2. CANDECOMP (CANonical DECOMPosition)

Page No.

1. OVERVIEW

Concisely: CANDECOMP (CANonical DECOMPosition) provides provides internal analysis of a 3- to 7-way data matrix of (dis)similarity matrices, by a weighted scalar product distance model using a linear transformation of the data.

Following the categorisation developed by Carroll and Arabie (1979) the program may be described as:

Data:	Three- to seven-way	Model:	Generalised Scalar products
	Two- to seven-mode		Two to seven sets of points
	Polyadic		Internal or External
	Linear		
	Complete		

1.1 ORIGIN, VERSIONS AND ACRONYMS

The present CANDECOMP program performs the analysis described in Carroll and Chang (1970) as canonical decomposition of N-way matrices. The original INDSCAL program performed both this N-way analysis and contained as a special case, the 3-way analysis which became known as the INDSCAL model (see 1.3). The present program is adapted from the original (1971) INDSCAL program.

1.2 CANDECOMP IN BRIEF

CANDECOMP takes as input a table of data values with between three and seven "ways". In the solution, each of these ways is represented by a configuration of points representing the elements of that particular way in a space of chosen dimensionality. Each data value is regarded as being the scalar product between the relevant elements. The program assumes that the data are at the interval level of measurement.

1.3 RELATION OF CANDECOMP TO OTHER MDS(X) PROGRAMS

CANDECOMP may be used to perform individual differences analysis if there are more than three ways (e.g. if the study involves replications).

The present program is a modified version of Carroll and Chang's original INDSCAL program. The so-called INDIFF option in that program (i.e. the special case when there were three ways and two modes in the data) became generally known, rather confusingly, as the INDSCAL model or, simply, "individual differences scaling". This INDIFF option now forms the INDSCAL-S program in the MDS(X) series, while this program provides the full range of options available in Carroll and Chang's original program.

2. DESCRIPTION

2.1 DATA

There are two basic forms of data input to CANDECOMP, which we will refer to as being applicable to

1. an "extended INDSCAL" analysis

and 2. the CANDECOMP analysis proper.

What we call the 'extended INDSCAL' refers to the case where two of the ways of the matrix refer to the same set of objects, that is one of the matrices is square and the row- and column-elements refer to the same set of objects. These objects will be represented by only one configuration in the output. By contrast all the ways of the CANDECOMP analysis are regarded as distinct.

2.1.1 The extended INDSCAL analysis

Users who wish to analyse three-way, two-mode data are referred to the INDSCAL-S program.

In an INDSCAL analysis of this sort we have a set of matrices obtained from a set of subjects. Each matrix is a matrix (dis)similarity coefficients of some sort between a set of stimuli. There will thus obviously be as many matrices as there are subjects and each matrix will have as many rows as there are stimuli. The INDSCAL model analyses the way in which the subjects differentially perceive the stimuli. Suppose that we are interested in extending this analysis to take account of the effect of other factors. We might, for instance, replicate a study, use different forms of data collection, split subjects into some rational groupings etc. etc., and wish to use the INDSCAL model to analyse the effects of these factors by the same model as we used to investigate the subjects in the original analysis.

If the user is analysing data of this type, then the parameter SET MATRICES should be given the value 1 on the PARAMETERS card. This tells the program that two of the ways of the matrix - those corresponding to the stimuli - are identical and should be set equal (see 2.2). The DATA TYPE parameter should also be given a suitable value. Users should read 2.1.3 for a description of the use of the SIZES parameter.

2.1.2 The CANDECOMP analysis

As we have noted, this 'extended INDSCAL' analysis is a special case of the general CANDECOMP analysis where two of the ways are identical. We now consider the general case, where all the ways are considered distinct. (They need not, of course, actually be distinct sets of entities, they will merely be regarded as such by the program).

Consider the paradigm case where a set of subjects has given numerical ratings to a set of stimuli on a number of criteria. Since the procedure is linear, the use of rankings is not recommended. The data consist of a set of matrices, one for each criterion, each of which contains as many rows as there are subjects and as many columns as there are stimuli. If such a study was replicated after a period of time, thus forming a fourth way, then the resulting data constitute another block of such matrices.

The default parameter values allow for this analysis.

2.1.3 The presentation of data to CANDECOMP

Data are read by the READ matrix card under the associated INPUT FORMAT card. The dimensions of the input matrix are given to the program by means of the SIZES control card which is peculiar to CANDECOMP. This card replaces the N OF SUBJECTS, N OF STIMULI cards which are not recognised by this program. The card takes as operand up to seven numbers, separated by commas each of which is the number of objects in one of the ways of the matrix. There are as many numbers as there are ways in the data.

