PSYCHOMETRIKA—VOL. 70, NO. 1, 45–69
MARCH 2005
DOI: 10.1007/s11336-002-1046-0

INTERPRETING DEGENERATE SOLUTIONS IN UNFOLDING BY USE OF THE VECTOR MODEL AND THE COMPENSATORY DISTANCE MODEL

K. VAN DEUN

CATHOLIC UNIVERSITY OF LEUVEN

P.J.F. GROENEN

ERASMUS UNIVERSITY ROTTERDAM

W.J. HEISER

LEIDEN UNIVERSITY

F.M.T.A. BUSING

LEIDEN UNIVERSITY

L. DELBEKE

CATHOLIC UNIVERSITY OF LEUVEN

In this paper, we reconsider the merits of unfolding solutions based on loss functions involving a normalization on the variance per subject. In the literature, solutions based on Stress-2 are often diagnosed to be degenerate in the majority of cases. Here, the focus lies on two frequently occurring types of degeneracies. The first type typically locates some subject points far away from a compact cluster of the other points. In the second type of solution, the object points lie on a circle. In this paper, we argue that these degenerate solutions are well fitting and informative. To reveal the information, we introduce mixtures of plots based on the ideal point model of unfolding, the vector model, and on the signed distance model. In addition to a different representation, we provide a new iterative majorization algorithm to optimize the average squared correlation between the distances in the configuration and the transformed data per individual. It is shown that this approach is equivalent to minimizing Kruskal's Stress-2.

Key words: unfolding, degeneracy, mixed plots, squared correlation, iterative majorization.

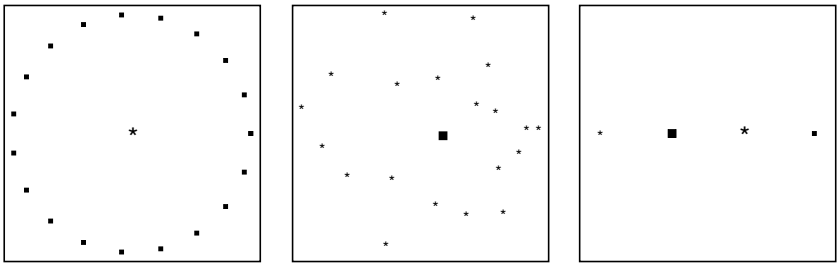
1. Introduction

Multidimensional unfolding is a data analysis technique for two-way two-mode proximity data. A well-known case is preference data concerning the preference orders of n subjects for m objects. The goal of multidimensional unfolding techniques is to represent such data in a low-dimensional space. In the distance-based unfolding models, coordinates for both the objects and the subjects are determined such that the distances of a subject to the objects reflects the subjects preference. These unfolding models are also called ideal point models and originate from the unidimensional unfolding model formulated by Coombs (1950).

Solutions to the distance-based unfolding problem have been sought within the framework of multidimensional scaling (Kruskal & Carroll, 1969). Unfolding is then a special case of multidimensional scaling (for a treatment of unfolding as multidimensional scaling, see, for example, Borg & Groenen, 1997, Section 12.1). Soon, it appeared that this approach resulted in degenerate

Requests for reprints should be sent to K. Van Deun, Department of Psychology, Catholic University of Leuven, Tiensestraat 102, B-3000 Leuven, BELGIUM. E-mail: katrijn.vandeun@psy.kuleuven.ac.be

© 2005 The Psychometric Society



(a) Objects Circle **(b)** Object Point **(c)** Two-plus-two

FIGURE 1.
Trivial solutions: the objects-circle (left panel), the object point (middle panel) and the two-plus-two configuration (right panel). Squares represent object points, stars subject points; small symbols indicate a single point, large multiple points.

solutions for the majority of cases. These are "extremely uninformative solutions with good or even perfect fit" (Heiser, 1989). A degenerate solution is characterized by the fact that the objects are at an (almost) equal distance to the subjects. The reason for the appearance of these solutions is the measurement level of the data and the badness-of-fit function used. Kruskal and Carroll (1969) and De Leeuw (1983) presented a detailed overview of such degeneracies. Here we summarize their main points.

Preference data are ordinal in nature so that monotone transformations per subject are allowed. Making use of Kruskal's monotone regression procedure (see Kruskal, 1964), in combination with a (simple) least-squares loss function, can result in a situation where all transformed data are equal. Even when using an interval transformation, equality of transformed data can occur (Busing, Groenen, & Heiser, in press). Note that unfolding with monotone transformations is called nonmetric unfolding while with interval and ratio transformations it is called metric unfolding. Many loss functions in multidimensional scaling take the form of a fraction where the numerator measures the deviation between the distances and the transformed data and the denominator equals a normalizing function. The different loss functions are distinguished by the normalizing function. For Stress-1, this function is the sum of squared distances. A trivial solution like the objects–circle (see Figure 1a) perfectly represents the transformed data: all objects are equidistant from the subjects, thus reflecting the equal transformed data and the sum of squared deviations equals zero. As the normalizing factor is non-zero, the solution has a zero badness-of-fit. This is a trivial and data-independent solution, but it has a perfect fit. The failure of Stress-1 as a badness-of-fit criterion in unfolding, led to the development of other stress measures. An historical overview, based on those given by Kruskal and Carroll (1969) and by De Leeuw (1983) is given next.

To avoid trivial solutions, normalizing functions were introduced that would equal zero for the trivial solution and as such make loss undefined at the trivial solution. A possible normalization function is the variance of the distances. In a first attempt, the variance over all subjects was used. This function led to another trivial and data-independent solution: the object point configuration (see Figure 1b). In this configuration, all objects are equidistant from each subject. Over the subjects, however, there is variation in the distances. As a result, the numerator of the loss function is zero while the denominator is non-zero giving a perfect but trivial solution. In a second attempt, the sum of the variances per subject was taken. The object point configuration could be thereby avoided but another trivial solution occurred, the two-plus-two solution (Figure 1c): let the small star on the left in Figure 1c denote the first subject and the small square on the right the stimulus the subject least prefers. All other subjects are represented by the large star and all other stimuli by the large square. This gives a trivial solution with perfect fit and with some dependence on the

data. Another attempt (Kruskal & Carroll, 1969) was made by defining the badness-of-fit as the root of the mean (over subjects) squared Stress with normalization on the variance. This function is called Stress-2. No trivial solution could be found in this case although *partially* degenerate solutions kept appearing in applications. With Stress-2, variance in the distances for each subject is required to avoid a loss that is undefined. Given the nature of the badness-of-fit function, these solutions should be data-dependent to a certain extent.

This short history shows that degenerate solutions form a heterogeneous class ranging from trivial data-independent solutions to partially degenerate solutions. De Leeuw (1983) showed that there are close connections between the object point trivial solution and the vector model and between the objects–circle trivial solution and the compensatory distance model. His conclusion is that it is not possible to avoid degeneracies using Stress-2 and hence this type of approach is a lost cause. We cite a part of the Discussion (De Leeuw, 1983): "[...], it is clear that in the neighborhood of a trivial point nonmetric unfolding algorithms will choose their paths according to the fit of the vector or compensatory distance model. [...] It seems to us that by using this partial degeneracy clever nonmetric unfolding programs will almost always find very good solutions, which tell us something about the data, but not very much." Contrary to his conclusion, the view of the present authors is that the solutions are informative just because of this relation. We suggest to look at the degenerate solutions from the perspective of the vector or compensatory distance model. In this paper, we claim that so-called degeneracies do contain useful information on the data. The problem is that the information is not adequately visualized in the ideal point solution. The optimization criterion that we will use is based on the squared correlation between the transformed data and the distances. An important property of the correlation is that both the distances and the disparities are normalized on their respective variances. Furthermore, it has a close connection to Stress-2, as was already noted by Kruskal and Carroll (1969). Given that the results of De Leeuw hold for Kruskal's Stress-2, this connection is desirable. In the appendix, we provide an iterative majorization (IM) algorithm for maximizing the squared correlation. This new algorithm has the important advantage over other algorithms that it yields a monotone non-decreasing sequence of the squared correlations throughout the alternating iterative procedure, whereas existing algorithms for Stress-2 cannot guarantee this.

This paper is organized as follows. First, the use of the squared correlation in the framework of unfolding is explained. Then, it is shown how the object point degeneracy is related to the vector model and how this connection can be used to come to an interpretation of the degenerate solution. In the next section, the connection between the objects-circle degeneracy and a signed version of the compensatory distance model is used to interpret such a degenerate solution. Both cases are illustrated with an empirical example. The solutions obtained here are based on an unfolding algorithm that maximizes the squared correlation. This algorithm is given in the appendix. We conclude with a discussion.

Note that for convenience of presentation, we limit ourselves in this paper to two-dimensional solutions only, even though most of what will be discussed below holds in higher dimensionality as well.

2. The Optimization Criterion

The goal of multidimensional unfolding is to find coordinates for the ideal points i, with $i = 1, ..., n$ and for the stimulus points j, with $j = 1, ..., m$ in a p-dimensional space with dimensions b, with $b = 1, ..., p$. The ideal point coordinates x_{ib} are collected in the $n \times p$ matrix $\mathbf{X}$ or in the np vector $\mathbf{x}$ while the stimulus coordinates y_{jb} are collected in the $m \times p$ matrix $\mathbf{Y}$ or in the mp vector $\mathbf{y}$. These coordinates have to be such that the distances from the ideal points to the stimulus points reflect the proximities δ_{ij}. The data are transformed according to their measurement level (ordinal, interval, ratio). Loss functions in unfolding are based on deviations

between the distances and the transformed data $\gamma_{ij} = f(\delta_{ij})$, called disparities. The disparities are collected in the $n \times m$ matrix $\mathbf{\Gamma}$ or in the nm vector $\boldsymbol{\gamma}$.

First, we will show how the correlation can be maximized by explicitly normalizing the disparities. In the second section, the equivalence of this loss function with Stress-2 is proved.

2.1. Maximizing the Squared Correlation

We propose to maximize the squared correlation between the distances and the disparities. The squared correlation is given by

$$r^2(\boldsymbol{\gamma}, \boldsymbol{d}) = \frac{(\boldsymbol{\gamma}'\boldsymbol{J}\boldsymbol{d})^2}{(\boldsymbol{\gamma}'\boldsymbol{J}\boldsymbol{\gamma})(\mathbf{d}'\mathbf{J}\mathbf{d})}, \tag{1}$$

with $\mathbf{J} = \mathbf{I} - \mathbf{1}\mathbf{1}'(\mathbf{1}'\mathbf{1})^{-1}$, $\mathbf{1}$ being a unit vector of length nm, $\mathbf{I}$ an identity matrix of size $nm \times nm$, and where $\mathbf{d}$ has elements $d_{ij} = (\sum_{b=1}^{p}(x_{ib} - y_{jb})^2)^{1/2}$, which is the Euclidean distance between ideal point i and stimulus point j.

We will now show that maximizing the squared correlation comes down to minimizing

$$\begin{aligned} L_{\mathrm{u}}(a, \mathbf{X}, \mathbf{Y}, \mathbf{\Gamma}) &= \|\boldsymbol{\gamma} - a\mathbf{d}\|_{\mathbf{J}}^2 \\ &= (\boldsymbol{\gamma} - a\mathbf{d})'\mathbf{J}(\boldsymbol{\gamma} - a\mathbf{d}) \\ &= \boldsymbol{\gamma}'\boldsymbol{J}\boldsymbol{\gamma} + a^2\mathbf{d}'\mathbf{J}\mathbf{d} - 2a\mathbf{d}'\mathbf{J}\boldsymbol{\gamma}, \end{aligned} \tag{2}$$

subject to $\boldsymbol{\gamma}'\boldsymbol{J}\boldsymbol{\gamma} = 1$, which is equivalent to requiring that the variance of the disparities is equal to $(\mathbf{1}'\mathbf{1})^{-1} = (nm)^{-1}$. We also see that the loss function is now defined by the sum of squares of the mean centered difference between $\boldsymbol{\gamma}$ and $a\mathbf{d}$.

