RR 88-02

THE CITY-BLOCK MODEL FOR THREE-WAY MULTIDIMENSIONAL SCALING

WILLEM J. HEISER

April 1988

DEPARTMENT OF
DATA THEORY

THE CITY-BLOCK MODEL FOR THREE-WAY MULTIDIMENSIONAL SCALING

The city-block representation of three-way proximity data, with diagonal dimension weights, is studied for four reasons: (1) since it is the model of choice in certain domains, (2) because it naturally enjoys a unique axes property, (3) a large class of city-block (dis)similarity coefficients exists, and (4) there are some algorithmic problems to solve. A convergent algorithm based on iterative majorization and alternating least squares is presented, and a number of special applications with Gower's (1971) general coefficient of similarity are discussed

1. INTRODUCTION

All of the currently available models for three-way multidimensional scaling are based on the (weighted) Euclidean metric. Such models relate K tables of order n x n, containing dissimilarities among n objects collected from K sources, by assuming the existence of a p-dimensional common configuration of points $\mathbf{x}_i$, and K linear transformations $\mathbf{T}_k$ such that the Euclidean distances among the points $\mathbf{x}_{ik} = \mathbf{T}_k\mathbf{x}_i$ reproduce the kth dissimilarity table as closely as possible. For up-to-date reviews and references see Arabie *et al.* (1987). The case in which all $\mathbf{T}_k$ are restricted to be diagonal is called the INDSCAL model. It is the most prominent case, both in theoretical studies and in applications.

A special property often attributed to the INDSCAL model is its *rotational determinacy*, or *unique axes property*. That is, one cannot rotate the common configuration (and keep the source weights in terms of the original axes) without diminishing the fit to the source dissimilarities. The interpretation of the unique axes property would be that the selected directions are somehow more fundamental for the process studied than the other directions in Euclidean space. But if this is the case one could wonder why the Euclidean model is used in the first place, rather than a distance model that is inherently rotationally determinate, such as the so-called *city-block*, or L_1 distance.

Micko and Fischer (1970) have described the the class of Minkowski power metrics as a differential attention model for psychological proximity judgments in the following way: "To every direction in R^n is assigned a nonnegative measure which reflects the importance of the corresponding attribute. Irrespective of the metric, an overall distance is obtained by projecting the distance vector in R^n on all directions of R^n, weighting the Euclidean lengths of these projections by the corresponding importance scores, and averaging the weighted projection lengths. A particular metric is determined by a particular distribution of importance scores over the directions of R^n. The city-block metric, e.g., assigns importance scores $\neq 0$ to a set of mutually orthogonal directions only, the Euclidean metric assigns equal importance scores to all directions in R^n." (*l.c.*, p. 118; *Note:* n is used here in the role of p in the present notation). So a model that accounts for individual differences by a mechanism of differential weighting of orthogonal axes would seem to be most naturally phrased as - or certainly not in conflict with - a metric that assigns a positive importance score only to these orthogonal axes. Psychological considerations that indicate the applicability of the city-block metric can be found in, e.g., Attneave (1950), Shepard (1964), Torgerson (1965), Eisler and Roskam (1977), and Borg *et al.* (1982).

Of course, applications of three-way MDS are not confined to the area of psychological scaling. Generally, the K tables may contain (dis)similarity coefficients of many kinds. One general coefficient that includes several other ones as special cases has been described by Gower (1971). It is a similarity coefficient measuring the resemblance between two objects based on either dichotomous characters, qualitative characters, quantitative characters, or on any mixture of these. Gower's coefficient S_{ij} is of particular interest here, because it is essentially a constant minus the city-block metric. Now, Gower (1971) also shows that the similarity matrix $\mathbf{S} = \{S_{ij}\}$ is positive semi-definite, which implies that we can always find an embedding of the objects in Euclidean space. This result corresponds to a result by Schoenberg (1937, 1938), stating that every city-block metric has a square root which is Euclidean. However, the fact that a Euclidean embedding is always possible under a simple transformation does not necessarily imply that it is also the most useful, simplest, or most revealing representation of the objects. Hence our interest in the study of city-block scaling itself. Some ideas for three-way scaling of **S** will be presented in the discussion section.

