

# Triadic distance models for the analysis of asymmetric three-way proximity data

Mark de Rooij\* and Willem J. Heiser

*Department of Psychology, Leiden University, The Netherlands*

Triadic distance models can be used to analyse proximity data defined on triples of objects. Three-way symmetry is a common assumption for triadic distance models. In the present study three-way symmetry is not assumed. Triadic distance models are presented for the analysis of asymmetric three-way proximity data that result in a simultaneous representation of symmetry and asymmetry in a low-dimensional configuration. An iterative majorization algorithm is developed for obtaining the coordinates and the representation of the asymmetry. The models are illustrated by an example using longitudinal categorical data.

## 1. Introduction

Recently there has been growing interest in generalizations of distance models for pairs of points to the case of triples of points. Important contributions have been made by Joly and Le Calvé (1995) and Heiser and Bennani (1997), who both gave an axiomatic framework for the study of triadic distances. Triadic distance models are suitable for analysing three-way one-mode dissimilarity data, where each element of a triple is treated equally. This notion is different from three-way two-mode data, where the index of the third dimension of the table refers to another set of objects, as in the INDSCAL model (Carroll & Chang, 1970). Generalized distance models try to represent triadic dissimilarities with triadic distances defined between three points. An important axiom in both papers mentioned above is that of three-way symmetry for the dissimilarities and the distances. As in two-way multidimensional scaling, this assumption will not always hold for observed dissimilarities. In the present paper, we will therefore propose some triadic distance models for the analysis of three-way dissimilarities that do not have the property of three-way symmetry.

A two-way dissimilarity ( $\delta_{ij}$ ) is defined as a lack of resemblance between objects taken two at a time ( $i, j = 1, \dots, K$ ): a large numerical value implies the two objects are dissimilar. Analogously, a three-way dissimilarity ( $\delta_{ijk}$ ) is defined as a lack of resemblance between objects taken three at a time ( $i, j, k = 1, \dots, K$ ). In general, we can additively decompose the dissimilarity into a symmetric part and a skew-symmetric part, or in matrix terms

$$\Delta = \mathbf{M} + \mathbf{N}. \quad (1)$$

In two-way data the symmetric matrix  $\mathbf{M}$  has elements  $m_{ij} = (\delta_{ij} + \delta_{ji})/2$ ; the skew-symmetric

\* Requests for reprints should be addressed to Mark de Rooij, Department of Psychology, University of Leiden, PO Box 9555, 2300 RB Leiden, The Netherlands (e-mail: rooijm@fsw.leidenuniv.nl).

matrix  $\mathbf{N}$  has elements  $n_{ij} = (\delta_{ij} - \delta_{ji})/2$ , where  $n_{ij} = -n_{ji}$ . Decomposition (1) has an interesting mathematical property: The sum of squares of the dissimilarities may be partitioned into two uncorrelated parts, one due to symmetry and another one due to skew-symmetry. It follows that, within the context of least squares, symmetry and skew-symmetry may be analysed separately without losing or doubling information (Gower & Zielman, 1998; Zielman & Heiser, 1996). The symmetric matrix, for example, could be analysed by some multidimensional scaling method (cf. Borg & Groenen, 1997), resulting in a low-dimensional representation of the dissimilarities through distances. The skew-symmetric matrix could be analyzed by, for example, a singular value decomposition resulting in a Gower diagram (Gower, 1977). An object  $i$  is said to dominate object  $j$  if it is observed that  $\delta_{ij} > \delta_{ji}$ .

In three-way data, the symmetric three-way table  $\mathbf{M}$  has elements  $m_{ijk} = (\delta_{ijk} + \delta_{ikj} + \delta_{jik} + \delta_{jki} + \delta_{kij} + \delta_{kji})/6$ . The residual,  $n_{ijk} = \delta_{ijk} - m_{ijk}$ , is the generalization of the skew-symmetric part. The sum over all permutations of the indices of the skew-symmetric part equals zero, that is,  $n_{ijk} + n_{ikj} + n_{jik} + n_{jki} + n_{kij} + n_{kji} = 0$ . In general it is not true that  $n_{ijk} = -n_{kji}$ , or that another specific relation exists among these six elements (except the sum to zero). The symmetric table  $\mathbf{M}$  can be analysed by a triadic distance model. In general, a PARAFAC decomposition (Harshman & Lundy, 1984) of a three-way skew-symmetric table  $\mathbf{N}$  does not have such nice properties as the singular value decomposition of the two-way skew-symmetric matrix. There are cases, however, in which the PARAFAC decomposition has a special form. If we observe that, for all  $i, j$  and  $k$ , there is an ordering of the six dissimilarities, that is,  $\delta_{ijk} \geq (\delta_{ikj}, \delta_{jik}) \geq (\delta_{jki}, \delta_{kij}) \geq \delta_{kji}$ , then the PARAFAC solution has some special structure. We will come back to this special structure in a later section. When we observe such an ordering for  $i, j$  and  $k$ , we will refer to that as dominance of  $i$  over  $j$  over  $k$ .

The models we propose in the present paper are for the analysis of the dissimilarities (symmetry plus asymmetry), and result in a simultaneous representation of the symmetry and asymmetry in the data. We first present a model for unrestricted asymmetry and then show how we can model special structures of three-way skew-symmetry through constraints on the latter model. The model of unrestricted asymmetry is in fact a three-way three-mode multidimensional unfolding model. In three-way unfolding we have triadic proximities between three sets of objects. Proximity measures for the objects within one set are lacking. A least squares algorithm is developed for finding a solution to the three-way three-mode unfolding problem. This problem could also be handled by a three-way one-mode algorithm on an extended matrix where every way consists of the union of the three sets, and intraset proximities are missing (Heiser & Bennani, 1997). This solution is not very efficient, since we then work with very large data matrices, and do not profit from knowing which entries are missing. On the other hand, our algorithm does make use of this knowledge. The restricted models presented here can be viewed as generalizations of the slide-vector model of Zielman and Heiser (1993).

An important source of data for the models presented are longitudinal studies with categorical responses. Having such data, we can form a contingency table with every way corresponding to a time point. The cells denote frequencies of transitions between categories. The number of transitions between categories is assumed to be a monotonically decreasing function of the dissimilarity between these categories. It is very unlikely in practice that the number of transitions from  $a$  to  $b$  to  $c$  is equal to the number of transitions between these categories in any other order. So it is very likely that we observe some asymmetry. When

there is an overall tendency of the individuals' transitions, we may call that a trend. A trend often results in dominance data as defined above.

First, we discuss triadic distances, and then briefly review the slide-vector model of Zielman and Heiser (1993). In Section 3 our asymmetric models are defined, together with the geometry of the model and an algorithm. An example is given in Section 4, and we conclude with a discussion.

## 2. Some background

In this section we give a short introduction to triadic distances and the slide-vector model, though algorithms for obtaining solutions are not discussed. For both problems a least squares loss function has been defined, which can be minimized by iterative majorization. For a general discussion of iterative majorization, the reader is referred to De Leeuw and Heiser (1980), Groenen (1993) or Heiser (1995). For majorization in a three-way context, see Heiser and Bennani (1997), and for the slide-vector model, see Zielman and Heiser (1993) or Heiser (1987).

### 2.1. Triadic distances

In this paper we focus on triadic distances that are a function of dyadic distances, especially on the  $L_p$  transform

$$d_{ijk} = \left[ d_{ij}^p + d_{jk}^p + d_{ik}^p \right]^{1/p}, \quad (2)$$

where  $i, j, k = 1, \dots, K$  denote the categories, and  $d_{ij}$  is the distance between categories  $i$  and  $j$ , satisfying the three metric axioms: symmetry ( $d_{ij} = d_{ji}$ , for all  $i$  and  $j$ ), minimality ( $d_{ij} \geq d_{ii} = 0$ , for all  $i$  and  $j$ ), and the triangle inequality ( $d_{ik} + d_{jk} \geq d_{ij}$ , for all  $i, j$ , and  $k$ ). The two cases that it is important to distinguish are the perimeter model ( $p = 1$ ) and the generalized Euclidean model ( $p = 2$ ).

We think the perimeter model should be used in conjunction with the city-block distance, while the generalized Euclidean model should be used in conjunction with the Euclidean distance. The reasoning is as follows. Using the perimeter model and the city-block distance, the three-way distance  $d_{ijk}$  is proportional to the sum of ranges over the dimensions. The range is defined as the maximum minus the minimum over  $i, j$  and  $k$  of the projections of these points on the dimensional axis.

