SPATIAL REPRESENTATIONS OF ASYMMETRIC PROXIMITIES

Berrie Zielman

Willem J. Heiser

Department of Data Theory

University of Leiden

Spatial Representations of Asymmetric Proximities

Berrie Zielman* and Willem J. Heiser

Department of Data Theory

University of Leiden

Abstract

Ordinary multidimensional scaling models are not appropriate for the analysis of square asymmetric matrices, because the Euclidean distance function is symmetric. A number of methods has been proposed to accommodate asymmetry, which can all be viewed as special cases of a general similarity-bias or hybrid model. In this paper asymmetry is viewed as a combination of symmetric similarity and dominance and the differences and similarities between the methods are revealed by applying a certain decomposition to the model parameters, clearly separating skew-symmetric dominance and symmetric similarity. The notion of skew-symmetry turns out to be an often seen element in modeling asymmetry, although sometimes in disguise and difficult to recognize.

There are some general methods for the analysis of rectangular tables that do not fit so easily in the developed framework. However, when applied to square tables these methods have interesting special cases: the DEDICOM model and the slide-vector model.

Key words: multidimensional scaling, asymmetry, skew-symmetry, similarity-bias model, DEDICOM, unfolding, feature matching model, distance density model, choice model

1. Asymmetries

Tables where the rows and columns classify the same set of entities occur in several situations. These tables are square and often asymmetric; object i is more often associated with or substituted by object j than the other way round. Other possible examples are:

brand switching counts   like/dislike judgments

counts of telephone calls among cities   citations among journals

sociometric choices   confusions of one stimulus with another

first choice-second choice connections   migration-rates

occupational mobility tables   communication and volume flows.

These stimulus comparison data are called by Coombs (1964, p403) symmetric proximity data if the table is symmetric and conditional proximity data if the table is asymmetric. In the case of symmetric proximity data all entries are comparable within the table. In the case of conditional proximity data the measures are only comparable within each row of the table.

Another type of situation in which asymmetries arise is the case of conjoint distances (Coombs, 1964, p44). Conjoint distances occur when pairs of pairs of stimuli are compared and both pairs have a stimulus in common. Examples of observational schemes for these data are: picking k out of n-1 stimuli that are most similar to the reference item, ordering k out of n-1 objects in terms of relative similarity to the reference item, the method of n-dimensional rank-order, the method of anchor-point ordering and the method of triads.

In the psychological literature, asymmetries are often regarded as response biases, context effects or sampling errors added to a symmetric structure (Holman and Marley, 1972). Researchers tend to preprocess their data to make them symmetric and then apply a

2

MDS model, acting in line with Beals et. al. (1968), who remark: "if asymmetries arise they must be removed by averaging or by an appropriate theoretical analysis that extracts a symmetric dissimilarity index". On the other hand one might assume that asymmetry may carry important information, for instance, Tversky (1977), who developed his feature matching model from the point of view that a similarity relation is asymmetric instead of symmetric.

It is remarkable that the majority of the models to be discussed in this paper represent asymmetric proximities by distinguishing a similarity component and dominance or preference component. In our experience this mixture of similarity and preference is often reflected in the observations. The entries in the table possibly indicate both a proximity relation and a dominance relation. Journals citing each other often can be regarded as similar; that journal that is cited more than it is citing is the dominant member and could be a prestigious journal. Asymmetry may have various meanings. For instance, when one is studying interactions among cities, large cities interact more often with small cities than the reverse. In market structure analysis, asymmetry in brand switching may indicate which brand attracts consumers from the other brands. In sociometric research, asymmetry can be a measure of the popularity of a person. Wish (1967) studied residuals from a multidimensional scaling analysis and found interesting order effects in a study of confusions among Morse code signals. Tversky (1977) found that the less prominent stimulus was more similar to the prominent stimulus than the prominent stimulus was to the less prominent one, in a study of judging similarity among nations. North Korea was judged more similar to China than the reverse. Nosofsky (1991) related other constructs such as salience, hierarchical status, good stimulus, easily encoded stimulus and high-frequency stimulus to the asymmetry between stimuli.

A popular method for representing these observations in a low-dimensional space is multidimensional scaling (MDS). MDS methods represent dissimilarity measures collected among n objects. The objects can be anything: journals, personality traits, brands, persons,

3

cities and so on. The only necessary ingredient is that we have a measure defined on pairs of objects indicating the dissimilarity or similarity between these two objects. Multidimensional scaling models search for a spatial representation of the objects or stimuli in a space of low dimensionality in such a way that the distances, denoted as $d_{ij}(\mathbf{X})$ with $(i=1,....n; j=1....n)$, among the n points approximate the dissimilarities, denoted as q_{ij}, as closely as possible. In this paper the Euclidean distance function is considered, by far the most commonly used MDS model, defined as:

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

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

The Euclidean distance function satisfies the following axioms:

$d_{ij} > d_{ii} = 0$ (minimality)

$d_{ij} = d_{ji}$ (symmetry)

$d_{ij} \leq d_{ik} + d_{jk}$ (triangle inequality)

The minimality axiom states that the distance between two objects should always be greater than or equal to zero; the distance between an object and itself should be zero. The symmetry axiom states that the distance from i to j should be equal to the distance from j to i. The triangle inequality states that the distance from i to j is smaller or equal to the distance from i to j if we travel via k. This paper studies the case where the symmetry axiom does not hold for the dissimilarities.

Although there are more methods available for analyzing these observations, for instance cluster analysis, we focus our attention primarily on MDS methods because this class of models can be regarded as a prototype of a symmetric model. Some of the decompositions we discuss can be used in combination with cluster analysis as well. For overviews of MDS we refer to Kruskal and Wish (1978), Carroll and Arabie (1982), Carroll and Wish (1982) and Coxon (1982).

4

This paper is built around two organizing themes, a general model proposed by Holman (1979) and a decomposition theorem from linear algebra. A psychological interpretation of this decomposition is a decomposition into a similarity and a dominance or preference component. The interrelations between the methods can be easily seen through these mirrors because it enables us to discuss asymmetric models in a single framework and the decomposition often suggests a reparameterization of the parameters into similarity and preference parameters that is more easily interpreted than the original parameters. The general model proposed by Holman (1979) will be called a hybrid model, a term borrowed from Carroll (1976). The hybrid model assumes proximity to be a function of symmetric similarity and asymmetric row and column bias. This general model was called a similarity-bias model by Nosofsky (1991), he also proposed a generalization of this model by including asymmetric-similarity components. All models in section 3 can be viewed as a hybrid model, a mixture of symmetric and asymmetric parameters.

We start our discussion with the decomposition of an asymmetric matrix. In section 2 we give an overview of existing methods for analyzing asymmetric tables. The interrelations between the methods are discussed by the decomposition theorem and the similarity-bias model. General methods for analyzing rectangular tables that can be specialized for the analysis of a square matrix are discussed in section 4.

2. Decomposition of asymmetric matrices into symmetric and skew-symmetric components

In this section we discuss decompositions of an asymmetric matrix: an additive decomposition, a multiplicative decomposition and a decomposition into an upper and lower triangle. It is shown that the additive and multiplicative decomposition yield a separation of the similarity and dominance aspects. In section 2.4 we discuss the singular

5

value decomposition of a special type of asymmetry which is called skew-symmetry or anti-symmetry. In section 2.5 we discuss other models for a skew-symmetry matrix.