2. THE CITY-BLOCK METRIC

For general theoretical results on the city-block metric the reader is referred to Fichet (1987) and Le Calvé (1987). We recall that for a configuration **X** of n points $\mathbf{x}_i$ in p dimensions, with coordinates x_{ia}, the city-block distance is determined by:

$$d(\mathbf{x}_i,\mathbf{x}_j) = \Sigma_a \mid x_{ia} - x_{ja} \mid , \qquad (1)$$

i.e., simply the sum of dimensionwise distances. In city-block space, the set of points with equal distance towards one central point is a square, and in general distances do not remain invariant under rotation of the coordinate system. For any two fixed points i and j, $d(\mathbf{x}_i,\mathbf{x}_j)$ is smallest when both points lie on one of the coordinate axes. If the coordinate system is rotated about $\mathbf{x}_i$, then $d(\mathbf{x}_i,\mathbf{x}_j)$ grows, reaches its maximum after a 45° rotation, and then shrinks again to the original values at 90°. In data analytical discussions of the city-block metric a recurrent worry is that, when the number of objects involved is small or when the dissimilarity data may monotonically transformed, the spatial representation is not unique or can become degenerate (cf. Shepard, 1974; Borg and Lingoes, 1987). Bortz (1974) has reported some examples of *partial isometries*, i.e. reflections of subsets of points around the 45° line that preserve all distances but alter the projections on the coordinate axes. Phenomena like these would certainly deserve more attention before city-block scaling can become a routine device.

For the three-way case expression (1) is generalized by substituting the INDSCAL relationship $\mathbf{x}_{ik} = \mathbf{W}_k\mathbf{x}_i$ between the common space points $\mathbf{x}_i$ and the individual source points $\mathbf{x}_{ik}$, where the notation $\mathbf{W}_k$ is used to indicate that the transformation is diagonal, with diagonal elements (dimension *weights*) w_{ka}. We obtain, for source $k = 1,\ldots,K$:

$$
\begin{aligned}
d(\mathbf{x}_{ik},\mathbf{x}_{jk}) &= \Sigma_a \mid w_{ka}x_{ia} - w_{ka}x_{ja} \mid \\
&= \Sigma_a \mid w_{ka} \mid \mid x_{ia} - x_{ja} \mid . \qquad (2)
\end{aligned}
$$

In the sequel we will omit the absolute value bars from the w_{ka}, remembering that they should be kept nonnegative. We note that, while in the Euclidean case there is a difference between either regarding the dimension weights as the salience of a coordinate *axis* or regarding them as the salience of coordinate *differences* (looking at, say, w_{ka} or w^2_{ka}), no such complication arises in the city-block case.

There are a number of special problems for city-block scaling, of two kinds.

(a) In the representation itself:
- the human eye tends to follow Euclid, not Manhattan;
- far more theoretical results exist on Euclidean spaces;
- possibilities for non-uniqueness, as mentioned above.

(b) In finding a city-block embedding:
- no algebraic method seems to be available;
- there are convergence problems in least squares fitting;
- and many local minima as well.

We will pay particular attention to the last two problems. As a preliminary to finding a good embedding with an algebraic method, Fichet (1987) has solved the L_1 *additive constant* problem by using a binary basis and by formulating it as a linear programming problem.

3. LEAST SQUARES FITTING OF THE THREE-WAY CASE

In this paper only least squares fitting will be considered; loss functions (partly) built upon least absolute residuals could be developed along the lines of Heiser (1988). Thus our aim is to minimize

$$\sigma^2(\mathbf{X},\mathbf{W}_1,...,\mathbf{W}_K) = \Sigma_i\Sigma_j\Sigma_k\, m_{ijk}\, (\delta_{ijk} - \Sigma_a\, w_{ka}\,|\, x_{ia} - x_{ja}\,|\,)^2\,, \tag{3}$$

where the m_{ijk} are fixed, given quantities called *masses*, indicating the relative importance, or precision, of each single dissimilarity δ_{ijk}. Masses of zero can be used to deal with missing data. We do not consider optimal transformations of δ_{ijk} here, but this feature could be incorporated without too much trouble. For convenience of presentation it is assumed that both the dissimilarities and the masses are symmetric for each source, i.e. $\delta_{ijk} = \delta_{jik}$ and $m_{ijk} = m_{jik}$; the consequences of dropping this restriction are quite straightforward. In order to make the loss function comparable across different sets of data, (3) is normalized through division by the total weighted sum of squares of the dissimilarities; the values to be reported below are the square root of the normalized loss. We minimize (3) iteratively by alternating minimizations over $\mathbf{X}$ for $\mathbf{W}_1,...,\mathbf{W}_K$ fixed at their current values with minimizations over $\mathbf{W}_1,...,\mathbf{W}_K$ for $\mathbf{X}$ fixed at its current values (in fact, we employ further partitionings of these subsets of parameters). For identification purposes we require in addition $1/K\ \Sigma_k\, w_{ka} = 1$ and $\Sigma_i\, x_{ia} = 0$ for each dimension a.

