MULTIDIMENSIONAL SCALING BY ANALYTICAL SOLUTION OF CUBIC EQUATIONS

C. J. M. van Brunschot *)

This brief article offers a description of a new computer program, SCALCANS, designed for the multidimensional scaling of (dis-)similarities data. The major merit of SCALCANS lies in the embodiment of a new principle for nonmetric multidimensional scaling: analytical solution of a set of cubic equations in unknown coordinates, whereby one coordinate at a time is estimated.

Introduction

In 1962 Roger N. Shepard published two articles in Psychometrika, presenting the first nonmetric multidimensional scaling algorithm (f, g). Since then prominent writers like Kruskal, Lingoes, Guttman, Young and many others have suggested and tested many procedures for nonmetric MDS. In nonmetric MDS we search for a set of inter-point distances that forms a representation of the set of dissimilarities between a number of

*) Institute for Health Care, Tilburg University. The author borrowed many of his insights from the stimulating cooperation with Dr. M. Croon of Psychology Department, Tilburg University.

entities. The systematic treatment of data as to be represented by relations between points in a space stems from Coombs (a). One looks for a scaling solution that has ratio features and is based upon data with ordinal features. Many writers have devoted attention to this interesting phenomenon of 'metric structures in ordinal data' (b, c, g, h, k). The more points are located in a space, the more surplus information is contained by the set of (dis-) similarities, and the more the final configuration shows metric determinacy, which means that points can hardly be moved in space without lowering the degree to which the solution corresponds to the information contained by the (dis-) similarities. In the remainder of this article we shall speak of dissimilarities, recognizing that similarities can very easily be transformed to dissimilarities bij reversing their order.
The relation between dissimilarities δ_{ij} and distances d_{ij} (both $n(n-1)/2$ in number, using n for the number of entities to be scaled) in non-metric multidimensional scaling usually is that of a monotonic transformation:

(1) $\delta_{ij} < \delta_{kl} \Rightarrow d_{ij} \leqslant d_{kl}$, or:

(2) $\delta_{ij} < \delta_{kl} \Rightarrow d_{ij} < d_{kl}$,

depending upon the assumption of weak of strong monotonicity. The deviation from monotonicity usually is expressed in a loss function. The aim is to minimize this loss function. In the method of steepest descent this is tried by moving all points (one by one, and in a series of iterations) in the direction of the negative gradient, using a step size to determine the variable distance with which to move the points. This means that the loss function F is differentiated with respect to each of m coordinates of all n points seperately. The n.m partial derivatives define the directions in which to move the n points in m dimensions with a variabele step size α to approximate the configuration for which the loss function is minimal. The process can be represented by:

(3) $x'_{ik} = x_{ik} - \alpha . \frac{\partial F}{\partial x_{ik}}$,

where $1 \leq i \leq n$ and $1 \leq k \leq m$. In one iteration all coordinates are estimated successively. After each iteration a test is carried out to determine if the loss function still descends. The method of steepest descent has some disadvantages:

* slow minimization process when the configuration approaches the optimum;
* the algorithm can get trapped in a local minimum;
* the determination of a step size. Because the computation of the really optimal step size would ask a lot of computing time, in these algorithms one always accepts some kind of an approximation to the optimal step size.

To overcome these disadvantages proposals have been made for alternative minimization techniques. One of these was put forward by Yoshio Takane, Forrest W. Young and Jan de Leeuw (i). In their article they mention their intention to investigate the loss function minimization process in case one coordinate at a time is estimated. In this case they see a possibility to regard a coordinate x_{ik} as the unknown in a cubic equation that can be

solved analytically. Using this procedure one does not compute step sizes, nor would one meet local minimum problems. This is the principle that found embodiment in SCALCANS.

'SCALCANS' stands for SCALing by ANalytical Solution of cubic equations. Successively attention will be focused on the answers to three questions: (1) how do we form these cubic equations, (2) how do we solve them, and (3) how does this procedure fit into an multidimensional scaling algorithm ? After this presentation of SCALCANS mention will be made of some test results of the program.