2.1   Averaging

A simple and very useful result is the following. Any square non-symmetric matrix Q with n rows and n columns can be additively decomposed into a symmetric and a skew-symmetric matrix,

$$Q = S + A, \tag{1}$$

where S is a symmetric matrix of averages $s_{ij} = \{q_{ij} + q_{ji}\}/2$ and A a skew-symmetric matrix with elements $a_{ij} = \{q_{ij} - q_{ji}\}/2$. The property $a_{ij} = -a_{ji}$ is called *skew-symmetry* and sometimes *anti-symmetry*. The matrix A describes the departures from symmetry, and can be viewed as the preference or dominance part of an asymmetric matrix; if $q_{ij} > q_{ji}$ then $a_{ij} > 0$. The matrix S describes the departures from symmetry, and can be viewed as a matrix with (dis)similarities. The matrices A and S are uncorrelated, the sum of squares of the matrix Q can be decomposed into sum of squares due to symmetry and sum of squares due to skew-symmetry:

$$\Sigma_i \Sigma_j\, q_{ij}^2 = \Sigma_i \Sigma_j\, s_{ij}^2 + \Sigma_i \Sigma_j\, a_{ij}^2.$$

Because of this split of sum of squares, the two components can be viewed independently.

The matrix S is the best symmetric approximation to the matrix Q in the least squares sense. A test for symmetry in a proximity matrix has been developed by Hubert and Baker (1979).

6

In the case of frequency data the matrix S is equal to the maximum likelihood estimator of the symmetric matrix under multinomial sampling; the fit of the symmetric matrix can be tested by the chi-square statistic (Bowker, 1948):

$$\chi^2 = \Sigma_i \, {}_{<}\Sigma_j \frac{(q_{ij} - q_{ji})^2}{q_{ij} + q_{ji}}.$$

This statistic follows a chi-square distribution with $n(n-1)/2$ degrees of freedom, where n is the number of objects in the table. Another diagnostic proposed by Carroll and Wish (1972), to assess the severity of the violations from the symmetry axiom is to study the rank order of the corresponding rows and columns. If the rank orders of the rows are unrelated to the rank orders of the columns then the symmetry axiom is untenable. If the asymmetry in the data is assumed to be noise, the matrix A can be ignored and the symmetric matrix S can be analyzed by an MDS program.

2.2   A multiplicative decomposition of a square matrix.

Another possibility of decomposing a square matrix is by writing this matrix as a product of the geometric mean (cf Arabie & Soli, 1978) and the square root of the odds,

$$q_{ij} = \xi_{ij} \, \omega_{ij},$$

where $\xi_{ij} = \sqrt{q_{ij} q_{ji}}$ denotes the geometric mean and $\omega_{ij} = \sqrt{\frac{q_{ij}}{q_{ji}}}$ denotes the square root of the odds, which is generally defined as the ratio of a probability and its complement. Suppose q_{ij} denotes the flow from category i to category j. Then, more specifically, we can speak of q_{ij} as the inflow, and q_{ji} as the outflow. If the odds is greater than one there is more inflow than outflow, object i dominates object j with respect to inflow. If the inflow equals the outflow the odds equals one; thus the odds represents the balance of the system.

We can still obtain an additive decomposition by taking the logarithm of the product of the odds and geometric mean:

$$\log \omega_{ij} = \tfrac{1}{2}(\log q_{ij} - \log q_{ji}) = \tfrac{1}{2}(\phi_{ij} - \phi_{ji}) = a_{ij}(\phi).$$

The logorithm of the odds is called the logit; the above result shows that the matrix with logits is skew-symmetric. A similar result can be obtained for the geometric mean:

$$\log \xi_{ij} = \log \sqrt{q_{ij} \, q_{ji}} = \tfrac{1}{2}(\phi_{ij} + \phi_{ji}) = s_{ij}(\phi).$$

Note that this reparameterization implies that the matrix $\mathbf{Q}$ has all elements greater than zero.

2.3   Decomposition into an upper and lower triangle

Suppose we perform one MDS analysis on the data elements below the diagonal and another on the data elements above the diagonal (cf. Laumann and Guttman,1966). This procedure implies that the proximity from i to j and from j to i are viewed as different processes, and are scaled twice: we obtain two configurations representing inflow and outflow. The two resulting configurations are compared by visual inspection or rotated toward each other by Procrustus rotation (Cliff, 1966). The procedure could be called multiplicative because the Procrustes procedure amounts to applying a weight matrix to the configuration.

Although intuitively plausible this idea may have unexpected consequences: if we interchange or permute some rows and their corresponding columns we obtain a different set of triangles. The analysis of two triangular matrices is not invariant of permutation of the rows and columns. A proposal by Gower (1977) to make the analysis invariant over permutations is to permute the rows and columns in such a way that the asymmetry is maximized, a possible technique for doing this is seriation (cf Huber, 1976). In the ideal

8

case the permutation results in a square matrix where all elements above (below) the diagonal are smaller than their corresponding elements below (above) the diagonal. This permutation yields a worst possible interpretation of an asymmetric matrix.

2.4   Decomposition of a skew-symmetric matrix

Gower (1977), Constantine and Gower (1978) and Gower and Digby (1981) studied the singular value decomposition of the skew-symmetric matrix **A**. The singular value decomposition is a bilinear method, which means that there are two sets of parameters, each of which forms a linear function with respect to the other. It decomposes any matrix **B** into a product of the form:

$$\mathbf{B} = \mathbf{W} \, \Lambda \, \mathbf{V'}.$$

Here **W** and **V** are both orthogonal matrices, i.e. $\mathbf{W'W} = \mathbf{I}$ and $\mathbf{V'V} = \mathbf{I}$, and Λ is a diagonal matrix with singular values. For a skew-symmetric matrix **A** the singular values come in pairs, i.e. Λ contains the singular values $\lambda_1, \lambda_1,..,\lambda_{n/2}, \lambda_{n/2}$, with the last singular value being equal to zero when n is odd. Due to this peculiarity, the singular value decomposition of a skew-symmetric matrix can be rewritten into a form that better expresses its fundamental structure:

$$\mathbf{A} = \mathbf{W} \, \Lambda \, \mathbf{J} \, \mathbf{W'},$$

where **W** and **J** are again orthogonal matrices and Λ is the diagonal matrix of singular value decomposition as defined above. The matrix **J** is a block diagonal matrix with 2 by 2 sub-matrices with zero's on the diagonal, 1 above the diagonal and -1 below the diagonal. When n is odd the last diagonal position is filled with a zero. The presence of **J** makes the