The consequence of using a centering operator in the loss function is that the addition of a constant to the distances or the disparities does not change the value of L_{u}. This property implies that the loss function is not influenced by the mean of the distances nor of the disparities.

To see that maximizing the squared correlation in (1) is equivalent to minimizing the least-squares loss function in (2), consider the following. At a minimum of (2), it must be true that the partial derivative of (2) to a equals zero, that is,

$$\frac{\partial L_{\mathrm{u}}}{\partial a} = 2a\mathbf{d}'\mathbf{J}\mathbf{d} - 2\mathbf{d}'\mathbf{J}\boldsymbol{\gamma} = 0, \text{ so that } a = \frac{\mathbf{d}'\mathbf{J}\boldsymbol{\gamma}}{\mathbf{d}'\mathbf{J}\mathbf{d}}. \tag{3}$$

Substituting a into (2) and using $\boldsymbol{\gamma}'\boldsymbol{J}\boldsymbol{\gamma} = 1$ gives

$$\begin{aligned} L_{\mathrm{u}}(a, \mathbf{X}, \mathbf{Y}, \mathbf{\Gamma}) &= \boldsymbol{\gamma}'\boldsymbol{J}\boldsymbol{\gamma} + \left(\frac{\mathbf{d}'\mathbf{J}\boldsymbol{\gamma}}{\mathbf{d}'\mathbf{J}\mathbf{d}}\right)^2 \mathbf{d}'\mathbf{J}\mathbf{d} - 2\frac{\mathbf{d}'\mathbf{J}\boldsymbol{\gamma}}{\mathbf{d}'\mathbf{J}\mathbf{d}}\mathbf{d}'\mathbf{J}\boldsymbol{\gamma} \\ &= \boldsymbol{\gamma}'\boldsymbol{J}\boldsymbol{\gamma} + \frac{(\mathbf{d}'\mathbf{J}\boldsymbol{\gamma})^2}{\mathbf{d}'\mathbf{J}\mathbf{d}} - 2\frac{(\mathbf{d}'\mathbf{J}\boldsymbol{\gamma})^2}{\mathbf{d}'\mathbf{J}\mathbf{d}} \\ &= \boldsymbol{\gamma}'\boldsymbol{J}\boldsymbol{\gamma} - \frac{(\mathbf{d}'\mathbf{J}\boldsymbol{\gamma})^2}{\mathbf{d}'\mathbf{J}\mathbf{d}} = 1 - \frac{(\mathbf{d}'\mathbf{J}\boldsymbol{\gamma})^2}{\mathbf{d}'\mathbf{J}\mathbf{d}\ \boldsymbol{\gamma}'\mathbf{J}\boldsymbol{\gamma}} = 1 - r^2(\boldsymbol{\gamma}, \boldsymbol{d}), \end{aligned}$$

showing that minimizing (2) is equivalent to maximizing the squared correlation between the distances and the disparities.

The optimal value for a can be interpreted as the regression weight when regressing the disparities on the distances. Multiplication of the distances with a certain constant results in the division of a by the same value (this does not apply to the disparities as they have a fixed variance). The overall result, given also the use of a centering operator, is that *linear transformations*

of the distances or addition of a constant to the disparities do not change the value of the loss function. This can be directly seen when reformulating the loss function as one minus the squared correlation between the distances and disparities given the length restriction on the disparities.

In the *row-conditional* case, we will maximize the mean squared correlation where the mean is taken over subjects:

$$R^2(\boldsymbol{\gamma}, \mathbf{d}) = \frac{1}{n}\sum_{i=1}^{n} r^2(\boldsymbol{\gamma}_i, d_i), \tag{4}$$

with $\mathbf{d}_i$ and $\boldsymbol{\gamma}_i$ m-sized vectors with as elements the distances of subject i to the objects and the optimally scaled data for subject i respectively. The same steps can be followed as in the unconditional case to show that maximizing the mean squared correlation is the same as mininizing

$$\begin{aligned} L_{\mathrm{c}}(\mathbf{a}, \mathbf{X}, \mathbf{Y}, \boldsymbol{\Gamma}) &= \frac{1}{n}\sum_{i} \|\boldsymbol{\gamma}_i - a_i\mathbf{d}_i\|^2_{\mathbf{J}} \\ &= \frac{1}{n}\sum_{i} (\boldsymbol{\gamma}_i - a_i\mathbf{d}_i)'\mathbf{J}(\boldsymbol{\gamma}_i - a_i\mathbf{d}_i), \end{aligned} \tag{5}$$

subject to $\boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i = 1$ for all i. Note that $\mathbf{J} = \mathbf{I} - \mathbf{1}\mathbf{1}'(\mathbf{1}'\mathbf{1})^{-1}$ is the centering matrix of size $m \times m$ since centering has to be applied to each subject separately. The variance of the disparities should be m^{-1} for each subject. The optimal a_i is obtained by $a_i = \mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i(\mathbf{d}_i'\mathbf{J}\mathbf{d}_i)^{-1}$. The loss function is now invariant under linear transformations of subject-related distances or addition of a constant to subject-related disparities.

2.2. *Implicit and Explicit Normalization*

Kruskal and Carroll (1969) proved that minimizing Stress-2 is equivalent to optimizing a variety of measures, including the correlation, but their proof was difficult to follow. Here, we present a simple proof for the equivalence of minimizing Kruskal's Stress-2 to maximizing the squared correlation in the unconditional case. The square of Kruskal's Stress-2 can be expressed as

$$\sigma_2^2(\boldsymbol{\gamma}, \mathbf{d}) = \frac{(\boldsymbol{\gamma} - \boldsymbol{d})'\boldsymbol{J}(\boldsymbol{\gamma} - \boldsymbol{d})}{\mathbf{d}'\mathbf{J}\mathbf{d}}. \tag{6}$$

Let us assume that the transformation can be scaled by a factor $\alpha > 0$, that is, if $\boldsymbol{\gamma}$ is an admissible transformation, so is $\alpha\boldsymbol{\gamma}$. This property holds for all common transformations used in MDS and unfolding, such as ordinal and interval transformations. Then, if $\sigma_2^2(\boldsymbol{\gamma}, \boldsymbol{d})$ is at its minimum, we may also minimize $\sigma_2^2(\alpha\boldsymbol{\gamma}, \boldsymbol{d})$ over α. Now,

$$\sigma_2^2(\alpha\boldsymbol{\gamma}, \mathbf{d}) = \frac{(\alpha\boldsymbol{\gamma} - \mathbf{d})'\mathbf{J}(\alpha\boldsymbol{\gamma} - \mathbf{d})}{\mathbf{d}'\mathbf{J}\mathbf{d}} = \alpha^2\frac{\boldsymbol{\gamma}'\mathbf{J}\boldsymbol{\gamma}}{\mathbf{d}'\mathbf{J}\mathbf{d}} + 1 - 2\alpha\frac{\boldsymbol{\gamma}'\mathbf{J}\mathbf{d}}{\mathbf{d}'\mathbf{J}\mathbf{d}}. \tag{7}$$

Setting the first derivative of $\sigma_2^2(\alpha\boldsymbol{\gamma}, \mathbf{d})$ with respect to α equal to zero yields the optimal α^* of $\alpha^* = \boldsymbol{\gamma}'\mathbf{J}\mathbf{d}/\boldsymbol{\gamma}'\mathbf{J}\boldsymbol{\gamma}$. Inserting this result into (7) gives

$$\begin{aligned} \sigma_2^2(\alpha^*\boldsymbol{\gamma}, \mathbf{d}) &= \frac{\boldsymbol{\gamma}'\mathbf{J}\boldsymbol{\gamma}}{\mathbf{d}'\mathbf{J}\mathbf{d}}\frac{(\boldsymbol{\gamma}'\mathbf{J}\mathbf{d})^2}{(\boldsymbol{\gamma}'\mathbf{J}\boldsymbol{\gamma})^2} + 1 - 2\frac{\boldsymbol{\gamma}'\mathbf{J}\mathbf{d}}{\boldsymbol{\gamma}'\mathbf{J}\boldsymbol{\gamma}}\frac{\boldsymbol{\gamma}'\mathbf{J}\mathbf{d}}{\mathbf{d}'\mathbf{J}\mathbf{d}} \\ &= 1 - \frac{(\boldsymbol{\gamma}'\mathbf{J}\mathbf{d})^2}{\mathbf{d}'\mathbf{J}\mathbf{d}\boldsymbol{\gamma}'\mathbf{J}\boldsymbol{\gamma}} = 1 - r^2(\boldsymbol{\gamma}, \mathbf{d}). \end{aligned} \tag{8}$$

Thus, (8) proves that unconditional unfolding also optimizes the correlation between $\mathbf{d}$ and $\boldsymbol{\gamma}$. Note that the correlation is independent of the scale, so that the distances found by minimizing Stress-2 will be a factor smaller or larger than those obtained by using our method.

For row-conditional unfolding, we can show the same equivalence if the transformations are row conditional.

3. Interpreting the Objects-point Degeneracy

Experience teaches that unfolding algorithms based on Stress-2 or on the correlation often give partially degenerate solutions characterized by a few distant (degenerate) points. We illustrate this phenomenon here with data about preference for family structures. Delbeke (1968) collected data about preference for family structure from 102 students of the University of Leuven. Here, 82 subjects will be considered. The family structures are composed by combining the number of sons (zero to three) with the number of daughters (zero to three). These 16 stimuli had to be ordered by subjects according to their preference. The solution is obtained as the best solution resulting from the unfolding algorithm that maximizes the squared correlation with 1,000 random starting configurations. These different starting configurations were used to deal with the problem of local minima. For this dataset, optimal scaling was based on interval transformations. To ensure proper convergence of the algorithm, the iterations are stopped when the difference in subsequent loss function values was smaller than 10^{-9} or after 10,000 iterations. The reason for such strict convergence criteria is that a degenerate solution may need many iterations before it occurs.

The ideal point configuration is shown in Figure 2. What we observe are a few outlying subjects and a tight cluster of points consisting of objects and subjects. Degeneracy occurs only for some subjects, namely the distant subjects. Therefore, this solution is partially degenerate and seems hardly interpretable. Furthermore, as the object points fall together with respect to the degenerate subjects, this type of degeneracy is of the object point type.

Despite the interpretational problem, we take a look at some measures of fit. In Table 1, four measures of fit are reported: the squared correlation (the optimization criterion used), the recov-

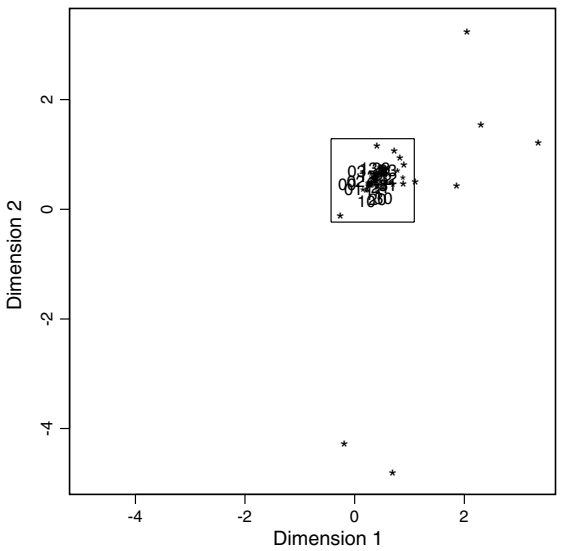


FIGURE 2.
Ideal point configuration for the Delbeke data.

TABLE 1.
Some measures of goodness-of-fit for the Delbeke data

Fit measure	Value
Average squared correlation	0.8630
Average recovered preference orders	0.9027
Average product–moment correlation	0.9277
Average Spearman rank correlation	0.9255

ered preference orders, the product–moment correlation and the Spearman rank correlation. The *recovered preference orders* measure the proportion of pairs of objects for which the preference order in the data is reproduced by the configuration. As the data are row-conditional, the measures are calculated for each subject separately and the mean value is reported.