The cubic equation

The loss function to be minimized in SCALCANS is:

$$(4)\quad F = \frac{2}{n\,(n-1)} \cdot \sum_{i<j}^{j=n} (d_{ij}^{\,2} - h_{ij})^2 ,$$

which measures the sum of squared deviations of squared distances from corresponding fitted values. The use of squared distances rather than distances is, as will become clear, dictated by the application of Cardan's analytical solution. The

fitted values form a transformation of squared distances that satisfies the requirement:

$$(5) \quad \delta_{ij} < \delta_{kl} \Longrightarrow h_{ij} \leq h_{kl},$$

an analogue to the requirement of weak monotonicity (1).

The sum of squares is multiplied by $2/n(n-1)$, or divided by $n(n-1)/2$, to correct the loss function for the number of points and, consequently, for the number of (dis-)similarities.

Minimization of F takes place in a number or cycle of iterations and each iteration can be divided into two phases. In the first phase the loss function is minimized by changing the fitted values, holding constant the n.m coordinates, by Kruskal's monotonic regression procedure with primary treatment of ties. So, the set of fitted values not only satisfies (5) but also is a conditional least squares approximation of the squared distances. We chose primary treatment of ties remembering Roskam's (d) warning that with secondary treatment of ties the user makes the bold assumption that each found minor difference between dissimilarities is meaning-

ful.

In the second phase minimization is achieved by changing the coordinates, holding constant the n (n-1)/2 fitted values. This job is done by solving a set of cubic equations. The present section will show how these equations are formed.

Minimization of the loss function (4) is equivalent to minimization of the uncorrected function:

$$(6)\quad F^{*} = \sum_{i<j}^{j=n} (d_{ij}^{2} - h_{ij})^{2}, \text{ whereby:}$$

$$(7)\quad d_{ij}^{2} = \sum_{t=1}^{t=m} (x_{it} - x_{jt})^{2}.$$

Essentially, a cubic equation is obtained by setting the partial derivative of F^{*} equal to zero:

$$(8)\quad \frac{\partial F}{\partial x_{ik}} = 0.$$

It can be shown that this equation is identical to:

$$(9)\quad 4 \sum_{\substack{j=1 \\ j\neq i}}^{j=n} \left\{ (x_{ik} - x_{jk})^{2} + q_{ijk} - h_{ij} \right\} (x_{ik} - x_{jk}) = 0,$$

whereby:

$$(10)\quad q_{ijk} = \sum_{\substack{t=1 \\ t\neq k}}^{t=m} (x_{it} - x_{jt})^2.$$

By some simple algebraic manipulations equation (8) can be rewritten as:

$$(11)\quad e_1 x_{ik}^3 + e_2 x_{ik}^2 + e_3 x_{ik} + e_4 = 0, \text{ whereby:}$$

$$(12)\quad e_1 = 4(n-1),$$

$$(13)\quad e_2 = 4 \sum_{\substack{j=1 \\ j\neq i}}^{j=n} (-3 x_{jk}),$$

$$(14)\quad e_3 = 4 \sum_{\substack{j=1 \\ j\neq i}}^{j=n} (q_{ijk} - h_{ij} + 3 x_{jk}^2),$$

$$(15)\quad e_4 = 4 \sum_{\substack{j=1 \\ j\neq i}}^{j=n} (h_{ij} x_{jk} - q_{ijk} x_{jk} - x_{jk}^3), \text{ and:}$$

$$(16)\quad q_{ijk} = \sum_{\substack{t=1 \\ t\neq k}}^{t=m} (x_{it} - x_{jt})^2.$$

Cardan's solution

By solving the cubic equation (11) we find the coordinate x_{ik} that minimizes F^*, other things being equal. This x_{ik} also minimizes the loss function F corrected for the number of points.

