RR 91-01

THREE-WAY SCALING OF ASYMMETRIC PROXIMITIES

Berrie Zielman

Department of Data Theory

University of Leiden

Table of Contents

Acknowledgement

This master thesis benefitted from the asymmetric interaction with several persons.
First of all I wish to thank Willem Heiser for his supervision and advise; Robert Tijsen for helping me with datacollection and Wim van der Kloot for his comments upon an earlier version of this paper.

Three-way scaling of asymmetric proximities

Abstract

This paper studies a method that extends the distance model with additional parameters to represent asymmetric data.The method is suitable to the analysis of several matrices simultaneously. It is assumed that asymmetry is common to all matrices, the differences between matrices are modeled by a regression weight. In addition weights are introduced into the loss function to handle missing data. These weights force us to estimate the parameters by a sophisticated algorithm. Two examples are presented.

key words: multidimensional scaling, asymmetry, missing data, skew symmetric functions

1. Introduction

One of the questions that may effect us deeply when meeting new people, is: will we like that person and will that person like us. An important characteristic of liking or interpersonal attraction is, that we may like a person more than that person feels attracted to us. This relationship is said to be asymmetric. We could ask a number of persons in a group how much he or she likes the other members of the group. The relationships can be collected in person by person matrix Δ. This matrix is square and the rows and columns refer to the same set of objects. These matrices are quite common in psychology and allied disciplines. Other examples are: sociometric data, confusions of one stimulus to another, migration-rates, frequencies of citations among scientific journals, first choice second choice data, economic input-output data, moving data, counts of telephone calls among cities or departments of a company, the retrieval of a coin in city i which was made in city j, occupational mobility tables and communication and volume flows. These matrices or tables may be asymmetric, the upper triangle differs from the lower triangle.

There are several situations where it might be reasonable to assume that the observations δ_{ij} in the table Δ are related to the distance between objects i and j in some hypothetical space. There is a problem with this type of analysis: distances are symmetric, the distance from the center of North Korea to the center of China equals the distance from the center of China to the center of North Korea. However, Tversky (1977) found in an experiment concerning the similarity of nations that North Korea was judged more similar to China than the reverse, thus symmetry is not a property of similarity. Tversky explained asymmetry by: "apparently, the direction of asymmetry is determined by the relative salience of the stimuli; the variant is more similar to the prototype than vice verse".

The fit of the distance model can be improved by extending the model with additional parameters modelling asymmetry. Gower (1977, 1984), Gower and Digby (1981), Constantine and Gower (1978) studied a multidimensional representation of asymmetry by a skew symmetric function, Weeks and Bentler (1982) and Okada (1988a,b) studied a one-dimensional representation of asymmetry; again by a skew symmetric function. A skew symmetric function has symmetric function values, except for sign. So far only two way matrices have been analyzed with skew symmetric functions, similar methods suitable for the analysis of three-way data do not exist yet. In three-way analysis there are more matrices to analyze simultaneously. The third way may correspond to replications, occasions, conditions or groups of persons, depending on the application.

The purpose of this study is to generalize the one-dimensional representation of asymmetry of Weeks and Bentler (1982) and Okada (1988a,b) into four important directions:

1. A three-way model representing asymmetry will be developed

2. Data weights will be introduced into the loss function

3. The asymmetry will be represented in more than one dimension

4. The model will be incorporated into the SMACOF scaling program.

Simultaneous scaling of more than one matrix is quite popular for symmetric matrices, so it is desirable to develop a method that can handle three-way asymmetric data. Missing data are likely to occur in these situations for instance if a journal no longer exists, or if citations are counted over a small period of time, or when brand-switching data are analyzed, and some brands have small market shares. If the matrices with the data weights are not symmetric the estimation of the parameters becomes cumbersome due to the possible occurrence of negative pseudo-distances. This problem is solved by applying the generalized SMACOF method proposed by Heiser (1991). More dimensions will in general be necessary for an adequate representation of asymmetry in three-way scaling. The proposed multidimensional representation of asymmetry will differ from the representation proposed by Gower (1977).

The symmetric part of the data will be approximated by distances, estimated by the SMACOF (Scaling by MAjorizing a COmplicated Function) method. The SMACOF method is used because a variety of individual differences models can be estimated and the method converges. The results of this study will be implemented into a computer program written in APL*.

* A Programming Language, by K.E. Iverson

2. Overview of Multidimensional Scaling

2.1. Introduction

The treatment of multidimensional scaling (MDS) will be brief; the reader is referred to Kruskal and Wish (1978) for a more detailed explanation of this subject. MDS models comprise a class of techniques that represent the dissimilarity between objects i and j, δ_{ij}, by distances $d_{ij}(\mathbf{X})$ computed among the rows of an n by p configuration matrix $\mathbf{X}$. The number of rows of $\mathbf{X}$ equal the number of objects n, and $\mathbf{X}$ has p columns which are usually called dimensions. Given the configuration matrix $\mathbf{X}$ the objects can be plotted in a low dimensional space. Usually the Euclidean distance is computed:

$$d_{ij}(\mathbf{X}) = \sqrt{\Sigma_s \, (x_{is} - x_{js})^2}, \tag{1}$$

where x_{is} is the coordinate of object i on dimension s.

2.2. Individual differences

Suppose there is more than one matrix to analyze, these replications may represent different points in time, different subjects or occasions. These matrices may be analyzed by an individual differences model. Individual differences models make a distinction between common variation due to all the matrices and specific variation due to a particular matrix. Individual differences scaling represents the m configurations by a single configuration, called the common space. This common space is allowed to vary in a limited sense, by applying a weight matrix to the common space. Individual differences are accounted for by individually stretching or shrinking the axes - this model is called INDSCAL (Carroll and Chang 1970), or by individually rotating and weighting the configurations; this model is called IDIOSCAL.

Special cases of these models are developed under the name reduced rank model by Young(1984) for the IDIOSCAL case and by Heiser and Stoop (1987) and Heiser (1988) for INDSCAL. Reduced rank models allow a high dimensionality of the common space but the individual spaces are of a lower dimensionality.

3. Other approaches to asymmetry in MDS

This section briefly discusses other methods to handle asymmetry in the data, more details can be found in Zielman (1991). The work of Okada (1988 a,b), Weeks and Bentler (1982) and Gower (1977) is discussed in the next section because it is closely related to our model.

Profile similarities can be computed among the rows or columns of any data table. This procedure yields a symmetric input matrix for a MDS program, but a choice has to be made wether the profile similarities are computed among the rows or the columns. There is a large number of coefficients that are potential candidates to serve as a similarity coefficient, one can think of the Pearson correlation coefficient, the chi-square distance or the Euclidean distance among others. Reviews of these coefficients can be found in Coxon (1982) and Everitt (1980).

A method discussed by Ekman (1963) and Levin and Brown (1979) is multiplicative rescaling of the rows or columns in such a way that the symmetry is maximized. This method is closely related to the choice model of Luce (1963), the difference is that in the choice model the rescaling parameters are part of the model and not operating on the data. The Choice model yields symmetric parameters, which can be analyzed by a MDS program. The choice model has been extended by Nakatani (1978) and Holman (1979), Heiser (1988) studied the relation with the quasi symmetry model (Caussinus, 1965).

Lauman and Guttman (1966) performed separate scaling of the lower and upper triangle; this procedure results in two configurations, representing outflow and inflow. These configurations can be visually compared or rotated to each other by a Procrustes procedure.

Young (1975, 1984) proposed the ASYMSCAL model, this model associates with every object i and dimension s a weight w_{is}. If asymmetry is present in the data the weights w_{is} and w_{js} will differ. Observe that this model doubles the number of parameters estimated, this will decrease the stability of the parameters.

