VIPSCAL: A combined vector ideal point model for preference data

K. Van Deun

Department of Psychology, Catholic University of Leuven,
Tiensestraat 102, B-3000 Leuven, Belgium, katrijn.vandeun@psy.kuleuven.ac.be

P. J. F. Groenen

Econometric Instititute, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, groenen@few.eur.nl

L. Delbeke

Department of Psychology, Catholic University of Leuven,
Tiensestraat 102, B-3000 Leuven, Belgium, Luc.Delbeke@psy.kuleuven.ac.be

20th January 2005

Econometric Institute Report EI 2005-03

Abstract

In this paper, we propose a new model that combines the vector model and the ideal point model of unfolding. An algorithm is developed, called VIPSCAL, that minimizes the combined loss both for ordinal and interval transformations. As such, mixed representations including both vectors and ideal points can be obtained but the algorithm also allows for the unmixed cases, giving either a complete ideal point analysis or a complete vector analysis. On the basis of previous research, the mixed representations were expected to be nondegenerate. However, degenerate solutions still occurred as the common belief that distant ideal points can be represented by vectors does not hold true. The occurrence of these distant ideal points was solved by adding certain length and orthogonality restrictions on the configuration. The restrictions can be used both for the mixed and unmixed cases in several ways such that a number of different models can be fitted by VIPSCAL.

Keywords: Unfolding, Vector Model, Ideal Point Model

1 Introduction

To describe the relation between a group of judges and the items they ordered according to their preference, models have been proposed that represent preference data in low-dimensional space. Among these, two are of particular interest here: the ideal point model of unfolding and the vector model. When one model should be used and when the other, is not obvious. Furthermore, it is plausible that, for a same data set, some judges would fit well an ideal point representation and some a vector representation. Therefore, a mixed model is of interest that combines both and where it is the model that determines which judges to represent by vectors and which by ideal points. Also, in a recent paper (Van Deun, Groenen, Heiser, Busing, & Delbeke, in press) it was shown how some frequently occurring degenerate ideal point solutions could be interpreted by representing some judges in the obtained solution by vectors. Their finding motivates the idea that at least some degeneracies can be avoided by a mixed model. The objective of this paper is to construct such a hybrid model and to provide an accompanying flexible algorithm. First, we discuss the ideal point and vector model. Then, we present a loss function that combines both models and we develop an alternating least squares and iterative majorization algorithm, named VIPSCAL, to minimize this loss. In an application of this algorithm, we show that the assertion that the vector model is an ideal point model with the ideal points at infinity (see Carroll, 1972; Coombs, 1975; Borg & Groenen, 1997) is not very useful: asymptotically, the assertion is true but in practice very distant ideal points often reflect a different preference order then a vector representation of the same point. As a result, the combined model still gives degenerate solutions. We show how restrictions can be incorporated in the algorithm easily to prevent the occurrence of distant points. The resulting algorithm allows for analyses under a number of different models, including, for example, a restricted ideal point model and the compensatory distance model. Some possible ways of analyzing data with the algorithm are illustrated with empirical data. Note that multidimensional unfolding is known to perform often well with simulated data while yielding degenerate solutions for empirical data (see Kruskal & Carroll, 1969). Therefore, the results we report depend on empirical data, not on simulated. As we believe that the algorithm is a very flexible one and, therefore, useful for others, it is developed here in quite some detail. The main merit of VIPSCAL is that it allows for a more thorough study and understanding of the relation between the distance and vector model.

2 The vector and ideal point model

Both the ideal point or distance model and the vector model map the preference data to a low-dimensional (often Euclidean) space: the judges and the items are located in this space such that the preference order given by a certain judge is reflected in the relation from the mapped judge to the mapped items. The difference between the two models lies in the function that is used to relate the

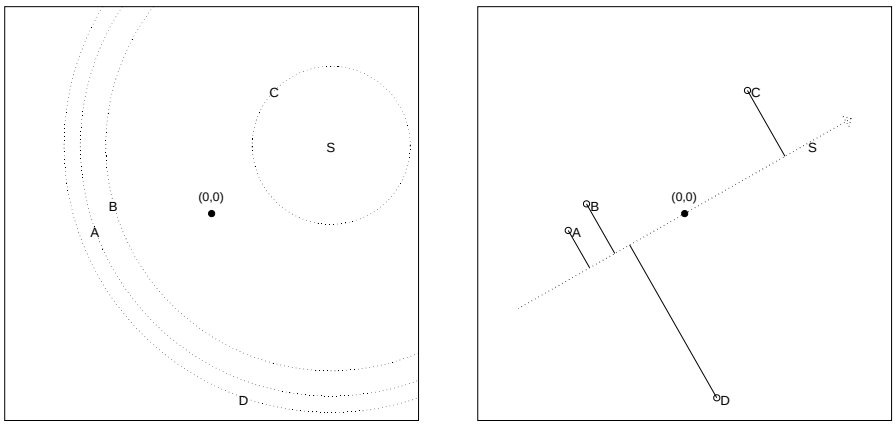


Figure 1: Ideal point and vector model representation of the relation between judge 'S' and the items 'A', 'B', 'C', and 'D'.