How can (11) be solved analytically ? for this problem we make use of Cardan's method (e). The cubic equation can, by dividing all terms by e_1 (2), be written as:

$$(17)\quad x_{ik}^3 - ax_{ik}^2 + bx_{ik} + c = 0.$$

The substitution $x_{ik} = y_{ik} - \frac{a}{3}$ reduces this equation to:

$$(18)\quad y_{ik}^3 + py_{ik} + q = 0, \text{ whereby:}$$

$$(19)\quad p = -\frac{a^2}{3} + b, \text{ and:}$$

$$(20)\quad q = 2\left(\frac{a}{3}\right)^3 - \frac{ab}{3} + c.$$

When we define:

$$(21)\quad A = \left(-\frac{q}{2} + Z^{\frac{1}{2}} \right)^{1/3},$$

$$(22)\quad B = \left(-\frac{q}{2} - Z^{\frac{1}{2}} \right)^{1/3}, \text{ and:}$$

$$(23)\quad Z = \left(\frac{p}{3}\right)^3 + \left(\frac{q}{2}\right)^2,$$

then the roots $y_{ik}^{(1)}$, $y_{ik}^{(2)}$, and $y_{ik}^{(3)}$ of the reduced equation (18) can be rewritten as:

$$(24)\quad y_{ik}^{(1)} = A + B,$$

$$(25)\quad y_{ik}^{(2)} = -\frac{A+B}{Z} + \frac{A-B}{Z} \cdot (-3)^{\frac{1}{2}}, \text{ and}$$

$$(26)\quad y_{ik}^{(3)} = -\frac{A+B}{Z} - \frac{A-B}{Z} \cdot (-3)^{\frac{1}{2}}.$$

Given the fact that $x_{ik} = y_{ik} - \frac{a}{3}$, the roots of the original cubic equation in x_{ik} can be found. One or three of these roots are real numbers; two of them may be imaginary numbers. To complement this

procedure there is a selection of that real x_{ik} that really renders F its minimimum value by simply substituting the roots in F and evaluating the resulting values of the loss function.
This procedure is the elaboration of the principle upon which SCALCANS is based. It amounts to the minimization of the loss function by computing one optimum coordinate holding constant all other coordinates and all fitted differences

The algorithm

In this section it will be shown how the forming and solving of cubic equations fits into an algorithm for nonmetric multidimensional scaling. For ease of survey the program SCALCANS has been arranged in five steps.

Step 1: Reading/writing/processing of input data

* text card;
* card with parameters for steering the program:
  - number of points ($1 \leqslant n \leqslant 30$);
  - number of dimensions ($1 \leqslant m \leqslant 10$);
  - similarities of dissimilarities as input ?

- input or computation of initial configuration ?
- punched output of final configuration required ?
- plot of final configuration required ?
- maximum number of iterations;
- class width for ties in (dis-) similarities;
- criterium for stabilization of loss function;

* (dis-) similarities. No missing data are permitted and only the upper triangular matrix of (dis-) similarities is read in.
* (if necessary) initial configuration.

In the first step this input is processed in three ways:

* if necessary, similarities will be transformed into dissimilarities by simply altering their sign;
* dissimilarities are ranked...
* ... and scanned for ties.

Step 2: Computation of initial configuration

Of course, this step will be skipped when the initial configuration is part of the input. Computation is based on a metric scaling procedure of Torgerson (j) that proved to be very succesful in suggesting initial estimates of coordinates.

Step 3: Initialization of process

* centration and standardization of configuration

* computation of squared euclidean distances between points
* ranking of these squared distances within ties detected in the first step. This is necessary for the primary treatment of ties
* computation of fitted values, using Kruskal's monotonic regression procedure with primary treatment of ties
* computation of loss function (4)

Step 4: Minimization of loss function in the specified maximum number of iterations. In each iteration the program moves all n.m coordinates (one by one) so that the loss function is actually minimized. After each estimation of a coordinate new squared distances and corresponding fitted values are computed before turning to the next coordinate. After each iteration a test is carried out to determine if the new loss function deviates more then the specified criterium from the loss function in the previous iteration. (Because of the analytical character of the second phase of the minimization process the loss function can never rise.) If so, the next iteration is started. If not, step 5 follows.