All goodness-of-fit values in Table 1 indicate a good fit of the ideal point solution to the data. Such high fit values indicate that the configuration reproduces the data very well. Therefore, we argue in this paper that it is worth giving such a "degenerate" solution a closer look.

Zooming in on the cluster gives the configuration in Figure 3. Subjects are represented by stars and the stimuli by the two-number codes. For this part of the data, a clearly interpretable structure appears, which is given at the end of this section. To come to an interpretation of the outlying points in Figure 2 that were not shown in Figure 3, we use some results of De Leeuw (1983). He showed that nonmetric unfolding algorithms minimizing Kruskal's Stress-2 choose a path in the neighborhood of the object point degeneracy that optimizes the fit of the vector model. Note that the model we are fitting is an ideal point model. The results of De Leeuw are based on two theorems. The first theorem shows that in the neighborhood of a trivial solution Stress-2 converges to a finite value. In the neighborhood of the object point degeneracy, this is Stress-2 for the vector model. The second theorem, which is based on directional derivatives states that "we can find points arbitrarily close to a trivial solution with arbitrary small derivatives." These theorems are

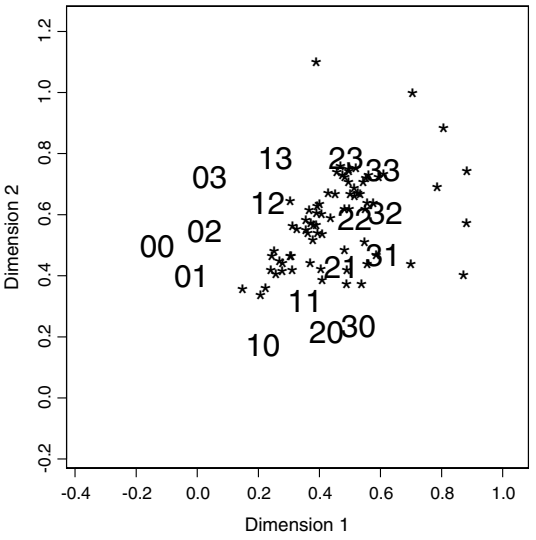


FIGURE 3.
Detail of the ideal point configuration for the Delbeke data. Subjects are indicated by stars, the stimuli by the two-number codes, where the first number indicates the number of sons and the second number the number of daughters.

proved by observing the behavior of unfolding algorithms in the neighborhood of trivial solutions: small perturbations δ are added to the trivial solution and Stress-2, or its directional derivative, and their limits for δ approaching zero were studied using l'Hospital's rule. So, what the result of De Leeuw tells us is that if the object point degeneracy occurs, then this solution is arbitrarily close to the local minimum of a vector model. A vector model is then fitted instead of an ideal point model. This holds when minimizing Stress-2 or when optimizing the squared correlation.

Let us now treat the vector model in more detail. The vector model goes back to Tucker (1960). In this model, subjects are represented by vectors and the reproduced data are constructed by the orthogonal projections of the objects on the subject vectors. Higher projections of objects on the subject vector indicate higher preference. The vector model is illustrated in Figure 4a. For the subject represented by the upward pointing vector, we see that the preference order is BEADC while this is ABCDE for the other subject. Carroll (1972) showed that the vector model is a special case of the ideal point model, namely, an ideal point model with the ideal points at infinity. This property for outlying ideal points can be understood by drawing its iso-preference contours: the contours become less curved and more like perpendicular lines (see Figure 4b).

In Figure 3, the distant subjects are not shown. To include those subjects, we make use of the vector model. First, we need to define an origin. Following Carroll (1972, p. 109), the origin is taken as the centroid of the object points. The vectors are then drawn from this origin in the direction of the ideal point representing the subject. We still need to define a rule to determine how distant a subject should be from the origin to represent it as a vector instead of an ideal point. Only subjects that are further from the origin than the most distant object are considered as candidate vectors. For these subjects, we propose the following rule: take the minimum distance, $d_{\min}$, for which the recovered preference based on a combination of ideal points (the points falling in or on the circle with radius $d_{\min}$) and vectors (the remaining points falling outside the circle with radius $d_{\min}$) has at least the same recovered preference orders as the original ideal point solution. The recovered preference orders for subjects represented by vectors are based on the vector model, that is, by orthogonally projecting the objects onto the vectors. To find this minimum distance, the following strategy can be followed. First, calculate the distances from the subjects (that are more

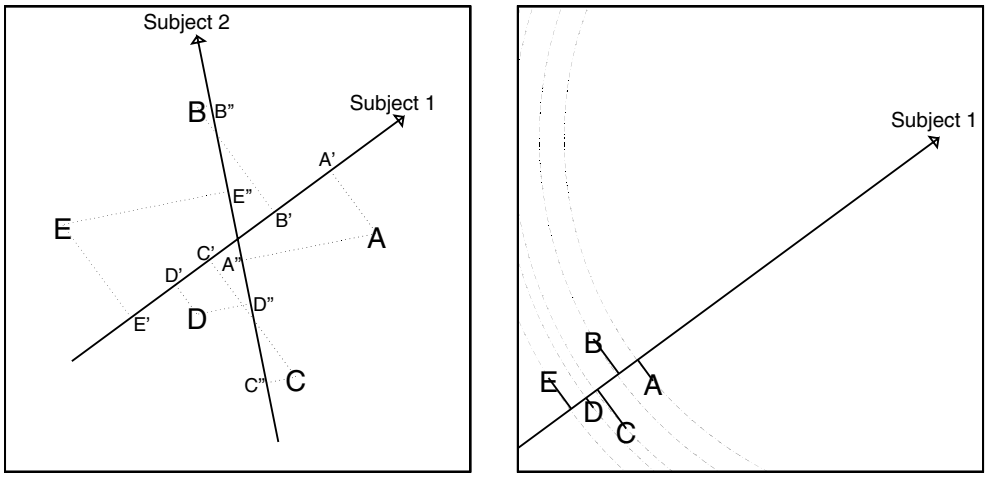


FIGURE 4.
Hypothetical example of the vector model for five objects and two subjects (a). Part b shows the ideal point representation where subject 1 is distant from the objects. Iso-preference contours are indicated by the dashed line, the projections according to the vector model by the solid lines.

distant than the most distant object) to the origin and order them from the smallest to largest. Set $d_{\min}$ equal to the smallest distance and check the rule. As long as the rule is not fulfilled, take the next distance in the sequence of ordered distances. A stepwise description of this search for vectors yields the strategy described below.

1. Measure the recovered preference order C_{IP} under the ideal point model of unfolding.
2. Set the origin $\mathbf{o}$ as the centroid of the objects, $\mathbf{o}' = m^{-1}\mathbf{1}'\mathbf{Y}$.
3. Calculate the distances of the objects to the origin $d_{jo} = \left(\sum_{b=1}^{p}(y_{jb} - o_b)^2\right)^{1/2}$ and of the subjects to the origin $d_{io} = \left(\sum_{b=1}^{p}(x_{ib} - o_b)^2\right)^{1/2}$.
4. Consider the subject points for which $d_{io} > \max_j d_{jo}$ and order their distances d_{io} from the smallest to largest. These ordered distances are denoted by $d_{(k)}$, $k = 1, ..., q$ and $q \leq n$.
5. Set $l := 1$
6. Calculate the recovered preference order C_{VIP} where the reproduced preference ordering for subjects with $d_{io} < d_{(l)}$ is based on the ideal point model and with $d_{io} \geq d_{(l)}$ on the vector model.
7. If $C_{\text{VIP}} < C_{\text{IP}}$, then $l := l + 1$; go to 6, else represent the subjects corresponding to $d_{(k)}$, $k = l, ..., q$ by vectors with length $d_{(l)}$ and terminate.

Applying this strategy to the Delbeke data shows that the eight most distant subjects can be represented by vectors without a decrease in the recovered preference orders. For this combination of vectors and ideal points, the recovered preference orders equal 0.9027. Note that only six of these eight subjects are located outside the square in Figure 2. The reason is that the bounding region that separates vectors from ideal points used in our strategy is circular, not square. The combined plot, called a vector-ideal point configuration is shown in Figure 5. A simultaneous interpretation of all objects and subjects is now possible. Considering the objects first, we see that the number of boys increases rightwards while the number of girls increases upwards. Four regions can be distinguished: a region of preference for more girls than boys (upper left), a region of preference for more boys than girls (lower right), a region of preference for few children (lower left), and a region of preference for many children (upper right). For most of the subjects, the interpretation is based on the ideal point model, but for eight subjects the interpretation is based

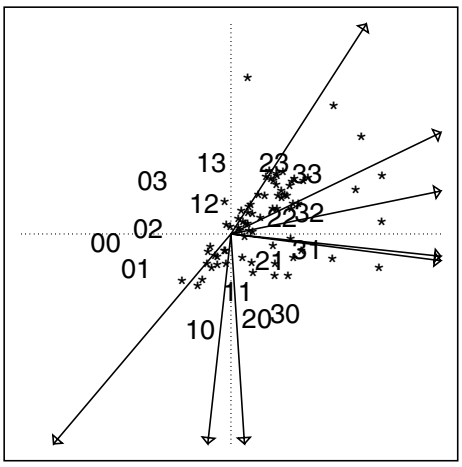


FIGURE 5.
Combined vector-ideal point configuration for the Delbeke data: the eight distant subjects are represented by vectors while the remaining subjects are indicated by stars.

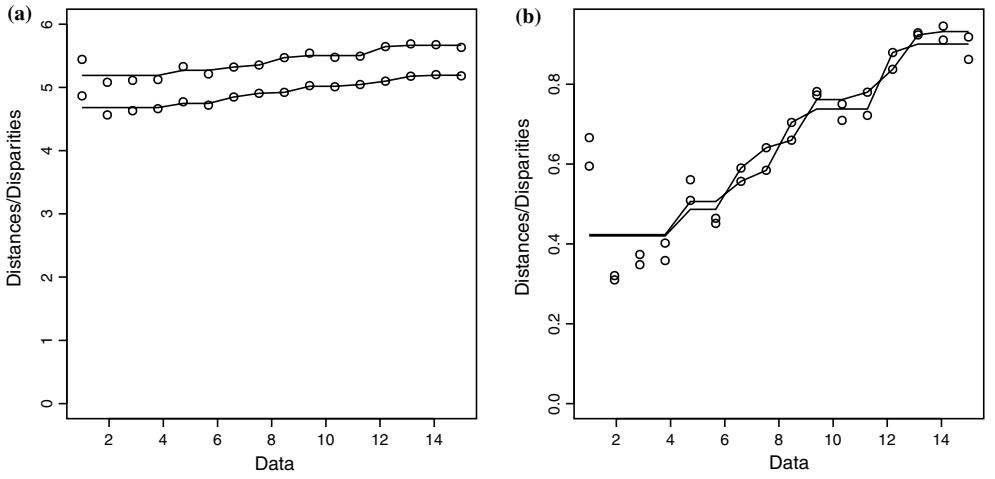


FIGURE 6.
Shepard plots of the distances (open circle) and disparities (lines) as a function of the data for the two most distant subjects. The left panel contains the Shepard plot for the ideal point configuration and the right panel for the combined vector-ideal point configuration.

on the vector model of unfolding. The preference orderings of the subjects represented as ideal points (the stars in Figure 5) are reflected in the order of the distances of the subject to the stimuli: the more distant the stimulus, the less the subject prefers the stimulus. Most subjects seem to prefer three to six children, consisting of an (almost) equal number of sons and daughters. For the subjects represented as vectors, preference is reflected in the orthogonal projections of the stimuli on the vector: three subjects have preference for many children (upper right region), three for an increasing number of sons (lower right region) and two for few children (lower left region). We believe that the combined vector-ideal point configuration in Figure 5 displays the unfolding solution in a non-degenerate manner that is easy to understand. In the ideal point configuration, there was almost no variation in the reconstructed distances for the distant subjects. This is no longer the case in the combined plot. The effect of representing the distant subjects as vectors on the variability in the reconstructed distances, as perceived, is illustrated in Figure 6 for the two most distant subjects. Both distances and disparities are plotted against the data in a so-called Shepard plot. In Figure 6a, these are the distances from the ideal point to the family structure points, while in Figure 6b these are the distances of the endpoint of the vector to the projection of the object point on the vector. Disparities are in both cases found by a monotone regression of the distances on the data.