3.1. Finding the weights for fixed coordinates

For a fixed common configuration of points the weights can be found by solving K independent, relatively small *nonnegative regression* problems, in much the same way as is done in ALSCAL (Takane *et al.*, 1977). Since ALSCAL is based on a loss function in terms of the squared Euclidean distances, it has to deal with the same type of terms $\Sigma_a\, w_{ka}d_{ija}$ - with d_{ija} the *squared* dimensionwise distances, rather than the absolute differences as the ones in (3). Suppose we collect the latter in the n^2xp matrix $\mathbf{D}$, the rows of which correspond to some fixed order of the pairs i,j, and the columns of which correspond to the dimensions. Let $\boldsymbol{\delta}_k$ denote the dissimilarities of source k, enumerated in the same fixed order, and let $\mathbf{M}_k$ denote the diagonal matrix of order n^2xn^2, containing the masses m_{ijk} strung out in the same way. Then for any source k the relevant part of (3) becomes

$$\sigma^2(\mathbf{w}_k) = (\boldsymbol{\delta}_k - \mathbf{D}\mathbf{w}_k)'\mathbf{M}_k(\boldsymbol{\delta}_k - \mathbf{D}\mathbf{w}_k)\,, \tag{4}$$

where $\mathbf{w}_k$ is the p-vector containing the diagonal elements of $\mathbf{W}_k$. Let us now denote the *unconstrained minimizer* of (4) by

$$\mathbb{w}_k = (\mathbf{D}'\mathbf{M}_k\mathbf{D})^{-1}\mathbf{D}'\mathbf{M}_k\boldsymbol{\delta}_k \,. \tag{5}$$

3.2. Keeping the weights nonnegative

Due to the orthogonality of least squares projection, loss function (4) can be split into two additive components:

$$\sigma^2(\mathbf{w}_k) = \sigma^2(\mathbb{w}_k) + (\mathbb{w}_k - \mathbf{w}_k)'\mathbf{D}'\mathbf{M}_k\mathbf{D}(\mathbb{w}_k - \mathbf{w}_k) \,. \tag{6}$$

It is clear that we now only need to minimize the second part of (6) over $\mathbf{w}_k \geq \mathbf{0}$, which is a constrained least distance problem in the metric $\mathbf{D}'\mathbf{M}_k\mathbf{D}$. There are several possibilities to solve this nonnegative least distance problem; since the order of the problem is p, which will in general be rather small, it turns out to be sufficient to use elementwise adjustments of $\mathbf{w}_k$; i.e., we may express the weights as $\mathbf{w}_k = \mathbf{p}_k + \alpha\mathbf{e}_a$, where $\mathbf{p}_k$ is some previous estimate satisfying nonnegativity, $\mathbf{e}_a$ is the ath column of the pxp identity matrix, and α is the adjustment parameter. Inserting the elementwise adjustment into (6), the restriction $\mathbf{w}_k \geq \mathbf{0}$ becomes $\alpha \geq -\mathbf{e}_a'\mathbf{p}_k$, from which it follows that we must choose α equal to $\mathbf{e}_a'\mathbf{D}'\mathbf{M}_k\mathbf{D}(\mathbb{w}_k - \mathbf{p}_k) / \mathbf{e}_a'\mathbf{D}'\mathbf{M}_k\mathbf{D}\mathbf{e}_a$ if this quantity satisfies the bound, and equal to $-\mathbf{e}_a'\mathbf{p}_k$ otherwise (making the adjusted weight exactly zero). This scheme can be repeated for $a = 1,\ldots,p$ (each time immediately updating $\mathbf{p}_k$). For large p the elementwise alternating least squares procedure is not recommended (it will become very slow), but other methods are available.