judge to the different items. In the ideal point representation, this function is the distance function while in the vector representation the scalar product is used. A graphical illustration is given in Figure 1 for a judge 'S' and four items 'A', 'B', 'C' and 'D' with the ideal point configuration depicted in the left panel and the vector representation in the right panel. The interpretation of the configuration in the left panel is based on the order of the distances between 'S' and the different items where the closest item represents the most preferred one. According to the ideal point solution, the preference order for judge 'S' is 'CBAD'. The circles around 'S' are called isopreference contours as objects lying on the same circle are equally preferred. In the right panel, the same coordinates are used as those of the left panel but now the judge is represented by an oriented vector. In this case, the preference order is reflected in the scalar product, or graphically, in the orthogonal projections of the items on the vector. The higher the projection falls on the vector, the more it is preferred. Here, the preference order is 'CDBA'. Equal preference in this model occurs for items that lie on the same line perpendicular to the vector.

This illustration shows that the reproduced rank orders, for the same set of coordinates, are in general different. It also shows that the underlying preference mechanisms are different: when we consider the underlying dimensions as the quantity of a certain attribute, then there are optimal amounts in case of the ideal point model with preference decreasing when one moves away from the ideal point in whatever direction. In case of the vector model, there is an optimal direction (called by Carroll, 1972, the *relative importance*) with preference increasing monotonically when one moves further and further away: here, a 'the more, the better' principle applies. The vector model has been described

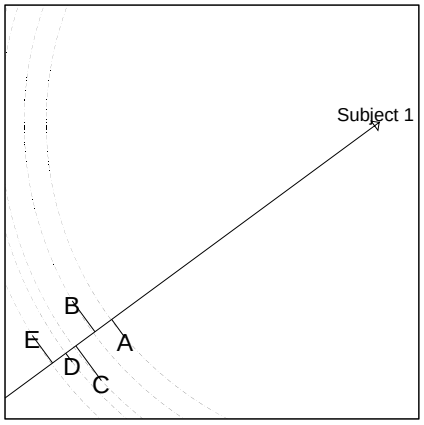


Figure 2: Ideal point at infinity

by Carroll (1972, pages 117-118) as an ideal point model with the ideal points at infinity, see Figure 2. In that case, the isopreference contours become almost equal to orthogonal projection lines (see also Section 4).

The ideal point model of unfolding has the advantage that its representations are easier to interpret but it is known to result in degenerate solutions in the majority of cases (Kruskal & Carroll, 1969; De Leeuw, 1983; Heiser, 1989; Kim, Rangaswamy, & DeSarbo, 1999; Busing, Groenen, & Heiser, in press and Van Deun et al., in press). Degenerate solutions fit well but are not interpretable, for example, because the judges cluster together in the center of a circle formed by the items. Van Deun et al. (in press) showed that often a part of the degeneracy is the result of a few distant subjects in the configuration. Following the results of De Leeuw (1983), they also showed how the configuration could be made interpretable by representing these distant subjects by vectors. Therefore, we propose to develop a model that combines the ideal point model and the vector model, hoping that such a model will give non-degenerate solutions as we expect the distant subjects to be represented by vectors. Note also that such a mixed model can be of psychological relevance considering that those judges that are represented by vectors follow another preference mechanism then those that are represented by ideal points. Interest in such a combined model has been expressed by DeSarbo and Carroll (1985). In case of known object coordinates (external unfolding), the PREFMAP program (Meulman, Heiser, & Carroll, 1986) already allows for mixed representations where some judges are represented by vectors and others by ideal points. However, in PREFMAP the user determines prior to the analysis how each subject should be represented.

3 A Combined Loss Function

Consider n judges $i = 1, ..., n$ who ordered m objects $j = 1, ..., m$ according to their preference. In the ideal point model, both are represented as points in low dimensional space such that the order of the distances of a judge to the different objects reflects the preference order in the nonmetric case or that the differences between the distances reflect the differences between the preference scores (metric case). Such a configuration can be obtained by minimizing the following loss function,

$$K(\mathbf{X}, \mathbf{Y}, \mathbf{\Gamma}) = n^{-1} \sum_{i=1}^{n} K_i(\mathbf{x}_i, \mathbf{Y}, \boldsymbol{\gamma}_i) = n^{-1} \sum_{i=1}^{n} \left[1 - r^2(\boldsymbol{\gamma}_i, \mathbf{d}_i)\right], \qquad (1)$$

with $\mathbf{d}_i$ the vector of distances for judge i and $\boldsymbol{\gamma}_i$ the transformed preference scores. Here, we will use Euclidean distances with the distance between judge i and object j given by $d_{ij} = [(\mathbf{x}_i - \mathbf{y}_j)'(\mathbf{x}_i - \mathbf{y}_j)]^{1/2}$ where $\mathbf{x}_i$ and $\mathbf{y}_j$ are the coordinate vectors for judge i and object j. As shown by Kruskal and Carroll (1969) and by Van Deun et al. (in press), minimizing (1) or, equivalently, maximizing the squared correlation between the transformed data and the distances is equivalent to minimizing Stress-2. Van Deun et al. (in press) showed that under the condition $\boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i = 1$, loss function (1) is equivalent to

$$K(\mathbf{X}, \mathbf{Y}, \mathbf{\Gamma}, \mathbf{a}) = n^{-1} \sum_{i=1}^{n} K_i(\mathbf{x}_i, \mathbf{Y}, \boldsymbol{\gamma}_i, a_i) = n^{-1} \sum_{i=1}^{n} \|\boldsymbol{\gamma}_i - a_i\mathbf{d}_i\|_{\mathbf{J}}^2, \qquad (2)$$

where a_i is a scale parameter and $\mathbf{J} = \mathbf{I} - m^{-1}\mathbf{1}\mathbf{1}'$ a centering operator. A similar proof is given below for the vector model. The a_i have to be restricted to be nonnegative in order to fit positive correlations between the distances and the transformed data as we want the distance from a judge to a preference item to increase when the item has a higher (rank) score.