4. Interpreting the objects-circle degeneracy

A second type of degeneracy that is often encountered when using Stress-2 or the squared correlation is the objects-circle. In this type of degeneracy, the objects lie on a circle and most of the subjects fall together in the center of the circle. We illustrate this type with results obtained by our algorithm for maximizing the squared correlation using ordinal transformations. The data come from Green and Rao (1972), who asked 21 Wharton School MBA students and their wives to rank 15 breakfast items according to their overall preference. The solution we will discuss below was obtained by selecting the best solution out of 1000 based on random starting configurations. We used the same convergence criteria as in Section 3.

The ideal point configuration is plotted in the left panel of Figure 7. Again, we observe a tight cluster of most of the subjects and the objects and a few outlying subjects. At first sight, we

seem to be dealing with a partial degenerate solution of the object point type as in the previous section. However, zooming in on the cluster reveals a configuration where all the objects lie on a circle (see the right panel of Figure 7) and 30 subjects fall together in the center of the circle (the fat star in the center). For these subjects, we are dealing with a degeneracy of the objects-circle type. In spite of the degeneracy, the measures in Table 2 indicate a reasonable or even good fit of the overall solution. Note that both the squared correlation and Stress-2 are undefined when the objects lie perfectly on a circle and at least one subject lies exactly in the center. Therefore, it cannot be the case that the subjects in Figure 7 lie exactly in the center and that the objects lie perfectly on a circle.

The ideal point solution in the left panel of Figure 7 takes the form of an object point degeneracy with a few distant subjects. As explained previously, the vector model should hold for these subjects, so we can apply the vector-seeking strategy developed in the previous section. However, for the subjects in the center of the circle, we use another result of De Leeuw (1983). He proved for Stress-2 that in the neighborhood of an objects-circle trivial solution, the unfolding algorithm fits a model closely related to the compensatory distance model.

The compensatory distance model was originally proposed by Coombs (1964), see also

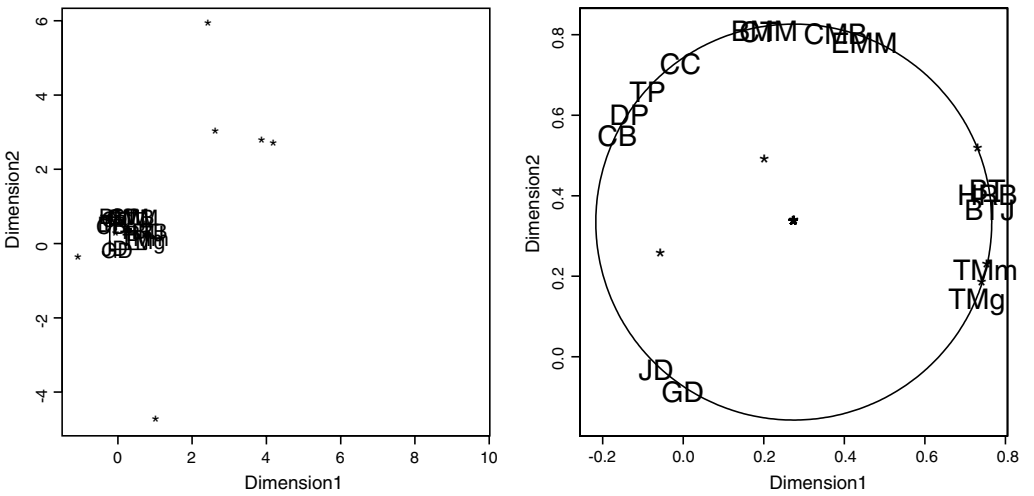


FIGURE 7.
Ideal point configuration for the breakfast data (left panel) and a detail (right panel). Subjects are indicated by stars, the stimuli by the letter codes: toast pop-up (TP), buttered toast (BT), English muffin and margarine (EMM), jelly donut (JD), cinnamon toast (CT), blueberry muffin and margarine (BMM), hard rolls and butter (HRB), toast and marmalade (TMm), buttered toast and jelly (BTJ), toast and margarine (TMg), cinnamon bun (CB), Danish pastry (DP), glazed donut (GD), coffee cake (CC), and corn muffin and butter (CMB).

TABLE 2.
Some measures of goodness-of-fit for the breakfast data

Fit measure	Value
Average squared correlation	0.8732
Average recovered preference orders	0.7231
Average product–moment correlation	0.6262
Average Spearman rank correlation	0.6186

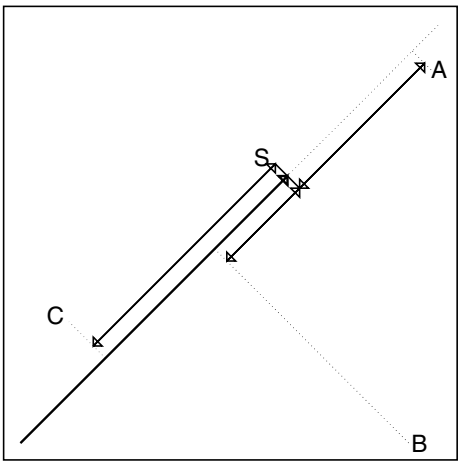


FIGURE 8.
Illustrative compensatory distance model configuration. S is a subject and the remaining points are objects.

Roskam (1968). In this model, distances are calculated by

$$d_{ij} = \left|\mathbf{x}_i^{'}(\mathbf{y}_j - \mathbf{x}_i)\right| = \left|\mathbf{x}_i^{'}\mathbf{y}_j - \mathbf{x}_i'\mathbf{x}_i\right|. \tag{9}$$

Equation (9) shows that the distance from an ideal point to the object is given by the distance from the ideal point to the orthogonal projection of the object on the subject-vector. An illustration of the compensatory distance model is given in Figure 8. Objects A, B, and C are projected on the subject vector and the distances of the projections to the ideal point S are considered. For this subject, the preference order given by the compensatory distance model is BAC.

De Leeuw (1983) proved for Stress-2 that in the neighborhood of an objects-circle trivial solution, the unfolding algorithm fits a model with signed distances given by

$$z_{ij} = -\mathbf{y}_j^{(0)\prime}(\mathbf{x}_i - \mathbf{y}_j), \tag{10}$$

where the $\mathbf{y}_j^{(0)}$ vectors refer to the trivial solution and the $\mathbf{y}_j$ vectors to the degenerate solution. From the results of De Leeuw (1983), we know that the object points in the degenerate solution come arbitrarily close to those in the trivial solution. Note that by definition all $\mathbf{y}_j^{(0)}$ have the same length, so that $\|\mathbf{y}_j^{(0)}\| = c$ for all j. However, it is not so that the lengths $\|\mathbf{y}_j\|$ are equal as the object points in the degenerate solution do not lie exactly on the circle.

The model in (10) differs in two respects from the original compensatory distance model of (9). First, the absolute signs are absent. Second, the role of the subjects and the objects is reversed. De Leeuw (1983) called (10) a signed version of the compensatory distance model. We will refer to this model as the signed distance model. Equation (10) can be rewritten as

$$z_{ij} = \mathbf{y}_j^{(0)\prime}\mathbf{y}_j - \mathbf{y}_j^{(0)\prime}\mathbf{x}_i. \tag{11}$$

To arrive at a representation that allows a better interpretation, we aim at representations where signed distances reflect the preference order. Monotone transformations of the distances are then allowed. For a sound interpretation of the signed distances, we need one more step. Without loss of generality, the distances z_{ij} in (10) can be divided by the norm of the $\mathbf{y}_j^{(0)}$ vectors,

that is,

$$\frac{z_{ij}}{\left\|\mathbf{y}_j^{(0)}\right\|} = \frac{\mathbf{y}_j^{(0)\prime}\mathbf{y}_j}{\left\|\mathbf{y}_j^{(0)}\right\|} - \frac{\mathbf{y}_j^{(0)\prime}\mathbf{x}_i}{\left\|\mathbf{y}_j^{(0)}\right\|} \approx \left\|\mathbf{y}_j\right\| - \frac{\mathbf{y}_j^{(0)\prime}\mathbf{x}_i}{\left\|\mathbf{y}_j^{(0)}\right\|}, \tag{12}$$

since $\|\mathbf{y}_j^{(0)}\| = c$ for all j by definition. The approximation sign is used in (12) because we assume that $\mathbf{y}_j^{(0)}$ has the same direction as $\mathbf{y}_j$ when $\mathbf{y}_j$ comes arbitrarily close to $\mathbf{y}_j^{(0)}$. Recall that the scalar product of two vectors divided by the norm of one of these vectors gives the length of the orthogonal projection. Signed distances in this version of the compensatory distance model are then calculated as the length of the $\mathbf{y}_j$ vector minus the length of the projection of the subject point on the object-vector. Note that (12) allows signed distances z_{ij} to be negative when the projection of the subject falls higher than the endpoint of the object vector. Clearly, this case does not occur in the objects-circle degeneracy.

An illustration of the signed distance model is given in Figure 9 for a subject S and three objects A, B, and C. The subject is represented by the letter S while the objects are represented by the three lines. These lines consist of three parts: a square, a dotted line segment, and a dashed line segment. The object point is indicated by the square. The signed distances are found by projecting the subject point onto the object vectors and then considering the distance of the projection to the square. This signed distance is positive when the projection falls on the dotted part and negative otherwise. A negative signed distance means higher preference than a positive distance. Therefore, the dashed lines are labeled by the "positive" object point labels and the dotted lines by the negative labels. So, the larger the distance from the projection on the dashed line to the object point, the more the object is preferred. Conversely, the larger the distance of the projection on the dotted line to the object point, the less the object is preferred. In Figure 9, it can be seen that the recovered preference rank order for subject S is ABC where the signed distance for object A is negative. To find the projection of a subject, we may draw a circle as in Figure 9. The center of the circle lies halfway on the line connecting the subject point to the origin and it has a radius such that the origin and the subject point lie on the circle. More formally, the orthogonal projections fall on a circle of radius $0.5\,\|\mathbf{x}_i\|$ with origin $0.5\mathbf{x}_i$. This can be proven using the sine and cosine rules for the double angle.

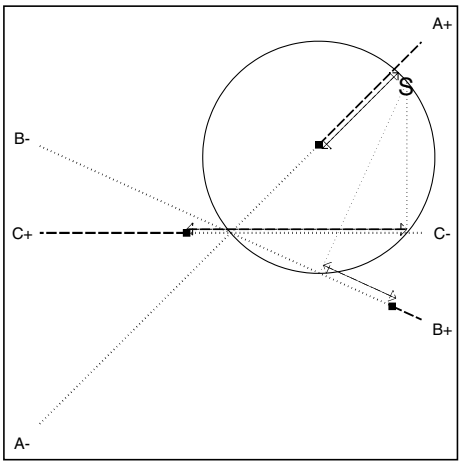


FIGURE 9.
Illustrative signed distance model configuration. S is a subject, the objects A, B, and C are represented by vectors.

We briefly go into the special case that all objects lie perfectly on a circle. Then, the vector model holds. This special case can be seen by noting that if all $\mathbf{y}_j$ have equal length, $\|\mathbf{y}_j\| = \|\mathbf{y}_j^{(0)}\| = c$, then (11) can be written as

$$z_{ij} = c - \mathbf{y}_j^{(0)\prime}\mathbf{x}_i. \tag{13}$$

The right-hand side term of (13) can be read as a constant minus the projection of the objects on the subject vector. The higher the projection falls on the subject vector, the smaller z_{ij} and thus the more the subject prefers the object. This representation is again of the vector model type. However, in practice we have never experienced that the objects fall so well on a circle that a vector representation holds. Note that if this exceptional case would occur, then no subject will lie exactly in the center as the correlation (and Stress-2) is then undefined. As the algorithm fits an ideal point model, in all these cases and others (for example, the complete sample is fitted by the signed compensatory distance model alone), the metric axioms are still satisfied.