Using the generalized Euclidean model in conjunction with the Euclidean distance, it can be shown that the triadic distance is proportional to the square root of the inertia of the three points considered, defined as the sum of squared distances to their centre of gravity (Joly & Le Calvé, 1995). In one dimension, the triadic distance is then proportional to the standard deviation of the three points. As in dyadic distances, both conjunctions have the main property that the defined distance is proportional to some measure of dispersion. This is not (as yet) established for other combinations.

In the present study, we focus on the combination of the generalized Euclidean model with the Euclidean distance metric. This choice is motivated by the fact that most common models for two-way asymmetric multidimensional scaling are based on the Euclidean distance (Zielman & Heiser, 1996). To gain some insight into this definition of triadic distance, we give an example in Fig. 1. A triadic distance according to the generalized Euclidean model is

equal to the square root of the sum of squared dyadic distances. Compare the three triadic distances  $d_{abc}$ ,  $d_{abd}$  and  $d_{ace}$ . Since the dyadic distances in the triple  $a, b, c$  are all smaller than or equal to the dyadic distances in the triple  $a, b, d$  the first triadic distance is smaller. The comparison of  $d_{abc}$  with  $d_{ace}$  is more difficult;  $d_{abc}$  is equal to  $\sqrt{17 + 8 + 5} = 5.48$ ,  $d_{ace}$  is equal to  $\sqrt{8 + 8 + 32} = 6.93$ , so the triadic distance of the triple  $a, b, c$  is the smallest, then the distance between  $a, c$  and  $e$ , and the largest distance is between  $a, b$  and  $d$ .

## 2.2. The slide-vector model

The slide-vector model represents symmetry in an observed dissimilarity matrix by distances in Euclidean space; the asymmetry is represented by a uniform shift in one direction imposed on this symmetric distance. (See Zielman & Heiser, 1993, for more details.) The model can be written in squared form as

$$d_{ij}^2(\mathbf{X}; \mathbf{u}) = \sum_m (x_{im} - x_{jm} + u_m)^2, \quad (3)$$

where  $i, j = 1, \dots, K$  denote the categories, the matrix  $\mathbf{X}$  contains the coordinates  $\{x_{im}\}$ , and  $m$  is the dimensionality index ( $m = 1, \dots, M$ ). When the slide-vector  $\mathbf{u}$  with elements  $\{u_m\}$  is equal to zero, the slide-vector distance simplifies to the Euclidean distance. If we define the matrix  $\mathbf{Y}$  with elements  $y_{jm} = x_{jm} - u_m$  it is clear that the slide-vector model is a special case of multidimensional unfolding, that is,  $d_{ij}^2(\mathbf{X}; \mathbf{u})$  may be written as

$$d_{ij}^2(\mathbf{X}; \mathbf{Y}) = \sum_m (x_{im} - y_{jm})^2. \quad (4)$$

In unfolding we find a configuration for both the rows ( $\mathbf{X}$ ) and the columns ( $\mathbf{Y}$ ) which are unrelated. In the case of the slide-vector the latter is a translation of the former.

An example of a configuration with a slide-vector is given in Fig. 2. Only the coordinates for the first way are shown; the coordinates for the second way can be obtained by subtracting the slide-vector from the coordinates of the first way. This is done in Fig. 3, where subscripts denote the corresponding way. Note that distances between coordinates with the same subscript are not related to data points. These distances have a meaning, however: they refer to the mean dissimilarity between two categories. If we carried out a multidimensional scaling on the symmetric dissimilarities, as discussed in the Introduction, this would be the solution. We can compare, for example, the distance  $d_{a_1, c_2}$  with  $d_{c_1, a_2}$  and see that the latter is a smaller distance, corresponding to a smaller dissimilarity than the first. In terms of a longitudinal study this effect would mean that there are more transitions from category  $c$  to category  $a$  than conversely. Category  $a$  is getting more popular, category  $c$  less. Note that  $a$  projects high onto the slide-vector; categories that project high onto the slide-vector dominate categories that project low onto the slide-vector.

## 3. Triadic distance models and several forms of asymmetry

Let us return to the generalized Euclidean model in squared form, where now each dyadic distance is a Euclidean distance denoted by  $d_{ij}(\mathbf{X})$ :

$$d_{ijk}^2(\mathbf{X}) = d_{ij}^2(\mathbf{X}) + d_{jk}^2(\mathbf{X}) + d_{ik}^2(\mathbf{X}). \quad (5)$$

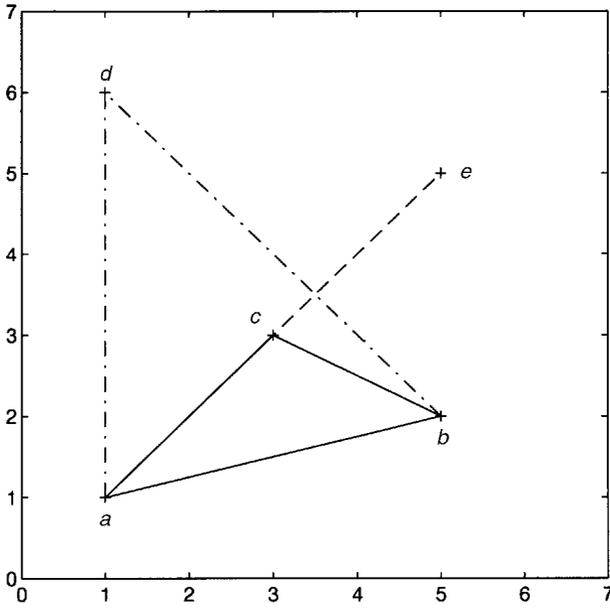


Figure 1. A configuration with three examples of triadic distances.

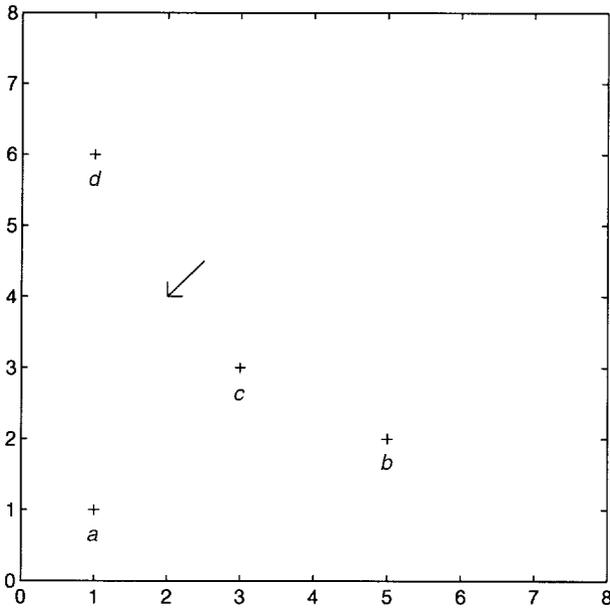
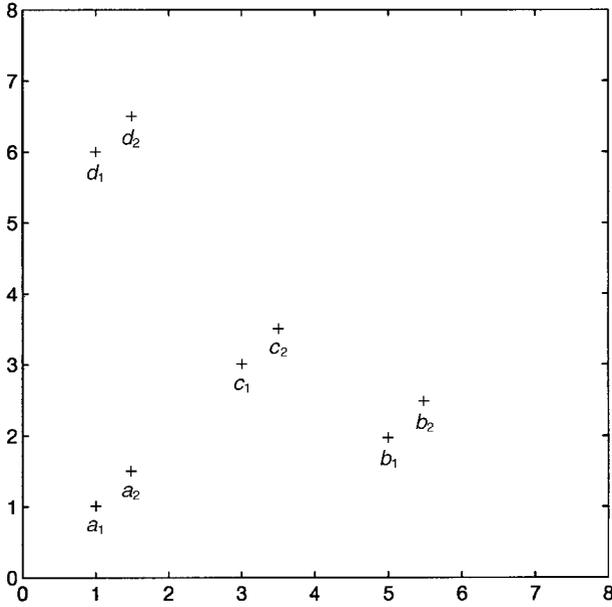
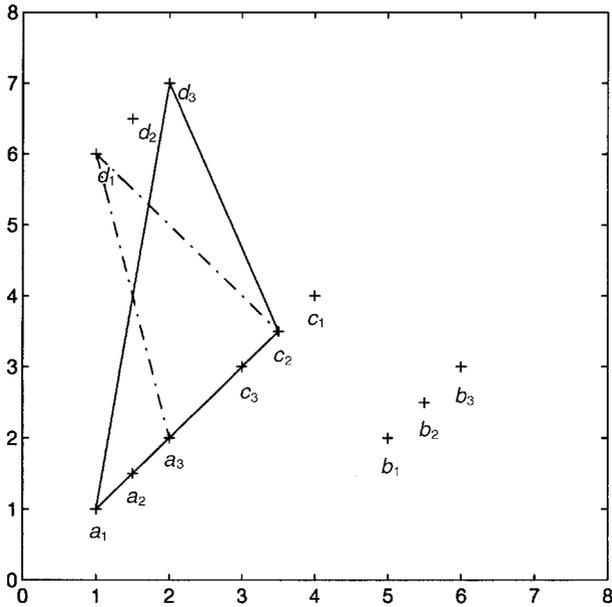


Figure 2. A configuration with a slide-vector.