In the vector model, the judges are represented as vectors and the preference order is reflected in the order of the scalar products between a judge and the different items. Let $\mathbf{Y}$ be the $m \times p$ matrix of coordinates of the items and $\mathbf{x}_i$ the $p \times 1$ vector with the judges' coordinates so that the scalar product vector is given by $\mathbf{Y}\mathbf{x}_i$. Then, loss for judge i can be measured as

$$\begin{aligned} L_i(\mathbf{x}_i, \mathbf{Y}, \boldsymbol{\gamma}_i, b_i) &= \|\boldsymbol{\gamma}_i - b_i\mathbf{Y}\mathbf{x}_i\|_{\mathbf{J}}^2 \\ &= \boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i + b_i^2\mathbf{x}_i'\mathbf{Y}'\mathbf{J}\mathbf{Y}\mathbf{x}_i - 2b_i\boldsymbol{\gamma}_i'\mathbf{J}\mathbf{Y}\mathbf{x}_i. \end{aligned} \qquad (3)$$

This loss function closely resembles the PRINCIPALS model of Young, Takane, and De Leeuw (1978) with loss $n^{-1}\sum_i \|\boldsymbol{\gamma}_i - \mathbf{Y}\mathbf{x}_i'\|^2$ and constraints $\boldsymbol{\gamma}_i'\mathbf{1} = 0$ and $\boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i = m$, which were imposed using explicit normalization. For a historical account of nonmetric principal components analysis, see Gifi (1990). Below we show that (3) is equivalent to

$$L_i(\mathbf{x}_i, \mathbf{Y}, \boldsymbol{\gamma}_i, b_i) = 1 - r^2(\boldsymbol{\gamma}_i, \mathbf{Y}\mathbf{x}_i), \qquad (4)$$

subject to $\boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i = 1$. The optimal b_i is found by setting

$$\frac{\partial L_i}{\partial b_i} = 2b_i\mathbf{x}_i'\mathbf{Y}'\mathbf{J}\mathbf{Y}\mathbf{x}_i - 2\boldsymbol{\gamma}_i'\mathbf{J}\mathbf{Y}\mathbf{x}_i = 0 \tag{5}$$

thus

$$b_i = \frac{\boldsymbol{\gamma}_i'\mathbf{J}\mathbf{Y}\mathbf{x}_i}{\mathbf{x}_i'\mathbf{Y}'\mathbf{J}\mathbf{Y}\mathbf{x}_i}. \tag{6}$$

Inserting this optimal value in (3) gives

$$\begin{aligned} L_i(\mathbf{x}_i, \mathbf{Y}, \boldsymbol{\gamma}_i) &= \boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i + \left(\frac{\boldsymbol{\gamma}_i'\mathbf{J}\mathbf{Y}\mathbf{x}_i}{\mathbf{x}_i'\mathbf{Y}'\mathbf{J}\mathbf{Y}\mathbf{x}_i}\right)^2 \mathbf{x}_i'\mathbf{Y}'\mathbf{J}\mathbf{Y}\mathbf{x}_i - 2\frac{\boldsymbol{\gamma}_i'\mathbf{J}\mathbf{Y}\mathbf{x}_i}{\mathbf{x}_i'\mathbf{Y}'\mathbf{J}\mathbf{Y}\mathbf{x}_i}\boldsymbol{\gamma}_i'\mathbf{J}\mathbf{Y}\mathbf{x}_i \\ &= 1 + \frac{(\boldsymbol{\gamma}_i'\mathbf{J}\mathbf{Y}\mathbf{x}_i)^2}{\mathbf{x}_i'\mathbf{Y}'\mathbf{J}\mathbf{Y}\mathbf{x}_i} - 2\frac{(\boldsymbol{\gamma}_i'\mathbf{J}\mathbf{Y}\mathbf{x}_i)^2}{\mathbf{x}_i'\mathbf{Y}'\mathbf{J}\mathbf{Y}\mathbf{x}_i} \\ &= 1 - \frac{(\boldsymbol{\gamma}_i'\mathbf{J}\mathbf{Y}\mathbf{x}_i)^2}{\mathbf{x}_i'\mathbf{Y}'\mathbf{J}\mathbf{Y}\mathbf{x}_i} = 1 - r^2(\boldsymbol{\gamma}_i, \mathbf{Y}\mathbf{x}_i), \end{aligned}$$

under the condition that $\boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i = 1$. Note that b_i should be restricted to be nonpositive for interpretational convenience: the higher the object projects on the vector representing the judge, the more the object is preferred (and thus a lower score), since preference orders are assumed to be dissimilarities. A low value indicates higher preference and a high value indicates a low preference.