3.3. Finding the coordinates for fixed weights

Fortunately, for fixed weights a reduction to a two-way city-block scaling problem is possible. Unfortunately, the two-way city-block problem is much harder than the usual Euclidean case. Eisler (1973) has shown how to obtain the coordinates by linear regression when the within-dimensional ranks of the points are known (or fixed). However, when we have to optimize over the within-dimensional ranks as well, two severe difficulties must be resolved (Hubert and Arabie, 1988):

(a) the usual gradient procedures show erratic convergence behaviour, and
(b) the number of local minima is extremely large.

For the first difficulty a solution will be offered here, based on a generalization of the majorization approach of De Leeuw and Heiser (1980), as described in Heiser (1987). For the second difficulty a suggestion by Hubert and Arabie (1988) will be followed, in slightly adjusted form.

Note that we may also reduce the problem straight away to a series of two-way one-dimensional scaling problems, as follows. For dimension a we form

$$\mathbb{\delta}_{ijk} = \delta_{ijk} - \Sigma_{b \neq a}\, w_{kb}\, | x_{ib} - x_{jb} | , \tag{7}$$

and collect these fixed quantities in the n^2-vector $\mathbb{\delta}_k$, in the same way as before. Now loss function (3) becomes

$$\begin{aligned} \sigma^2(\mathbf{x}_a) &= \Sigma_i \Sigma_j \Sigma_k\, m_{ijk}\, (\mathbb{\delta}_{ijk} - w_{ka}\, | x_{ia} - x_{ja} |)^2 \\ &= \Sigma_k\, (\mathbb{\delta}_k - w_{ka}\, \mathbf{d}_a)'\mathbf{M}_k(\mathbb{\delta}_k - w_{ka}\, \mathbf{d}_a) , \end{aligned} \tag{8}$$

where $\mathbf{d}_a$ is the ath column of $\mathbf{D}$. Again we first compute the unconstrained minimizer $\mathbb{d}_a$ as

$$\mathbb{d}_a = \mathbf{M}_a^{-1}\, \Sigma_k\, w_{ka}\mathbf{M}_k\mathbb{\delta}_k, \quad \text{with} \tag{9}$$

$$\mathbf{M}_a = \Sigma_k\, w^2_{ka}\mathbf{M}_k , \tag{10}$$

in which $\mathbf{M}_a$ is the diagonal matrix of aggregated masses, weighted by the *squares* of the dimension weights. Slight adjustments in (9) and (10) are necessary when some of the aggregated masses are zero. Now (8) can be split into the two components

$$\sigma^2(\mathbf{x}_a) = \Sigma_k\, (\mathbb{\delta}_k - w_{ka}\, \mathbb{d}_a)'\mathbf{M}_k(\mathbb{\delta}_k - w_{ka}\, \mathbb{d}_a) + (\mathbb{d}_a - \mathbf{d}_a)'\mathbf{M}_a(\mathbb{d}_a - \mathbf{d}_a) , \tag{11}$$

the first part of which is constant, while the second part constitutes a one-dimensional scaling problem with respect to $\mathbb{d}_a$ regarded as dissimilarities, and with masses $\mathbf{M}_a$. This scheme can be repeated for all p axes.

3.4. Lack of nonnegativity of the unconstrained minimizer

As remarked above, the usual gradient procedures for minimizing the second part of (11) show erratic convergence behavior. The reason suggested here is the fact that the quantities $\mathbb{d}_a$ as defined in (9) need not be nonnegative, because they are based on the quantities $\mathbb{\delta}_k$, as defined in (7), which in turn need not be nonnegative. Now nonnegativity of the "data" of the scaling problem (in the present case the data are $\mathbb{d}_a$) is an important condition in the convergence proof of the usual procedure (De Leeuw and Heiser, 1980). So what we have to do in the present case is using a *generalized* procedure that does cope properly with negative elements in $\mathbb{d}_a$.

4. A CONVERGENT ONE-DIMENSIONAL SCALING PROCEDURE WHEN SOME DATA ELEMENTS ARE NEGATIVE

Let us register the elements of $\mathbb{d}_a$ that are negative in the index set $S(N)$, and express them, rearranged in double subscripted form, explicitly as $\mathbb{d}_{ij} = -\,|\,\mathbb{d}_{ij}\,|$ for $i,j \in S(N)$, dropping reference to the particular dimension we are working at. The remaining index pairs are collected in the index set $S(P)$, and the loss function for the one-dimensional scaling problem becomes

$$\begin{aligned}\sigma^2(\mathbf{x}) &= \Sigma_i\Sigma_j \, m_{ij} \, (\mathbb{d}_{ij} - |x_i - x_j|\,)^2 \\ &= \Sigma_i\Sigma_j \, m_{ij} \, \mathbb{d}^2_{ij} + \Sigma_i\Sigma_j \, m_{ij} \, (x_i - x_j)^2 \\ &\quad - 2 \, \Sigma_{i,j \in S(P)} \, m_{ij} \, \mathbb{d}_{ij}|x_i - x_j| + 2 \, \Sigma_{i,j \in S(N)} \, m_{ij} \, |\,\mathbb{d}_{ij}\,| \, |x_i - x_j| \, . \end{aligned} \tag{12}$$