**Figure 3.** The same configuration as in Figure 2 but now with the categories for both ways.



**Figure 4.** The same configuration as in Fig. 2 but now with the categories for all three ways.

Let  $i$  correspond to the first way,  $j$  to the second way, and  $k$  to the third way of the three-way table.

We assume there is asymmetry between the first and second way, and asymmetry between the second and third way. These two asymmetries do not need to be the same. In any case, the asymmetry between the first and third way is equal to the sum of the two asymmetries. Each asymmetry is modelled by a shift on the Euclidean distance between the two ways. The asymmetry between the first and second way is modelled by a shift  $\mathbf{u}$ , the asymmetry between the second and third way by a shift  $\mathbf{v}$ , and the asymmetry between the first and third way by the shift  $\mathbf{u} + \mathbf{v}$ . If we now substitute the distance as defined by (3), for every dyadic distance in (5) we obtain

$$d_{ijk}^2(\mathbf{X}; \mathbf{u}; \mathbf{v}) = d_{ij}^2(\mathbf{X}; \mathbf{u}) + d_{jk}^2(\mathbf{X}; \mathbf{v}) + d_{ik}^2(\mathbf{X}; \mathbf{u} + \mathbf{v}). \quad (6)$$

Of course, we could also consider a constant shift, in which case  $\mathbf{u} = \mathbf{v}$ , and the shift between  $i$  and  $k$  is twice this shift. Now defining  $y_{jm} = x_{jm} - u_m$  and  $z_{jm} = y_{jm} - v_m$ , so that  $z_{jm} = x_{jm} - u_m - v_m$ , we can formulate a squared three-way unfolding distance

$$d_{ijk}^2(\mathbf{X}; \mathbf{Y}; \mathbf{Z}) = d_{ij}^2(\mathbf{X}; \mathbf{Y}) + d_{jk}^2(\mathbf{Y}; \mathbf{Z}) + d_{ik}^2(\mathbf{X}; \mathbf{Z}). \quad (7)$$

Model (7) without the constraints of the slide-vector will be called the unrestricted asymmetric model. Solutions for the slide-vector model with different shifts (6) can be found by constraining the solution of the unrestricted asymmetric model; this model will be called the slide-2 model. For equal shifts ( $\mathbf{u} = \mathbf{v}$ ), further constraints need to be applied; the latter model will be called the slide-1 model. The number of parameters for the unrestricted asymmetry model is  $3KM$ ; this number is substantially reduced for the slide-2 model ( $(K + 2)M$ ), and the slide-1 model ( $(K + 1)M$ ).

### 3.1. Geometry

For the unrestricted asymmetric model we find separate coordinates for each way. We can now just look at triples of points, and infer from their triadic distances the dissimilarities. It is important to note that distances between categories of one way are not related to observations.

The geometry of the slide-1 and slide-2 model needs a more detailed discussion. We will only discuss the slide-1 model. The interpretation of the slide-2 is analogous. The slide-1 model gives a simultaneous representation of symmetry and asymmetry in a low-dimensional configuration. The symmetry is modelled by the distances between the points, and represents the mean dissimilarity between the corresponding categories. The asymmetry is represented by a vector, called the slide-vector, that is linked to the dimensions in the same configuration. The slide-vector gives the direction of dominance. In other words, categories that project high onto the slide-vector are dominant. In terms of longitudinal data, categories that project high onto the slide-vector are getting more popular, categories that project low onto the slide-vector are losing popularity. We will illustrate this with two examples.

Let us look at Fig 2 again, and assume it is a solution obtained with the slide-1 model to a three-way problem. The points shown are the categories for the first way. The coordinates for the categories of the second way can be found by subtraction of the slide-vector from the points of the first way. For the third way, the slide-vector is again subtracted from the points of the second way. This reparameterization is shown in Fig. 4, in which we also compare two triadic distances  $a_1, c_2, d_3$  and  $d_1, c_2, a_3$  (subscripts denote ways). The former triadic distance

is larger, implying a larger dissimilarity. In terms of a longitudinal study, more transitions are made from  $d$  to  $c$  to  $a$  than the other way around. Looking again at Fig. 2, we see that  $a$  projects highest onto the slide-vector. As we expected,  $a$  is getting more popular compared to  $d$  which projects low onto the slide-vector.

Let us now take a one-dimensional numerical example. Suppose we have three points with coordinates  $a = 1, b = 2$  and  $c = 3$ . The slide-vector is equal to  $u = 2$ . In Table 1 the corresponding squared triadic distances according to the slide-1 model are given. Note that  $c$  projects highest onto the slide-vector, then  $b$  and last  $a$ . We should thus expect that the distance from  $c$  to others is larger than the distance from others to  $c$ . To show this is actually true, compare for example  $d_{abc}^2 = 6$  (squared distance from  $a$  to  $b$  to  $c$ , defined by (6)) and  $d_{cba}^2 = 54$ . Note the dominance as discussed in the Introduction of  $d_{cba} \geq (d_{cab}, d_{bca}) \geq (d_{bac}, d_{acb}) \geq d_{abc}$ . If the points are not equidistant, as they are in this example, a complete ordering can be formulated. Now look at the triadic distances of form  $d_{iji}$ , and note the following equalities  $d_{iji} = d_{ijj}, d_{iji} = d_{jji}$ . In terms of longitudinal data, these equalities imply that according to the slide-1 model, it makes no difference at what time point you make a change, or in which order changes are made. Of course, for the slide-2 model this would make a difference. To finish this example, note that the distances on the main diagonal are not equal to zero.

### 3.2. Algorithm

We first develop the least squares algorithm for the unrestricted asymmetry model. Later we will show how to incorporate the constraints for the slide-1 and slide-2 model. The coordinates for the unrestricted asymmetry model can be found by minimizing the loss function.

$$\sigma^2(\mathbf{X}; \mathbf{Y}; \mathbf{Z}) = \sum_i \sum_j \sum_k w_{ijk} (\delta_{ijk} - d_{ijk}(\mathbf{X}; \mathbf{Y}; \mathbf{Z}))^2, \tag{8}$$

**Table 1.** Expected squared distance matrix corresponding to a hypothetical example of a one-dimensional solution with three points  $a = 1, b = 2$  and  $c = 3$  and a slide-vector  $u = 2$

		Way 3		
Way 1	Way 2	$a$	$b$	$c$
$a$	$a$	24	14	8
	$b$	26	14	6
	$c$	32	18	8
$b$	$a$	38	26	18
	$b$	38	24	14
	$c$	42	26	14
$c$	$a$	56	42	32
	$b$	54	38	26
	$c$	56	38	24

in which the  $w_{ijk}$  are predefined weights, which can be used, for example, to code missing data. The  $\delta_{ijk}$  are the given three-way dissimilarities, and the  $d_{ijk}(\mathbf{X}; \mathbf{Y}; \mathbf{Z})$  are the three-way distances. Define the three-way matrices  $\mathbf{W}$  with elements  $w_{ijk}$ , and  $\mathbf{A}$  with elements  $a_{ijk} = w_{ijk}\delta_{ijk}/d_{ijk}(\mathbf{X}; \mathbf{Y}; \mathbf{Z})$  if  $d_{ijk}(\mathbf{X}; \mathbf{Y}; \mathbf{Z}) \neq 0$ , and otherwise 0. Using the definition of the generalized unfolding distance (7), we obtain