9

left singular vectors **W** a permutation and reflection of the right singular vectors **WJ'**. The two-dimensional model is given by the typical elements

$$a_{ij} = \lambda_1(w_{i1}w_{j2} - w_{i2}w_{j1}). \tag{2}$$

When the objects o_i with coordinates (w_{i1}, w_{i2}) are plotted in a two dimensional space, the area of the triangle with vertices at the two points and the origin O is an approximation of the element a_{ij}. This diagram is also called a Gower diagram. The areas of the triangles O o_i o_j and O o_j o_i are equal, but they have opposite sign, thus modelling skew-symmetry. Two points may be far apart while there is still a perfect symmetric relation, this happens when two points are located on a line that passes through the origin. The representation of points by vectors will be more useful; this is illustrated in Figure 1.
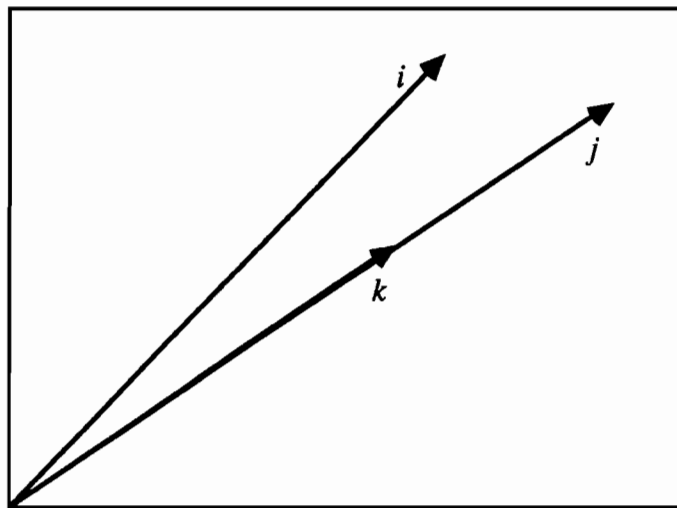

Figure 1: Representation of asymmetry by the bilinear model

Three objects are depicted in Figure 1; the relation between objects j and k is symmetric, the relations of these two objects to object i are asymmetric. The greater the angle between two

10

vectors, the larger the asymmetry is. Figure 1 says nothing about the symmetric similarity between the points; this information must be displayed in an additional plot.

If all points are collinear on a line (or almost collinear) the asymmetric part of the data can be modeled by a linear skew-symmetric function. This special case has been studied by Weeks and Bentler (1982), Okada (1988 a,b), Takane and Shibayama (1985) and Holman (1979) and - in the context of Thurstone case V scaling - by Mosteller (1951) Gulliksen (1956) and Torgerson (1958). This simple linear form of skew-symmetry can be written as:

$$a_{ij}(\mathbf{r}) = r_i - r_j.$$

Note that this equation is obtained from equation (1) if we substitute $w_{i2} = w_{j2} = 1$ and, $r_i = \lambda_1 w_{i1}$. Observe that if this simple model is true, we can permute the rows and columns of the matrix $\mathbf{Q}$ by the order of $\mathbf{r}$, this yields a seriated matrix (see section 2.3).

The linear model predicts asymmetry from the differences between the row and column marginals. The model discards information from the individual cells. This can be shown inserting the definition of a skew-symmetric matrix into the equation for computing the r_i parameters; which amounts to taking the row means of $\mathbf{A}$:

$$\mathbf{r} = 1/n \; \mathbf{A}\mathbf{e} = 1/n \; (\mathbf{Q}-\mathbf{Q'})\mathbf{e} = 1/n \; (\mathbf{Q}\mathbf{e} - \mathbf{Q'}\mathbf{e}),$$

where $\mathbf{e}$ denotes an n-vector with unities. So only the row and column sums of $\mathbf{Q}$ are relevant for determining the least squares estimates of $\mathbf{r}$. This linear form is the basis of most asymmetric models (cf section 3.1, 3.3, 3.4, 3.5, 3.6, 3.7, 3.9).

2.5 Other models for a skew-symmetric matrix

Up to a monotonic transformation of the data the skew-symmetric part of the linear model is equal to the Thurstone case V scaling model. It might be interesting to fit other Thurstonian models to the skew-symmetric part of the data. A general approach to Thurstonian scaling, subsuming a large number of models is given by Takane (1980) and Heiser and De Leeuw (1981). They proposed the non-linear model:

$$a_{ij} = \frac{r_i - r_j}{d_{ij}(\mathbf{X}^*)}$$

Here the skew-symmetric part of the original model is divided by the distance between points in the space $\mathbf{X}^*$. When fitting the Heiser-de Leeuw-Takane model to the skew-symmetric part, the configuration $\mathbf{X}^*$ will in general differ from the configuration $\mathbf{X}$ obtained from the MDS method on the symmetric part. If this configuration matrix $\mathbf{X}^*$ is restricted to be a diagonal matrix the Thurstone Case III scaling model is obtained.

A restricted version of the model, called the wandering vector model has been proposed by Carroll (1981) and De Soete and Carroll (1986). In the wandering vector model the skew-symmetry is depicted as a direction of increasing dominance in the multidimensional space.

In this section we have explained that any square matrix can be decomposed into a symmetric matrix and a skew-symmetric matrix and that these matrices can be interpreted as similarities and preferences (or dominances). This theorem will return as an organizing theme in the next section, which starts with an introduction of the similarity-bias model, our second organizing theme.

3. Methods for analyzing square tables

This section starts with a discussion of the similarity-bias model; in subsequent sections it will be shown how various other models are related to the similarity-bias model. The decomposition into a symmetric part and a skew-symmetric part shows the similarity and dominance aspects of the various models.

3.1   The similarity-bias or hybrid model

Holman (1979) proposed a general linear model for the analysis of asymmetry. This model is, especially in stimulus identification experiments, relatively easy to interpret in terms of symmetric similarity and asymmetric bias for the rows and columns. The proximity of stimulus i to stimulus j is given by:

$$\delta_{ij} = F(s_{ij} + r_i + c_j)$$

where F is a general monotonic function, s_{ij} is a symmetric similarity function and r_i and c_j are bias functions on the rows and columns. The similarity-bias model is a hybrid model (Carroll, 1976; Carroll and Pruzansky, 1981) because it allows a mixture of models: the similarity function can be continuous and the bias functions can be discrete. The model can represent differential self similarity and asymmetry if $r_i \neq c_j$.

The bias components of the model, $r_i + c_j$, can be decomposed using the theory of section 2.1 into a symmetric similarity and skew-symmetric preference part by defining new parameters $u_i = (r_i + c_i)/2$ and $a_i = (r_i - c_i)/2$. The asymmetric part or bias components of the model can now be written as the sum of a skew-symmetric part and symmetric component

$$r_i + c_j = (u_i + u_j) + (a_i - a_j),$$

13

where $u_i + u_j$ is symmetric and can be thought of as unique dimensions (Bentler and Weeks, 1978) or a star-tree (Carroll, 1976); this star-tree can accommodate high centrality or nearest neighbor data (Tversky and Hutchinson, 1986). If object i is a nearest neighbor in the set, the object is the most similar object to all the other objects. Multidimensional scaling imposes a bound on the number of objects that can be near to an object. If object i is the nearest neighbor the corresponding u constant is the smallest. The dominance term $a_i - a_j$ is skew-symmetric: $a_i - a_j = -(a_j - a_i)$. This is the simplest form of skew-symmetry; the points should lay on a straight line in the Gower diagram (cf section 2.4). A number of models are special cases of this general hybrid model, for instance Weeks and Bentler (1982), Saito (1986) and Okada (1988 a,b) are special cases with the Euclidean distance function as a similarity function and a linear skew-symmetric function to accommodate asymmetry.

Nosofsky (1991) proposed an extension of the similarity-bias model. In addition to bias-related asymmetries he proposed similarity-related asymmetries:

$$\delta_{ij} = F(s_{ij} + r_i + c_j + m_{ij})$$

where m_{ij} is a similarity-related asymmetry or dominance component. For identification purposes we require s_{ij} to be symmetric, m_{ij} to be skew-symmetric and r_1 and c_1 to be zero. Other constraints could be imposed on the bias-parameters, for instance that their sums are zero.