Here m_{ij} denotes the elementwise mass of the two-way problem, for dimension a defined by (10), and $\mathbb{d}_{ij}$ denotes the elementwise dissimilarity, for dimension a defined by (9). Reference to dimension a has also been dropped from the scale values $\mathbf{x} = \{x_i, i = 1,...,n\}$. The key result in the *majorization* approach (De Leeuw and Heiser, 1980) is the observation that the third term in (12), including the -2, is a *concave* function of **x**, which can be majorized by a suitably chosen *linear* function. Together with the second term, which is quadratic in **x**, this enables us to majorize the entire loss function with an auxiliary quadratic function that is easy to minimize. Starting at any feasible point, the auxiliary function is constructed so that it precisely touches the loss function at that point and is larger everywhere else. Then the minimizer of the auxiliary function defines a new feasible point with diminished loss function value.

4.1. Majorization of a convex function

In the present situation we have to deal also with the fourth term in (12), which is a *convex* function of **x**. Whereas we cannot hope to majorize a convex function with a linear one, we can use a quadratic here. The reader is referred to Heiser (1987) for details. The basic inequality to be used now is

$$2 \, m_{ij} \, |\,\mathbb{d}_{ij}\,| \, |x_i - x_j| \;\leq\; m_{ij} \, |\,\mathbb{d}_{ij}\,| \, |y_i - y_j| + \{m_{ij} \, |\,\mathbb{d}_{ij}\,| \, / \, |y_i - y_j|\} \, (x_i - x_j)^2 \, , \tag{13}$$

the right-hand side being a quadratic function of **x** for a given set of coordinates **y**. When some of the elements of **y** are equal (this almost surely happens when some of the data elements are negative!), special precautions are required to avoid division by zero in (13). Together with the linear function for majorizing the third term in (12) we now have an auxiliary function that is at most quadratic and majorizes the original function over its entire domain. So a monotonically decreasing series of steps can be constructed, from which it follows that convergence to a local minimum is guaranteed.

4.2. Avoiding local minima

We now follow a suggestion by Hubert and Arabie (1988) to use *pairwise interchanges* of any initial estimate of the scale values for avoiding local minima in one-dimensional scaling. Thus the next iteration scheme is used:

```
(i)   SEED  ← initial estimate
(ii)  TRY   ← generalized 1-dim scaling
      while still improvements found do
            for all pairs i,j do
(iii)             interchange i and j in SEED
(iv)              TRY ← generalized 1-dim scaling
                  if loss improves then
(v)                     set SEED ← TRY
                  end if
            end for
      end while
```

This scheme, although computationally quite intensive, is easy to program and it turns out to work remarkably well. The generalized one-dimensional scalings do not converge as fast as ordinary ones, but usually the number of iterations is acceptable (10 to 20), also due to the fact that good initial estimates are used. The city-block scaling program constructed along these lines has been called LONESCALE (actually, it is a set of APL functions).

5. SIMPLE STRUCTURE PATTERNS IN THE WEIGHTS

With respect to dimensionality the following trade-off situation frequently occurs. On the one hand one supposes that in each individual space dimensionality is low, while on the other hand one has to accept a common space of high dimensionality if the data is to be described reliably at all, by which is meant that (almost) all variability is accounted for. Therefore it is perhaps useful to point out the existence of a special case of the INDSCAL model, first discussed in the Euclidean least squares case by Heiser and Stoop (1986), in which some of weights are restricted to be zero. This type of restriction diminishes the number of parameters while still keeping the basic idea of *simple structure* intact. Under the city-block model only very minor adjustments suffice to fit this version of three-way multidimensional scaling.