The DEDICOM model (DEcomposition into DIrectional COMponents) proposed by Harshman e.a. (1982) describes the data by a set of orthogonal dimensions or aspects and a matrix that describes the relations between the dimensions. Dimension s may be stronger related to dimension q than the reverse.

Cunningham (1978) proposed free trees as a representation of the dissimilarities. A free tree is a set of nodes and a set of links between the nodes in which every pair of nodes is connected by a unique sequence of links. The free tree can be adjusted by weighting the links between the nodes to represent asymmetry.

The unfolding model associates with each object i a row point $\mathbf{x}_i$ and a column point $\mathbf{y}_j$. The model predicts symmetry if the points $\mathbf{x}_i$ and $\mathbf{y}_i$ coincide. The unfolding model considers the symmetric an asymmetric part of the data as inseparable. See Coombs (1964) for a discussion of the unfolding model.

Tversky (1977) criticized the use of distance models with similarity data and developed a set theoretical approach to similarity. The similarity between two objects is a linear function of the features shared by the objects and the features unique to object i and the features unique to object j. The feature matching model predicts asymmetry if the unique features are given unequal weight. Krumhansl (1978) extended the distance model in a response to Tversky's criticism with parameters representing spatial density of the points in the configuration.

Another possibility is to transform the asymmetry away by (monotone) regression. This can be done by transforming the data row or column conditional. Values within rows are regarded as comparable with each other, the values among the rows are regarded as incomparable. In the case of linear transformations this approach is related to the work of Levin and Brown (1979), because the target (the distance matrix) is symmetric, so the rescaling will optimize the symmetry of the transformed data as well. If the data are linearly transformed the regression weights can be interpreted as bias parameters, or the tendency to favour some responses over others; this interpretation is not possible when monotone regression is applied.

4. Modelling asymmetry by skew symmetric functions

4.1. The model

Any square matrix can be additively decomposed into a symmetric matrix and a skew symmetric matrix:

$$\Delta = \{\Delta + \Delta'\}/2 + \{\Delta - \Delta'\}/2 = \mathbf{S} + \mathbf{A} . \qquad (2)$$

The matrix Δ' denotes the transpose of Δ. The symmetric matrix **S** is a matrix of averages and has elements $s_{ij} = s_{ji} = \{\delta_{ij} + \delta_{ji}\}/2$, the matrix **A** contains the residuals from symmetry and has elements $a_{ij} = -a_{ji} = \delta_{ij} - s_{ij}$. The matrix **A** is called a skew symmetric matrix, because of the property $a_{ij} = -a_{ji}$. The sum of squares of the matrix Δ can be decomposed as a sum of squares due to symmetry and a sum of squares due to asymmetry, the cross-product vanishes. Therefore the asymmetric part of the data may be viewed independently of the symmetric part of the data. If it is assumed that the asymmetry in the data is due to noise or measurement error the matrix **A** is ignored and the matrix **S** can be analyzed by a MDS program. But if the researcher thinks the asymmetry may contain meaningful information, the matrix **A** can be modelled by additional parameters. Because of the skew symmetric property of **A** a skew symmetric function is an obvious choice.

Okada (1988 a,b) added a linear skew symmetric function to the distances; the model can be written as:

$$m_{ij}(\mathbf{X};\mathbf{r}) = d_{ij}(\mathbf{X}) + (r_i - r_j), \tag{3}$$

where $m_{ij}(\mathbf{X};\mathbf{r})$ is the predicted value of the observed δ_{ij} , $d_{ij}(\mathbf{X})$ is the Euclidean distance function. The r_i parameters represent the asymmetry. If we write $c_{ij} = r_i - r_j$ the r_i parameters form a skew symmetric function c_{ij} because this function has the property $c_{ij} = - c_{ji}$. A quantity $|c_{ij}|$ is subtracted from the distance if δ_{ij} is smaller than the distance, the same quantity is added to the distance if the dissimilarity is larger than the distance.

The model assumes that the diagonal is zero: this means that the minimality axiom must hold for the data. The minimality axiom states that the distance between an object and itself equals zero. When analyzing brand switching data the diagonal entries in the table corresponds with the stayers within a category; frequently one is not interested in the stayers but in the switchers, so this is not a problem in many applications.

The Weeks-Bentler (1982) model is the same as (3), except that their model allows for an additive constant and the dimensions may be oblique. Their computer program to estimate the parameters of the model allows for constraints on these parameters, e.g. proportionality and fixed value constraints. For a discussion of constraints in MDS the interested reader is referred to Bentler and Weeks (1978), Weeks and Bentler (1982), De Leeuw and Heiser (1980), Heiser and Meulman (1983) and Meulman and Heiser (1984).

4.2. Relations to other models

Gower (1977, 1984), Gower and Digby (1981) and Constantine and Gower (1978) studied the singular value decomposition (SVD) of skew symmetric matrices. SVD is a bilinear model; the model of Okada (1988 a,b) and Weeks and Bentler (1982) is a linear model. The SVD approximates the skew symmetry by triangles in two dimensional spaces. The SVD is equivalent to the skew symmetric part of (3) when all points are located on a straight line.

Van der Heijden (1987) and Van der Heijden e.a.(1989) studied the Escofier generalization of correspondence analysis. This type of correspondence analysis was used to study residuals from log linear models, in particular symmetry and quasi symmetry models. This model is also equivalent to the skew symmetric part of (3) when the points are located on a line and the appropriate row and column metrics are chosen.

The model (3) is also a special case of the distance density model of Krumhansl (1978). The distance density model assumes similarity to be a function of the distance between object i and j and the density of points surrounding objects i in the configuration and the density of points surrounding object j.

Takane and Shibayama (1986) showed the relation of the choice model by Luce (1963) with the model of Krumhansl (1978) by a logarithmic transformation. If model (3) is transformed by an exponential function the choice model is obtained.

4.3. Interpretation of the parameters

If the asymmetry in the data is structural the additional parameters will improve the fit of the model. But how are the parameters interpreted? The r_i parameters can be represented as a one dimensional scale. Objects with close values on this scale are almost symmetric. Objects that are located far apart on the scale are asymmetric representing inflow from category i to category j and outflow from category j to category i. If the asymmetry parameters are in deviation from the mean, they can be interpreted as the average gain or loss of the object to all other objects, depending on the sign of the parameter. Positive sign indicates more inflow than outflow, negative values of the parameters indicate more outflow than inflow. The statement of Okada (1988 b) that the larger r_i there is a greater tendency of outflow is refined. This is especially important in marketing research where it is interesting to identify winners and losers. Brands with positive scale values are on the average winners, brands with negative scale values are viewed as less attractive by consumers. To return to Tversky's (1977) criticism: in some cases it might be possible to speak of a prototype if the object has a small r_i value. However, this is a theoretical question; it cannot be answered by data analysis.

To improve the possibilities for interpretation, Okada (1988a,b) offered a natural way of displaying the parameters in the MDS configuration. This idea of joint representation is illustrated in Figure 1.