2.1.3.1 The order of the SIZES card

The order in which the ways are entered on the SIZES card is crucial.

The number of columns in the data matrix should be specified as the third number on the SIZES card.

The number of rows in the basic matrix should be the second number on the card.

The number of matrices in the third way is the first number.

The number of elements in the fourth, fifth, sixth and seventh ways is given by the fourth, fifth, sixth and seventh numbers respectively.

In the case of the extended INDSCAL analysis, the first and second ways are identical, thus the second and third numbers on the SIZES card must be equal.

2.1.3.1.1 Example

Suppose we are interested in assessing the sound-quality of stereo amplifiers,* and that we have ten different makes of equipment. We gather together say twenty listeners and proceed in the following way. A tape containing extracts of different types of music and speech is

---

*Thanks are due to S.P. Thomas and Q. Deane of the Consumers Association for suggesting this application and describing the basic form of the experiment.

played to the listeners using each of the amplifiers in turn. Before each of the amplifiers is used the tape is played through a 'reference' machine. The listeners are asked to assess each of the sets on, say, five criteria (e.g. distortion, frequency response and channel separation.

This assessment is done on a nine-point scale in comparison with the reference set which is scored as an arbitrary 5. Thus, so far we have a three-way data matrix, listeners x amplifiers x criteria. Since it is possible that some of the criteria may be influenced by the characteristics of, say, the speakers used in the reproduction of the tape, a further way might be added by playing the tape through each amplifier, say, four times, each time through a different set of speakers. Replications in say, three rooms of different acoustic properties might constitute a fifth way, and if we were foolhardy and/or rich enough to repeat the whole procedure, without serious revolt from the listeners, we might add a sixth way. Thus we have 20 listeners, 10 sets, 5 criteria, 4 speakers, 3 rooms and 2 replications.

Arranging the data so that the sets (in which we are primarily interested form the rows of the matrix (see 2.2)) our data look like this.

Each matrix has ten rows and five columns, this being the set of ratings given to each of the sets on each of the criteria by <u>one</u> of the listeners and there will be twenty such matrices corresponding to the twenty listeners. (i.e. (20 x 10) = 200 lines in all, since the matrices follow each other without break). There will then be another three such blocks of 200 lines (making four blocks, 800 lines in all) corresponding to the different speaker types. Each of the three rooms will have provided 800 lines in this way, making 2400 lines and since there are two replications there will be in all 4800 lines, each of five columns in the data matrix. The SIZES card corresponding to this matrix would be

col 1	col 16
SIZES	20, 10, 5, 4, 3, 2

2.2. THE MODEL

The CANDECOMP program generates one configuration for each way of the analysis and the number of points in each configuration will be the number of elements in the corresponding way of the matrix. In the extended INDSCAL analysis however (i.e. when SET MATRICES (1)) matrices two and three - those corresponding to the second and third numbers on the SIZES card - are set equal when the algorithm has converged. One more iteration is then performed and only one configuration then produced for this way of the data (see INDSCAL-S).

The axes of the solution space are identical in each configuration and the solution should be interpreted in relation to these axes which it has usually been found, yield readily to substantive interpretation. Each configuration then reflects the differential importance of the properties represented by the axes in the following way. Each point in each configuration is properly considered as the terminus of a vector drawn from the origin of the space and for each vector the ratio between its coordinate on axis a and on axis b reflects the differential importance of the properties represented by those axes in the judgement of that subject and analysis should focus on this patterning.

All the configuration are normed so that the sum of squares of the coordinates on each axis is unity except for matrix 1. This means that strictly speaking the patterning of weights (coordinates) is comparable across 'ways'. It is not, however, clear how this is to be interpreted in the general case. The first matrix, being un-normed, will tend to show greater dispersion among the vectors and it is recommended that the 'way' in which the user wishes to concentrate forms the first way of the data. (i.e. the second element on the SIZES card).

2.2.1. The algorithm

1. The input data matrices are converted into matrices of scalar products.

2. The scalar products between the elements in the input configuration input by the user or generated by the program are calculated to serve as initial estimates of the solution.

3. Each scalar product is assumed to be the result of the vector multiplication of as many vector coordinates as there are ways in the data matrix. At each iteration, all but one of these is held constant while the remaining parameter (coordinate) is estimated.