6. APPLICATION: COLA DATA

As an example the "Cola" data will be analysed, which are given in Schiffman *et al.* (1981); this book contains extensive Euclidean analyses of these data, performed with the programs MINISSA, POLYCON, KYST-2, INDSCAL, SINDSCAL, ALSCAL, and MULTISCALE. It concerns a tasting experiment with 10 different brands of "cola" beverages (with the bubbles removed) at room temperature. The 45 pairwise dissimilarity jugments were made over a 5-day period, by 10 judges.

Table 1. Stimulus scale values for 3-dim city-block solution of Cola data.

Cola	dim.1	dim.2	dim.3
Diet Pepsi	1.71	2.21	0.56
RC Cola	- 1.48	1.85	0.22
Yukon	- 0.68	- 0.32	2.78
Dr. Pepper	- 2.56	- 3.08	0.56
Shasta	- 1.96	.56	- 0.64
Coca-Cola	- 0.68	0.03	- 1.82
Diet Dr. Pepper	1.63	- 3.18	0.00
Tab	2.97	- 0.38	- 1.28
Pepsi-Cola	- 1.70	1.34	- 0.99
Diet Rite	2.74	0.98	0.61

Table 2. Subject dimension weights for 3-dim city-block solution of Cola data.

Subject	dim.1	dim.2	dim.3
01	1.15	.95	.95
02	.45	1.52	1.09
03	.32	1.57	1.74
04	1.43	.65	.89
05	1.69	.16	.56
06	1.61	.31	.44
07	.71	1.58	1.50
08	.73	1.28	.52
09	1.37	.50	1.21
10	.53	1.48	1.11

So this yields a 10 x 10 x 10 dissimilarity table, which is symmetric for each subject. The three-way city-block solutions in two, three, and four dimensions have loss function values of .343, .295, and .278, respectively. This would seem to indicate a three-dimensional common space;

the scale values of the stimuli on the three axes are given in Table 1, while the subject dimension weights are given in Table 2. Although MDS results are usually presented in a series of plots - a trademark that has sometimes overshadowed the whole enterprise of distance modelling - the temptation to do the same thing is resisted here, since the city-block metric is an outstanding example of a model in which only the coordinate axes are meaningful. There is no need to inspect other directions. There is no "rotation problem", i.e. no invariance of results under rotations of the coordinate system. What one could do is to plot the dimensions separately next to each other (an "inside-out" plot of the table of coordinates, Ramsay, 1980).

The first axis clearly gives the contrast between the diet colas and the nondiet colas, while the second axis separates the cherry-flavoured colas from the regular ones. On the third axis Yukon stands out, but it is less clear how the others are ordered. There are more judgment data from the same subjects that could be used to account for the stimulus differences, but it would lead us too far to attempt this now. The dimension weights, Table 2, indicate considerable individual variation among subjects. Subjects 2, 3, and 10 appear insensitive to the diet-nondiet distinction, whereas subjects 5 and 6 appear to experience most taste differences along this axis; analogous remarks can be made about other subjects and dimensions. Compared to the analyses in Schiffman *et al.* (1981), the LONESCALE results appear to be in quite close agreement with the MULTISCALE results.

7. DISCUSSION

The three-way scaling model discussed in this paper could be extended with features like unique stimulus axes, analogous to the Winsberg and Carroll contribution to this volume, and optimal transformations of the dissimilarities in an obvious manner. An interesting possibility that suggest itself starting from the city-block metric would be to fit monotonic splines rather than merely positive dimension weights to the dimensionwise coordinate differences.

If one would hypothesize that different sources associate different importances to *different directions* in space, it would be better to use a Euclidean model with *individual reduced rank metrics* (i.e., the matrices $\mathbf{T}_k$ - although of order p x p - are constrained to be of rank r_k, as e.g. in Easterling, 1987), rather than to try to connect this idea with the city-block metric. Note that if $\mathbf{T}_k$, in additon to being of reduced rank, is restricted to diagonal form one gets the individual differences model discussed in section 5.

In case the dissimilarity data arise from using Gower's similarity coefficient matrices $\mathbf{S}_k$ interesting applications can be suggested. For instance, suppose each $\mathbf{S}_k$ contains the similarity coefficients merely corresponding to one variable in a group of K variables. Then the present analysis really is a principal components analysis on a distance scale. Would we *also* optimise over monotonic transformations of the similarities we would obtain a method for analyzing *ordered metric scales*, which are prominent in classical unfolding theory (Coombs, 1964). Next, suppose one variable is qualitative and coded in $\mathbf{S}_1$, whereas the other variables constitute $\mathbf{S}_2$. Then we obtain *discriminant analysis in the city-block metric*. Finally, suppose the $\mathbf{S}_k$ matrices are arbitrary, but we restrict the configuration $\mathbf{X}$ to some known binary structure. Then we would obtain yet another interesting form of *generalized discriminant analysis* of K sets of variables.