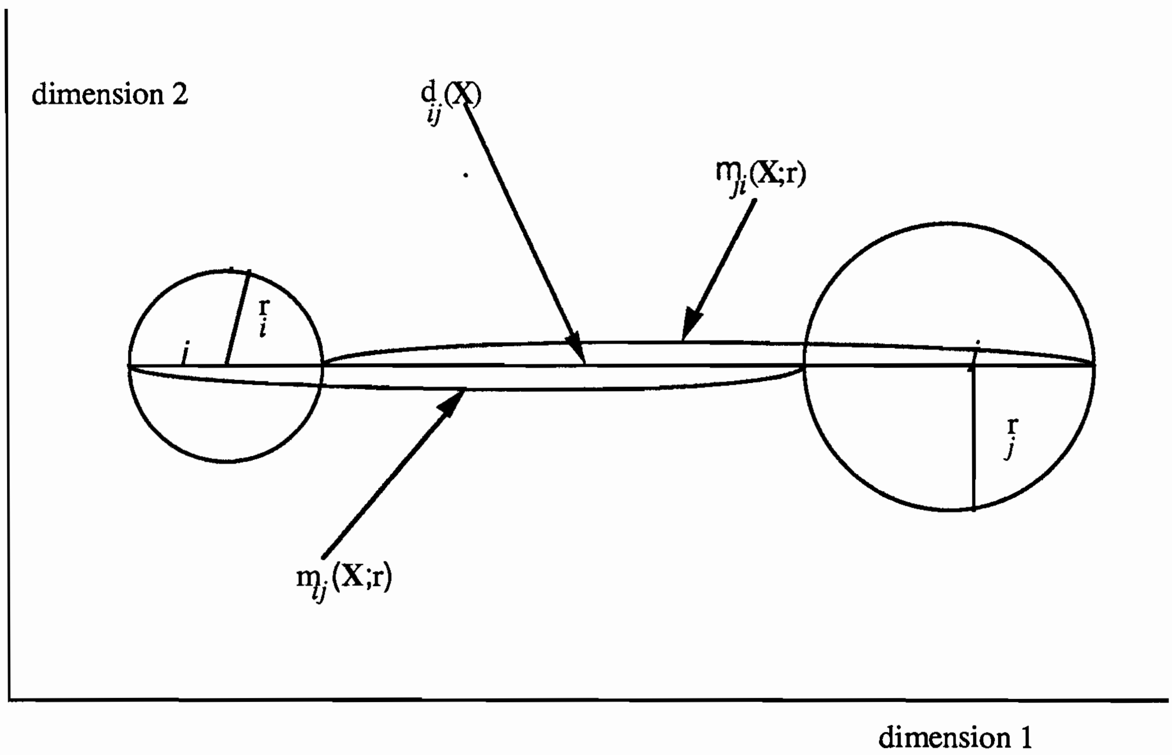

Figure 1: Joint representation of asymmetry and distances

In Figure 1 two points are drawn in a two dimensional space. A circle (sphere, hypersphere) is drawn around each point, with radius equal to the value of r_i. In this figure the flow from object i to object j is smaller than the flow from object j to object i. The asymmetry parameters are also called radii.

To incorporate the skew symmetric parameters into the configuration, the parameters can be temporarily re-scaled in such a way that the smallest value equals zero. This rescaling can be done without loss of generality, this will be explained in the next section.

4.4. Identification

There is some indeterminacy in the asymmetric part of the model that needs special attention. An arbitrary constant 1 can be added to the asymmetry parameters without changing the predicted values of the model. We may write:

$$m_{ij}(\mathbf{X};\mathbf{r}) = d_{ij}(\mathbf{X}) + (r_i - r_j) = d_{ij}(\mathbf{X}) + ((r_i + 1) - (r_j + 1)) \,. \qquad (4)$$

For identification purposes a parameter or a function of the parameters has to be fixed. Weeks and Bentler set the first value equal to zero in the unconstrained case, Okada(1988a,b) fixes the lowest value of the scale. This paper uses $n^{-1} \sum_j r_i = 0$; the parameters are in deviation from their mean, for mathematical convenience and to facilitate the interpretation. To display the scale values as radii the scale values are temporarily re-scaled in such a way that their lowest value becomes zero.

4.5. The three-way model

The model of Okada (1988 a,b) which can handle only two-way data will be generalized in such a way that several matrices can be analyzed simultaneously and that a multidimensional representation of asymmetry is possible. The proposed model is:

$$m_{ijk}(\mathbf{X}_k;\mathbf{R};\mathbf{U}_k) = d_{ij}(\mathbf{X}_k) + \Sigma_s u_{ks}(r_{is} - r_{js}). \qquad (5)$$

The asymmetry is modeled by a multidimensional skew symmetric function, the functions are common to all the k slices of the data matrix, and a regression weight u_{ks} indicating the relative importance of the asymmetry dimension s for slice k. The skew symmetric part of (5) corresponds with the vector model of preference data (Carroll, 1972). The vector model positions the objects in a p-dimensional space; the subjects are represented by vectors. The projections of the stimuli on the subject vector indicates the rank-order of the objects for a particular subject. In model (5) the objects are similarly displayed, but the third-way is represented by vectors. These vectors may now be called occasion or points in time vectors, depending on the application.

There are two possibilities of displaying the additional asymmetry parameters. The asymmetric dimensions can be represented by adding several circles (or spheres; hyper-spheres) to the points in the configuration. If the researcher chooses two dimensions for the asymmetry an object is represented by a point and two circles. There is compensation if two objects have the same circles, but where these circles corresponds to different dimensions. Note that the dimensionality of the configuration matrix can be chosen independently of the dimensionality of the asymmetric scales. If a weight u_{ks} is negative it means that the direction of asymmetry is reversed for slice k.

A convenient way of displaying the asymmetry parameters is a joint plot of the asymmetry weights and asymmetry scale values. The matrix **R** gives coordinates for the objects on the asymmetry scale and the coordinates for the third way vectors can be found in the rows of the matrix **U**. Such a representation is illustrated in Figure 2.

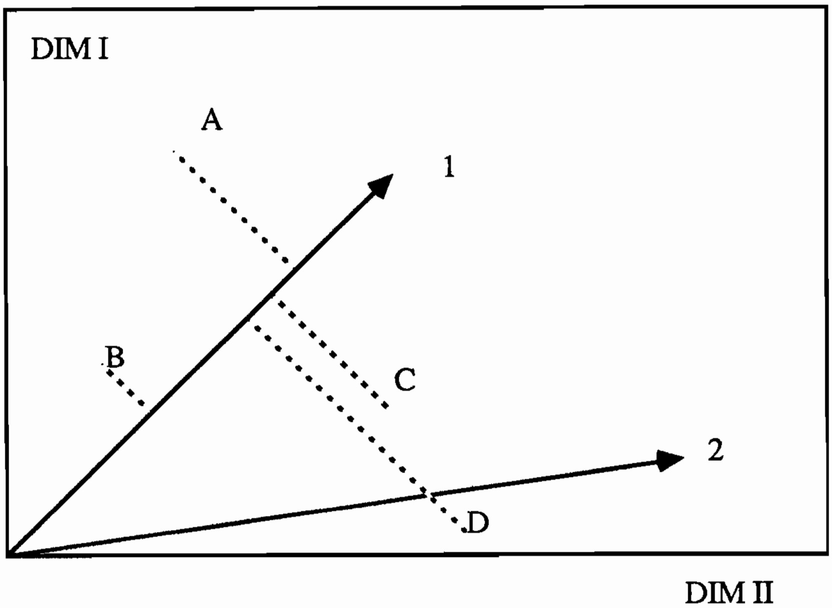

Figure 2: Joint representation of asymmetry scales and asymmetry weights

In Figure 2 four objects A,B,C,D are depicted together with two third way vectors marked by numbers; the dashed lines are the projections of the objects on the third way vector denoted by 1. The reader is encouraged to draw the projection lines for vector 2. The order of outflow for vector 1 is ACDB and for vector 2 DCAB.

The distances are computed among the rows of the configuration matrix $\mathbf{X}_k$. This configuration matrix can also be restricted to the form $\mathbf{Z}\mathbf{H}_k$, where $\mathbf{Z}$ is the common space and $\mathbf{H}_k$ is a weight or transformation matrix of the common space.

5. The algorithm

This section discusses the algorithm to estimate the parameters of the model. In the next subsection an improved solution for the asymmetric component in Okada's algorithm will be indicated.