If we analyze an asymmetric matrix by a model that is the sum of a Euclidean distance and a two dimensional singular value decomposition of the skew-symmetric matrix, that is,

$$\delta_{ij} = d_{ij}(X) + \lambda_1(w_{i1}w_{j2} - w_{i2}w_{j1}),$$

14

we obtain an asymmetric similarity model with no bias components. Similarly, if we analyze the symmetric matrix by a distance model and the skew-symmetric matrix by the model proposed by Takane (1980) and Heiser and De Leeuw (1981) (see section 2.4) we obtain an asymmetric similarity model.

We conclude this section with the remark that the bias-components r_i and c_j were initially incorporated to accommodate asymmetry, but they influence the symmetry as well. This need not be an unfortunate state of affairs because the symmetric part of the bias components adds a star-tree to the model that can be interpreted as centrality bias.

3.2   Feature matching model

Tversky (1977) challenged the dimensional-metric assumptions that underlie the geometrical approach to the analysis of similarity. From a set theoretical viewpoint the feature matching model was developed. The feature matching model assumes that each object is characterized by a set of features. The similarity between objects i and j is expressed as a function of their common and distinctive features. The additive version of the model is called the contrast model. In terms of similarities we have:

$$q_{ij} = \theta \, f \, (i \cap j) - \alpha \, f \, (i - j) - \beta \, f(j - i),$$

where $(i \cap j)$ is the number of features shared by objects i and j, $(i - j)$ are the set of features unique to object i with respect to object j, $(j - i)$ is the set of unique features belonging to object j with respect to object i and f is a measure function of the features. This model is said to differ from the other models described in this paper by assuming a psychological hypothesis instead of a mathematical hypothesis. The psychological content of the model is discussed in Tversky and Gati (1978). The parameters θ, α, β are assumed to be positive; they must be estimated from the data, which can be done by linear regression

techniques. The function f measures the contribution of the individual features to the similarity between the objects. If the values of α and β are different the model is capable of representing asymmetry. In addition to representing asymmetry, the model describes differences in self-similarities of the objects as well. The feature matching model implies that if i is more similar to j it must be true that j is more self similar than i (Nosofsky, 1991).

The feature matching model is an example of an asymmetric similarity model because the number of unique features of an object is a relational term. The number of unique features depends on the feature set of the comparison object. In the special case that the measure function f is additive, the model can be viewed as a similarity-bias model (Nosofsky, 1991; Holman, 1979).

The major problem of the feature matching model is that a suitable feature set has to be defined. This problem is sometimes relatively easy to solve, for instance in case of a letter confusion matrix (Keeren and Baggen, 1981), and sometimes difficult to solve, for instance in the case of a matrix with sociometric interactions.

3.3   The distance-density model

The distance-density model (Krumhansl, 1978, 1982, 1988) extends the ordinary distance function with additional components. Dissimilarity is modelled as a function of the inter-point distance and the local density of points in the configuration. The formal structure of the model is:

$$d_{ij}(\mathbf{X}, \alpha, \beta) = d_{ij}(\mathbf{X}) + \alpha v_i + \beta v_j,$$

where v_i is a measure of density of points surrounding point i in the configuration. The weights α and β must be estimated from the data. The distance-density model assumes that within dense subregions finer discriminations are made than within less dense subregions.

16

Two points within dense subregions have smaller similarities than two points of equal interpoint distance within less dense subregions. This modified distance function need not satisfy the minimality and symmetry axiom. If the weights α, β are unequal, the model is capable of modelling asymmetry. As in the feature matching model the distance-density model also represents the diagonal of a proximity matrix. The model predicts the opposite relation between self-similarities and asymmetries compared to Tversky's model. If i is more similar to j then it is also the case that i is more similar to itself than j.

Krumhansl (1978) proposed three measures of density: first, the self similarities (the diagonal of the original table); second, the weighted sum of the distances from an object to the other points, weighted in such a way that the small distances contribute more to the density than the large distances; and thirdly the number of points within a fixed radius of the point. Other measures of density can be obtained from cluster analysis. Examples are: the number of nodes to pass in a hierarchical cluster diagram, or the number of objects sharing the same cluster.

The distance-density model is a special case of the similarity-bias model (Nosofsky, 1991), with a Euclidean distance as the similarity function, and the density of points as a bias function.

The symmetric elements of the distance-density model are $d_{ij}(X) + (\alpha+\beta)(v_i + v_j)$, and skew-symmetric elements $(\alpha-\beta)(v_i - v_j)$

Krumhansl (1982) studied the application of the distance-density model and the feature matching model in the analysis of letter confusion matrices and concluded that the models had surface similarities. The distance-density model has been extended to tree models by De Sarbo et. al. (1990).

3.4   Quasi-symmetry model

The model of quasi-symmetry (Caussinus, 1965) supposes a symmetric model for pairs but includes parameters for the row categories α_i and column categories β_j which are allowed to be different, and which enter the model multiplicatively:

$$q_{ij} = \mu \, \alpha_i \, \beta_j \, \eta_{ij}.$$

Since the η_{ij} parameters are symmetric, they can be analyzed by a multidimensional scaling program. To identify the parameters of the model, we have to specify some constraints, a discussion of possible constraints can be found in Constantine and Gower (1982). The quasi-symmetry model can be written as a similarity-bias model (Holman, 1979) with an exponential function and $s_{ij} = \log \eta_{ij}$, $r_i = \log \alpha_i$ and $c_j = \log \beta_j$. This reformulation as a hybrid model can be further simplified by reparameterizing into symmetric and skew-symmetric bias terms.

Using the odds it can be shown that the model of quasi-symmetry has a linear form of skew-symmetry:

$$\log \, \omega_{ij} = \tfrac{1}{2}(\log\alpha_i - \log\beta_i) - \tfrac{1}{2}(\log\alpha_j - \log\beta_j)= \phi_i - \phi_j,$$

were $\phi_i = \tfrac{1}{2}(\log\alpha_i - \log\beta_i)$, a result that can be found in Agresti (1990).

To compare the model of quasi-symmetry to the additive models we rewrite the model in terms of symmetric and skew-symmetric components. The skew-symmetric elements of the quasi-symmetry model are of the form:

$$a_{ij} = (\alpha_i \, \beta_j - \beta_i \, \alpha_j)\eta_{ij}.$$

18

The first term on the right-hand side is of a form similar to a two-dimensional singular value decomposition of a skew-symmetric matrix. This suggests an elegant way to interpret the row parameters α and the column parameters β. These parameters can be plotted in a two dimensional space and this space can be interpreted in terms of areas and collinearities, as discussed in section 3.3. We must bear in mind that the areas between points are blown up or shrunk by the symmetry parameters of the model. This formulation shows that the skew-symmetry in the data is modelled as a product of the skew-symmetric and symmetric parameters of the model.