$$\begin{aligned} \sigma^2(\mathbf{X}; \mathbf{Y}; \mathbf{Z}) &= SSQ_\delta \\ &+ \sum_i \sum_j \sum_k w_{ijk} \{d_{ij}^2(\mathbf{X}; \mathbf{Y}) + d_{ik}^2(\mathbf{X}; \mathbf{Z}) + d_{jk}^2(\mathbf{Y}; \mathbf{Z})\} \\ &- 2 \sum_i \sum_j \sum_k a_{ijk} \{d_{ij}^2(\mathbf{X}; \mathbf{Y}) + d_{ik}^2(\mathbf{X}; \mathbf{Z}) + d_{jk}^2(\mathbf{Y}; \mathbf{Z})\}, \end{aligned} \quad (9)$$

where  $SSQ_\delta$  means the weighted sum of squares of the dissimilarities. Rewriting (9),

$$\begin{aligned} \sigma^2(\mathbf{X}; \mathbf{Y}; \mathbf{Z}) &= SSQ_\delta \\ &+ \sum_i \sum_j w_{ij+} d_{ij}^2(\mathbf{X}; \mathbf{Y}) + \sum_i \sum_k w_{i+k} d_{ik}^2(\mathbf{X}; \mathbf{Z}) + \sum_j \sum_k w_{+jk} d_{jk}^2(\mathbf{Y}; \mathbf{Z}) \\ &- 2 \sum_i \sum_j a_{ij+} d_{ij}^2(\mathbf{X}; \mathbf{Y}) - 2 \sum_i \sum_k a_{i+k} d_{ik}^2(\mathbf{X}; \mathbf{Z}) - 2 \sum_j \sum_k a_{+jk} d_{jk}^2(\mathbf{Y}; \mathbf{Z}), \end{aligned} \quad (10)$$

where the  $+$  in the indices means summing over the corresponding index. We see through this formulation that the minimization of the loss function is only dependent on dyadic distances. Moreover, we see that the minimization basically consists of three smaller minimization problems, which are two-way unfolding problems.

Let us take a more detailed look at one of these, the minimization over  $d_{ij}^2(\mathbf{X}; \mathbf{Y})$ . First define the  $K \times K$  diagonal matrices

$$\begin{aligned} \mathbf{A}_i &= \text{diag}(a_{i++}), \\ \mathbf{W}_i &= \text{diag}(w_{i++}), \end{aligned} \quad (11)$$

and similarly  $\mathbf{A}_j$ ,  $\mathbf{A}_k$ ,  $\mathbf{W}_j$  and  $\mathbf{W}_k$ . Further define the  $K \times K$  matrices

$$\begin{aligned} \overline{\mathbf{A}}_{ij} &= \{a_{ij+}\}, \\ \overline{\mathbf{W}}_{ij} &= \{w_{ij+}\}, \end{aligned} \quad (12)$$

and analogously  $\mathbf{A}_{ik}$ ,  $\mathbf{A}_{jk}$ ,  $\mathbf{W}_{ik}$  and  $\mathbf{W}_{jk}$ .

For the minimization over  $d_{ij}^2(\mathbf{X}; \mathbf{Y})$ , define a  $2K \times 2K$  matrix  $\mathbf{F}$  with the matrix  $-\mathbf{W}_{ij}$  in the upper right-hand corner and its transpose in the lower left-hand corner. In the upper left-hand corner place the matrix  $\mathbf{W}_i$ , and in the lower right-hand corner place the matrix  $\mathbf{W}_j$ . Also define a  $2K \times 2K$  matrix  $\mathbf{D}$  with the matrix  $-\mathbf{A}_{ij}$  in the upper right-hand corner and its transpose in the lower left-hand corner. In the upper left-hand corner place the matrix  $\mathbf{A}_i$ , and in the lower right-hand corner place the matrix  $\mathbf{A}_j$ . Finally, define the matrix  $\mathbf{S}^0$  with initial estimates for  $\mathbf{X}$  and  $\mathbf{Y}$  on top of each other. According to the SMACOF unfolding theory (Heiser, 1981, 1987), an update for  $\mathbf{S}$  is found by

$$\mathbf{S}^+ = \mathbf{F}^{-1} \mathbf{D} \mathbf{S}^0, \quad (13)$$

where  $\mathbf{F}^-$  is a generalized inverse of  $\mathbf{F}$ . Iterate until convergence by taking the updates as initial estimates. The matrices  $\mathbf{F}$  and  $\mathbf{D}$  are quite large, and therefore difficult to handle. Since  $\mathbf{F}$  and  $\mathbf{D}$  have the special structure as defined above, we can rewrite (13) as

$$\begin{aligned}\mathbf{W}_i\mathbf{X}^+ - \mathbf{W}_{ij}\mathbf{Y}^+ &= \mathbf{A}_i\mathbf{X}^0 - \mathbf{A}_{ij}\mathbf{Y}^0, \\ \mathbf{W}_j\mathbf{Y}^+ - \mathbf{W}'_{ij}\mathbf{X}^+ &= \mathbf{A}_j\mathbf{Y}^0 - \mathbf{A}'_{ij}\mathbf{X}^0.\end{aligned}\quad (14)$$

This system of equations can easily be solved by elimination (for details, see Heiser, 1981, 1987).

For the three-way problem, we would like to solve the three minimization problems at once. This can again be done by defining extended matrices  $\mathbf{F}$  and  $\mathbf{D}$ , now of order  $3K \times 3K$ . The matrix  $\mathbf{F}$  has the form

$$\mathbf{F} = \begin{pmatrix} 2\mathbf{W}_i & -\mathbf{W}_{ij} & -\mathbf{W}_{ik} \\ -\mathbf{W}'_{ij} & 2\mathbf{W}_j & -\mathbf{W}_{jk} \\ -\mathbf{W}'_{ik} & -\mathbf{W}'_{jk} & 2\mathbf{W}_k \end{pmatrix}.\quad (15)$$

The matrix  $\mathbf{D}$  has the same form but now defined on the matrices  $\mathbf{A}$  instead of  $\mathbf{W}$ . The matrix  $\mathbf{S}$  is a matrix with  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$  on top of each other. An update is given by (13). Again, the matrices  $\mathbf{F}$  and  $\mathbf{D}$  are very large and difficult to handle. Therefore we rewrite the problem in two steps, making use of the special structure of  $\mathbf{F}$  and  $\mathbf{D}$ .

For the first step, auxiliary matrices  $\mathbf{X}^*$ ,  $\mathbf{Y}^*$  and  $\mathbf{Z}^*$  are computed from initial estimates  $\mathbf{X}^0$ ,  $\mathbf{Y}^0$  and  $\mathbf{Z}^0$  as follows:

$$\begin{aligned}\mathbf{X}^* &= 2\mathbf{A}_i\mathbf{X}^0 - \mathbf{A}_{ij}\mathbf{Y}^0 - \mathbf{A}_{ik}\mathbf{Z}^0, \\ \mathbf{Y}^* &= 2\mathbf{A}_j\mathbf{Y}^0 - \mathbf{A}'_{ij}\mathbf{X}^0 - \mathbf{A}_{jk}\mathbf{Z}^0, \\ \mathbf{Z}^* &= 2\mathbf{A}_k\mathbf{Z}^0 - \mathbf{A}'_{ik}\mathbf{X}^0 - \mathbf{A}'_{jk}\mathbf{Y}^0.\end{aligned}\quad (16)$$

In the second step, the updates  $\mathbf{X}^+$ ,  $\mathbf{Y}^+$  and  $\mathbf{Z}^+$  can be found by solving the system of equations:

$$\begin{aligned}2\mathbf{W}_i\mathbf{X}^+ - \mathbf{W}_{ij}\mathbf{Y}^+ - \mathbf{W}_{ik}\mathbf{Z}^+ &= \mathbf{X}^*, \\ 2\mathbf{W}_j\mathbf{Y}^+ - \mathbf{W}'_{ij}\mathbf{X}^+ - \mathbf{W}_{jk}\mathbf{Z}^+ &= \mathbf{Y}^*, \\ 2\mathbf{W}_k\mathbf{Z}^+ - \mathbf{W}'_{ik}\mathbf{X}^+ - \mathbf{W}'_{jk}\mathbf{Y}^+ &= \mathbf{Z}^*.\end{aligned}\quad (17)$$