We combine both models defined by (2) and (3) in a weighted sum,

$$\begin{aligned} L(\mathbf{X}, \mathbf{Y}, \boldsymbol{\Gamma}, \mathbf{a}, \mathbf{b}) &= n^{-1}\sum_{i=1}^{n}\left[g_iK_i + (1-g_i)L_i\right] \\ &= n^{-1}\sum_{i=1}^{n}\left[g_i\|\boldsymbol{\gamma}_i - a_i\mathbf{d}_i\|_{\mathbf{J}}^2 + (1-g_i)\|\boldsymbol{\gamma}_i - b_i\mathbf{Y}\mathbf{x}_i\|_{\mathbf{J}}^2\right], \\ &= n^{-1}\sum_{i=1}^{n}\left[1 + g_ia_i^2\mathbf{d}_i'\mathbf{J}\mathbf{d}_i - 2g_ia_i\mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i + (1-g_i)b_i^2\mathbf{x}_i'\mathbf{Y}'\mathbf{J}\mathbf{Y}\mathbf{x}_i\right. \\ &\quad \left. -2(1-g_i)b_i\mathbf{x}_i'\mathbf{Y}'\mathbf{J}\boldsymbol{\gamma}_i\right], \end{aligned} \tag{7}$$

where the weights g_i take values zero or one, $a_i \geq 0$, $b_i \leq 0$ and $\boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i = 1$. Note that the use of dummy weights g_i implies that a subject is either represented by a vector or an ideal point.

We will minimize (7) by alternating least squares using the algorithm presented here and named VIPSCAL.

1. Choose initial $\mathbf{X}^{(0)}$, $\mathbf{Y}^{(0)}$, $\mathbf{a}^{(0)}$, $\mathbf{b}^{(0)}$ and $\mathbf{g}^{(0)}$.
2. $l := 0$.
3. $l := l + 1$.
4. Update the person coordinates $\mathbf{X}$ given $\mathbf{Y}^{(l-1)}$, $\boldsymbol{\Gamma}^{(l-1)}$, $\mathbf{a}^{(l-1)}$, $\mathbf{b}^{(l-1)}$ and $\mathbf{g}^{(l-1)}$.

5. Update the object coordinates $\mathbf{Y}$ given $\mathbf{X}^{(l)}$, $\mathbf{\Gamma}^{(l-1)}$, $\mathbf{a}^{(l-1)}$, $\mathbf{b}^{(l-1)}$ and $\mathbf{g}^{(l-1)}$.
6. Compute the transformation update $\mathbf{\Gamma}$ given $\mathbf{X}^{(l)}$, $\mathbf{Y}^{(l)}$, $\mathbf{\Gamma}^{(l-1)}$, $\mathbf{a}^{(l-1)}$, $\mathbf{b}^{(l-1)}$ and $\mathbf{g}^{(l-1)}$.
7. Update $\mathbf{a}$ and $\mathbf{b}$ based on $\mathbf{X}^{(l)}$, $\mathbf{Y}^{(l)}$, $\mathbf{\Gamma}^{(l)}$ and $\mathbf{g}^{(l-1)}$.
8. Update $\mathbf{g}$ based on $\mathbf{X}^{(l)}$, $\mathbf{Y}^{(l)}$, $\mathbf{\Gamma}^{(l)}$, $\mathbf{a}^{(l)}$ and $\mathbf{b}^{(l)}$
9. If $L^{(l)} - L^{(l-1)} < \epsilon$ or $l = l_{max}$ stop. Else return to 3.

Suitable updates can be easily found for Steps 7 and 8. The more difficult parts are the Steps 4, 5 and, due to the constraint $\boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i = 1$, also 6. The update of the coordinates (Steps 4 and 5) will be based on majorization procedures. Note that this algorithm can be used to fit a full ideal point model by setting $\mathbf{g}^{(0)} = \mathbf{1}$ and dropping Step 8. In a similar way, when setting $\mathbf{g}^{(0)} = \mathbf{0}$, a full vector model will be fitted. Below we develop each update.

3.1 Updating the coordinates

In this section, we will derive expressions that update the coordinates $\mathbf{X}$ and $\mathbf{Y}$. As the loss function is complicated, we will replace it by majorizing functions that are easy to minimize (see De Leeuw, 1994, Heiser, 1995, and Lange, Hunter, & Yang, 2000 for a general introduction to iterative majorization). These functions are never lower than the original function and touch the original function at a supporting point. By taking the previous estimate as the supporting point in the current iteration, it is guaranteed that the loss function is nonincreasing with each new update. In practice, the values of the loss function converge to a local minimum. When the original function has a bounded Hessian, it can be majorized by a quadratic function. Kiers (2002) introduced one that can be easily updated both in the constrained and unconstrained case. This function is of the following form

$$g(\mathbf{Z}) = k + \text{tr}\mathbf{A}\mathbf{Z} + \sum_q \text{tr}\mathbf{B}_q\mathbf{Z}\mathbf{C}_q\mathbf{Z}', \tag{8}$$

where k includes all terms that are constant in $\mathbf{Z}$ with $\mathbf{Z}$ denoting the matrix we want to update (this is in our case, $\mathbf{x}_i$ or $\mathbf{Y}$). So, we will look for a function that majorizes (7) and that is of the form (8).