3.5   Choice model

The choice model assumes that the ratios of the probabilities of choosing a stimulus do not change if the number of stimuli increases, or if different stimuli are included in the choice set. The model is related to the Bradley-Terry model (Luce, 1963) and can be written as:

$$q_{ij} = \frac{\beta_j \, \eta_{ij}}{\Sigma_k \beta_k \, \eta_{jk}} .$$

The β parameters reflect the tendency to favor some responses over others, and they will account for at least part of the asymmetry in the data. Usually they are interpreted as bias parameters. For identification purposes we may require $\Sigma_k \beta_k = 1$ and $\eta_{ii} = 1$ for all i. The η parameters are symmetric; these parameters can be further analyzed by a MDS program. Note that the normalizing term in the denominator, which ensures that the row sums of $\mathbf{Q}$ are one, can also be seen as a row-specific factor accounting for asymmetry, since even if $\beta_j = 1$ for all j, $\mathbf{Q}$ will generally not be symmetric. It is the weighted average similarity of object i towards all other objects. The choice model is a product-multinomial model, which

regards the row totals as fixed. If it is not assumed that the row totals are fixed the model of quasi-symmetry is obtained (Heiser, 1988). The logit of the choice model can be written as:

$$\log \omega_{ij} = \phi_j - \phi_i,$$

where $\phi_j = \beta_j - \Sigma_k \beta_k \, \eta_{jk}$. The model is a special case of the similarity-bias model (Nosofsky, 1991) with a similarity function η_{ij}, a row bias term $\log \beta_i$ and a column bias term $\log \Sigma_k \beta_k \, \eta_{jk}$. Rewriting the choice model as a hybrid model we can distinguish symmetry and skew-symmetry parameter. The choice model has been extended by Nakatani (1972), Van Putten (1982) and Holman (1979). Takane and Shibayama (1986) showed the relation of the choice model to the distance-density model.

3.6 Scaling to symmetry

A method proposed by Levin and Brown (1979) is multiplicative scaling of the rows or columns in such a way that the symmetry of the rescaled datamatrix is maximized. This procedure has the problem of the choice of the side constraints on the scaling factors, since different constraints lead to different solutions. The analysis yields rescaling coefficients, which can be interpreted as a tendency to favor some particular response over others. The rescaled data matrix can be analyzed by a MDS method. The rescaling coefficients operate on the data; this in contrast to the choice model and the quasi-symmetry model. Implicitly the procedure assumes a model of the form:

$$QV = S,$$

where V is a diagonal matrix of rescaling coefficients. When a perfect rescaling is possible in the sense that S is indeed symmetric, we may rewrite this equation as $Q = SV^{-1}$, then we

apply the decomposition into a symmetric and skew-symmetric component from section 2.1 to the matrix Q, and the skew-symmetric component in the data is modeled by:

$$a_{ij} = (\alpha_i - \alpha_j) s_{ij},$$

where $\alpha_i = 1/v_{ii}$. In terms of the decomposition the skew-symmetric component of the data is weighted by the symmetric part of the model. For objects with different rescaling coefficients the model predicts severe asymmetries if the symmetry is large. If we assume that the symmetry parameters s_{ij} are inversely related to the distance we obtain the Heiser, De Leeuw and Takane extension of the linear model from section 2.5. These parameters α can also be obtained by the linear model discussed in section 2.5 when the elements of the asymmetric matrix Q are transformed by the logarithm. This model is a special case of the similarity-bias model where the column bias components are restricted to one and the bias components for the rows are estimated.

A more general solution of $QV = S$ is possible by requiring V to be orthogonal. The matrix V can be found by $V = LM'$ from the singular value decomposition of

$$Q = L\Sigma M'.$$

The matrix S has the structure $L\Sigma L'$. A method for studying the matrix LM' is given in Gower (1977).

3.7   Row conditional transformations

Another possibility for dealing with asymmetry is to transform the asymmetry away by monotone or some other form of regression simultaneously with a multidimensional scaling program (Kruskal, 1964). This can be done row-conditionally which means that values *within* rows are regarded as comparable with each other, and values *among* the rows

21

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 that 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. One could, of course, also transform column-conditionally, in which case the interpretation would have to be in terms of stimulus bias instead of response bias (Nosofsky, 1991).

3.8   ASYMSCAL

Young (1975, 1984, 1987) proposed the following weighted Euclidean distance model to represent asymmetry:

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

where the weights w_{is} are specific for each stimulus i and dimension s. When the weights are unity the model reduces to the Euclidean distance model. This is the simplest model; Young also indicates a rotation model and a reduced rank model but these are not further discussed in the literature. The asymmetry is accounted for by row or stimulus weights operating on the dimensions of the configuration. The resulting configuration can be interpreted as a "birds eye view" (Collins, 1984). For each stimulus the configuration must be adjusted by shrinking or stretching the axes. Thus for every stimulus the analysis yields p row weights resulting in n by p additional parameters. This makes the configuration very difficult to interpret. A remedy to this problem is to average the weights over meaningful clusters of objects and then plot spaces for groups of objects using the averaged weights.

The ASYMSCAL procedure has skew-symmetric elements:

$$\tfrac{1}{2}\Sigma_s(w_{is} - w_{js})(x_{is} - x_{js})^2,$$

22

which are similar to the skew-symmetric elements of the rescaling method discussed in section 3.6, except that the object weights or rescaling coefficients are now related to the dimensions. The differences in skew-symmetry between the object weights are larger for dissimilar objects, because these differences are weighted by $(x_{is} - x_{js})^2$.

The ASYMSCAL model can be viewed as the sum of p similarity-bias models in multiplicative form with a symmetric similarity function $\log(x_{is} - x_{js})^2$, a row bias $\log w_{is}$ and the column bias equal to one.

3.9   Bidirectional trees

Additive trees (Sattath and Tversky, 1977) represent objects as "leaves" on a tree in such a way that distances calculated between the leaves on the tree correspond as closely as possible to the dissimilarities. Additive trees are also known under the name "free" tree (Cunningham, 1978) and "pathlength tree" (Carroll, 1976). In trees we can distinguish visible and invisible nodes; visible nodes represent objects and invisible nodes represent clusters of objects. Cunningham (1978) generalized the tree to a bidirectional tree by allowing differential weighting of the pathlenths corresponding to different directions to represent asymmetry. The length of a sequence of links from i to j will in general differ from the length from j to i. Carroll and Pruzansky (1975) have shown that an additive tree can be decomposed into a ultrametric tree and a star tree. A star tree is a tree with one internal node connecting all objects and this structure can also be represented by the linear form $u_i + u_j$. In the case of a bidirectional tree this decomposition yields a asymmetric star-tree that can be decomposed into $u_i + u_j + a_i - a_j$. This parameter structure is identical to the asymmetric part of the similarity-bias model and thus the bidirectional tree can be viewed as a special case of Holman (1979) additive model with an ultrametric tree as a similarity function and row and column bias, where it is convenient to decompose these bias-parameters into a skew-symmetric and symmetric part.

4. Joint representation of rows and columns.

The methods discussed in this section are methods for analyzing rectangular tables. A square table is a special case of a rectangular table, and therefore square asymmetric tables can be analyzed by these more general methods. These methods, components analysis and unfolding, do not easily fit in the hybrid framework. However, the DEDICOM model and the slide-vector model can be viewed as constrained versions of these general models that do fit in the framework of the previous section and we restrict the discussion to these two methods.