For the interpretation of the degenerate subjects in (extremely near to) the center of the objects-circle degeneracy, we rely on the signed distance model. As we want the distances to reflect only the preference order, monotone transformations and thus linear transformations (with a positive regression weight) of the distances are allowed. Therefore, we shall allow an extension of the signed distance model by subtraction of a constant v, that is,

$$\frac{z_{ij}}{\left\|\mathbf{y}_j^{(0)}\right\|} - v \approx \left(\left\|\mathbf{y}_j\right\| - v\right) - \frac{\mathbf{y}_j^{(0)\prime}\mathbf{x}_i}{\left\|\mathbf{y}_j^{(0)}\right\|}. \tag{14}$$

This adaption means that all object vectors are shortened by the same length v. The value of the recovered preference orders remains unchanged because rank orders are invariant to the addition of a constant. This extension has the advantage that the interpretation becomes easier: by shortening the $\mathbf{y}_j$ vectors considerably, their difference in length will become visible.

The interpretation of Figure 7 is done in several steps. First, the outlying points (which are of the object-point degeneracy type) are represented by vectors. The same vector- seeking strategy is followed as for the Delbeke data but with the center of the circle as the origin. From the results of De Leeuw (1983), we know that this is the origin of the signed distance model which should hold for the objects-circle degeneracy. The center of the circle is found by use of the IM procedure explained in Appendix B.

To identify the subject points that can be represented by a vector in Figure 7, we apply the vector-seeking strategy of Section 3. This strategy indicates that all subjects lying outside the objects circle can be represented by vectors without diminishing the average recovered preference order. The recovered preference order for this combination of vectors and ideal points equals 0.7231, the same value as the recovered preference order of the full ideal point representation (0.7231). Figure 10 shows the solution for these seven subjects as vectors together with the five subjects that do not lie in the center and are not outlying. This is a combined vector-ideal point configuration that is interpreted as explained in Section 3. With respect to the breakfasts, three clusters can be discerned: a cluster of toast items, a cluster of the donuts, and a cluster of soft items (and TP). The interpretation for subjects represented by a vector is based on the vector model. For example, the four vectors pointing upwards in the upper-right part of Figure 10 represent subjects with a preference for the items with butter or margarine. The preference of subjects represented by a star should be interpreted according to the ideal point model. The subject lying between TMm and TMg, for example, has preference for the toast items (except for TP).

To interpret the preference of the subjects lying in or very near to the center of the circle, the signed version of the compensatory distance model is used. First, it is checked if the signed distance representation indeed yields the same preference rank orders as the objects circle ideal

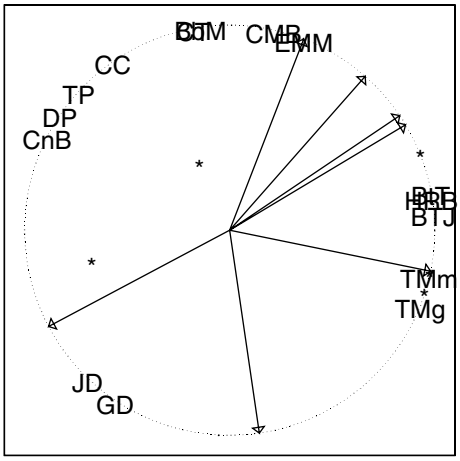


FIGURE 10.
Combined vector-ideal point solution for the 12 subjects not falling in the center of the objects circle.

point solution. In this case, the recovered preference order was different for each of the three types of subjects for the seven outlying subjects, the vector model was used; for five subjects, the ideal point model was used; and for the remaining subjects, the signed distance model was used. The total recovered preference order was the same as under the ideal point model (0.7231). Next, the object vectors are all shortened with the same length. Here, this length is chosen so that the least distant object falls as far from the origin as the least distant subject. The resulting configuration is given in Figures 11a and 11b.

We propose to interpret the signed distance configuration in two steps. In a first step, the object points and the subject points are considered. Figure 11a shows these object points (the squares) and the ideal points (the stars). We first consider the breakfast items. Toast pop-up (TP) has a remarkable position: it is an outlying point with respect to the other breakfast items and it

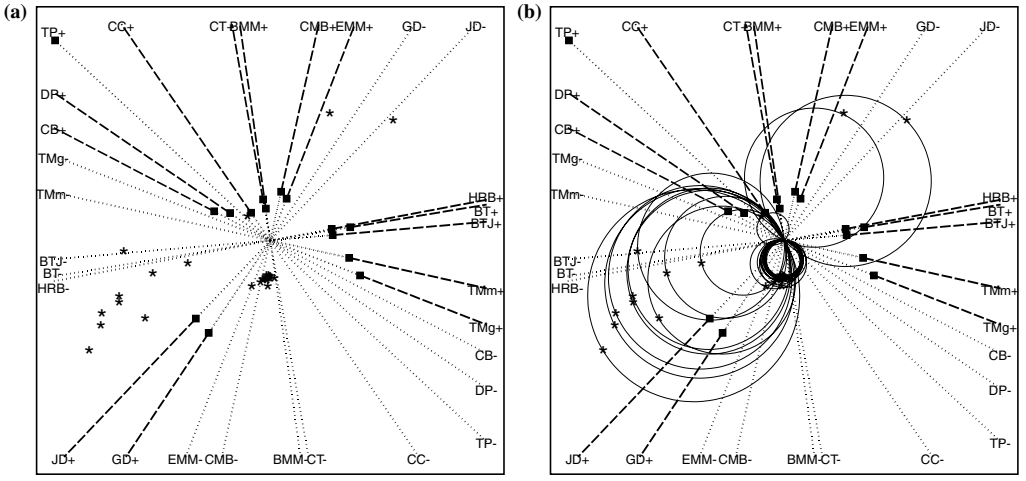


FIGURE 11.
Signed compensatory distance model solution. Distances can be found by projecting the subjects on the object vectors and considering the distance of the projection to the object. These distances are negative if the projection falls on the dashed line. In the right panel, circles for each subject are plotted to find the projection of a subject onto an object vector.

lies opposite to the other toast items. Three clusters can be observed. A cluster of the hard items (with the exception of cinnamon toast), a cluster of the soft items, and a cluster of the donuts. Next, we turn our attention to the ideal points. Here too, we find a few clusters, with one being very tight. Clustered subjects will have similar preference. In the next step, the preference structure of the subjects is considered. This structure is given by the individual projection circles in Figure 11b. The preference ordering is reflected in the distance of the intersections of the circle with the object vector to the object point. Intersections with a dotted part will give positive distances (lower preference), while these will be negative for the dashed part of the vector (higher preference). So, a distinction between more and less liked items can be made quickly. For example, the subjects in the lower left part of Figure 11b have more preference for donuts (JD and GD), cinnamon buns (CB), danish pastries (DP), and coffee cake (CC) than for the toast items (BT, BTJ, TMm, TMg, TP, CT), hard rolls (HRB), and blueberry muffin (BMM). It seems that they have more preference for the soft than for the hard items. A finer discrimination can be made by comparing the distances of the projection to the square (the object point), keeping in mind that larger distances on dashed lines mean higher preference while the opposite is true for the dotted parts.

The effect of representing the distant subjects as projection circles on the perceived variability in the reconstructed distances is illustrated in Figure 12 for two central subjects. Here, both distances and disparities are plotted against the data in a Shepard plot. In Figure 12a, these are the distances from the ideal point to the breakfast points while these are the signed distances of the breakfast point to the projection of the subject on the vector in Figure 12b. Disparities are, in both cases, found by a monotone regression of the distances on the data.

In Figure 11, fifteen subjects are clustered together and except that their projection falls far from the TP point, it is difficult to see what their preference order would be for the other 14 breakfasts. For this reason, a separate configuration is constructed where the object vectors are shortened even more (see Figure 13). As most breakfast points in Figure 11a fall opposite the cluster of subjects, it would be best (for reasons of visibility) to transform the signed distances z_{ij} so that the transformed object points lie as much as possible on the same side as the subject points. By taking v in equation (14) larger than $\min_j(\|\mathbf{y}_j\|)$, this was realized in Figure 13. In the plot, this transformed 'negative length' is accounted for by extending the dashed lines. The

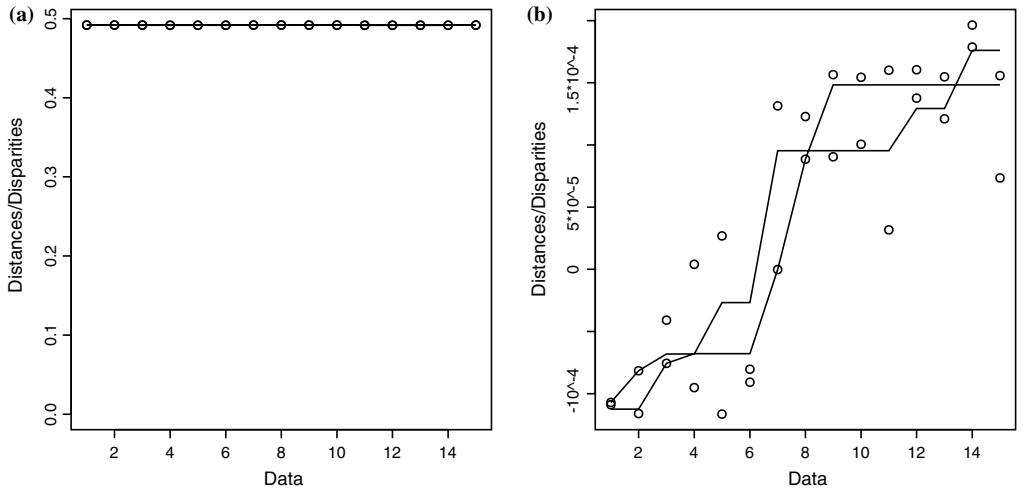


FIGURE 12.
Shepard plots of the distances (open circle) and disparities (lines) as a function of the data for two central subjects. The left panel contains the Shepard plot for the ideal point configuration and the right panel for the signed distance configuration.

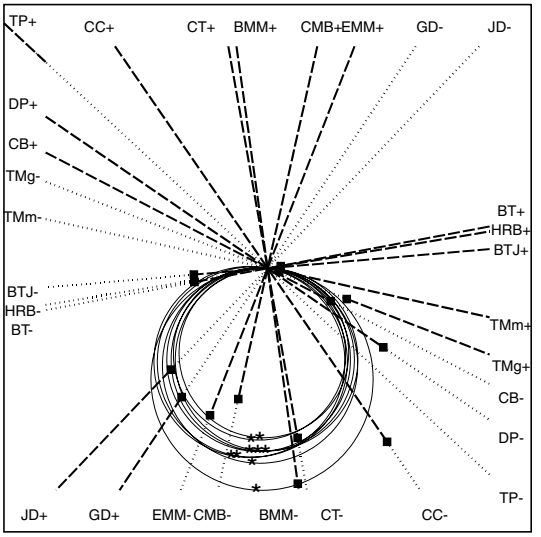


FIGURE 13.
Signed compensatory distance model solution enlarged for a cluster of subjects. As in Figure 10, distances can be found by projecting the subjects on the object vectors and by considering the distance of the projection to the object. These distances are negative if the projection falls on the dashed line.

interpretation remains as before: the distance of a projection that falls on a dashed line segment is negative (higher preference), while it is positive when the projection falls on a dotted segment (lower preference). Negative distances occur for the items CC, BTJ, DP, BMM, BT, TMm, GD, JD, and CB with the largest distances for most of the subjects for the first items (most preferred items). The remaining distances are all positive, with the largest distances for most of the subjects for TP, HRB, CMB, TMg, EMM, and CT (least preferred items). A substantive interpretation for these subjects is hard to find.