As explained by Kiers (2002), when the $\mathbf{C}_q$ are positive semi-definite, the minimization of (8) is itself based on the majorizing function

$$g(\mathbf{Z}) \leq k + \text{tr}\mathbf{F}\mathbf{Z} + \sum_q \lambda_q \text{tr}\mathbf{C}_q\mathbf{Z}'\mathbf{Z}, \tag{9}$$

with λ_q the largest eigenvalue of $\left(2^{-1}\mathbf{B}_q + 2^{-1}\mathbf{B}_q'\right)$ and with $\mathbf{F} = \mathbf{A} + \sum_q \left(\mathbf{C}_q\mathbf{Z}_0'(\mathbf{B}_q + \mathbf{B}_q') - 2\lambda_q\mathbf{C}_q\mathbf{Z}_0'\right)$. Without constraints, the optimal $\mathbf{Z}$ is given by

$$\mathbf{Z}^+ = -\frac{1}{2}\mathbf{F}'\left(\sum_q \lambda_q\mathbf{C}_q\right)^{-1}. \tag{10}$$

The function that we want to minimize is given by (7), that is,

$$\begin{aligned} L(\mathbf{X},\mathbf{Y}) &= n^{-1}\sum_{i=1}^{n}\left[1+g_i a_i^2\mathbf{d}_i'\mathbf{J}\mathbf{d}_i - 2g_i a_i\mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i + (1-g_i)b_i^2\mathbf{x}_i'\mathbf{Y}'\mathbf{J}\mathbf{Y}\mathbf{x}_i\right. \\ &\quad \left. -2(1-g_i)b_i\mathbf{x}_i'\mathbf{Y}'\mathbf{J}\boldsymbol{\gamma}_i\right]. \end{aligned} \tag{11}$$

To minimize this function, we will construct a quadratic majorizing function of the basic form (8). As majorization is closed under summation, we will majorize each of the terms that are not a linear or quadratic function of the coordinates. These are the two terms involving distances, namely the terms $g_i a_i^2\mathbf{d}_i'\mathbf{J}\mathbf{d}_i$ and $-2g_i a_i\mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i$.

Note that we have to update two sets of coordinates $\mathbf{X}$ and $\mathbf{Y}$ where the judges enter the loss function by summing over all n judges. This allows us to update the coordinates $\mathbf{x}_i$ for each of the judges separately (the sum is minimized by minimizing each of its components). Next, we will show how each of the terms can be expressed in the standard form (8), and this both for the object coordinate matrix $\mathbf{Y}$ and the subject coordinate vector $\mathbf{x}_i$. For the sake of clarity, we will use a notation that follows closely the notation in (8) with numbered and lettered subscripts: we will use a subscript x and a subscript y when the term is expressed in function of respectively $\mathbf{x}_i$ and $\mathbf{Y}$. The numbers count the different terms that are in the same form. For example, k_{2x} would be the second term that is constant in $\mathbf{x}_i$.

First, we show how the terms that are already linear or quadratic in the coordinates can be arranged in the desired form. Recall the following properties: (1) a scalar and the trace of this scalar are equivalent, (2) trace functions are invariant under transposition, and (3) trace functions are invariant under cyclic permutations. Then, (11) in function of $\mathbf{x}_i$ and with the two last terms expressed in the form (8) gives,

$$\begin{aligned} L(\mathbf{x}_i) &= k_{1x} + g_i a_i^2\mathbf{d}_i'\mathbf{J}\mathbf{d}_i - 2g_i a_i\mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i \\ &\quad +\mathrm{tr}\mathbf{B}_{1x}\mathbf{x}_i\mathbf{C}_{1x}\mathbf{x}_i' - 2\mathrm{tr}\mathbf{A}_{1x}\mathbf{x}_i, \end{aligned} \tag{12}$$

with $k_{1x} = \boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i$, $\mathbf{B}_{1x} = (1-g_i)b_i^2\mathbf{Y}'\mathbf{J}\mathbf{Y}$, $\mathbf{C}_{1x} = 1$, and $\mathbf{A}_{1x} = (1-g_i)b_i\boldsymbol{\gamma}_i'\mathbf{J}'\mathbf{Y}$. In the same way we get in function of the object coordinates $\mathbf{Y}$,

$$\begin{aligned} L(\mathbf{Y}) &= k_{1y} + \sum_{i=1}^{n}\left[g_i a_i^2\mathbf{d}_i'\mathbf{J}\mathbf{d}_i - 2g_i a_i\mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i\right] \\ &\quad +\mathrm{tr}\mathbf{B}_{1y}\mathbf{Y}\mathbf{C}_1 y\mathbf{Y}' - 2\mathrm{tr}\mathbf{A}_{1y}\mathbf{Y}, \end{aligned} \tag{13}$$

with $k_{1y} = \sum_i \boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i$, $\mathbf{B}_{1y} = \mathbf{J}$, $\mathbf{C}_{1y} = \sum_i (1-g_i)b_i^2\mathbf{x}_i\mathbf{x}_i'$, and $\mathbf{A}_{1y} = \sum_i (1-g_i)b_i\mathbf{x}_i\boldsymbol{\gamma}_i'\mathbf{J}'$.