4.1 The DEDICOM model

The DEDICOM (DEcomposition into DIrectional COMponents) model was proposed by Harshman et. al. (1982). The model takes a non-spatial approach first but afterwards the parameters can be graphically displayed. To show how the DEDICOM model is related to components analysis, we write the general decomposition of a datamatrix:

$$Q = LH',$$

where L is an n by p matrix with component scores and H is an n by p matrix with component loadings. The DEDICOM model is a special case of factor analysis or components analysis in the sense that the factor loadings for the rows are a linear transformation of the factor loadings of the columns, and can be written as:

$$Q = L R L'$$

24

The matrix **L** denotes an *n* by *p* matrix of loadings of the *n* observed objects onto *p* dimensions or aspects of the objects. The matrix **R** of order *p* by *p* is an asymmetric matrix describing the directional relationships among the components or dimensions. The first component may be more strongly related to the second component or dimension than the reverse. That this is a constrained version of the components model can be seen by writing **H=LR'**.

Applying the general tactic again, the matrix **R** can be decomposed into a symmetric part **C** and a skew-symmetric part **T**. As a result of the distributive properties of matrix multiplication the model can be decomposed into a symmetric part and a skew-symmetric part:

$$Q = L\,R\,L' = L\,C\,L' + L\,T\,L'.$$

It follows that the DEDICOM model is an additive scalar product model; the model is the sum of an oblique factor model for the similarity part and a skew-symmetric function of the factors for the dominance part.

In Nosofsky's terminology, the model can be viewed as an asymmetric similarity model with a scalar product similarity function as the symmetric component and an asymmetric similarity function for the skew-symmetric part without row and column bias.

There is a rotational indeterminacy in the model. This rotational indeterminacy may be used to rotate the matrix with loadings to obtain a simple diagram. The matrix **L** may be rotated it we pre- and postmultiply **R** by the inverse of the chosen rotation matrix. A convenient choice of a rotation matrix is the matrix with singular vectors of the SVD of the matrix **T**. If the matrix **L** is rotated by this rotation matrix the diagram of two dimensions can be interpreted as follows: the angle of two vectors multiplied with the obliqueness of the axis corresponds to the symmetric part of the matrix **Q**, and the area of the triangle corresponds to the skew-symmetric part of the model as in the SVD method of section 2.4.

If we choose this rotation matrix, the matrix **R** is symmetric except in its 2 by 2 diagonal blocks. The first dimension is asymmetrically related to the second dimension and not to any other dimension. Of course the SVD of the matrix **T** remains interesting in its own right because a plot of the singular vectors shows graphically how the dimensions are related (see section 2.4)

If we have a two-dimensional solution there is another useful rotation. This rotation method shows the close relationship with the SVD of section 2.4. First, we compute the eigen decomposition of the matrix $\mathbf{C} = \mathbf{K \Lambda K'}$. Second, assuming the inverse $\Lambda^{-1/2}$ exists, the matrix **L** is rotated with the scaled eigen vectors $\mathbf{K \Lambda}^{-1/2}$ of the symmetric matrix **C**. The matrix **T** is a 2 by 2 skew-symmetric matrix that can be written as $\tau\mathbf{J}$, where τ is a scalar and $\mathbf{J} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. The matrix **T** is rescaled by this rotation with values $\lambda_1\lambda_2$. Using this rotation the matrix **C** reduces to the identity matrix. When the objects are plotted using the coordinates of $\mathbf{LK\Lambda}^{-1/2}$, the area between the points and the origin corresponds to the skew-symmetric part, while the angle between points corresponds between to the symmetric part.

The two-dimensional DEDICOM model has been proposed under the name ASYMSCAL by Chino (1978). The **R** matrix in Chino's method has a very special structure:

$$\mathbf{R} = \alpha\mathbf{I} + \beta\mathbf{J},$$

where α,β are parameters to be estimated, **I** is the identity matrix order 2 by 2 and **J** is a 2 by 2 skew-symmetric matrix with one above the diagonal; the cell below the diagonal is filled with -1. After the rotation of the two-dimensional DEDICOM described previously, the **R** matrix of the DEDICOM model has similar structure with $\alpha=1$. The models are identical because we may rescale the matrix with loadings, if we adjust the **R** matrix by the inverse of the rescaling factor.

Algorithms for fitting the DEDICOM model are given by Kiers (1989) and Kiers et al.(1990). An algorithm for fitting the off-diagonal DEDICOM model is given by Ten Berge and Kiers (1989). This algorithm is nearly identical to Chino's (1978) algorithm, the difference between these two algorithms is that the **R** matrix is estimated in different ways.

4.2   The slide vector model

The model to be discussed in this section can be viewed as a constrained version of the unfolding model, which associates with each object i a row point x_i and a column point y_i. The unfolding model is a general model for rectangular data with two configurations **X** and **Y,** for rows and columns respectively. The model can also be applied to square data and as a consequence the model uses twice the number of parameters to account for the asymmetry in the data. In particular, the dissimilarities are modeled by:

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

where only the distances between points of different sets are compared; the distances within each of the two sets are only implicitly defined. The model predicts symmetry if the points x_i and y_i coincide. The unfolding model considers the symmetric and the asymmetric part of the data as inseparable; see Coombs (1964) or Heiser (1981) for a discussion of the unfolding model.

The *slide-vector model* (Kruskal 1973, cf De Leeuw and Heiser, 1982) is a relatively unknown model that at first sight seems to be an extended symmetric Euclidean model. The asymmetry is represented by adding a vector to the dimensionwise differences; as a consequence the quasi-distance becomes asymmetric. This vector corresponds to a *shift* or *translation* of the points in one direction. The squared quasi-distance, q_{ij}^2, is written as:

$$q_{ij}^2 = \Sigma_S \, (\, x_{is} - x_{js} + z_s)^2,$$

where z_s are the elements of the slide vector z. It is not hard to verify that the model is a special case of the unfolding model with $y_{js}=x_{js} - z_s$. The construction of an asymmetric distance is illustrated in Figure 2.
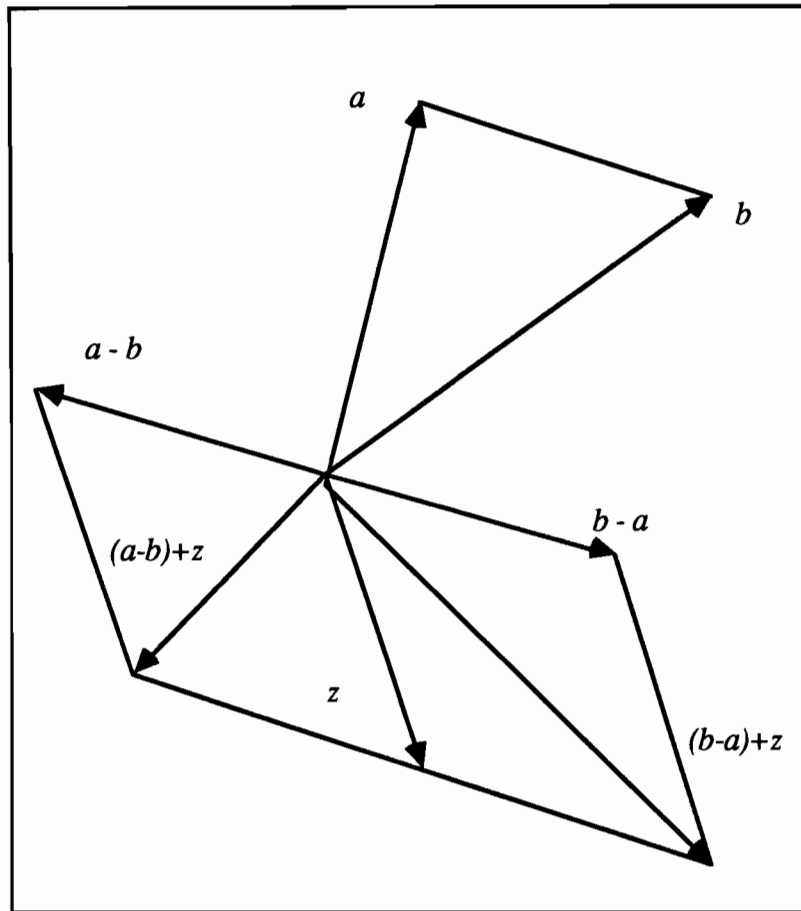