REFERENCES

Arabie, Ph., Carroll, J.D., and DeSarbo, W.S. (1987). *Three-way scaling and clustering.* Newbury Park, CA: SAGE Publications.

Attneave, F. (1950). Dimensions of similarity. *The American Journal of Psychology*, **63**, 516-556.

Borg, I., Schönemann, P.H. and Leutner, D. (1982). Merkmalsüberdeterminierung und andere Artefakte bei der Wahrnehmung einfacher geometrische Figuren. *Zeitschrift für experimentelle und angewandte Psychologie*, **24**, 531-544.

Borg, I. and Lingoes, J. (1987). *Multidimensional Similarity Structure Analysis.* New York: Springer-Verlag.

Bortz, J. (1974). Kritische Bemerkungen über den Einsatz nichteuklidischer Metriken im Rahmen der multidimensionalen Skalierung. *Archiv für Psychologie*, **126**, 194-212.

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

De Leeuw, J. and Heiser, W.J. (1980). Multidimensional scaling with restrictions on the configuration. In *Multivariate Analysis, Vol V* (P.R. Krishnaiah, ed.). Amsterdam: North-Holland.

Easterling, D.V. (1987). Using the generalized Euclidean model to study ideological shifts in the U.S. Senate. In *Multidimensional Scaling: History, Theory, and Applications* (F.W. Young and R.M. Hamer, eds.). Hillsdale, NJ: Erlbaum.

Eisler, H. (1973). The algebraic and statistical tractability of the city-block metric. *British Journal of Mathematical and Statistical Psychology*, **26**, 212-218.

Eisler, H. and Roskam, E.E. (1977). Multidimensional similarity: an experimental and theoretical comparison of vector, distance, and set theoretical models (II. Multidimensional analyses: the subjective space). *Acta Psychologica*, **41**, 335-363.

Fichet, B. (1987). The role played by L1 in data analysis. In *Statistical Data Analysis based on the L1-norm and related methods* (Y. Dodge, ed.). Amsterdam: North-Holland.

Gower, J.C. (1971). A general coefficient of similarity and some of its properties.*Biometrics*, **27**, 857-874.

Heiser, W.J. (1987). Majorizing STRESS when some dissimilarities are negative. *Internal Report RR-87-17*, Department of Data Theory, Leiden, The Netherlands.

Heiser, W.J. (1988). Multidimensional scaling with least absolute residuals. In *Classification and Related Methods of Data Analysis* (H.H. Bock, ed.). Amsterdam: North-Holland.

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

Hubert, L., and Arabie, Ph. (1988). Relying on necessary conditions for optimization: unidimen-sional scaling and some extensions. In *Classification and Related Methods of Data Analysis* (H.H. Bock, ed.). Amsterdam: North-Holland.

Le Calvé, G. (1987). L1-embeddings of a data structure (I,D). In *Statistical Data Analysis based on the L1-norm and related methods* (Y. Dodge, ed.). Amsterdam: North-Holland.

Micko, H.C. and Fischer, W. (1970). The metric of multidimensional psychological spaces as a function of the differential attention to subjective attributes. *Journal of Mathematical Psychology*, **7**, 118-143.

Ramsay, J.O. (1980). *Inside-out displays and more*. Paper presented at the Symposium "Multivariate Data Display", Psychometric Society Meeting, Iowa City.

Schiffman, S.S., Reynolds, M.L. and Young, F.W. (1981). *Introduction to Multidimensional Scaling*. New York: Academic Press.

Schoenberg, I.J. (1937). On certain metric spaces arising from Euclidean space by a change of metric and their imbedding in Hilbert space. *Ann. Math.*, **40**, 787-793.

Schoenberg, I.J. (1938). Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.*, **44**, 522-536.

Shepard, R.N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, **1**, 54-87.

Shepard, R.N. (1974). Representation of structure in similarity data: problems and prospects. *Psychometrika*, **39**, 373-421.

Takane, Y., Young, F.W. and De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika*, **42**, 7-67.

Torgerson, W.S. (1965). Multidimensional scaling of similarity. *Psychometrika*, **30**, 379-393.