Solutions are found through the usual elimination method, resulting in the following equations:

$$\begin{aligned}\mathbf{Z}^+ &= \{\mathbf{B}_k - (\mathbf{W}'_{jk} + \mathbf{W}'_{ik}(2\mathbf{W}_i)^{-1}\mathbf{W}_{ij})(\mathbf{B}_j)^{-1}(\mathbf{W}_{jk} + \mathbf{W}'_{ij}(2\mathbf{W}_i)^{-1}\mathbf{W}_{ik})\}^{-1} \\ &\quad \times \{(\mathbf{Z}^* + \mathbf{W}'_{ik}(2\mathbf{W}_i)^{-1}\mathbf{X}^*) + [(\mathbf{W}'_{jk} + \mathbf{W}'_{ik}(2\mathbf{W}_i)^{-1}\mathbf{W}_{ij})(\mathbf{B}_j)^{-1}(\mathbf{Y}^* + \mathbf{W}'_{ij}(2\mathbf{W}_i)^{-1}\mathbf{X}^*)]\}, \\ \mathbf{Y}^+ &= \mathbf{B}_j^{-1}\{(\mathbf{Y}^* + \mathbf{W}'_{ij}(2\mathbf{W}_i)^{-1}\mathbf{X}^*) + (\mathbf{W}_{jk} + \mathbf{W}'_{ij}(2\mathbf{W}_i)^{-1}\mathbf{W}_{ik})\mathbf{Z}^+\}, \\ \mathbf{X}^+ &= (2\mathbf{W}_i)^{-1}\{\mathbf{X}^* + \mathbf{W}_{ij}\mathbf{Y}^+ + \mathbf{W}_{ik}\mathbf{Z}^+\},\end{aligned}\quad (18)$$

where  $\mathbf{B}_l = 2\mathbf{W}_l - \mathbf{W}'_{il}(2\mathbf{W}_i)^{-1}\mathbf{W}_{il}$ . Use the updates again as initial estimates and iterate until convergence. Convergence is proved in general by Heiser and Bennani (1997). Our algorithm is a special case of theirs, with the difference that we make use of the knowledge of missing data in the extended matrix.

To incorporate the restrictions of the slide-1 and slide-2 model, in every iteration we have to solve a metric projection problem. De Leeuw and Heiser (1980) show that it is very easy to incorporate constraints in the SMACOF algorithm. Zielman and Heiser (1993) did this for the two-way slide-vector model. First, we have to define matrices  $\mathbf{E}$  for the constraints of the slide-vector models. For the slide-1 model,  $\mathbf{E}$  is defined as

$$\mathbf{E}_1 = \begin{pmatrix} \mathbf{I} & \mathbf{1} \\ \mathbf{I} & \mathbf{0} \\ \mathbf{I} & -\mathbf{1} \end{pmatrix}. \quad (19)$$

For the slide-2 model, it is defined as

$$\mathbf{E}_2 = \begin{pmatrix} \mathbf{I} & \mathbf{1} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & -\mathbf{1} \end{pmatrix}. \quad (20)$$

The metric projection problem is the following

$$L(\mathbf{C}) = \text{tr}(\mathbf{S} - \mathbf{E}\mathbf{C})' \mathbf{F}(\mathbf{S} - \mathbf{E}\mathbf{C}), \quad (21)$$

where  $\mathbf{S}$  is the matrix with the coordinates  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$  on top of each other. Now (21) needs to be minimized over  $\mathbf{C}$ . The solution is given by

$$\mathbf{C} = (\mathbf{E}'\mathbf{F}\mathbf{E})^{-} \mathbf{E}'\mathbf{F}\mathbf{S}, \quad (22)$$

where  $(\mathbf{E}'\mathbf{F}\mathbf{E})^{-}$  is a generalized inverse of  $(\mathbf{E}'\mathbf{F}\mathbf{E})$ . The configuration of points is found as the first  $K$  rows of  $\mathbf{C}$ . For the slide-1 model, the slide-vector is given in the last row of  $\mathbf{C}$ . For the slide-2 model, the last two rows of  $\mathbf{C}$  give the slide-vectors.

### 3.3. More about asymmetry

Zielman and Heiser (1996) demonstrated that not only can data be additively decomposed into a symmetric and a skew-symmetric part, but also the model parameters for many asymmetric models. For the squared slide-vector model the decomposition can be written as

$$d_{ij}^2(\mathbf{X}; \mathbf{u}) = \sum_m (x_{im} - x_{jm})^2 + \sum_m u_m^2 + 2 \sum_m u_m (x_{im} - x_{jm}). \quad (23)$$

The term  $2\sum_m u_m (x_{im} - x_{jm})$  is skew-symmetric.

For the generalized slide-vector model, we can write

$$\begin{aligned} d_{ijk}^2(\mathbf{X}; \mathbf{u}; \mathbf{v}) &= \sum_m (x_{im} - x_{jm})^2 + \sum_m u_m^2 + 2 \sum_m u_m (x_{im} - x_{jm}) \\ &+ \sum_m (x_{jm} - x_{km})^2 + \sum_m v_m^2 + 2 \sum_m v_m (x_{jm} - x_{km}) \\ &+ \sum_m (x_{im} - x_{km})^2 + \sum_m (u_m + v_m)^2 + 2 \sum_m (u_m + v_m)(x_{im} - x_{km}). \end{aligned} \quad (24)$$

The terms  $2\sum_m u_m (x_{im} - x_{jm})$ ,  $2\sum_m v_m (x_{jm} - x_{km})$  and  $2\sum_m (u_m + v_m)(x_{im} - x_{km})$  correspond to

the skew-symmetric part. Rewriting this we obtain, for the skew-symmetric part,

$$2 \sum_m (2u_m x_{im} + v_m x_{im} - u_m x_{jm} + v_m x_{jm} - u_m x_{km} - 2v_m x_{km}), \quad (25)$$

and when  $\mathbf{u} = \mathbf{v}$ , this becomes

$$\sum_m (6u_m x_{im} - 6u_m x_{km}) = 6 \sum_m u_m (x_{im} - x_{km}). \quad (26)$$

In two-way analysis, we can decompose the data into a symmetric part and a skew-symmetric part. A singular value decomposition on the skew-symmetric data gives a nice representation of this skew-symmetry. The skew-symmetry between category  $i$  and  $j$  is approximated by twice the area of the triangle of the two points and the origin  $O$  (Gower, 1977). The areas of the triangles  $Oij$  and  $Oji$  are equal but different in sign, representing the skew-symmetry. A generalization of the singular value decomposition is the PARAFAC model (Harshman & Lundy, 1984). In PARAFAC the three-way data  $q_{ijk}$  are modelled by

$$q_{ijk} = \sum_m \lambda_m g_{im} h_{jm} l_{km}, \quad (27)$$

where  $g_{im}$ ,  $h_{jm}$  and  $l_{km}$  are elements of the component loading matrices  $\mathbf{G}$  for the first way,  $\mathbf{H}$  for the second way, and  $\mathbf{L}$  for the third way, respectively. The component loading matrices are constrained to be of unit length. The generalized singular values ( $\lambda_m$ ) are elements on the superdiagonal of a three-way core array, with the other elements of this array equal to zero.

As already mentioned in the Introduction, when a dominance relation exists in the data, the PARAFAC solution of the matrix  $\mathbf{N}$  has a special form. We will just give an illustration; a theoretical justification (if it exists) for the results obtained is beyond the scope of the present paper and will be left for future research. Table 2 gives a skew-symmetric table with dominance relations. We computed distances according to the slide-1 model in one dimension; the categories are  $a = 1$ ,  $b = 2$ ,  $c = 3$ ,  $d = 4$  and  $e = 5$ , and the slide-vector is equal to  $u = 2$ . From these distances, the symmetric part is subtracted; the residuals, or skew-symmetric part, are shown in Table 2.

In a PARAFAC analysis of the data in Table 2, only two components are needed, and fitting more components results in degeneracies. With two components, 99% of the sum of squares of the data is accounted for. The generalized singular values ( $\lambda_m$ ) are equal for the two components.