5. Discussion

In the unfolding literature, degenerate solutions have been looked upon as uninformative. In this paper, though, we showed that the frequently occurring object point and objects-circle degeneracies are not only well fitting, but also subject to a sound interpretation. The relations between the object-point degeneracy and the vector model, and between the objects-circle degeneracy and the signed distance model have been used to construct new representations that reveal the information contained in the degenerate solutions. The case of interval scaling has been illustrated with an empirical data set that yielded a degeneracy of the object-point type. For this degeneracy, a combined vector-ideal point configuration was constructed. A second empirical data set was used to illustrate the case of ordinal scaling. The resulting objects-circle degeneracy was made interpretable by constructing a representation that uses signed distances.

The results discussed here hold not only for degeneracies obtained when using Stress-2 as a loss function, but also when optimizing the squared correlation. We gave a simple proof for the equivalence of this optimization criterion with Stress-2, both in a row-conditional and an unconditional approach. Furthermore, we developed an iterative unfolding algorithm that guarantees a nondecreasing sequence of the squared correlation.

In this paper, we illustrate the well-known fact that degenerate solutions cannot be avoided by normalizing on the variance, as was shown by De Leeuw (1983) for ordinal transformations. The present paper also includes interval transformations. As to the occurrence of degenerate solutions

with ordinal and interval transformations, we refer to Busing, Groenen, andHeiser (in press): there it is proved for raw Stress and Stress-1 that the presence of an intercept in the transformation is a necessary and sufficient condition for a degeneracy in unfolding.

The object-point degeneracy raises questions about the nature of the outlying subjects. These distant points have a large impact on the way the configuration looks. Therefore, one could consider them not only as outliers but also as influential data points that are better discarded from the unfolding analysis. Two remarks can be made here. First, due to the scaling invariance, the loss function is not more sensitive to these outlying subjects in comparison with the more centrally located subjects (in contrast to known problems in regression analysis with respect to outliers) and second, discarding them from the analysis does not guarantee that no other subjects become distant. In fact, the results of De Leeuw (1983) indicate that this phenomenon is likely to happen. Some experimentation with empirical data has confirmed the occurrence of this phenomenon.

In this paper, the representation of a subject as an ideal point, a vector, or a projection circle was based on the degenerate unfolding solution. However, it could be of interest to explicitly model a mixed representation, and to develop an algorithm that "could accommodate mixtures of vector and unfolding (ideal point) representations" (see DeSarbo & Carroll, 1985). In this way, no arbitrary search strategy is needed to determine if a subject is to be represented as a vector instead of an ideal point. In theory, even for very distant subjects it could be the case that the recovered preference orders are worse under a vector representation then under an ideal point representation. Then, the approach proposed here would not work. By explicitly modeling mixed representations, this situation can be avoided. We intend to investigate these possibilities in future research. If the objective is to obtain a solution where the data are interpreted by distances between ideal points and object points, then another approach is called for such as the penalty approach of Busing et al. (in press). Of course, unfolding methods that use ratio transformations can also be used, such as Kim, Rangaswamy, and DeSarbo (1999). A comparison of these approaches can be found in Busing et al. (in press).

Appendix A. A Convergent Algorithm to Maximize the Squared Correlation

We now describe a monotonically convergent algorithm to maximize the squared correlation. In each iteration t, the disparities $\mathbf{\Gamma}$ and the coordinates $\mathbf{X}$ and $\mathbf{Y}$ are updated. Below, we present the algorithm. Let $\mathbf{d}^{(l)}$ denote the vector of distances between the coordinates of $\mathbf{X}^{(l)}$ and $\mathbf{Y}^{(l)}$ at iteration l. Also, let $L^{(l)}$ abbreviate $L(a^{(l)}, \mathbf{X}^{(l)}, \mathbf{Y}^{(l)}, \mathbf{\Gamma}^{(l)})$. Note that if row-conditional transformations are used, then $L = L_{\mathrm{c}}$, for the unconditional case $L = L_{\mathrm{u}}$. Now, the algorithm may be expressed schematically as follows.

1. Choose an initial configuration $\mathbf{X}^{(0)}$, $\mathbf{Y}^{(0)}$.
2. Compute the transformation update $\mathbf{\Gamma}^{(0)}$, given the data and given $\mathbf{X}^{(0)}$ and $\mathbf{Y}^{(0)}$.
3. Calculate the regression weight $a^{(0)}$.
4. Set $L^{(0)} = 1 - r^2(\boldsymbol{\gamma}, \mathbf{d})$.
5. Set $l := 0$.
6. $l := l + 1$.
7. Update the coordinates $\mathbf{X}^{(l)}$, $\mathbf{Y}^{(l)}$ given $\mathbf{\Gamma}^{(l-1)}$ and $a^{(l-1)}$.
8. Compute the transformation update $\mathbf{\Gamma}^{(l)}$ given $\mathbf{X}^{(l)}$, $\mathbf{Y}^{(l)}$, and $a^{(l-1)}$.
9. Calculate $a^{(l)}$.
10. If $L^{(l)} - L^{(l-1)} < \epsilon$ or $l = l_{\max}$ stop.
11. Go to 6.

Note that Step 8 can be skipped in case of interval or ratio scaling.

5.1. Updating the Coordinates using Iterative Majorization

We will describe here how the coordinates can be updated, given the regression weight and the disparities using IM. For a general description of IM in the context of multidimensional scaling, we refer to De Leeuw and Heiser (1977) and Borg and Groenen (1997). Extensions needed in this paper can be found in Busing et al. (in press). The basic principle of IM is to minimize a function by the use of an auxiliary function, the majorizing function. This function should be easier to minimize, should touch the original function at the current estimate, and should never be lower than the original function. From these properties, it can be derived that the loss function is nonincreasing at each update. In our case, we will have quadratic majorizing functions.

5.1.1. The Unconditional Case

The way in which we will proceed is to find a quadratic majorizing function that can be minimized using standard derivation techniques (De Leeuw, 1994; Appendix B in Groenen, Heiser, & Meulman, 1999). The minimum of this quadratic form can be found analytically. As the results here are based on developments in the framework of multidimensional scaling, we need to reformulate the unfolding problem as a special case of multidimensional scaling.

We start with the following formulation of the loss function in terms of the vectors of distances $\mathbf{d}$ and disparities $\boldsymbol{\gamma}$,

$$L_{\mathrm{u}}(a, \mathbf{X}, \mathbf{Y}, \boldsymbol{\Gamma}) = \boldsymbol{\gamma}'\mathbf{J}\boldsymbol{\gamma} + a^2\mathbf{d}'\mathbf{J}\mathbf{d} - 2a\mathbf{d}'\mathbf{J}\boldsymbol{\gamma}. \tag{15}$$

$\mathbf{d}$ is the vector of distances d_{ij} and $\mathbf{J} = \mathbf{W} - (\mathbf{w}'\mathbf{w})^{-1}\mathbf{w}\mathbf{w}'$ is a weighted centering operator. The matrix $\mathbf{W}$ is a $(n+m)^2 \times (n+m)^2$ diagonal matrix with the vector $\mathbf{w}$ on the diagonal. The weights w_{ij} equal zero when the corresponding dissimilarity is missing and one otherwise. Let $\mathbf{Q} = [\mathbf{X}'\ \mathbf{Y}']'$ be the matrix of coordinates. Note that $\mathbf{d}$ contains both within set distances (distances between an ideal point and another one or between an object point and another object point) and between set distances (distances between an ideal point and an object point). The weights of within set objects are zero. In our derivation, we represent terms that do not include the distances (or the coordinates) by an indexed constant k. Replacing $\mathbf{J}$ by its components gives

$$\begin{aligned} L_u(a, \mathbf{X}, \mathbf{Y}, \boldsymbol{\Gamma}) &= a^2\mathbf{d}'(\mathbf{W} - (\mathbf{w}'\mathbf{w})^{-1}\mathbf{w}\mathbf{w}')\mathbf{d} - 2a\mathbf{d}'(\mathbf{W} - (\mathbf{w}'\mathbf{w})^{-1}\mathbf{w}\mathbf{w}')\boldsymbol{\gamma} + k_1 \\ &= a^2\mathbf{d}'\mathbf{W}\mathbf{d} - a^2(\mathbf{w}'\mathbf{w})^{-1}\mathbf{d}'\mathbf{w}\mathbf{w}'\mathbf{d} - 2a\mathbf{d}'\mathbf{W}\boldsymbol{\gamma} \\ &\quad + 2a(\mathbf{w}'\mathbf{w})^{-1}\mathbf{d}'\mathbf{w}\mathbf{w}'\boldsymbol{\gamma} + k_1, \end{aligned} \tag{16}$$

with $k_1 = \boldsymbol{\gamma}'\mathbf{J}\boldsymbol{\gamma}$. Equation (16) contains terms of squared distances, minus squared distances, minus the distances, the distances, and a constant term. The majorizing function that we are searching for is of the form

$$\mathbf{d}'\mathbf{F}\mathbf{d} - 2\mathbf{d}'\mathbf{g} + k \tag{17}$$

where $\mathbf{F}$ is diagonal and both $\mathbf{F}$ and $\mathbf{g}$ are nonnegative. Note that (17) has the standard form of unfolding and multidimensional scaling problems minimizing Stress (De Leeuw, 1977), for which an update is readily obtained. However, (16) involves two more terms than this standard form. These two terms are the one involving minus the squared distances $-a^2(\mathbf{w}'\mathbf{w})^{-1}\mathbf{d}'\mathbf{w}\mathbf{w}'\mathbf{d}$ and the one representing plus a linear function of the distances $2a(\mathbf{w}'\mathbf{w})^{-1}\mathbf{d}'\mathbf{w}\mathbf{w}'\boldsymbol{\gamma}$.

Let $\mathbf{d}_0$ denote the distances obtained from the previous iteration and $\mathbf{R}$ the corresponding matrix of coordinates. We develop a majorizing function for both $-a^2(\mathbf{w}'\mathbf{w})^{-1}\mathbf{d}'\mathbf{w}\mathbf{w}'\mathbf{d}$ and $2a(\mathbf{w}'\mathbf{w})^{-1}\mathbf{d}'\mathbf{w}\mathbf{w}'\boldsymbol{\gamma}$ using the inequality

$$c(\mathbf{d} - \mathbf{d}_0)'\mathbf{P}(\mathbf{d} - \mathbf{d}_0) \geq 0 \tag{18}$$

for $c \geq 0$ and $\mathbf{P}$ being positive semi-definite. Note that for some majorizing inequalities, we also use $d_{ij} \geq 0$, $\boldsymbol{\gamma}_{ij} \geq 0$ and $a \geq 0$.

First, we majorize $-a^2(\mathbf{w}'\mathbf{w})^{-1}\mathbf{d}'\mathbf{w}\mathbf{w}'\mathbf{d}$. Application of (18), substituting c with $a^2(\mathbf{w}'\mathbf{w})^{-1}$ and $\mathbf{P}$ with $\mathbf{w}\mathbf{w}'$, leads to a linear majorizing function of a minus quadratic function (Busing et al., in press, p. 21), that is,

$$-a^2(\mathbf{w}'\mathbf{w})^{-1}\mathbf{d}'\mathbf{w}\mathbf{w}'\mathbf{d} \leq a^2(\mathbf{w}'\mathbf{w})^{-1}(\mathbf{d}'_0\mathbf{w}\mathbf{w}'\mathbf{d}_0 - 2\mathbf{d}'\mathbf{w}\mathbf{w}'\mathbf{d}_0). \tag{19}$$

We continue by majorizing $2a(\mathbf{w}'\mathbf{w})^{-1}\mathbf{d}'\mathbf{w}\mathbf{w}'\boldsymbol{\gamma}$. Equation (18) written in scalar form gives $c(d_{ij} - d_{ij0})^2 \geq 0$ or $cd_{ij} \leq (\frac{c}{2})(d_{ij}^2 d_{ij0}^{-1} + d_{ij0})$. Recognizing that $\mathbf{w}'\boldsymbol{\gamma}$ is a scalar and setting $c = 2a(\mathbf{w}'\mathbf{w})^{-1}(\mathbf{w}'\boldsymbol{\gamma})$, we see that $2a(\mathbf{w}'\mathbf{w})^{-1}\mathbf{d}'\mathbf{w}\mathbf{w}'\boldsymbol{\gamma}$ is a weighted sum of cd_{ij}'s. Each of the terms for which the weight is one is majorized and the terms with weight zero are kept equal to zero. This leads to the following quadratic majorizer:

$$2a(\mathbf{w}'\mathbf{w})^{-1}(\mathbf{w}'\boldsymbol{\gamma})\mathbf{d}'\mathbf{w} \leq \frac{2a(\mathbf{w}'\mathbf{w})^{-1}(\mathbf{w}'\boldsymbol{\gamma})}{2}(\mathbf{d}'(\mathbf{D}_0^{-1}\mathbf{W})\mathbf{d} + \mathbf{w}'\mathbf{d}_0), \tag{20}$$

with $\mathbf{D}_0^{-1}$ a diagonal matrix with the inverse of the current distances on the diagonal. If $d_{ij_0} < \epsilon$, than d_{ij_0} is set equal to ϵ, with ϵ some small number. Although, strictly speaking, the majorization requirements are not fulfilled for $\epsilon \geq 0$, we may choose ϵ small enough, so that the algorithm remains monotone nonincreasing.