5.1. Improved estimation for Okada's algorithm

Okada(1988a,b) estimated the parameters, the configuration and the radii, by a steepest descent algorithm, in fact a modified version of Kruskal's (1964 a,b) algorithm. Weeks and Bentler (1982) employed the Gauss-Newton algorithm. A least squares loss function was minimized in both algorithms. Both algorithms suffer from the problem of local minima. It can be shown that for the estimation of the radii an iterative method is not necessary, an analytical solution exists. We have to minimize:

$$\sigma(\mathbf{X},r) = \sum_i \sum_j (\delta_{ij} - d_{ij}(\mathbf{X}) - (r_i - r_j))^2. \quad (6)$$

This loss function is Kruskal's (1964 a,b) raw STRESS extended with the additional components for modelling asymmetry. The loss function will be split up into two orthogonal components by inserting (2) into (6).

$$\sigma(\mathbf{X},r) = \sigma(\mathbf{X}) + \sigma(r) = \sum_i \sum_j (s_{ij} - d_{ij}(\mathbf{X}))^2 + \sum_i \sum_j (a_{ij} - (r_i - r_j))^2. \quad (7)$$

The cross-products $d_{ij}(\mathbf{X})(r_i - r_j)$, $(r_i - r_j)s_{ij}$, $s_{ij}\, a_{ij}$ and $a_{ij}\, d_{ij}(\mathbf{X})$ sum to zero. The first part of the loss function is an MDS problem,the right hand side of (7) can be minimized under the identification condition $\sum_j r_j = 0$ by $r_i = n^{-1} \sum_j a_{ij}$, this is the mean of the row elements of the matrix A. This solution is derived from Torgerson (1958), Mosteller (1951) and Gulliksen (1956). They used it to estimate scale values of stimuli from paired comparisons. Up to a monotonic transformation the skew symmetric part of (3) is identical to the Thurstone case V scaling model. It is easily seen that a multidimensional representation of asymmetry with a linear model is not possible with a two way matrix; this is the case because the residual quantities sum row and column wise to zero.

In the Weeks-Bentler approach decomposition (7) can still be used, but the solution of the skew symmetric parameters, the mean of the rows of **A**, is only valid if there are no restrictions on the r_i parameters.

5.2. The algorithm for three-way scaling

We now turn to the three-way model; the loss function to be minimized is:

$$\sigma(\mathbf{Z};\, \mathbf{H}_k;\, \mathbf{r}\,;\mathbf{U}) = \Sigma_i \Sigma_j \Sigma_k \underline{w}_{ijk}(\,\delta_{ijk} - d_{ij}(\mathbf{ZH}_k) - \Sigma_s u_{ks}(r_{is} - r_{js})\,)^2. \tag{8}$$

The weights $\underline{w}_{ijk}$ are introduced into the loss function to code missing data; the corresponding elements of $\underline{\mathbf{W}}_k$ are set equal to one if the dissimilarity is observed otherwise the elements are made zero. Loss function (8) can be decomposed for every slice k in the same way as (7) if the weight matrices $\underline{\mathbf{W}}_k$ are symmetric. Analogously to the two-way case (8) the loss function for the three-way case can be decomposed as:

$$\sigma(\mathbf{Z};\, \mathbf{H}_k;\, \mathbf{R}\,;\mathbf{u}_k) = \sigma(\mathbf{Z};\, \mathbf{H}_k) + \sigma(\mathbf{R};\, \mathbf{u}_k). \tag{9}$$

This convenient property of the loss function does not hold if the weight matrix is asymmetric; for instance, if one dissimilarity of the pair i,j is missing. To handle asymmetric weight matrices an Alternating Least squares algorithm is used. This algorithm uses two major steps; in the first step the distances are kept fixed and the asymmetric parameters are estimated; in the second step the asymmetric parameters are kept fixed at their current values and the distance part of the model is improved. These two steps are alternated until the loss function value cannot be decreased anymore. If both dissimilarities are missing there is no problem at all, the decomposition (9) can still be used.

If there are asymmetric weight matrices the additive properties of the model are use to generate pseudo data values. In the first step the distance is subtracted from the known data value and the missing element across the diagonal is set equal to minus one times the difference between data and distance. In the second step the distances are estimated by subtracting the skew symmetric part of the model from the known data; this yields pseudo distances of which some may be negative because a quantity is subtracted from the dissimilarities. Positivity of the pseudo distances is required for the SMACOF method to converge properly. Fortunately this non-convergence can be avoided by using the generalized SMACOF method developed by Heiser (1991). The scaling problem will be discussed in section 7; the estimation of the asymmetry parameters is discussed in the next section.

6. Estimation of the asymmetric component

The asymmetric component is estimated via an Alternating Least Squares algorithm: in the first step the weights are kept fixed at their current values, and the radii are estimated. In the second step the radii are fixed, and the regression weights are estimated. The skew symmetric part of the model is estimated with the skew symmetric part of the data.

Decomposition (2) is used for that part of the data which is not missing. If missing data is present the skew symmetric part is estimated by $a_{ijk} = \delta_{ijk} - d_{ij}(\mathbf{X}_k)$. The twin sister of a_{ijk}; which is the unknown or missing a_{jik} is set at $-a_{ijk}$. The weight matrix $\underline{\mathbf{W}}_k$ is symmetrized; this matrix is denoted by $\mathbf{W}_k$. The loss function that has to be minimized is:

$$\sigma(\mathbf{r}; u_{ks}) = \Sigma_i \Sigma_j \Sigma_k w_{ijk}(a_{ijk} - \Sigma_s u_{ks}(r_{is} - r_{js}))^2. \qquad (10)$$

6.1. Estimation of the asymmetry parameters for fixed weights

The parameters r_{is} model the asymmetry in the data, the u_{ks} indicate the relative importance of the asymmetric component in slice k. The origin of the scale of the r_{is} parameters is defined to be $\Sigma_j \Sigma_k w_{ijk} \Sigma_s r_{js} u_{ks} = 0$. To make the r_{is} parameters suitable for graphical representation, the r_{is} parameters are re-scaled after the analysis in such a way that the smallest value is zero. The scale values will be computed dimension after dimension. For dimension s we compute the following quantities:

$$\underline{a}_{ijk} = a_{ijk} - \Sigma_{s \neq q} u_{ks}(r_{is} - r_{js}), \tag{11}$$

where q is a subscript indicating dimensions estimated in previous cycles. Loss function (10) is minimized dimensionwise by inserting the quantities according to (11) into the loss function. Setting the partial derivative of loss function (10) with respect to r_{is} equal to zero yields the stationary equation:

$$\Sigma_j \Sigma_k w_{ijk} r_{is} u_{ks}^2 - \Sigma_j \Sigma_k w_{ijk} \underline{a}_{ijk} u_{ks} = 0. \tag{12}$$

Which can be expressed in matrix notation as:

$$\mathbf{V}\mathbf{r}_s = \mathbf{Q}\,\mathbf{e}. \tag{13}$$

The matrix **V** has elements:

$$v_{ij} = m^{-1} \Sigma_k w_{ijk} u_{ks}^2; \tag{14a}$$

$$v_{ii} = m^{-1} \Sigma_k \Sigma_{i \neq j} w_{ijk} u_{ks}^2. \tag{14b}$$

It is the matrix of average weights weighted by the squares of the dimension weights.The matrix **Q** has elements:

$$q_{ij}=m^{-1} \Sigma_k w_{ijk} \underline{a}_{ijk} u_{ks}. \tag{15}$$

The matrix **Q** is a matrix with averages over the k slices weighted by the dimension weights. The solution of $\mathbf{r}_s$ becomes:

$$\mathbf{r}_s = \mathbf{V}^+\mathbf{Q}\mathbf{e}, \tag{16}$$

where **e** is the vector consisting of n ones. The inverse of the matrix **V** does not exist , because it is not a matrix of full rank. The generalized inverse $\mathbf{V}^+$ of **V** is used, which has the property $\mathbf{V}\mathbf{V}^+\mathbf{r} = \mathbf{r}$ .This matrix $\mathbf{V}^+$ can be written as $\{\mathbf{V} + \mathbf{e}\mathbf{e}'/n\}^{-1} - \mathbf{e}\mathbf{e}'/n$ because **e** is the unique vector from the null space of **V**.