Table 3 gives the component loadings for the three ways. The two component matrices for the first and third way have a similar structure. The component loadings for the first way are a translation of the component loadings of the third way. In mathematical terms  $\mathbf{L} = \mathbf{J}_1 \mathbf{G} \mathbf{J}_2$ , where  $\mathbf{J}_1$  is a square interchange permutation matrix with 1s on the lower left to upper right diagonal, and  $\mathbf{J}_2$  is a square  $2 \times 2$  matrix with 0s on the diagonal, 1 above, and  $-1$  below the diagonal. The first column of the component loading matrix  $\mathbf{H}$  is equal to  $-\mathbf{J}_1$  times the second column of this matrix. So, with the component loading matrix for the first way, and the first column of the component loading matrix for the second way, we have (almost) all information available in the skew-symmetric table. We can see this more clearly when we

**Table 2.** Skew-symmetric three-way matrix with dominance relations

Way 1	Way 2	Way 3				
		<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
<i>a</i>	<i>a</i>	0.00	-1.26	-2.49	-3.47	-4.01
	<i>b</i>	0.10	-1.26	-2.76	-4.22	-4.98
	<i>c</i>	0.33	-0.96	-2.49	-4.21	-6.09
	<i>d</i>	0.56	-0.53	-1.89	-3.47	-4.98
	<i>e</i>	0.65	-0.23	-1.20	-2.65	-4.01
<i>b</i>	<i>a</i>	1.16	0.10	-0.96	-1.89	-2.65
	<i>b</i>	1.16	0.00	-1.26	-2.49	-3.47
	<i>c</i>	1.27	0.10	-1.26	-2.76	-4.22
	<i>d</i>	1.44	0.33	-0.96	-2.49	-4.22
	<i>e</i>	1.48	0.56	-0.53	-1.89	-3.47
<i>c</i>	<i>a</i>	2.16	1.27	0.33	-0.53	-1.20
	<i>b</i>	2.14	1.16	0.10	-0.96	-1.89
	<i>c</i>	2.16	1.16	0.00	-1.26	-2.49
	<i>d</i>	2.24	1.27	0.10	-1.26	-2.76
	<i>e</i>	2.39	1.44	0.33	-0.96	-2.49
<i>d</i>	<i>a</i>	2.91	2.24	1.44	0.56	-0.23
	<i>b</i>	2.97	2.16	1.27	0.33	-0.53
	<i>c</i>	2.97	2.14	1.16	0.10	-0.96
	<i>d</i>	2.91	2.16	1.16	0.00	-1.26
	<i>e</i>	2.88	2.24	1.27	0.10	-1.26
<i>e</i>	<i>a</i>	3.36	2.88	2.39	1.48	0.65
	<i>b</i>	3.50	2.91	2.24	1.44	0.56
	<i>c</i>	3.70	2.97	2.16	1.27	0.33
	<i>d</i>	3.50	2.97	2.14	1.16	0.10
	<i>e</i>	3.36	2.91	2.16	1.16	0.00

write the PARAFAC model for the three-way skew-symmetric matrix  $\mathbf{N}$  in matrix terms. We obtain

$$\begin{aligned}
 {}_{IJ}\mathbf{N}_K &= \mathbf{G}\Lambda(\mathbf{H} \otimes \mathbf{L}') \\
 &= \mathbf{G}\Lambda(\mathbf{H} \otimes (\mathbf{J}_1\mathbf{G}\mathbf{J}_2)') \\
 &= \mathbf{G}\Lambda(\mathbf{H} \otimes \mathbf{J}_2'\mathbf{G}'\mathbf{J}_1),
 \end{aligned} \tag{28}$$

**Table 3.** Component loadings for the PARAFAC analysis of Table 2

	<b>G</b>		<b>H</b>		<b>L</b>	
<i>a</i>	0.66	0.48	0.46	-0.35	-0.38	-0.60
<i>b</i>	0.32	0.48	0.50	-0.42	-0.42	-0.34
<i>c</i>	-0.03	0.46	0.50	-0.50	-0.46	-0.03
<i>d</i>	-0.34	0.42	0.42	-0.50	-0.48	0.32
<i>e</i>	-0.60	0.38	0.35	-0.46	-0.48	0.66

where  ${}_{IJ}\mathbf{N}_K$  is a matrix version of the three-way array  $\mathbf{N}$ , and  $\otimes$  is the Kronecker product. This is not exactly the same as in the standard formulation, but now the matrix is the same as those given in the different tables presented.

Because of the structure described above, we can also describe the skew-symmetric structure in terms of the first component vectors. Another formulation of the PARAFAC model is

$${}_{IJ}\mathbf{N}_K = \sum_m \lambda_m \mathbf{g}_m \otimes \mathbf{h}_m \otimes \mathbf{l}'_m, \quad (29)$$

where  $\mathbf{g}_m$  is the  $m$ th column of  $\mathbf{G}$ . In our analysis we only used two components, so we obtain

$$\begin{aligned} {}_{IJ}\mathbf{N}_K &= \lambda_1 \mathbf{g}_1 \otimes \mathbf{h}_1 \otimes \mathbf{l}'_1 + \lambda_2 \mathbf{g}_2 \otimes \mathbf{h}_2 \otimes \mathbf{l}'_2 \\ &= \lambda_1 \mathbf{g}_1 \otimes \mathbf{h}_1 \otimes \mathbf{l}'_1 + \lambda_2 (\mathbf{J}_1 \otimes \mathbf{J}_1) (\mathbf{l}_1 \otimes \mathbf{h}_1 \otimes \mathbf{g}'_1) \mathbf{J}_1, \end{aligned} \quad (30)$$

the latter expression resulting from using the permutation matrices.

If we analyse the residuals we obtain exactly the same solution. Analysing the residuals of the latter analysis again produces the same solution. Only the fit of the subsequent analyses declines. In this sense we constructed something similar to what are called ‘bimensions’ or ‘hedron’ (Gower & Zielman, 1998) in two-way analysis of skew-symmetric matrices.

The structure defined above is very strong. Other dominant skew-symmetric structures exist, for which the PARAFAC solution is not as nice as above. However, most often we find two strong components. Furthermore, the components for the first and third way strongly resemble each other. They give the direction of the slide-vector. The component loadings for the second way are most often found in one quadrant of the component space. If we find such a structure when analysing data, we could conclude that there is something like a unidimensional dominance or growth.

#### 4. Data analysis

In this section we apply the models to an empirical data set, analysing the data using the unrestricted asymmetric model, the slide-2 model and the slide-1 model. Moreover, we also carry out an analysis with a symmetric constraint, where the matrix  $\mathbf{E}$  consists of three identity matrices on top of each other. By comparing our models with the symmetric model, we obtain a measure of how much of the asymmetry is explained by the different asymmetric models.

We will compare different models by means of stress values and a measure called ‘percentage dispersion accounted for’ (%DAF). The latter is analogous to the diagnostic ‘variance accounted for’ in regression analysis (Heiser & Bannani, 1997). A basic optimality property for weighted least squares is that when a stationary point is reached, the following equality holds:  $\sum_{ijk} w_{ijk} \delta_{ijk} d_{ijk}(\mathbf{X}; \mathbf{Y}; \mathbf{Z}) = \sum_{ijk} w_{ijk} d_{ijk}^2(\mathbf{X}; \mathbf{Y}; \mathbf{Z})$ . Then, rewriting (8), we obtain  $\sigma^2(\mathbf{X}; \mathbf{Y}; \mathbf{Z}) = \sum_{ijk} w_{ijk} \delta_{ijk}^2 - \sum_{ijk} w_{ijk} d_{ijk}^2(\mathbf{X}; \mathbf{Y}; \mathbf{Z})$ . We can evaluate a solution by the percentage of dispersion accounted for, that is

$$\%DAF = 100 \times \frac{\sum_{ijk} w_{ijk} d_{ijk}^2(\mathbf{X}; \mathbf{Y}; \mathbf{Z})}{SSQ_\delta}.$$

## 4.1. Swedish politics data

As an illustration we look at a data set obtained from Upton (1978, p. 128). Swedish respondents were asked how they voted in three consecutive elections (1964, 1968, 1970). The data are given in Table 4. The four political parties are the Social Democrats (SD), the Centre party (C), the People's party (P), and the Conservatives (Con); this ordering is from left- to right-wing parties. The table gives the frequency of 64 possible sequences between these four parties at the three time points.

Transition frequencies are measures of similarity. First, we need to transform them to dissimilarities. We added  $1/64$  (the number of cells) to the frequencies to deal with zero frequencies, and then divided them by 1652 (total number of respondents plus 1). These observed relative frequencies ( $p_{ijk}$ ) were transformed to dissimilarities by the inverse of the Gaussian transform

$$p_{ijk} = \exp\{-\delta_{ijk}^2\}. \quad (31)$$

Such a transformation is often used in stimulus recognition data, or stimulus generalization data (Nosofsky, 1985; Shepard, 1957), where a similarity parameter is transformed into a distance.