Inserting (19) and (20) in (16) and working out the products leads to the following result:

$$\begin{aligned} L_{\mathrm{u}}(a, \mathbf{X}, \mathbf{Y}, \boldsymbol{\Gamma}) &\leq a^2\mathbf{d}'\mathbf{W}\mathbf{d} - 2a^2(\mathbf{w}'\mathbf{w})^{-1}\mathbf{d}'\mathbf{w}\mathbf{w}'\mathbf{d}_0 + a^2(\mathbf{w}'\mathbf{w})^{-1}\mathbf{d}'_0\mathbf{w}\mathbf{w}'\mathbf{d}_0 - 2a\mathbf{d}'\mathbf{W}\boldsymbol{\gamma} \\ &\quad + a(\mathbf{w}'\mathbf{w})^{-1}(\mathbf{w}'\boldsymbol{\gamma})\mathbf{d}'(\mathbf{D}_0^{-1}\mathbf{W})\mathbf{d} + a(\mathbf{w}'\mathbf{w})^{-1}(\mathbf{w}'\boldsymbol{\gamma})\mathbf{w}'\mathbf{d}_0 + k_1 \\ &= \mathbf{d}'[a^2\mathbf{W} + a(\mathbf{w}'\mathbf{w})^{-1}(\mathbf{w}'\boldsymbol{\gamma})(\mathbf{D}_0^{-1}\mathbf{W})]\mathbf{d} \\ &\quad - 2\mathbf{d}'(a\mathbf{W}\boldsymbol{\gamma} + a^2(\mathbf{w}'\mathbf{w})^{-1}\mathbf{w}\mathbf{w}'\mathbf{d}_0) + k_2, \end{aligned} \tag{21}$$

with $k_2 = k_1 + a^2(\mathbf{w}'\mathbf{w})^{-1}\mathbf{d}'_0\mathbf{w}\mathbf{w}'\mathbf{d}_0 + a(\mathbf{w}'\mathbf{w})^{-1}(\mathbf{w}'\boldsymbol{\gamma})\mathbf{w}'\mathbf{d}_0$.

We re-express this inequality using the notation

$$L_{\mathrm{u}}(a, \mathbf{X}, \mathbf{Y}, \boldsymbol{\Gamma}) \leq \mathbf{d}'\mathbf{F}\mathbf{d} - 2\mathbf{d}'\mathbf{g} + k_2, \tag{22}$$

where

$$\mathbf{F} = a^2\mathbf{W} + a(\mathbf{w}'\mathbf{w})^{-1}(\mathbf{w}'\boldsymbol{\gamma})(\mathbf{D}_0^{-1}\mathbf{W}), \tag{23}$$

and

$$\mathbf{g} = a\mathbf{W}\boldsymbol{\gamma} + a^2(\mathbf{w}'\mathbf{w})^{-1}\mathbf{w}\mathbf{w}'\mathbf{d}_0. \tag{24}$$

Equation (22) is in the standard multidimensional scaling form.

The final quadratic form is found by using the results of De Leeuw (1977) for multidimensional scaling (see also Borg & Groenen, 1997, Section 8.5.). There, it is explained that $\sum_{i<j} f_{ij}d_{ij}^2 = \mathrm{tr}\mathbf{Q}'\mathbf{A}(\mathbf{R})\mathbf{Q}$, with $\mathbf{f}$ the vector of the diagonal elements of $\mathbf{F}$, and with $\mathbf{A}(\mathbf{R})$ having elements

$$a_{ij} = \begin{cases} -f_{ij} & \text{if } i \neq j, \\ \sum_{j,\, i \neq j} f_{ij} & \text{if } i = j, \end{cases} \tag{25}$$

and that $-2\sum_{i<j} g_{ij}d_{ij}(\mathbf{Q}) \leq -2\text{tr}\mathbf{Q}'\mathbf{B}(\mathbf{R})\mathbf{R}$ with $\mathbf{B}(\mathbf{R})$ having elements

$$b_{ij} = \begin{cases} 0 & \text{if } i \neq j \text{ and } d_{ij}(\mathbf{R}) = 0, \\ -g_{ij}d_{ij}^{-1}(\mathbf{R}) & \text{if } i \neq j \text{ and } d_{ij}(\mathbf{R}) > 0, \\ \sum_{j,\, i\neq j} g_{ij}d_{ij}^{-1}(\mathbf{R}) & \text{if } i = j. \end{cases} \tag{26}$$

Putting these results together, we get

$$\begin{aligned} L_{\text{u}}(a, \mathbf{X}, \mathbf{Y}, \mathbf{\Gamma}) &\leq \sum_{i<j} f_{ij}d_{ij}^2(\mathbf{Q}) - 2\sum_{i<j} g_{ij}d_{ij}(\mathbf{Q}) + k_2 \\ &\leq \text{tr}\mathbf{Q}'\mathbf{A}(\mathbf{R})\mathbf{Q} - 2\text{tr}\mathbf{Q}'\mathbf{B}(\mathbf{R})\mathbf{R} + k_2, \end{aligned} \tag{27}$$

The right-hand side of (27) can be minimized analytically using standard derivation techniques. The update of the coordinates is then given by

$$\mathbf{Q}^+ = \mathbf{A}(\mathbf{R})^-\mathbf{B}(\mathbf{R})\mathbf{R}, \tag{28}$$

with $\mathbf{A}(\mathbf{R})^-$ the Moore-Penrose inverse as $\mathbf{A}(\mathbf{R})$ is not of full rank. The Moore-Penrose inverse is given by $\mathbf{A}(\mathbf{R})^- = (\mathbf{A}(\mathbf{R}) + \mathbf{1}\mathbf{1}')^{-\mathbf{1}} - n^{-2}\mathbf{1}\mathbf{1}'$ (Borg & Groenen, 1997, p. 154) provided that the weights w_{ij} do not allow for the subjects and objects to be split into smaller independent unfolding problems. The update of the coordinates of the subject points $\mathbf{X}^+$ is then given by the first n rows of $\mathbf{Q}^+$ while the update of the coordinates of the object points $\mathbf{Y}^+$ is given by the last m rows of $\mathbf{Q}^+$.

5.1.2. The Conditional Case

To update the coordinates in the row-conditional case, we note that each term in summation (5) is an unconditional case,

$$L_{\text{c}}(\mathbf{a}, \mathbf{X}, \mathbf{Y}, \mathbf{\Gamma}) = \frac{1}{n}\sum_i L_{\text{u}}(a_i, \mathbf{x}_i, \mathbf{Y}, \boldsymbol{\gamma}_i), \tag{29}$$

with $\mathbf{x}_i$ the vector of coordinates of subject i. The same line of reasoning used in the unconditional case can be followed here. Let $\mathbf{Q}_i = [\mathbf{x}_i^{'}\ \mathbf{Y}']'$. Then, this reasoning leads, in equivalence with (27), to

$$L_{\text{c}}(\mathbf{a}, \mathbf{X}, \mathbf{Y}, \mathbf{\Gamma}) \leq \frac{1}{n}\sum_i \left[\sum_j f_{ij}d_{ij}^2(\mathbf{Q}_i) - 2\sum_j g_{ij}d_{ij}(\mathbf{Q}_i) + k_{3i}\right] \tag{30}$$

$$= \frac{1}{n}\left[\sum_{i<j} f_{ij}d_{ij}^2(\mathbf{Q}) - 2\sum_{i<j} g_{ij}d_{ij}(\mathbf{Q}) + \sum_{i=1}^{n} k_{3i}\right] \tag{31}$$

$$\leq \frac{1}{n}\left(\text{tr}\mathbf{Q}'\mathbf{A}(\mathbf{R})\mathbf{Q} - 2\text{tr}\mathbf{Q}'\mathbf{B}(\mathbf{R})\mathbf{R} + \sum_{i=1}^{n} k_{3i}\right), \tag{32}$$

with $k_{3i} = \boldsymbol{\gamma}_i^{'}\mathbf{J}\boldsymbol{\gamma}_i + a_i^2(m)^{-1}\mathbf{d}_{0_i}^{'}\mathbf{w}\mathbf{w}^{'}\mathbf{d}_{0_i} + a(m)^{-1}\boldsymbol{\gamma}_i^{'}\mathbf{w}\mathbf{w}^{'}\mathbf{d}_{0_i}$ and with $\mathbf{A}(\mathbf{R})$ and $\mathbf{B}(\mathbf{R})$ being defined as previously, however, taking into account that for the row-conditional case

$$f_{ij} = a_i w_{ij} + a_i m^{-1} w_{ij}\frac{\sum_{j=1}^{m+n} w_{ij}\boldsymbol{\gamma}_{ij}}{d_{0_{ij}}} \tag{33}$$

and

$$g_{ij} = a_i w_{ij} \gamma_{ij} + a_i^2 m^{-1} w_{ij} \left(\sum_{j=1}^{m+n} w_{ij} d_{0_{ij}} \right), \tag{34}$$

with i being related to a subject row entry. As before d_{0ij} is the distance of point i to point j based on the previous estimate of the coordinates. The remaining values are found by symmetry or equal zero when $w_{ij} = 0$.

5.2. *Transformation Update*

In this section, we explain how to update the disparities $\boldsymbol{\gamma}$. Usually, a least-squares regression with or without an intercept is used to obtain the disparities for data with an interval or ratio-level of measurement, respectively.