Step 5: Output

After rotation to principal axes, centration and standardization SCALCANS gives the following output:

* number of iterations
* loss function;
* squared euclidean distances;
* fitted values;
* final configuration;
* (if required) plot of the final configuration;
* (if required) punched output of final configuration.

Some test results

Out of all tests of SCALCANS we selected one series of tests that appears very clear through its geographic interpretation. On a map we measured (as the crow flies) the distances between 8 cities in the province of Noord-Brabant, namely Bergen op Zoom, Roosendaal, Breda, Tilburg, Eindhoven, Oss, 's-Hertogenbosch and Waalwijk. This resulted in a matrix of dissimilarities that served as input to SCALCANS. An initial configuration in two dimensions appeared to be a perfect solution. This configuration served as input to 8 runs of the program, whereby several cities

changed places (with a varying number of dimensions). This was done in order to get indications of the possibility that:

1. a bad initial configuration prevents the program to find a satisfying solution, i.e. the algorithm gets trapped in a local minimum;
2. solutions in one dimension show varying values of the loss function, i.e. some solutions appear to be local minimum solutions.

In each run a maximum of 25 iterations was used. These runs made the following conclusions possible:

1. All solutions in 2 dimensions were excellent
2. The one-dimensional solutions showed practically identical coordinates and loss functions.
3. The geographic intercity structure was clearly recognizable in all two-dimensional solutions.
4. From 1) to 3) we can conclude that no local minimum problems were encountered.
5. After, say, 5 iterations the minimization process slows down considerably, but this hardly is a problem because of the high speed of minimization in the first few iterations.

Combined with all the other test results, we arrive at the following general conclusions: SCALCANS is a sound computer program for the scaling of stimuli by using the rank order of (dis-) similarities. The solution of cubic equations as the fundamental principle of the algorithm ensures a decline of the loss function. No local minimum problems were encountered. The minimization process goes very fastly in the first few iterations but very slowly thereafter. The procedure for the calculation of an initial configuration proved very sound.

One disadvantage of SCALCANS is that it makes use of squared distances, rather than distances. This disadvantage is minor, however, because a rank order of distances corresponds perfectly with the rank order of squared distances.

Literature

(a) Coombs, C.H.: A Theory of Data. New York (1964).

(b) Kruskal, J.B.: Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. Psychometrika 29(1964), pp. 1-27.

(c) Kruskal, J.B. & R.N. Shepard: A Nonmetric Variety of Linear Factor Analysis. Psychometrika 39 (1974), pp. 123-157.

(d) Roskam, E.E.: Metric Analysis of Ordinal Data in Psychology. Voorschoten (1968).

(e) Selby, S.M.: Standard Mathematical Tables. Cleveland (1972).

(f) Shepard, R.N.: The Analysis of Proximities: multidimensional scaling with an unknown distance function I. Psychometrika 27 (1962-I), pp. 125-140.

(g) Shepard, R.N.: The Analysis of Proximities: multidimensional scaling with an unknown distance function II. Psychometrika 27 (1962-II), pp. 219-246.

(h) Shepard, R.N.: Metric Structures in Ordinal Data, Journal of Mathematical Psychology 3(1966), pp. 287-315.

(i) Takane, Y., F.W. Young & J. de Leeuw: Nonmetric Individual Differences Multidimensional Scaling: an alternative least squares method with optimal scaling features. Psychometrika 42 (1977), pp. 7-67.

(j) Torgerson, W.S.: Theory and Methods of Scaling. New York/London (1958).

(k) Young, F.W.: Nonmetric Multidimensional Scaling: recovery of metric information. Psychometrika 35 (1970), pp. 455-473.

Tilburg, january 1980.