How the terms $g_i a_i^2\mathbf{d}_i'\mathbf{J}\mathbf{d}_i$ and $-2g_i a_i\mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i$ can be majorized by functions that are quadratic in the coordinates, is explained in appendix.

3.2 Updating X

We replace the second and third terms of equation (12) by their majorizing functions consisting of the sum of (31), (35), (39), and (42) developed in appendix.

The following majorizing function is then obtained for the $\mathbf{x}_i$

$$g(\mathbf{x}_i) = k_x + tr\mathbf{A}_x\mathbf{x}_i + \sum_{q=1}^{3} \text{tr}\mathbf{B}_{qx}\mathbf{x}_i\mathbf{C}_{qx}\mathbf{x}_i' \tag{14}$$

with $k_x = k_{1x}+k_{2x}+k_{3x}+k_{4x}+k_{5x}$ and $\mathbf{A}_x = -2(\mathbf{A}_{1x}+\mathbf{A}_{2x}+\mathbf{A}_{3x}+\mathbf{A}_{4x}+\mathbf{A}_{5x})$. This majorizing function can in turn be majorized as done by Kiers (2002) (see equation (9)),

$$g(\mathbf{x}_i) \leq k_x + \text{tr}\mathbf{F}_x\mathbf{x}_i + \sum_{q=1}^{3} \lambda_{qx}\text{tr}\mathbf{C}_{qx}\mathbf{x}_i'\mathbf{x}_i, \tag{15}$$

with λ_{qx} the highest eigenvalue of $\left(2^{-1}\mathbf{B}_{qx} + 2^{-1}\mathbf{B}_{qx}'\right)$ and with $\mathbf{F}_x = \mathbf{A}_x + \sum_q \left[\mathbf{C}_{qx}\mathbf{x}_{i0}'(\mathbf{B}_{qx} + \mathbf{B}_{qx}') - 2\lambda_{qx}\mathbf{C}_{qx}\mathbf{x}_{i0}'\right]$. As $\mathbf{B}_{2x}$ and $\mathbf{B}_{3x}$ are scalars, λ_{2x} and λ_{3x} equal these and only λ_{1x} has to be found by an eigenvalue decomposition. Note that all $\mathbf{C}_{qx}$ are positive semi-definite such that we could use (9) as a majorizing function. When there are no constraints, the optimal $\mathbf{x}_i$ is given by

$$\mathbf{x}_i^+ = -\frac{1}{2}\mathbf{F}_x' \left(\sum_q \lambda_{qx}\mathbf{C}_{qx}\right)^{-1}. \tag{16}$$

A relaxed update $\mathbf{x}_i^{++}$ can be used to accelerate convergence (see De Leeuw & Heiser, 1980 and Heiser, 1995),

$$\mathbf{x}_i^{++} = 2\mathbf{x}_i^+ - \mathbf{x}_i. \tag{17}$$

The relaxed update still guarantees a nonincreasing sequence of loss values. As the formulas connected to the update of $\mathbf{x}_i$ do not involve another judge i', the order in which the judges are updated does not influence the outcome.

3.3 Updating Y

The majorizing function for the $\mathbf{Y}$ coordinates is given by replacing the second and third terms of equation (13) by the sum of (32), (36), (40), and (43) of the appendix yielding

$$g(\mathbf{Y}) = k_y + \text{tr}\mathbf{A}_y\mathbf{Y} + \sum_{q=1}^{3} \text{tr}\mathbf{B}_{qy}\mathbf{Y}\mathbf{C}_{qy}\mathbf{Y}', \tag{18}$$

with $k_y = k_{1y}+k_{2y}+k_{3y}+k_{4y}+k_{5y}$ and $\mathbf{A}_y = -2(\mathbf{A}_{1y}+\mathbf{A}_{2y}-\mathbf{A}_{3y}+\mathbf{A}_{4y}+\mathbf{A}_{5y})$. The update is then given by

$$\mathbf{Y}^+ = \frac{1}{2}\mathbf{F}_y \left(\sum_q \lambda_{qy}\mathbf{C}_{qy}\right)^{-1}, \tag{19}$$

with $\mathbf{F}_y = \mathbf{A}_y + \sum_q \left(\mathbf{C}_{qy}\mathbf{Y}_0'(\mathbf{B}_{qy} + \mathbf{B}_{qy}') - 2\lambda_{qy}\mathbf{C}_{qy}\mathbf{Y}_0'\right)$ and λ_{qy} the highest eigenvalue of $\left(2^{-1}\mathbf{B}_{qy} + 2^{-1}\mathbf{B}_{qy}'\right)$. Given the special form of the $\mathbf{B}_{qy}$ matrices,

the eigenvalues can be set equal to respectively 1, $\sum_i c_{1i}$ and the maximal value on the diagonal of $\sum_i c_{3i}\mathbf{D}_{i0}^{-1}$. Here, too, the number of iterations can be approximately halved using the relaxed update:

$$\mathbf{Y}^{++} = 2\mathbf{Y}^{+} - \mathbf{Y}. \tag{20}$$

3.4 Updating $\boldsymbol{\Gamma}$

The loss for judge i can be written as,

$$L_i(\boldsymbol{\gamma}_i) = \|\boldsymbol{\gamma}_i - [g_i a_i \mathbf{d}_i + (1-g_i) b_i \mathbf{Y}\mathbf{x}_i]\|_{\mathbf{J}}^2. \tag{21}$$