4. When this process has converged, the two matrices referring to the symmetric matrix are set equal (if SET MATRICES (1)), the appropriate normalisation performed (see 2.3.1) and the solution output.

2.3 FURTHER FEATURES

2.3.1 Normalisation options

Two different questions of normalisation arise: over the input data and over the solution.

2.3.1.1 Normalisation of the data input

If the program is being used to perform a higher-way INDSCAL analysis, then the input matrices are normalised so that the influence of each subject is equalised in the analysis before the data are converted to scalar products. When a set of covariances or correlations are input the program does not convert to scalar products (since both covariances and correlations are scalar products) and, in the case of correlations, neither does it normalise. It is therefore important that data of this type be announced to the program by means of the relevant DATA TYPE parameter value.

In the case of the general CANDECOMP analysis the data are not normalised and differences in magnitude between subjects' judgements will affect the analysis. It is recommended, however, that the data for a CANDECOMP analysis be centred before the analysis proceeds both to provide a common origin for the various 'ways' and to eliminate consensual effects which often overwhelm fine structural detail.

2.3.1.2 Normalisation of the solution

Each of the configurations except that referring to the subjects of the solution is normalised as noted above (2.2). It is therefore recommended that the way in which the user wishes more variation to be concentrated form the first way (rows) of the input matrix.

It should, however, be noted that differences in the magnitude of scales needed by different subjects will affect the length of the vectors (the distance of a particular point from the origin) in this space and it is more than ever important to concentrate on the ratio between the coordinates on the respective axes.

2.3.2 Initial configuration

An initial configuration, which provides the initial estimates for the iterative procedure, is normally generated by the program from a pseudo-random distribution. CANDECOMP is prone to suboptimal solutions and users are recommended to make a number of runs with different starting configurations. A series of similar (preferably identical) solutions will usually indicate that a global minimum has been found.

2.3.2.1 Initial configuration for the extended INDSCAL option

If the CANDECOMP program is being used to perform the extended INDSCAL analysis (i.e. SET MATRICES(1)) then the user may choose to input an initial configuration of the points represented by the symmetric matrix (the stimulus matrix). This may be an a priori guess at the solution or the result of a MINISSA analysis in which the averaged judgements have been analysed. In this case the configuration is input after the READ CONFIG control card. It consists of the coordinates of the stimulus points in the maximum dimensionality requested. These are read according to the associated INPUT FORMAT card.

2.3.3 External analysis

Users may wish to use CANDECOMP to perform an "external" INDSCAL analysis by holding constant a known configuration and estimating the configurations of subjects etc. This may be done only if SET MATRICES(1). A configuration is input by the user as described above and the FIX POINTS parameter is set to 1 on the PARAMETERS card. The program will then estimate only the remaining matrices.

3. INPUT PARAMETERS

3.1 LIST OF PARAMETERS

Keyword	Default	Function
DATA TYPE	0	0: An N-way table is input. 1: Lower triangle similarity matrix. 2: Lower triangle dissimilarity matrix. 3: Lower triangle matrix of distances. 4: Lower triangle correlation matrix. 5: Lower triangle covariance matrix. 6: Full symmetric similarity matrix. 7: Full symmetric dissimilarity matrix.
RANDOM	12345	(Any positive integer) Seed for pseudo; random number generator.
SET MATRICES	0	0: The CANDECOMP analysis is performed. 1: The performed extended INDSCAL analysis is performed (matrix 2 and 3 are set equal.
FIX POINTS	0	0: Iterate and solve for all matrices. 1: One matrix is held constant (external analysis).
CRITERION	0.005	(values between 0 and 1) Sets improvement level for terminating iterations.
CENTRE	0	0: No action. 1: If an N-way table is input (DATA TYPE (0)) it will be centred by subtracting the 'row means' in each of the N-ways (see section 2.3.1).

3.2 NOTES

1. The control card SIZES is obligatory for CANDECOMP.

2. The control cards $\begin{matrix} \text{N} \\ \text{NO} \end{matrix}$ OF $\begin{matrix} \text{SUBJECTS} \\ \text{STIMULI} \end{matrix}$ are not valid with CANDECOMP.

3. When DATA TYPE takes values 1 through 5 no diagonal is input. For values 6 and 7 the diagonals are input but ignored.

4. In the parameters SET MATRICES and FIX POINTS the spaces are significant characters.

5. Program Limits

Maximum no. of dimensions	=	10
Maximum no. of elements per way	=	100
Way 1 × Way 2 × Way 3	$\leq$	1800

3.3. PRINT, PLOT AND PUNCH OPTIONS

The general format for printing, plotting and punching options are as follows. We let n denote the number of ways in the analysis $(3 \leq n \leq 7)$, m the number of modes $(2 \leq m \leq 7)$.