6.2 Finding weights for fixed radii

The next step is to solve for u_{ks} while holding **R** fixed at its current value. Setting the partial derivative of the loss function with respect to u_{ks}equal to zero yields the stationary equation:

$$\Sigma_i\Sigma_j(r_{is}- r_{js})^2 \; w_{ijk} u_{ks} \; - \Sigma_i \Sigma_j w_{ijk} a_{ijk} (r_{is} - r_{js}) = 0. \tag{17}$$

The solution of u_{ks} is:

$$u_{ks}= (\mathbf{p}'\mathbf{M}_k\mathbf{p})^{-1}\mathbf{p}'\mathbf{M}_k \, q_k \, , \tag{18}$$

where $\mathbf{M}_k$ is a diagonal matrix of order n^2 by n^2 with the elements w_{ijk} on the diagonal, **p** is a n^2 vector with the estimates r_{is} - r_{js}, $\mathbf{q}_k$ is also an n^2 vector with elements a_{ijk} ; these elements of $\mathbf{q}_k$ corresponds with the elements in **p**. The dimension weight u_{ks} can be interpreted as a regression weight. For identification purposes the weights u_{ks} are re-scaled so that their sum of squares becomes unity. By iterating between 16 and 18 solutions for both sets of parameters a sought until no further improvement of loss as measured by (10) is possible.

7. The SMACOF theory

The SMACOF algorithm minimizes a loss function called STRESS by repeatedly minimizing a more simple function: the majorizing function. The algorithm is a modified version of Guttman's (1968) algorithm, the algorithm is discussed in detail in De Leeuw and Heiser (1980, 1982) and De Leeuw (1988). The STRESS measures the departure of the distances, $d_{ij}(\mathbf{X}_k)$, from the adjusted dissimilarities, s_{ijk}. In out case these adjusted dissimilarities are defined by the matrix **S** in decomposition (2).

If in (8) the symmetric part cannot be estimated from the data due to missing elements, the current model estimate of the asymmetric part of the model is substracted from the data; the quantities s_{ijk} and s_{jik} are set equal to δ_{ijk} - $\Sigma_s u_{ks}(r_{is} - r_{js})$. This definition of s_{ijk} has the consequence that these quantities may become negative. In the case of negative dissimilarities the SMACOF estimation method will not converge properly. Therefore the generalized SMACOF method (Heiser, 1991) will be employed. The STRESS can be written as:

$$\sigma(\mathbf{X}_1,..,\mathbf{X}_k) = \Sigma_i\Sigma_j\Sigma_k w_{ijk}(s_{ijk} - d_{ij}(\mathbf{X}_k))^2 =$$
$$\Sigma_i\Sigma_j\Sigma_k w_{ijk}\, s_{ijk} - 2\, \Sigma_i\Sigma_j\Sigma_k w_{ijk}\, s_{ijk}\, d_{ij}(\mathbf{X}_k) + \Sigma_i\Sigma_j\Sigma_k w_{ijk}\, d_{ij}^2(\mathbf{X}). \qquad (19)$$

Which can be written in matrix notation as:

$$\sigma(\mathbf{X}_k) = c_k - 2\mathrm{tr}\mathbf{X}_k'\mathbf{B}(\mathbf{X}_k)\mathbf{X}_k + \mathrm{tr}\ \mathbf{X}_k\mathbf{V}_k\mathbf{X}_k, \tag{20}$$

where c_k is the sum of squared dissimilarities. The matrix $\mathbf{B}(\mathbf{X}_k)$ is the sum of two matrices, an off-diagonal matrix and a diagonal matrix. The off-diagonal part has elements:

$$b_{ijk} = \frac{-(w_{ijk}s_{ijk} + w_{jik}s_{jik})}{d_{ij}(\mathbf{X}_k)} \qquad \text{if } d_{ij}(\mathbf{X}_k) \neq 0, \tag{21a}$$

$$b_{ijk} = 0 \qquad \text{if } d_{ij}(\mathbf{X}_k) = 0. \tag{21b}$$

The diagonal part has elements equal to the row sums of the off-diagonal part multiplied by -1. The matrix $\mathbf{V}_k$ is also build up from a diagonal matrix and an off-diagonal matrix. The off-diagonal matrix has elements:

$$v_{ijk} = -(w_{ijk} + w_{jik}). \tag{22}$$

The diagonal elements of the matrix $\mathbf{V}_k$ are equal to the row sums multiplied by -1.

Loss function (20) can be decreased by minimizing a more simple function called the *majorizing function* :

$$\sigma(\mathbf{X}_k,\mathbf{Y}_k) = c_k - 2\mathrm{tr}\mathbf{X}_k'\mathbf{B}(\mathbf{Y}_k)\mathbf{Y}_k + \mathrm{tr}\ \mathbf{X}_k\mathbf{V}_k\mathbf{X}_k. \tag{23}$$

This function is minimized by computing the *Guttman transform* (De Leeuw and Heiser, 1980)

$$\mathbf{X}_k = \mathbf{V}_k^{+} \, \mathbf{B}(\mathbf{Y}_k)\mathbf{Y}_k, \tag{24}$$

where $\mathbf{V}_k^{+}$ denotes the generalized inverse of $\mathbf{V}_k$. The matrix $\mathbf{V}_k^{+}$ is constructed in the same way as the matrix in section 6. The algorithm is quite simple: set $\mathbf{Y}_k = \mathbf{X}_k$ and compute again a new configuration by the Guttman transform. A proof of convergence of this algorithm can be found in De Leeuw and Heiser (1980) and De Leeuw(1988). This proof of convergence is only valid for positive pseudo-distances; as remarked in section 5.2 the pseudo-distances may in some cases become negative. The generalized SMACOF method proposed by Heiser (1991) assures convergence in the presence of negative pseudo-distances by adjusting the **B** and the **V** matrices.

The previously computed individual configurations do not satisfy the restrictions imposed by the particular individual differences model that has to be estimated. The common space and the individual weight matrices can be found by solving the metric projection problem (De Leeuw and Heiser 1980).

$$\mathrm{L}(\mathbf{Z};\mathbf{H}_k) = m^{-1} \, \Sigma_k \mathrm{tr} \, (\mathbf{X}_k - \mathbf{Z}\mathbf{H}_k)'\mathbf{V}_k(\mathbf{X}_k - \mathbf{Z}\mathbf{H}_k). \tag{25}$$

In (25) $\mathbf{X}_k$ is the Guttman-transform of the previous iteration, $\mathbf{Z}$ is the common space, $\mathbf{H}_k$ is the individual weight matrix This matrix $\mathbf{H}_k$ can be diagonal (INDSCAL model), a full matrix (IDIOSCAL model), or a full matrix of rank $r < p$ (reduced rank model). To find $\mathbf{Z}$ the following system of equations has to be solved :

$$m^{-1} \, \Sigma_k \, \mathbf{V}_k \, \mathbf{Z} \, \mathbf{H}_k \, \mathbf{H}_k' = m^{-1} \, \Sigma_k \mathbf{V}_k \mathbf{X}_k \mathbf{H}_k \, . \tag{26}$$

This can be done in a number of ways depending on the structure of the dataweight matrices and the model estimated. A detailed treatment of these solutions can be found in Heiser and Stoop (1986). The matrix $\mathbf{H}_k$ can be found by :

$$\mathbf{H}_k = (\mathbf{Z}'\mathbf{V}_k\mathbf{Z})^{-1}\, \mathbf{Z}'\mathbf{V}_k\mathbf{X}_k \,. \tag{27}$$

If an INDSCAL solution is desired the weights can be found by solving the metric projection problem dimension after dimension. After the common space **Z** and the weight matrix $\mathbf{H}_k$ have been found a new Guttman-transform is computed by setting $\mathbf{Y}_k = \mathbf{ZH}_k$. Steps 24, 26 and 27 are computed repeatedly until the convergence criterion is met.

