analyse. Leiden, The Netherlands: Department of Data Theory, University of Leiden, 1980 (preliminary edition).

Jennrich, R. A generalization of the multidimensional scaling model of Carroll & Chang (Working Papers in Phonetics No. 22). Los Angelos: University of California, 1972.

Jones, L.E. & Young, F.W. Structure of a social environment: longitudinal individual differences scaling of an intact group. Journal of Personality and Social Psychology, 1972, 24, 108-121.

Kroonenberg, P.M. & De Leeuw, J. TUCKALS2: A principal component analysis of three mode data (Research Bulletin RB 001-77). Leiden, The Netherlands: Department of Data Theory, University of Leiden, 1977.

Kroonenberg, P.M. & De Leeuw, J. Principal component analysis of three-mode data by means of alternating least-squares algorithms. Psychometrika, 1980, 45, 69-97.

Kroonenberg, P.M., Scaling of input data for three-mode principal component analysis (WEP-reeks, WR 81-21-EX). Leiden, The Netherlands: Vakgroep Wijsgerige en Empirische Pedagogiek, Rijksuniversiteit Leiden, 1981.

Kruskal, J.B. & Wish, M. Multidimensional Scaling (Sage University paper nr. 11). Beverly Hills, Calif.: Sage Publications, 1978.

Levelt, W.J.M., Van de Geer, J.P. & Plomp, R. Triadic comparisons of musical intervals. British Journal of Mathematical and Statistical Psychology, 1966, 19, 163-179.

Levin, J. Three-mode factor analysis. (Unpublished doctoral thesis). Urbana, III: University of Illinois Press, 1963.

Levin, J. Three-mode factor analysis. Psychological Bulletin, 1965, 64, 442-452.

Osgood, C.E., Suci, G.J. & Tannenbaum, T.H. The measurement of meaning. Urbana, III: University of Illinois Press, 1957.

Rutishauser, H. Computational aspects of F.L. Bauer's simultaneous iteration method. Numerische Mathematik, 1969, 13, 4-13.

Snyder, F.W. & Wiggins, N. Affective meaning systems: a multivariate approach. Multivariate Behavioral Research, 1970, 5, 454-468.

Takane, Y., Young, F.W. & De Leeuw, J. Non-metric individual differences multi-dimensional scaling: An alternating least squares method with optimal scaling features. Psychometrika, 1977, 42, 7-67.

Tucker, L.R. Implications of factor analysis of three-way matrices for measurement of change. In C.W. Harris (Ed.) Problems in measuring change. Madison, Wis.: University of Wisconsin Press, 1963, 122-137.

Tucker, L.R. The extension of factor analysis to three-dimensional matrices. In Gulliksen, H. and Frederiksen, N. (Eds.) Contributions to Mathematical Psychology, New York: Holt, Rinehart & Winston, 1964.

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

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

Van de Geer, J.P. Toepassing van drieweg-analyse voor de analyse van multiple tijdreeksen. In C.J. Lammers, Groei en ontwikkeling van de ziekenhuisorganisaties in Nederland (interimrapport). Leiden, The Netherlands: Sociologisch Instituut, Rijksuniversiteit Leiden, 1974.

Wolters, M. Dimensions of Dutch Parties (Research Bulletin, RB 003-75). Leiden, The Netherlands: Department of Datatheory, University of Leiden, 1975.