Figure 2: the slide-vector model

Two objects, a and b are depicted with their difference vectors and the distance is computed by first subtracting the vectors (*a-b*; *b-a*); these difference vectors are of the same length but with opposite sign. By adding the slide-vector, indicated by the vector z in the Figure, to these difference vectors, and then establishing their length, an asymmetric quasi-distance

28

d_{ba} and d_{ab} indicated by $(b-a)+z$ and $(a-b)+z$ is obtained. From Figure 1 it follows that objects located in a direction similar to the slide vector dominates the other objects. The sum vectors shows how the asymmetry is obtained in the quasi-distance.

If the square in the above equation for q_{ij} is expanded and if terms are rearranged, it becomes clear that this model distinguishes a symmetric and a skew-symmetric part:

$$q_{ij}^2 = \Sigma_s (x_{is} - x_{js})^2 + \Sigma_s z_s^2 + 2\Sigma_s z_s (x_{is} - x_{js}).$$

From this decomposition it follows that the model assumes points laying far apart to be more asymmetric than points laying close together on a dimension. The term

$$\Sigma_s z_s (x_{is} - x_{js})$$

is skew-symmetric, a property that is difficult to recognize in original form of the model. The skew-symmetric part of the model is *compensatory*, because the total difference may vanish if large differences on the first dimension are compensated for by differences with opposite sign on the other dimensions.

This skew-symmetric part of the model corresponds with the vector model for preference data proposed by Tucker (1960). It implies that points with the highest projections on this vector are preferred over the other objects. For this reason the slide vector could as well be called a *preference vector*. In the context of asymmetry it means that the objects with high projections are more often associated with the other objects than the reverse. Figure 3 tries to illuminate this aspect of the slide vector model.
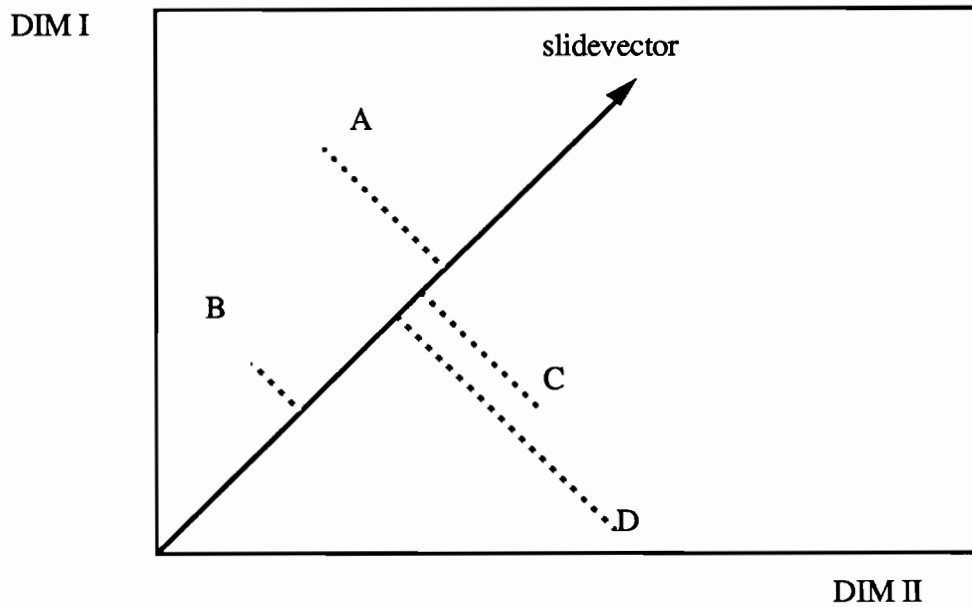
29

Figure 3: Joint representation of the symmetry and asymmetry

In Figure 3 four objects A, B, C, D are depicted as points in a two-dimensional space. The dashed lines in the Figure correspond to the projections of the objects on the slide vector. Objects with high projections dominate objects with low projections. In this example object A dominates the other objects. The distances among the points can be interpreted as the similarity or resemblance of the objects; object C is more similar to object D than to object A.

The slide vector has the desirable property that the asymmetry is related to the dimensions of the configuration. In Figure 3 the asymmetry is related to both dimensions of the scaling solution. The slide vector model is a special case of the similarity-bias model with the squared Euclidean distance and a constant as a similarity function and with row and column bias parameters that are a linear combination of the dimensions.

5. Discussion

Asymmetries in proximity data can be modeled in a different number of ways, and this paper has shown that a number of these methods are special cases of a general hybrid model. More in particular, these models were discussed in the similarity-bias framework provided by Nosofsky (1991) and it was shown that other models could be added to this framework (Weeks and Bentler, 1982; Okada, 1988 a,b; Saito, 1986; Levin and Brown, 1979; Cunningham (1978) and Young, 1984). The importance of asymmetric similarities was already recognized in psychometrics (Tversky, 1977; Harshman, 1982; Chino, 1978) although not always under that name. Asymmetric similarity models generally are multidimensional, whereas similarity-bias models are in general one-dimensional. Although we focused our attention on asymmetries, it was noted that the general hybrid model accommodates high centrality data as well. This can be an advantage over the existing symmetric MDS models.

Our point of view that asymmetric proximities can be regarded as a function of similarity and dominance is often only implicit in the model; the decomposition into symmetric and skew-symmetric components revealed that many different models assume dominance to arise from a linear function. This is a rather unfortunate state of affairs because the skew-symmetric part of the data may be as complex as the symmetric part. Although we used the term dominance as the generic name for skew-symmetry, in applications it can frequently be interpreted as a concept reflecting certain other aspects of the stimuli. Examples from identification confusion experiments are prototype, easily encoded stimulus, good stimulus and so on. Some models have been defined where dominance is explicitly related to theoretical concepts as is the case in the distance density model and the feature matching model.

The decomposition of an asymmetric matrix into a symmetric and a skew-symmetric matrix thus provides a good rationale for decomposing model parameters. The

similarities and differences between the methods in terms of the untransformed data are more easily seen from this point of view, and, in a number of cases the data can be more easily interpreted. Some models represent skew-symmetry in the data by symmetry and skew-symmetry parameters jointly, while a logarithmic transformation of the odds showed that the symmetric parameters in the skew-symmetric matrix vanished. Note that this is only a matter of transformations, and the choice of the specific decomposition. The decomposition justifies separate analysis of symmetry and skew-symmetry from a data analysis point of view. Theoretically it may be interesting to combine information from separate analysis afterwards. However, the linkage between symmetry and skew-symmetry may or may not exist, this is an empirical question.