8. Examples

8.1. Scientometric transaction matrices

In this section the method will be illustrated by analyzing journal to journal citation data, these type of data are also known as scientometric transaction matrices (Tijssen et. al. 1986).

For the purposes of the present study the journals were selected from the social sciences with the restriction that they pay attention to quantitative methods. The number of given and received citations were extracted from the Journal Citation Reports. The number of citations were accumulated over the last ten years, for instance Psychological Bulletin cites in 1988 to publications in 1979 - 1988. The reason to count only the last ten years is to correct for age of the different journals. A journal that exists for several decades has a greater probability to be cited than a journal that exists for a few years. The citations were for each year collected in a citation matrix where the rows refer to the citing journals and the columns to the cited journals. Journal Citation reports only mentions the number of times the journal APM is cited by the other journals; the citing behavior of this journal is not reported. This leads to the situation that for every citation matrix one row is completely missing.

The selected journals are listed in Table 1.

Table 1. Journals in the analysis with their plot labels

Journal	label
Applied Psychological Measurement	APM
British Journal of Mathematical and Statistical Psychology	BJ
Educational Psychological Measurement	EPM
Multivariate Behavioral Research	MBR
Psychological Bulletin	PB
Psychological Review	PR
Journal of Mathematical Psychology	JM
Psychometrika	PMK

The table lists primarily journals devoted to psychology, the reason is that other journals that obey the criteria (such as Sociological methods and research and American journal of sociology) are not reported in the Journal Citation Reports. The data were collected for eight years (1981 - 1988). Before 1981 the Journal Citation Reports only mentions the journals Psychometrika, Psychological Bulletin and Psychological Reports.

The resulting 8 citation matrices are not appropriate as input for the present program. The first thing that has to be done is to convert the proximities into dissimilarities; and the second step is to correct for differences in size of the journals. A journal publishing a large number of papers will have a greater probability to be citing another journal and also to be cited by another journal.

These two problems can be solved by assuming that the following relationship holds between frequencies, n_{ij} , and dissimilarities, δ_{ij},:

$$n_{ij} = \frac{k\, p_i\, p_j}{\delta_{ij}^2}, \tag{28}$$

where p_i is the mean of the rows, p_j is the mean of the columns and k is an unknown scaling factor. In the ideal case this scaling constant should be estimated, in this study some reasonable values were inserted. Formula (28) is borrowed from physics were it is used to describe the attraction between two objects for instance the attraction between the moon and the earth; in data analysis it was used by Thurstone (1929) and Tobler and Wineburg (1971). Equation (28) is inverted to give approximated distances

$$\delta_{ij} = \sqrt{\frac{k\ p_i\ p_j}{n_{ij}}} \tag{29}$$

The influence of the size of a journal has been corrected for by taking the marginals of the data matrix into account. These quantities are suitable as input for the program. Another way of looking at the model is that the deviations from the independence model are described by the modified distances model. Because we cannot divide by zero a small positive constant is added to the data table. This method of dealing with empty cells has in the present application the consequence that zero cells with large marginals are assumed to be further apart than cells with small marginals, this is perhaps not unreasonable.

To determine the number of dimensions neccessary for a good fit of the model, the number dimensions was varied from two through five. The number of asymmetry scales fixed at one dimension. The STRESS values are respectively 0.3993, 0.3834, 0.3760, 0.3702. Increasing the dimensionality for the distance part of the model to three or more dimensions is not worth the trouble of interpreting a three or higher dimensional configuration. Keeping the dimensionality of the distance part of the model fixed at two dimensions the number of asymmetry scales was increased, this did not yield a substantial better fit of the model. The STRESS values are 0.3858, 0.3740, 0.3630. The Shepard plot, a plot of model estimates vs dissimilarities suggested a logarithmic relationship between the model and the data. Transforming dissimilarities by a logarithmic function was also asummed by Shepard (1957) and Heiser (1988). The data were reanalyzed by a logoratihmic transformation. The STRESS for the two dimensional distance model with one asymmetry scale was 0.082; squared correlation between transformed data and model estimates was 0.92. By trial and error the scaling factor was set equal to the number of objects. The above described process was repeated with the transformed data. The model with a two dimensional distance model and a one dimensional asymmetry scale was chosen as a reasonable model. The common space coordinates and the asymmetry scale are reported in Table 2; the common space is displayed in Figure 3.

Table 2. Common space coordinates and asymmetry scale

Journal	dim I	dim II	asymmetry
APM	-0.063	-0.047	0.052
BJ	0.079	-0.133	0.002
EPM	-0.277	-0.120	0.016
MBR	-0.197	0.028	0.012
PB	-0.043	0.246	-0.055
PR	0.212	0.272	-0.031
JM	0.279	-0.016	0.010
PMK	0.010	-0.228	-0.005

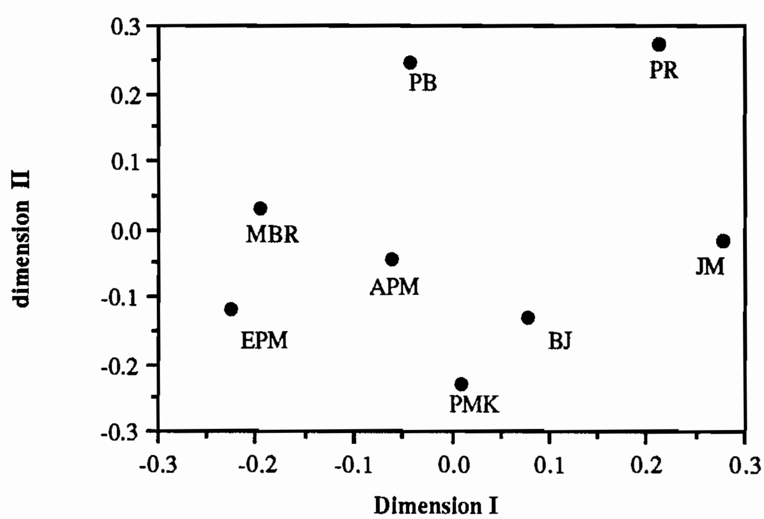

Figure 3: Common space journal citation data

The first dimension is hard to interpret, it seems that this dimension has something to do with the amount of paying attention to test-theoretic issues. The journals EPM and APM score low on this dimension, the journals PR and JM which seldom pay attention to these questions are at the right side of the dimension. The second dimension makes a distinction between mathematical versus psychological content. It seem that there is little interaction between the primarily psychological journals like PB and PR and the other more mathematical journals.

The asymmetry parameters are displayed in the third column of Table 2. The journals PMK, PB and PR were cited by the other journals more times than they were citing those journals. The papers published in these journals were used in the other journals. The journal APM was more citing the other journals than that it was cited. The dimension weights and asymmetry weights are reported in Table 3.