This function can be minimized by (monotone) regressing the $\boldsymbol{\gamma}_i$'s on $g_i a_i \mathbf{d}_i + (1-g_i) b_i \mathbf{Y}\mathbf{x}_i$ and this under a nonnegativity constraint $\gamma_{ij} \geq 0$ and a length constraint $\boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i = 1$. Let $\boldsymbol{\delta}_i$ be the rank scores for judge i, then fixed disparities of the following form are obtained for data with an interval level of measurement (see also Van Deun et al., in press):

$$\boldsymbol{\gamma}_i = (\boldsymbol{\delta}_i'\mathbf{J}\boldsymbol{\delta}_i)^{-1/2}\boldsymbol{\delta}_i. \tag{22}$$

Equation (22) clearly shows that the update is independent from the coordinate values. Therefore, the $\boldsymbol{\gamma}_i$'s can be fixed right away. In the nonmetric case, as shown by Gifi (1990), the length restriction can be imposed after the monotone regression by dividing the updated disparities by $\boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i^{1/2}$. In case all γ_{ij} are equal, Groenen, Os, and Meulman (2000) show how the optimal disparities are given by $\boldsymbol{\gamma}_i = \mathbf{S}\mathbf{b}_{opt}$ with $\mathbf{S} = \mathbf{JLM}$ where $\mathbf{M}$ is a matrix of ones on and below the diagonal and zeroes above it, $\mathbf{L}$ is a matrix composed of rows with zeroes and a one in the r-th column of row j where r is the rank value of d_{ij}. Let $\mathbf{G} = \mathbf{S}'\mathbf{S}$, then $\mathbf{b}_{opt}$ is a vector of all zeroes except one which equals $(g)_{ii}^{-1}$ (see also Van Deun et al., in press). The nonzero value is chosen such that $\mathbf{d}'\mathbf{d} + 1 - 2\mathbf{b}'\mathbf{S}'\mathbf{d}$ is minimal. An additive constant can be added without loss of generality and this property can be used to make all disparities nonnegative (since $\mathbf{S} = \mathbf{JLM}$ is centered). Nonnegative γ_{ij}'s are important as the majorizing functions for $\mathbf{X}$ and $\mathbf{Y}$ depend on the assumption of nonnegative γ_{ij}'s.

3.5 Updating a and b

The optimal value for a_i can be found by setting the partial derivative of (7) with respect to a_i equal to zero:

$$\frac{\partial L_i}{\partial a_i} = 2g_i a_i \mathbf{d}_i'\mathbf{J}\mathbf{d}_i - 2g_i \mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i = 0. \tag{23}$$

Under the constraint that $a_i \geq 0$, the following update is obtained,

$$a_i = \begin{cases} \mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i(\mathbf{d}_i'\mathbf{J}\mathbf{d}_i)^{-1} & \text{if } a_i \geq 0 \\ 0 & \text{if } a_i < 0 \end{cases}. \tag{24}$$

In the same way, we find that the update for b_i is given by

$$b_i = \begin{cases} \boldsymbol{\gamma}_i'\mathbf{J}\mathbf{Y}\mathbf{X}_i(\mathbf{x}_i'\mathbf{Y}'\mathbf{J}\mathbf{Y}\mathbf{x}_i)^{-1} & \text{if } b_i \leq 0 \\ 0 & \text{if } b_i > 0 \end{cases}. \tag{25}$$

3.6 Updating G

Each g_i is evaluated by taking the difference $L(g_i = 1, \mathbf{x}_i, \mathbf{Y}, \boldsymbol{\gamma}_i, a_i, b_i) - L(g_i = 0, \mathbf{x}_i, \mathbf{Y}, \boldsymbol{\gamma}_i, a_i, b_i)$. If the difference is positive or zero, g_i is set equal to zero and else to one. So, in case of equal fit by both models, priority is given to the vector model.

3.7 A rational start

Good initial values $\mathbf{X}^{(0)}$ and $\mathbf{Y}^{(0)}$ can be found by a classical multidimensional scaling of the augmented data matrix. The augmented data matrix is obtained as described in Van Deun, Heiser, and Delbeke (2004) with the following steps,

1. Expand the $n \times m$ data with an $m \times m$ matrix that has ones on the diagonal and $1 + 2^{-1}m$ elsewhere.
2. Center each row.
3. Scale the additional m rows by multiplying them by $3^{-1/2}(m+1)^{1/2}$.
4. Derive the $(n+m) \times (n+m)$ matrix of dissimilarities by calculation of the Euclidean distance between all rows.
5. Add a constant to the distances measured between the first n and last m rows such that the mean resulting distance equals the mean off diagonal distance measured among the first n rows.