We analysed the dissimilarities derived from Table 4 using the unrestricted asymmetric model, the slide-2 model, the slide-1 model and the symmetric model, using weights equal to one. The stress values and the % DAF values for the analyses in one, two, and three dimensions are shown in Table 5. It is clear that for all analyses (symmetric, slide-1, slide-2

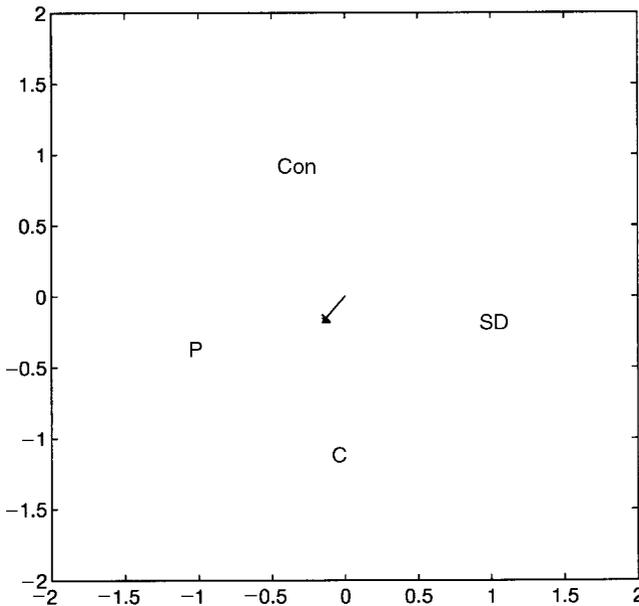
**Table 4.** Transition frequency table for Swedish elections for the years 1964, 1968 and 1970 (SD = Social Democrats; C = Centre party; P = People's party; Con = Conservatives)

		1970			
1964	1968	SD	C	P	Con
SD	SD	812	27	16	5
	C	5	20	6	0
	P	2	3	4	0
	Con	3	3	4	2
C	SD	21	6	1	0
	C	3	216	6	2
	P	0	3	7	0
	Con	0	9	0	4
P	SD	15	2	8	0
	C	1	37	8	0
	P	1	17	157	4
	Con	0	2	12	6
Con	SD	2	0	0	1
	C	0	13	1	4
	P	0	3	17	1
	Con	0	12	11	126

**Table 5.** Results for the analysis of Table 4

Model	Measure	1 dim	2 dim	3 dim
Symmetric	$\sigma^2$	55.51	18.93	12.84
	%DAF	87.50	95.74	97.11
Slide-1	$\sigma^2$	42.87	13.52	6.16
	%DAF	90.34	96.95	98.61
Slide-2	$\sigma^2$	42.79	12.97	5.39
	%DAF	90.36	97.08	98.79
Asymmetric	$\sigma^2$	12.25	5.33	2.55
	%DAF	97.24	98.80	99.42

and unrestricted asymmetry), two dimensions seem to give the best representation. The slide-2 model does not add much information to the results of an analysis with the slide-1 model. The unrestricted asymmetric model adds some information, especially in the one-dimensional case. The preferred analysis is the two-dimensional slide-1 model. Figure 5 shows the results. The smallest triadic distance is between the Social Democrats, the Centre party and the People's party, so most transitions occur between these three parties. The largest triadic distance is between the Social Democrats, the Centre party and the Conservatives; just a few transitions occur between these three parties. This effect seems to be intuitively correct because it cannot be expected that as many people go from a right-wing party to a left-wing party, as go from a right-wing or left-wing party to a party in the middle of the political

**Figure 5.** Solution for Swedish politics data.

spectrum. The slide-vector points in the direction of the Centre party and the People's party, so these parties are gaining votes over the years. The dyadic distances have a meaning, too. The square root of twice the squared dyadic distance represents the number of people who make a change once, or go back and forth between two parties. Since that is a monotonic function, we can compare the dyadic distances to gain some insight into the number of people who make transitions between two parties. Most transitions occur between the Centre party and the People's party. The largest dyadic distance is between the Centre party and the Conservatives, implying a small number of transitions between these two parties.

As we can see in Table 4, most of the respondents stay with their first choice. Only a few respondents switch twice. As is often done in log-linear analysis, we will have a closer look at the so-called movers, by giving zero weight to the dissimilarities on the main diagonal, and zero weight to the cells with structure  $iji$ ,  $ijj$  and  $ijj$ .

Table 6 gives, from the people who switched twice, the number of respondents who leave a specific party, the number of respondents who go to a specific party, and the net gain or loss, for both transitions from 1964 to 1968 and from 1968 to 1970.

The stress values and the %DAF values of the analysis in one, two and three dimensions for the four models are reported in Table 7. In three dimensions we do not gain much compared to two dimensions. With the unrestricted asymmetry model, we are clearly overfitting the data, since, in two dimensions the dispersion accounted for is over 99.9%. The number of parameters of the latter model is equal to the number of observations. The unrestricted asymmetric model in three dimensions fits more parameters than observations, so we did not try to fit this model. The slide-1 model explains 54% of the stress not accounted for by the symmetric model, while the slide-2 model explains 69%. This is a reasonable gain, compared to the small extra number of parameters that need to be fitted.

We will discuss the solution of the slide-2 model in more detail. Figure 6 shows the solution. Compare first the four points. Generally, there are the most transitions between the Social Democrats, the Centre party and the People's party, and the triadic distance is the smallest; the largest is the triadic distance between the Social Democrats, Centre party and Conservatives. The dyadic distances do not have a meaning in this analysis, because we gave all dyadic dissimilarities zero weight. There are two slide-vectors, the first being the solid vector and the second the dash-dotted vector. The slide-vectors point in the direction of dominance, or, in other words, in the direction of the categories that are gaining votes. Projecting the points onto the first slide-vector, we find that the Centre party projects highest, then the Conservatives, then the People's party, and last the Social Democrats. This is exactly

**Table 6.** Number of 'movers' who leave and go to a party (SD = Social Democrats; C = Centre party; P = People's party; Con = Conservatives)

	1964–1968			1968–1970		
	Leave	Go to	Net	Leave	Go to	Net
SD	16	3	-13	3	1	-2
C	1	8	7	8	13	5
P	5	6	1	6	12	6
Con	4	9	5	9	0	-9

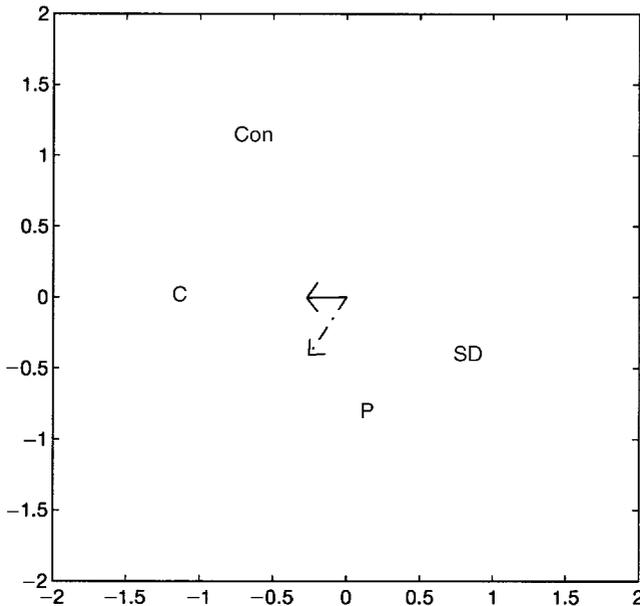
**Table 7.** Results for the analysis of the ‘movers’ of Table 4

Model	Measure	1 dim	2 dim	3 dim
Symmetric	$\sigma^2$	3.09	3.02	3.02
	%DAF	98.64	98.67	98.67
Slide-1	$\sigma^2$	3.07	1.38	1.38
	%DAF	98.65	99.39	99.39
Slide-2	$\sigma^2$	2.99	0.93	0.93
	%DAF	98.69	99.59	99.59
Asymmetric	$\sigma^2$	0.89	0.13	****
	%DAF	99.60	99.94	****

the same ordering as the net gain in Table 3. The same reasoning holds for the second slide-vector, with only the Centre party and People’s party interchanged compared to their net gain. The difference, however, in net gain is only one transition. So the model appears to give the appropriate answers.

### 5. Comparison and discussion

In this last section we discuss results obtained with log-linear models, compare them to our results, and show the advantages of both approaches. This section ends with some general remarks about the proposed methodology.