Table 3. Dimension weights and asymmetry weights

year	dim I	dim II	asymmetry
1981	0.399	0.282	0.266
1982	0.344	0.365	0.330
1983	0.386	0.327	0.235
1984	0.370	0.334	0.344
1985	0.323	0.383	0.451
1986	0.307	0.398	0.410
1987	0.364	0.342	0.331
1988	0.325	0.384	0.406

From Table 3 it can be concluded that there is no trend in the data in the sense that some dimensions become more and more important. It seems that the asymmetry was more important for the last four years than it was in the first four years. This may mean that current research in psychometrics solves problems coined by psychologists or that researchers illustrate their papers with realistic examples instead of synthetic data.

8.2. Marriages between ethnic groups

The following data were taken from a study by Pagnini and Morgan (1990). They examined the assortive mating among European immigrants and native whites at the turn of the century. The conclusion of the study was that there was a strong tendency to marry within the same ethnic group. They presented two data sets; a sample of marriages in the United States and a sample of marriages in New York city. The authors scored the eight groups on an interval scale according to Bogardus (1928) social ordering of the groups. This scoring was able to account for the pattern of intermarriages in the United States sample but not in the New York City sample. The New York City data set will be reanalyzed in this section to examine the social distances between the groups. The rows of the data table classify men according to their ethnic origin and the columns classify the women by their ethnic origin. The ethnic categories are listed in Table 4; the order of the categories listed corresponds to Bogardus (1928) rankorder of the groups.

Table 4. Ethnic categories New York city data

Country	label
British	Br
Irish	Ir
Scandinavian	Scan
German	Germ
Italian	It
Polish	Pol
Central European Jewish	CEJ
Eastern European Jewish	EEJ

For a definition of the categories see Pagnini and Morgan (1990). It is known wether the men and women were immigrants or born in the United States. Four types of marriages were distinguished; an immigrant man marrying an immigrant woman, an immigrant man marrying a native woman, a native man marrying an immigrant woman and a native man marrying an immigrant woman. These different types of marriages were regarded as replications. The data were converted into pseudo distances by the the gravity model.

The STRESS for the two dimensional distance model with a one dimensional skew symmetric function was 0.44. Adding more parameters to the distance model or to the skew symmetric part of the model did not yield a considerable reduction of the STRESS. The Shepard plot suggested a logarithmic transformation of the data. The data were reanalyzed with a logarithmic transformation of the data. The STRESS for the two dimensional distance model with one asymmetry scale was 0.04 (squared correlation data vs predicted values: 0.951). The common space coordinates and the asymmetry parameters are reported in Table 5 and graphically displayed in Figure 4.

Table 5. Common space coordinates and asymmetry parameters

Country	dim I	dim II	asymmetry
BR	-0.010	0.030	-0.007
IR	-0.056	0.162	-0.018
SC	-0.043	0.060	-0.011
GERM	0.033	0.119	0.000
IT	0.177	-0.043	0.045
POL	0.134	0.097	-0.009
CEJ	-0.145	-0.169	0.004
EEJ	-0.090	-0.256	-0.006

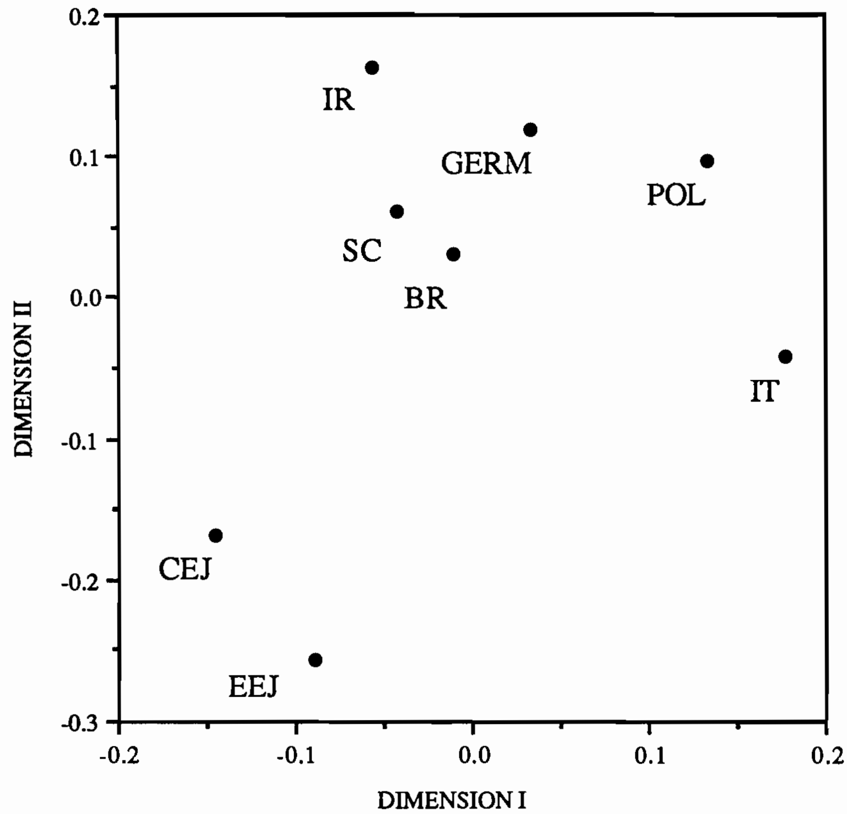

Figure 4: Common space New York city data

The two dimensional solution shows a clustering of the northern European countries with the Italians, Poles and the Jews clearly separated from them. These northern Europeans are known as the old immigrants; immigrants which live for some generations in the United States. The first dimension makes separates the Poles and Italians from the other ethnic groups. It seems that there is no dimension which is related to catholic or protestant religion: Italy, Ireland and Poland are catholic countries, the Scandinavian countries, England and Germany are predominantly protestant countries. The second dimension separates the two Jewish groups from the other groups. The dimension weights are reported in Table 6.

Table 6. Weights for dimensions and asymmetry

Generation	dim I	dim II	asymmetry
immigrant-immigrant	0.545	0.473	0.669
immigrant-native	0.528	0.483	0.621
native-immigrant	0.432	0.532	0.285
native-native	0.487	0.510	0.294

An immigrant man choses a wife from the nearest ethnic category on the second dimension. A native man choses a wife from the nearest ethnic category on the first dimension; men born in the United States only care about marrying a Jewish or a non-Jewish spouse. It can be concluded that second generation Italians become better integrated into the society of New York.

The asymmetry parameters are reported in the third column of Table 5. From the asymmetry scale we see that the Italian men are more likely to marry a woman from another ethnic category. This tendency is stronger for Italian immigrants than Italians that were borne in America. Another way of saying this is that Italian females were not allowed to marry immigrants from a different ethnic origin. The Britains, Irish, Germans and Scandinavians are almost symmetric. The central European Jews are the least popular husbands. For the second generation these effect are less important.

9. Discussion

Extending the distance function with a linear skew symmetric function might reveal interesting aspects of the stimuli. This gain of information has been attained with relatively few parameters. The present method can analyze numerical data only; monotone regression was not yet incorporated into the program, but this should not be very complicated. A problematic aspect of the dimension weights is that these weights may point to different processes. If for instance an asymmetry weights decreases it may mean that the scale becomes less appropriate as a description of the skew symmetric part of the model or that the skew symmetric part in general decreases. If the missing data pattern is symmetric this problem can be solved by separate normalization of the symmetric part and the skew symmetric part of the data.

An interesting application of the multidimensional skew symmetry model could be the extension of the weighted additive difference model for preference data (Takane, 1982) to a multidimensional model. This model has also skew symmetric elements. The possibility of such a generalization might be studied in the future.

Another line of future research could be the generalization of other models than the Thurstone Case V model to three way models and then incorporate them into a MDS program.

References

Bentler, P.M. & Weeks, D.G. (1978). Restricted multidimensional scaling models. *Journal of Mathematical Psychology*, 17, 138-151.

Bishop, Y.M.M., Fienberg, S.E. & Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. M.I.T press, Cambridge, MA.