3.3.1 PRINT options

Option	Form	Description
INITIAL	n matrices will be printed.	The initial estimates of the configurations are printed. Each matrix contain the coordinates of the points on the required dimension. If the user has input an initial configuration, then the second two matrices will be identical.
FINAL	m matrices are printed.	The solution configurations are printed. Each matrix contains the coordinates of the relevant number of points on the axes of the space. These are followed by the correlations between each subject's data and solution. The matrix of cross-products between the dimensions is printed.
HISTORY		The overall correlation at each iteration is printed. The unnormalised matrices at convergence are also printed (there will be n of these).

By default only the FINAL matrices and the overall correlation at convergence are printed.

3.3.2 PLOT options

Option	Description
INITIAL	The initial configuration may be plotted as r(r-1)/2 plots only if one has been input by the user.
CORRELATIONS	The overall correlation at each iteration is plotted in the form of a histogram.
WAY1 WAY2 WAY3 WAY4 WAY5 WAY6 WAY7	r(r-1)/2 plots are produced for each way specified.

3.3.3 PUNCH options

Option	Description
FINAL	The configuration of points for each way in the chosen dimensionality is punched in a fixed format.
CORRELATIONS	The overall correlation at each iteration is output in a fixed format.

4. EXAMPLES

4.1 TEST RUN

```
col 1                col 16

RUN NAME             EXAMPLE FROM SEC. 2.1
TASK NAME            LISTENING TESTS AB NAUSEAM
DIMENSIONS           4 TO 2
SIZES                20,10,5,4,3,2
INPUT MEDIUM         DISC
INPUT FORMAT         (5F3.0)
READ MATRIX

   <all the data follow here>

COMPUTE
FINISH
```

BIBLIOGRAPHY

Bloxom, B. (1965) Individual differences in multidimensional scaling, Princeton University Educational Testing Service Research Bulletin, 68-45.

Carmone, F.J., P.E. Green and P.J. Robinson (1968) TRICON: an IBM 360/65 program for the triangularisation of conjoint data, Journal of Marketing Research, 5, 219-20.

Carroll, J.D. (1974) Some methodological advances in INDSCAL, mimeo, Psychometric Society, Stanford.

Carroll, J.D. and P. Arabie (1979) Multidimensional scaling, in M.R. Rozenzweig and L.W. Porter (eds.) 1980 Annual Review of Psychology, pp 607-649, Palo Alto Ca., Annual Reviews.

Carroll, J.D. and J.J. Chang (1970) Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition, Psychometrika, 35, 283-319.

Carroll, J.D. and M. Wish (1974) Multidimensional perceptual models and measurement methods, in E.C. Carterette and M.P. Friedman Handbook of Perception, Vol.2, New York: Academic Press (Ch. 5 Individual differences in perception).

Carroll, J.D. and M. Wish (1975) Models and methods for three way multidimensional scaling, in R.C. Atkinson, D.H. Krantz, R.D. Luce and P Suppes (eds.), Contemporary Methods in Mathematical Psychology, San Francisco: Freeman.

Coxon, A.P.M. and C.L. Jones (1974) Applications of multidimensional scaling techniques in the analysis of survey data, in C.J. Payne and C.O'Muircheartaigh, Survey Analysis, London: Wiley.

Gower, J.C. ( ) The analysis of three-way grids, in P. Slater (ed.) Dimensions of Intrapersonal Space (Vol.2), London: Wiley.

Horan, C.B. (1969) Multidimensional scaling: combining observations when individuals have different perceptual structure, Psychometrika, 34, 2, pt.1, 139-165.

Jackson, D.N. and S.J. Messick (1963) Individual differences in social perception, British Journal of Social Clinical Psychology, 2, 1-10.

Kruskal, J.B. (1972) A brief description of the 'classical' method of multidimensional scaling, Bell Telephone Laboratories, mimeo.

Tagg, S.K. (1979) The analysis of repertory grids using MDS(X), MDS(X) Project working paper.

Torgerson, W.S. (1958) Theory and methods of scaling, New York: Wiley.

Tucker, L.R. (1960) Intra-individual and inter-individual multidimensionality, in H. Gulliksen and S. Messick (eds.), Psychological scaling: Theory and applications, New York: Wiley.