**Figure 6.** Solution for the analysis of the ‘movers’ of the Swedish politics data.

Upton (1978), Fingleton (1984) and Lindsey (1993) analysed the Swedish politics data using log-linear models. Upton concluded that a loyalty-distance model provides an adequate fit to the data (with a likelihood ratio statistic ( $LR$ ) of 55.01 with 40 degrees of freedom). The model can be written as

$$v_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} + \lambda_k^C + \sum_e \lambda_e^D \pm \lambda_1^X \pm \lambda_1^Y, \quad (32)$$

where  $v_{ijk}$  is the log of the expected frequency, and  $\mu$ ,  $\lambda_i^A$ ,  $\lambda_j^B$ ,  $\lambda_k^C$  and  $\lambda_{ij}^{AB}$  are the usual log-linear parameters for the mean, the three main effects (here  $A$  denotes 1964,  $B$  denotes 1968 and  $C$  denotes 1970) and an interaction effect between the first and second time point. The parameters  $\lambda_e^D$  are distance parameters defined by Upton only for the transition from 1968 to 1970. The distances are unidimensional and additive, so distance parameters are only defined for consecutive parties (on a given order), and the distance between, for example, the Social Democrats and the People's party is equal to the distance from the Social Democrats to the Centre party plus the distance from the Centre party to the People's party. The parameters  $\lambda_1^X$  and  $\lambda_1^Y$  represent loyalty, the first for the transition from the first to the third time point, the second for the transition from the second to the third time point. If there is loyalty, a value is added; if there is no loyalty, this value is subtracted. Fingleton (1984) concluded that the same model fits the data well, and argues that the distance parameter between the Centre party and the People's party might be set equal to zero, fitting one parameter less.

The distance concept in these models is rather *ad hoc* (distances might even be negative) and can only be unidimensional. In our models, we have a better-defined distance. The loyalty in our models is described by the length of the slide-vector. When more people are loyal to their party, the slide-vector will become shorter. Our analysis shows that there is more than a strict left-right political spectrum, and two dimensions are needed to represent the data. The number of parameters for the loyalty-distance model is 24; the number of parameters for our model for all the data (not only the movers) is 10, representing a substantial gain.

Lindsey (1993) fitted a first-order Markov chain, with the likelihood ratio statistic equal to 207.33 with 36 degrees of freedom. The model does not fit the data: the vote in 1970 is clearly dependent on the vote of 1964. He goes on to try fitting a quasi-symmetry model for the pairs of transitions 1964–1968 and 1968–1970, which does not fit the data either ( $LR = 73.49$ ,  $df = 22$ ).

Using a 'normal' hierarchical log-linear model approach, we find that the model with all first-order interactions fits the data reasonably well ( $LR = 34.46$ ,  $df = 27$ ). The model fits a lot of parameters (i.e., 37). This analysis does give us totally different information compared to the analysis with our models. It tells us that votes in 1970 are dependent on the votes of 1964, even if we adjust for the association between the second and third time point. Our models tell us that the Social Democrats and the Conservatives are getting less popular, and that there is a general tendency to vote for the parties in the so-called centre. Moreover, our models tell us, for those who wish to switch, the party to which they will most probably go.

The proposed models for the analysis of three-way proximity data may give useful insight into the structure of the three-way data. In many cases it is not necessary to fit the unrestricted asymmetry model and the data can be fitted by either the slide-1 or slide-2 model; this reduces the number of parameters significantly and gives us a sharp image of the symmetry and asymmetry in the data. It also enables us to easily grasp the dominance relations in the data. The unrestricted asymmetric model can always be applied to data with some success. To fit

the slide-vector models successfully, we must expect an overall tendency in the individuals' transitions. When each person makes his or her transitions based on individual preferences, and these preferences are diverse, we cannot expect the slide-vector models to fit well.

Even for count data, as in the example, where the least squares method does not seem to be very appropriate, the models do give us useful information. Moreover, with large tables having sparse cells the proposed methodology can still be used, but one should be cautious about interpreting the results of a log-linear analysis because of the zero entries (Krzanowski & Marriott, 1995). Finally, log-linear analysis merely gives insight into the interaction on the variable level, whereas our models give insight into the interaction on the category level.

In examples of log-linear models one often finds dichotomized variables to ensure the frequencies in the cells are large enough to base a model on. Our models can also be used to analyse large tables with sparse cells, as long as there is enough variation in the cell frequencies. Our models should not be applied to  $3 \times 3 \times 3$  or smaller tables. One main advantage of our models is the reduction in the number of parameters. This is especially true when the number of categories is large.

Our models do not, as do generalized linear models, decompose the data into additive parts for main effects, first-order interaction effects and second-order interaction effects. The models fit to the data, and can be interpreted directly by looking at the configuration. In other words, our models describe the data, and not an isolated interaction effect present in the data.

The main difference between our models and log-linear models is that our models focus on the category level, rather than on the variable level. This focus on the category level allows a more detailed analysis than is possible through an omnibus test at variable level.

In the present paper we have only discussed an application to three-way one-mode count data. Our models can also be used with three-way one-mode data that are not count data, but where the data can in some way be understood as (dis)similarities. An example is sociometric data (cf. Bond, Horn, & Kenny, 1997) in which every person in a group is asked to rate how much he or she thinks each other member of a group likes another member of a group. It is well known that in sociometric interaction data, some form of asymmetry often exists.

The three-way unfolding model can also be used to analyse three-way three-mode data. Heiser and Bennani (1997) and Joly and Le Calvé (1995) give examples. Heiser and Bennani analysed the data using an extended data matrix with missing intraset proximities. Our algorithm, however, is more efficient with respect to computer memory. We do not have to deal with matrices of the form described in (15), and so we do not need to compute a generalized inverse for this matrix.

### Acknowledgements

We are grateful to the Netherlands Organization for Scientific Research (NWO) for funding this project. This research was conducted while the first author was supported by a grant of the NWO Foundation for Behavioural and Educational Sciences (575-30-006).

### References

- Bond, C. F., Horn, E. M., & Kenny, D. A. (1997). A model for triadic relations. *Psychological Methods*, 2, 79–94.
- Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling: theory and applications*. New York: Springer-Verlag.

- 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.
- De Leeuw, J., & Heiser, W. J. (1980). Multidimensional scaling with restrictions on the configuration. In P. Krishnaiah (Ed.), *Multivariate analysis 5* (pp. 501-522). Amsterdam: North-Holland.
- Fingleton, B. (1984). *Models of category counts*. Cambridge: Cambridge University Press.
- Gower, J. C. (1977). The analysis of asymmetry and orthogonality. In J. Barra, F. Brodeau, G. Romer, & B. van Cutsem (Eds.), *Recent developments in statistics* (pp. 109-123). Amsterdam: North-Holland.
- Gower, J. C., & Zielman, B. (1998). Orthogonality and its approximation in the analysis of asymmetry. *Linear Algebra and its Applications*, *278*, 183-193.
- Groenen, P. J. F. (1993). *The majorization approach to multidimensional scaling*. Leiden: DSWO.
- Harshman, R. A., & Lundy, M. E. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H. G. Law, C. W. Snyder, J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 122-215). New York: Praeger.
- Heiser, W. J. (1981). *Unfolding analysis of proximity data*. Unpublished doctoral dissertation, Leiden University.
- Heiser, W. J. (1987). Joint ordination of species and sites: The unfolding technique. In P. Legendre & L. Legendre (Eds.), *Developments in numerical ecology* (pp. 189-221). Berlin: Springer-Verlag.
- Heiser, W. J. (1995). Convergent computation by iterative majorization: Theory and applications in multidimensional data analysis. In W. J. Krzanowski (Ed.), *Recent advances in descriptive multivariate analysis* (pp. 157-189). Oxford: Oxford University Press.
- Heiser, W. J., & Bennani, M. (1997). Triadic distance models: Axiomatization and least squares representation. *Journal of Mathematical Psychology*, *41*, 189-206.
- Joly, S., & Le Calvé, G. (1995). Three-way distances. *Journal of Classification*, *12*, 191-205.
- Krzanowski, W. J., & Marriott, F. H. C. (1995). *Multivariate statistics. Part 2: Classification, covariance structures and repeated measures*. London: Edward Arnold.
- Lindsey, J. K. (1993). *Models for repeated measurements*. Oxford: Oxford University Press.
- Nosofsky, R. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception and Psychophysics*, *38*, 415-432.
- Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325-345.
- Upton, G. J. G. (1978). *The analysis of cross-tabulated data*. Chichester: Wiley.
- Zielman, B., & Heiser, W. J. (1993). The analysis of asymmetry by a slide-vector. *Psychometrika*, *58*, 101-114.
- Zielman, B., & Heiser, W. J. (1996). Models for asymmetric proximities. *British Journal of Mathematical and Statistical Psychology*, *49*, 127-146.

Received 2 March 1999; revised version received 20 August 1999