We conclude our paper with the remark that Tversky's (1977) challenge of the standard MDS model has been very influential and it has helped viewing asymmetry as a systematic phenomenon.

References

Agresti, A. (1990). *Categorical data analysis*. Wiley, New York.

Arabie, P., & Soli, S. D. (1982). The interface between the types of regression and methods of collecting proximity data. In: Golledge, R. G. and Rayner, J. N. (Eds.) *Proximity and preference: problems in the multidimensional analysis of large data sets*. University of Minnesota Press, MN, pp 90 - 115.

Beals, R., Krantz, D. H., & Tversky, A. (1968). The foundations of multidimensional scaling. *Psychological Review, 75*, 127-128.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and Practice*. MIT press, Cambridge, MA.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs I. the method of paired comparisons. *Biometrika, 39*, 324-345.

Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association, 43*, 572-574.

Carroll, J. D. (1976). Spatial, non-spatial and hybrid models for scaling. *Psychometrika, 41*, 439-463.

Carroll, J. D. (1981). Models and methods for multidimensional analysis of preferential choice (or other dominance) data. In E.D. Lantermann & H. Feger (Eds), *Similarity and Choice*. Huber, Bern, 234-289.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. In: M.R. Rosenzweig and L. W. Porter (Eds), *Annual Review of Psychology, 31*, 607-649.

Carroll, J. D., & Pruzansky, S. (1981). Discrete and hybrid scaling models. In E.D. Lantermann & H. Feger (Eds), *Similarity and Choice*. Huber, Bern, 108-139.

Carroll, J. D., & Wish, M. (1974). Multidimensional perceptual models and measurement methods. In: E. C. Carterette and M. P. Friedman (Eds), *Handbook of perception*. Academic press, New York, 391-447.

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

Chino, N. (1978). A graphical technique for representing the asymmetric relationships between *N* objects. *Behaviormetrika, 5*, 23-40.

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika, 31*, 33-42.

Collins, L. M. (1987). Sociometry: deriving sociograms via asymmetric multidimensional scaling. In: F. W. Young and R. M. Hamer (Eds). *Multidimensional scaling: history, theory and applications*. Lawrence Erlbaum associates. Hillsdale, New Jersey, 179-196.

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

Constantine, A. G., & Gower, J. C. (1982). Models for the analysis of interregional migration. *Environment and Planning A, 14*, 477-497.

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 LTD, London.

Cunningham, J. P. (1978). Freetrees and bidirectional 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, 501-522.

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, 285-316

DeSarbo, W. S., Manrai, A. K., & Burke, R. R. (1990). A non spatial methodology for the analysis of two way proximity data incorporating the distance density hypothesis. *Psychometrika, 55*, 229-253.

De Soete, G., & Carroll, J. D. (1986). A maximum likelihood method for fitting the wandering vector model. *Psychometrika, 48*, 553-566.

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

Gower, J. C. (1977). The analysis of asymmetry and orthogonality. In: J. R. Barra, F. Brodeau, G. Romer, & B. van Cutsem (Eds.), *Recent developments in statistics*. North Holland, Amsterdam, 109-123.

Gower, J. C., & Digby, P. G. (1981). Expressing complex relationships in two dimensions. In: V. Barnett (Ed), *Interpreting multivariate data*. Wiley, New York, 83-118.

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

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. (1981). Unfolding analysis of proximity data. Doctoral thesis, Department of Data Theory, University of Leiden, The Netherlands.

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., & De Leeuw, J. (1981). Multidimensional mapping of preference data. *Mathematiques et Sciences humaines, 19*, 39-96.

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

Holman, E. W., & Marley, A. A. J. (1974). Stimulus and response measurement. In: E. C. Carterette and M. P. Friedman (Eds), *Handbook of perception*. Academic press, New York, 173-207.

Hubert, L. J. (1976). Seriation using asymmetric proximity measures. *British Journal of Mathematical and Statistical Psychology, 29*, 32-52.

Hubert, L. J., & Baker, F. B. (1979). Evaluating the symmetry of a proximity matrix. *Quality and Quantity, 43*, 81-91.

Keeren, G., & Baggen, S. (1981). Recognition models of alphanumeric characters. *Perception and psychophysics, 29*, 234-246.

Kiers, H. A.L. (1989). An alternating least squares algorithm for fitting the two and three way DEDICOM model and the IDIOSCAL model. *Psychometrika, 54*, 515-521.

Kiers, H. A. L., Ten Berge, J. M. F., Takane, Y., & De Leeuw, J. (1989). A generalization of Takane's algorithm for DEDICOM. *Psychometrika, 55*, 151-159.

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review, 84*, 445-463.

Krumhansl, C. L. (1982). Density versus feature weights as predictors of visual identifications. Comments on Appelman and Mayzner. *Journal of experimental psychology: general, 111*, 101-108.

Krumhansl, C. L. (1988). Testing the density hypothesis: comment on Corter. *Journal of experimental psychology: general, 117*, 101-104.

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., & Gutman, 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, 103-189.

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.

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, *23*, 94-140.

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

Okada, A. (1988 b). An analysis of intergenerational occupational mobility by asymmetric multidimensional scaling. *Procedings of the SMABS 88 conference*. University of Groningen, 1-15.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika*, *42*, 319.

Saito, T. (1986). Multidimensional scaling to explore complex aspects in dissimilarity judgment. *Behaviormetrika*, *20*, 35-62.

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

Ten Berge, J. M. F., & Kiers, H. A. L. (1989). Fitting the off-diagonal DEDICOM model in the least-squares sense by a generalization of the Harman and Jones minres procedure of factor analysis. *Psychometrika, 54*, 333-337.

Takane, Y. (1980). Maximum likelihood estimation in the generalized case of Thurstone's model of comparitive judgment. *Japanese Psychological Research, 22*, 188-196.

Takane, Y., & Shibayama, T. (1986). Comparison of models for stimulus recognition data. In: J. De Leeuw e.a. (Eds), *Multidimensional data analysis*, DSWO press, Leiden, 119 - 134.

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

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

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

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

Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch & S. B. B. Lloyd (Eds), *Cognition and categorization.* Lawrence Erlbaum associates. Hillsdale, New Jersey, 79-98.

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

Wish, M. (1967). A model for the perception of morse code like signals. *Human factors, 9*, 529-539.

Wish, M., & Carroll, J. D. (1974). Applications of Individual differences scaling to studies of human perception and judgment. In: E. C. Carterette and M. P. Friedman (Eds), *Handbook of perception.* Academic press, New York, 449-491.

Wish, M., & Carroll, J. D. (1982). Multdimensional scaling and its applications. In P. R. Krisnaiah and L. N. Kanal (Eds.), *Handbook of statistics,* Vol 2. North Holland, Amsterdam, 317 - 345.

Young, F. W. (1975). An asymmetric Euclidean model for multiprocess asymmetric data. Paper presented at the US-Japan seminar on multidimensional scaling, University of California, San Diego, La Jolla.

Young, F. W. (1984). The general Euclidean model. In: H. G. Law et al. (Eds.), *Research methods for multimode data analysis.* Praeger, New York, 440-470.

Young, F. W. (1987). *Multidimensional scaling.* Lawrence Erlbaum associates. Hillsdale, New Jersey.