Bogardus, E. S. (1928). *Immigration and Race Attitudes*. Heath, New York.

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

Carroll, J. D. (1972). Individual differences and multidimensional scaling. In: R.N. Shepard e.a.(Eds). *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*. Seminar press, New York.

Caussinus, H. (1965). Contribution a l'analyse statistique des tableaux de correlation. Annals of the Faculty of Science, University of Toulouse, 29, 77-182.

Constantine, A.G. & Gower, J.C.(1978). Graphical representation of asymmetric matrices. *Journal of the Royal Statistical Society (series C),* 27, 297-304.

Coombs, C.H. (1964). *A Theory of Data* Wiley, New York.

Coxon, A.P.M. (1982). *The User's Guide to Multidimensional Scaling*. Heinemann Educational Books L.T.D, Londen.

Cunningham, J.P. (1978). Freetrees and bidirictional trees as representations of psychological distance. *Journal of Mathematical Psychology* 17:165-188.

De Leeuw, J. & Heiser, W.(1980). Multidimensional scaling with restrictions on the configuration. In: P.R. Krisnaiah (Ed.). *Multivariate analysis-V*. North Holland, Amsterdam.

De Leeuw, J. & Heiser, W.(1982). Theory of multidimensional scaling. In P.R. Krisnaiah and L.N. Kanal (Eds.), *Handbook of Statistics*,Vol 2. North Holland, Amsterdam.

De Leeuw, J. (1988). Convergence of the majorization method for multidimensional scaling. *Journal of Classification* , 5, 163-180.

Ekman, G. (1963). A direct method for multidimensional ratio scaling. *Psychometrika*, 28, 33-41.

Everitt, B.(1980). *Cluster Analysis*. Heinemann Educational books, London.

Gower, J.C. (1977). The analysis of asymmetry and orthogonality. In: J.R.Barra, F. Brodeau, G. Romer & B. van Clutsem (Eds.), *Recent Developments in Statistics.* North Holland, Amsterdam.

Gower, J.C.(1984). Multidimensional scaling displays In: H.G. Law et al. (Eds.), *Research Methods for Multimode Data Analysis.* Praeger, New York.

Gower, J.C. & Digby, P.G.(1981).Expressing complex relationships in two dimensions. In: V. Barnett (Ed.), *Interpreting Multivariate Data.* Wiley, New York.

Gulliksen, H. (1956). A least squares solution for paired comparisons with incomplete data. *Psychometrika*, 20, 125-134.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika* , 33, 469-506.

Harshman, R.A., Green, P.E., Wind, Y.,Lundy, M.E.,(1982). A model for the analysis of asymmetric data in marketing research. *Marketing Science* , 1, 204-242.

Heiser, W.J. (1988). Selecting a stimulus set with prescribed structure from emperical confusion matrices. *British Journal of Mathematical and Statistical Psychology.* 41, 37-53.

Heiser, W.J.(1989). The city block model for three way multidimensional scaling. In: R. Coppi and S. Bolasco (Eds.). *Multiway data analysis.* North Holland, Amsterdam.

Heiser, W.J. (1991). A generalized majorization method for least squares multidimensional scaling of pseudodistances that may be negative. *Psychometrika* (In press).

Heiser, W.J. & De Leeuw, J. (1986). SMACOF-I. *User's guide UG-86-02,* Department of Data Theory, Leiden .

Heiser, W. J. & Meulman, J. (1983). Constrained multidimensional scaling, including confirmation. *Applied Psychological Measurement*, 7, 381-404.

Heiser, W.J. & Stoop, I. (1986). *Explicit SMACOF algorithms for individual differences scaling.* Internal report RR 86-14, Department of Data Theory, Leiden.

Holman, E.W. (1979). Monotonic models for asymmetric proximities. *Journal of Mathematical Psychology* , 20, 1-15.

Krumhansl, C.L. (1978). Concerning the applicability of geometric models to similarity data: the interelationship between similarity and spatial density. *Psychological Review*, 85, 445-463.

Kruskal, J.B. (1964 a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* , 29, 115-129.

Kruskal, J.B. (1964 b).Nonmetric multidimensional scaling: a numerical method. *Psychometrika* , 29, 115-129.

Kruskal, J.B. & Wish,M. (1978). Multidimensional scaling. *Sage university paper series on quantitative applications in the social sciences*, 07-011,.Beverly Hills and London: Sage Publications.

Laumann, E.O. & Guttman, L.(1966). The relative associational contigiuty of occupations in an urban setting. *American Sociological Review*, 31, 169-178.

Levin, J. & Brown, M. (1979). Scaling a conditional proximity matrix to symmetry. *Psychometrika* , 44, 239-243.

Luce, R.D.(1963). Detection and recognition. In: R.D. Luce et al (Eds.), *Handbook of Mathematical Psychology*. Vol I. Wiley, New York.

Meulman, J. & Heiser, W. J. (1983). Constrained multidimensional scaling: more directions than dimensions. In: T. Havranek et al. (Eds.), *Compstat 1984, Proceedings in computational Statistics*. Physica Verlag, Vienna.

Mosteller, F. (1951). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, 16, 3-9.

Nakatani, L.H. (1972). Confusion choice model for multidimensional psychophysics. *Journal of Mathematical Psychology* , 9, 104-109.

Okada, A. (1988 a). Asymmetric multidimensional scaling of car switching data. In W. Gaul & M. Schader (Eds.), *Data, Expertknowledge and decisions* (pp. 279-290). Springer verlag, Berlin.

Okada, A. (1988 b). An analysis of intergenerational occupational mobility by asymmetric multidimensional scaling . In H. van Schuur (Ed), *The Many Faces of Multivariate Analysis*. Procedings of the SMABS 88 conference, Rion, Groningen.

Pagnini, D.L. & Morgan, S.P. (1990). Intermarriage and social distance among U.S. immigrants at the turn of the century. *American Journal of Sociology*, 96, 405-432.

Shepard, R.N. (1957). Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22,325-345.

Takane, Y. (1978). Maximum likelihood additivity analysis. *Psychometrika*, 47, 225-241.

Thurstone, L.L. (1929). Theory of Attitude measurement. *Psychological Review*, 36, 222-241.

Tijssen, R., De Leeuw, J. & Van Raan, T. (1986). Quasi correspondence analysis of scientometric transaction matrices. *Internal report RR-86-11*, Leiden.

Tobler, W.R. & Wineburg, S. (1971). A Cappadocian speculation. *Nature*, 231,39-42.

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

Tversky, A. (1977). Features of similarity. *Psychological Review,* 84, 327-252.

Van de Geer, J.P. (1986). *Introduction to Lineair Multivariate Data Analysis,* Vol 1. D.S.W.O. Press, Leiden.

Van der Heijden, P.G.M. (1987). *Correspondence Analysis of Longitudinal Categorical Data*. D.S.W.O. Press, Leiden.

Van der Heijden, P.G.M., De falguerolles, A. & De Leeuw, J. (1989). A combined approach to contingency table analysis using correspondence analysis and loglineair analysis. *Journal of the Royal Statistical Society (series C)*.38, 249-292.

Weeks D.G. & Bentler, P.M. (1982). Restricted multidimensional scaling models for asymmetric proximities. *Psychometrika* , 47, 201-208.

Young, F.W. (1975). An asymmetric Euclidean model for multi-process asymmetric data. Proceedings of the U.S.-Japan Seminar, University of California, San Diego.

Young, F.W. (1984). The general Euclidean model. In: H.G. Law et al. (Eds.), *Research methods for multimode data analysis*. Praeger, New York.

Zielman, B. (1991). *Asymmetry in Euclidean multidimensional scaling*. Research Report, Department of Data Theory, Leiden.