Wish, M. and J.D. Carroll (1974) Applications of individual differences scaling to studies of human perception and judgment, in Carterette and Friedman (1974): see Carroll and Wish 1974 above.

Wold, H. (1966) Estimation of principal components and related models by iterative least squares, in P. Krishnaiah (ed.), International Symposium on multivariate analysis, New York: Academic Press.

APPENDIX 1

No other known programs perform the CANDECOMP type of analysis

APPENDIX 2

This appendix is based on Carroll and Chang (1970) which is used with permission.

A2.1 The three way model

The model in the three way case is a generalisation of the INDSCAL model (see INDSCAL-S) except that we seek

$$z_{ijk} \cong \sum_{a}^{r} w_{ia} x_{ja} y_{ka} \tag{1}$$

where

$i = 1,\ldots,n_1$
$j = 1,\ldots,n_2$
$k = 1,\ldots,n_3$

Given then initial estimates of the x's and y's we may derive least-squares estimates for the w's by standard multiple regression procedures

$$\underset{\sim}{W} = \underset{\sim}{P}\,\underset{\sim}{Q}^{-1} \tag{2}$$

where

$$\underset{\sim}{W} = w_{ia}$$

and

$$\underset{\sim}{P} \equiv \{p_{ia} = \sum_{j}^{n_2} \sum_{k}^{n_3} z_{ijk} x_{ja} y_{ka}\} \tag{3}$$

and

$$\underset{\sim}{Q} = \{q_{ab} = \sum_{j}^{n_2} \sum_{k}^{n_3} (x_{ja} y_{ka}) (x_{jb} y_{kb}) = (\sum_{j}^{n_2} x_{ja} x_{jb}) (\sum_{k}^{n_3} y_{ka} y_{kb}) \} \tag{4}$$

where

$a = b = 1,\ldots,r$

We may see that these equations do give a least-squares estimate of $\underset{\sim}{W}$ if we rewrite (1) as

$$z^*_{is} \cong \sum_a^r w_{ia}\, g_{sa} \qquad s = j.k \tag{5}$$

or in matrix rotation

$$\underset{\sim}{Z}^* \cong \underset{\sim}{W}\, \underset{\sim}{G}' \tag{6}$$

To determine the least-squares estimate of $\underset{\sim}{W}$ we get

$$\hat{\underset{\sim}{W}} = \underset{\sim}{Z}^*\, \underset{\sim}{G}\, (\underset{\sim}{G}'\underset{\sim}{G})^{-1} \tag{7}$$

This is equivalent to (2) if we identify

$$\underset{\sim}{W} = \hat{\underset{\sim}{W}}$$

$$\underset{\sim}{P} = \underset{\sim}{Z}^*\, \underset{\sim}{G}$$

$$\underset{\sim}{Q} = \underset{\sim}{G}'\underset{\sim}{G}$$

A2.2 Generalisation to the multiway case

Generalisation to higher ways discussed in Carroll and Wish (1975, pp 98-99) makes (1)

$$z_{i_1}, z_{i_2}, \ldots, z_{i_N} = \sum_{a}^{r} x_{i_1 a}^{(1)} \; x_{i_2 a}^{(2)} \ldots x_{i_N a}^{(N)} \qquad (8)$$

while (2) becomes

$$\underset{\sim}{W}_\ell = \underset{\sim}{P}_\ell \, \underset{\sim}{Q}_\ell^{-1} \qquad \ell = 1, \ldots, N \qquad (9)$$

where

$$\underset{\sim}{W}_\ell \equiv \{ w_{i_\ell a}^{(\ell)} \}$$

$$\underset{\sim}{P} \equiv \{ p_{i_\ell a}^{(\ell)} = \sum_{i_1} \sum_{i_2} \ldots \sum_{i_{\ell-1}} \sum_{i_{\ell+1}} \ldots \sum_{i_N} z_{i_1 i_2 \ldots i_\ell \ldots i_N} \, (w_{i_1 a}^{(1)} \; w_{i_2 a}^{(2)} \ldots w_{i_{\ell-1} a}^{(\ell-1)} \; w_{i_{\ell+1} a}^{(\ell+1)} \ldots w_{i_N a}^{(N)} \} \qquad (10)$$

and

$$\underset{\sim}{Q} \equiv \{ q_{aa'}^{(\ell)} = \prod_{\ell' \neq \ell} \; \sum_{i_{\ell'}} w_{i_{\ell'} a}^{(\ell')} \; w_{i_{\ell'} a'}^{(\ell')} \} \qquad (11)$$