For data with an ordinal level of measurement, Kruskal's (1965) monotone regression can be used. These results still hold as the problem is a least-squares problem, which can be seen in the following formulation

$$\min_{\boldsymbol{\gamma}} L_{\mathrm{u}}(a, \mathbf{X}, \mathbf{Y}, \boldsymbol{\Gamma}) = \min_{\boldsymbol{\gamma}} \left[\|\boldsymbol{\gamma} - a\mathbf{d}\|_{\mathbf{J}}^2 \right]. \tag{35}$$

The length restriction $\boldsymbol{\gamma}'\mathbf{J}\boldsymbol{\gamma} = 1$ can be imposed after the (monotone) regression by dividing the updated disparities by $(\boldsymbol{\gamma}'\mathbf{J}\boldsymbol{\gamma})^{1/2}$ (see Groenen, Os, & Meulman, 2000). However, the situation can occur where all optimally scaled data are equal. The solution for this case can be obtained by formulating the monotone regression problem as a length constrained nonnegative least-squares problem. Let $\boldsymbol{\gamma} = \mathbf{Sb}$ where $\mathbf{S} = \mathbf{JLM}$ defines the (centered) monotone transformation base with $\mathbf{M}$ a triangular matrix of ones with zeroes in the upper triangular part and $\mathbf{L}$ a permutation matrix such that $\mathbf{L}'\boldsymbol{\gamma}$ is ordered increasingly. Furthermore, let $\mathbf{b}$ be constrained so that $b_i \geq 0$ for all i and that $\mathbf{b}'\mathbf{S}'\mathbf{Sb} = 1$. Then, the monotone regression problem in (35) is equivalent to minimizing the length-constrained nonnegative least-squares problem $\|\mathbf{Sb} - a\mathbf{Jd}\|^2$ subject to $b_i \geq 0$ for all i and to $\mathbf{b}'\mathbf{S}'\mathbf{Sb} = 1$. Let $\mathbf{G} = \mathbf{S}'\mathbf{S}$. As shown in Groenen et al. (2000), the optimal $\mathbf{b}$, $\mathbf{b}_{\mathrm{opt}}$, in this case, has a nonzero element $b_i = (g_{ii})^{-1/2}$ and all other $b_{j\neq i} = 0$ for the i for which $\mathbf{d}'\mathbf{d} + 1 - 2a\mathbf{b}'\mathbf{S}'\mathbf{d}$ is minimal. In principle, $0 \leq i \leq m$ so that m different values of $\mathbf{d}'\mathbf{d} + 1 - 2a\mathbf{b}'\mathbf{S}'\mathbf{d}$ need to be checked. However, if $i = 1$, then $\boldsymbol{\gamma}$ would consist of all values equal. Therefore, i needs to vary only from 2 to m, so that only m-1 comparisons are needed. Indeed, this solution corresponds to one of the edges of the cone of monotone transformations. Note also that the obtained disparities are centered. To make them nonnegative, we add a constant.

Next, we will show that the disparities are fixed for the interval (and ratio) level of the measurement. We recall the fact that the loss function is invariant under the addition of a constant. This invariance justifies the centering of both the data and the distances, thus fixing the intercept to zero (this does not affect the slope). The disparities may then be expressed as $\boldsymbol{\gamma}_{\mathrm{c}} = b_1\boldsymbol{\delta}_{\mathrm{c}}$, where the subscript c indicates the centering and with b_1 the slope parameter. The length constraint $\boldsymbol{\gamma}'\mathbf{J}\boldsymbol{\gamma} = 1$, which equally holds for the centered disparities, fixes the slope parameter: $b_1 = (\boldsymbol{\delta}'_{\mathrm{c}}\mathbf{J}\boldsymbol{\delta}_{\mathrm{c}})^{-1/2}$, so that the disparities depend on the data only by

$$\boldsymbol{\gamma}_{\mathrm{c}} = (\boldsymbol{\delta}'_{\mathrm{c}}\mathbf{J}\boldsymbol{\delta}_{\mathrm{c}})^{-1/2}\boldsymbol{\delta}_{\mathrm{c}}. \tag{36}$$

Remark that the regression weight is positive and non-zero. Equality of the disparities can clearly not occur so that in this case the loss values are nonincreasing. After having obtained the centered disparities, a constant should be added to make them nonnegative. This does not change the loss due to the centering by $\mathbf{J}$ in the loss function.

5.3. Updating the Regression Weight

The update of the regression weight a has already been developed in Section 2 as $a = \mathbf{d}'\mathbf{J}\boldsymbol{\gamma}(\mathbf{d}'\mathbf{J}\mathbf{d})^{-1}$. Since our loss function is based on the *squared* correlation between distances and disparities, the sign of the correlation is irrelevant. However, from the interpretational point of view, we wish for a solution where an increase in the distance corresponds to a larger dissimilarity or a smaller similarity value. Therefore, negative correlations between the distances and the disparities are undesirable. Apart from the interpretational problem, a negative relation between the distances and the disparities results in negative a's while the update of the coordinates is based on the assumption of nonnegative $a's$. To solve both the interpretational and the algorithmic problem, the calculation of a should be adapted in the following way:

$$a = \begin{cases} \mathbf{d}'\mathbf{J}\boldsymbol{\gamma}(\mathbf{d}'\mathbf{J}\mathbf{d})^{-1} & \text{if } a \geq 0 \ , \\ 0 & \text{if } a < 0 \ . \end{cases} \tag{37}$$

In the row-conditional case, it should be checked for each subject that a_i is nonnegative. If this is not the case, a_i should be replaced by 0.

We remark that monotone convergence of the algorithm is still guaranteed and that the algorithm now is forced to find a solution for which distances and disparities are positively related. In our experience, negative $a's$ disappear rather quickly in the iterative process, if they appeared at all.

Appendix B. Finding the Origin of Points Almost on a Circle

In this section, it is shown how the origin of points that lie almost on a circle can be found. To do this, we will make use of IM. A point that lies in the origin is almost equidistant to m points lying almost on a circle. Stated differently, the variance of the distance of this point to the points on the circle should be very small. Therefore, we must minimize

$$\begin{aligned} L(\mathbf{c}) &= \sum_{j=1}^{m} \left(d_j - \overline{d}_j\right)^2 \\ &= \sum_{j=1}^{m} d_j^2 - \frac{1}{m}\left(\sum_{j=1}^{m} d_j\right)^2 \end{aligned} \tag{38}$$

where d_j is the Euclidean distance between point j and the origin whose coordinates are given by the vectors $\mathbf{y}_j$ and $\mathbf{c}$, respectively, that is, $d_j = \|\mathbf{y}_j - \mathbf{c}\|$. The d_j's are the distances of the center to the m object points that lie almost on a circle. Equation (38) can be majorized by applying the scalar form of inequality (18) to the last term of (38), that is,

$$L(\mathbf{c}) \leq \sum_{j=1}^{m} d_j^2 - \frac{2}{m}\left(\sum_{j=1}^{m} d_j\right)\left(\sum_{j=1}^{m} d_{j0}\right) + k_1, \tag{39}$$

with d_{j0} the distances based on the previous estimation of $\mathbf{c}$ and $k_1 = (\sum_j d_{j0})^2$. The function $-d_j$ appearing in the right-hand side of (39) can be majorized again by using the Cauchy-Schwarz inequality, that is,

$$-d_j = -\|\mathbf{y}_j - \mathbf{c}\| \leq -\frac{(\mathbf{y}_j - \mathbf{c})'(\mathbf{y}_j - \mathbf{c}_0)}{\|\mathbf{y}_j - \mathbf{c}_0\|} = -\frac{(\mathbf{y}_j - \mathbf{c})'(\mathbf{y}_j - \mathbf{c}_0)}{d_{j0}}. \tag{40}$$

Inserting this result into (39) and noting that $\sum_j d_j^2 = \sum_j \|\mathbf{y}_j - \mathbf{c}\|^2$ gives

$$\begin{aligned} L(\mathbf{c}) &\leq \sum_{j=1}^{m} d_j^2 - \frac{2}{m}\left(\sum_{j=1}^{m} d_{j0}\right)\sum_{j=1}^{m}\frac{(\mathbf{y}_j - \mathbf{c})'(\mathbf{y}_j - \mathbf{c}_0)}{d_{j0}} + k_1 \\ &= \sum_{j=1}^{m}\|\mathbf{y}_j\|^2 + m\|\mathbf{c}\|^2 - 2\mathbf{c}'\left(\sum_{j=1}^{m}\mathbf{y}_j\right) \\ &\quad -\frac{2}{m}\left(\sum_{j=1}^{m} d_{j0}\right)\sum_{j=1}^{m}\frac{(\mathbf{y}_j - \mathbf{c})'(\mathbf{y}_j - \mathbf{c}_0)}{d_{j0}} + k_1 \\ &= m\mathbf{c}'\mathbf{c} - 2\mathbf{c}'\left[\sum_{j=1}^{m}\mathbf{y}_j + \frac{1}{m}\left(\sum_{j=1}^{m} d_{j0}\right)\sum_j \frac{(\mathbf{y}_j - \mathbf{c}_0)}{d_{j0}}\right] + k_2, \end{aligned} \tag{41}$$

with

$$k_2 = k_1 + \sum_j \|\mathbf{y}_j\| - \frac{2}{m}\left(\sum_j d_{j0}\right)\sum_j \frac{\mathbf{y}_j'(\mathbf{y}_j - \mathbf{c}_0)}{d_{j0}}.$$

It follows from (41) that the update of the origin is given by

$$\mathbf{c}^+ = \frac{1}{m}\sum_j \mathbf{y}_j - \frac{1}{m^2}\left(\sum_{j=1}^{m} d_{j0}\right)\sum_j \frac{\mathbf{y}_j - \mathbf{c}_0}{d_{j0}}.$$

References

Borg, I., & Groenen, P.J.F. (1997). *Modern Multidimensional Scaling*. New York: Springer-Verlag.

Busing, F.M.T.A., Groenen, P.J.F., & Heiser, W.J. (in press). Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika*.

Carroll, J.D. (1972). Individual differences and multidimensional scaling. In R.N. Shepard, A.K. Romney, & S.B. Nerlove (Eds.), *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences* (Vol. 1, pp. 105–155). New York: Seminar Press.

Coombs, C.H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, *57*, 148–158.

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

Delbeke, L. (1968). *Construction of Preference Spaces: An Investigation into the Applicability of Multidimensional Scaling Models*. Leuven: Leuvense universitaire uitgaven.

De Leeuw, J. (1977). Applications of convex analysis to multidimensional scaling. In J.R. Barra, G. Romier, & B. van Cutsem (Eds.), *Recent Developments in Statistics* (pp. 133–145). Amsterdam, The Netherlands: North-Holland.

De Leeuw, J. (1983). *On Degenerate Nonmetric Unfolding Solutions* (Tech. Rep.). Department of data theory, FSW/RUL.

De Leeuw, J. (1994). Block relaxation algorithms in statistics. In H.H. Bock, W. Lenski, & M.M. Richter (Eds.), *Information Systems and Data Analysis* (pp. 308–325). Berlin: Springer-Verlag.

De Leeuw, J., & Heiser, W.J. (1977). Convergence of correction-matrix algorithms for multidimensional scaling. In J.C. Lingoes, E.E. Roskam, & I. Borg (Eds.), *Geometric Representations of Relational Data* (pp. 735–752). Ann Arbor, MI: Mathesis Press.

DeSarbo, W.S., & Carroll, J.D. (1985). Three-way metric unfolding via alternating least squares. *Psychometrika*, *50*, 275–300.

Green, P.E., & Rao, V.R. (1972). *Applied Multidimensional Scaling: A Comparison of Approaches and Algorithms*. New York: Holt, Rinehart and Winston.

Groenen, P.J.F., Heiser, W.J., & Meulman, J.J. (1999). Global optimization in least-squares multidimensional scaling by distance smoothing. *Journal of Classification*, *16*, 225–254.

Groenen, P.J.F., Os, B.-J., & Meulman, J.J. (2000). Optimal scaling by alternating length-constrained nonnegative least squares with application to distance-based analysis. *Psychometrika*, *65*, 511–524.

Heiser, W.J. (1989). Order invariant unfolding analysis under smoothness restrictions. In G. De Soete, H. Feger, & K.C. Klauer (Eds.), *New Developments in Psychological Choice Modelling* (pp. 3–31). Amsterdam: Elsevier Science.

Kim, C., Rangaswamy, A., & DeSarbo, W.S. (1999). A quasi-metric approach to multidimensional unfolding for reducing the occurrence of degenerate solutions. *Multivariate Behavioral Research*, *34*, 143–180.

Kruskal, J.B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*, 115–129.

Kruskal, J.B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society B*, *27*, 251–563.

Kruskal, J.B., & Carroll, J.D. (1969). Geometrical models and badness-of-fit functions. In P. R. Krishnaiah (Ed.), *Multivariate Analysis* (vol 2, pp. 639–671). New York: Academic Press.

Roskam, E.E.C.I. (1968). *Metric Analysis of Ordinal Data in Psychology*. Voorschoten: VAM.

Tucker, L.R. (1960). Intra-individual and inter-individual multidimensionality. In H. Gulliksen & S. Messick (Eds.), *Psychological Scaling: Theory and Applications* (pp. 155–167). New York: Wiley.

Manuscript received 20 DEC 2002
Final version received 2 DEC 2003