As $\boldsymbol{\Gamma}^{(0)}$ we take the one used with interval transformations given by (22). Then, optimal values can be obtained for $\mathbf{a}^{(0)}$, $\mathbf{b}^{(0)}$ and $\mathbf{g}^{(0)}$ by (24), (25) and as explained in Section 3.6.

4 The difference between outlying ideal points and vectors representation

We apply the combined vector ideal point or VIPSCAL model to the journal data collected by Roskam (1968): he asked 39 staff members to rank ten psychological journals according to their preference for consultation. Here, we chose an interval level of measurement. To account for the problem of local minima, a multistart procedure was used: 50 solutions were computed based on random initial configurations and we retained the solution with the lowest loss value. For each of the 50 analyses, the iterative procedure was stopped either when the difference in loss was less then 10^{-9} or when the number of iterations exceeded 1000. In Figure 3, the resulting combined vector ideal point configuration is depicted where the right panel is a detail of the left panel. Contrary to our expectations, it seems that not all distant ideal points are representable by vectors (see Figure 3a). As a result, we consider the solution as being degenerate. In the right panel of Figure 3, a detail of the left panel is shown in which the outlying subjects are plotted by vectors. Here, an interpretation of the preference

Table 1: Labels used for the journal data.

Labels for the journals	
JEXP	Journal of Experimental Psychology
JAPP	Journal of Applied Psychology
JPSP	Journal of Personality and Social Psychology
MuBR	Multivariate Behavioral Research
JCONS	Journal of Consulting
JEDU	Journal of Education
Pmet	Psychometrika
HuRe	Human Relations
PsBu	Psychological Bulletin
HumDev	Human Development
Labels for the judges	
1	Social psychology
2	Educational and developmental psychology
3	Clinical psychology
4	Mathematical psychology and psychological statistics
5	Experimental psychology
6	Cultural psychology and psychology of religion
7	Industrial psychology
9	Physiological and animal psychology

for most of the judges is possible: judges represented by an ideal point prefer a journal more the closer it is located to its ideal point, while judges represented by the dotted vectors prefer a journal more the higher it projects on its vector. No interpretation is possible on the basis of the configurations in Figure 3 for the distant judges (represented by ideal points in the left panel and by vectors in the right panel). Consider the projections for outlying judge '3' which are drawn in the right panel. Two pairs of projections fall close together: these are the pair 'JCONS' and 'HuRe' and the pair 'JEDU' and 'JAPP'. 'JCONS' lies closer to the ideal point then 'HuRe' but it is 'HuRe' that projects higher on the vector. Similarly, 'JAPP' is closer to the ideal point then 'JEDU' while it is the latter that projects higher on the vector. In other words, the reproduced preference orders under the vector and ideal point model are not the same for this distant subject.

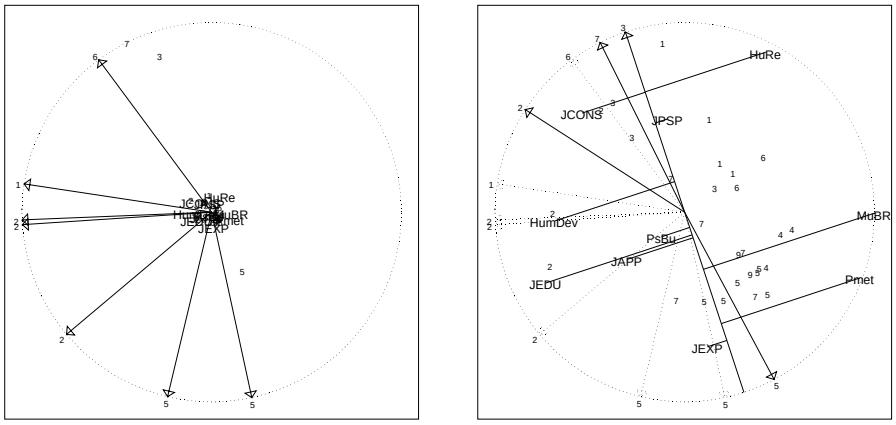
To understand what has happened, consider the situation with two points A and B and a judge S depicted in Figure 4. The squared Euclidean distance from S to A is given by $d^2(S,A) = x^2+y^2$ and also $d^2(S,B) = (x+a)^2 = x^2+a^2+2ax$. In the vector model, the distance from S to the projection is considered so $d_v^2(S,A) = x^2$ and $d_v^2(S,B) = (x+a)^2 = x^2+a^2+2ax$. The preference order is inferred by comparing the distances, for the ideal point model:

$$d^2(S,A) - d^2(S,B) = y^2 - a^2 - 2ax \tag{26}$$

and for the vector model

$$d_v^2(S,A) - d_v^2(S,B) = -a^2 - 2ax. \tag{27}$$

Then, if $y^2 > a^2 + 2ax$ the preference orders are reversed when comparing both models. The reversal can be accomplished with small y when $x \to \infty$ by taking



(a) (b)

Figure 3: VIPSCAL solution for the journal data (interval analysis). The right panel is a detail of the left panel with the distant judges represented by the full vectors and the judges fitting a vector representation by the dotted vectors. The journals are labelled with the letter codes and the subjects with a number (see Table 1).

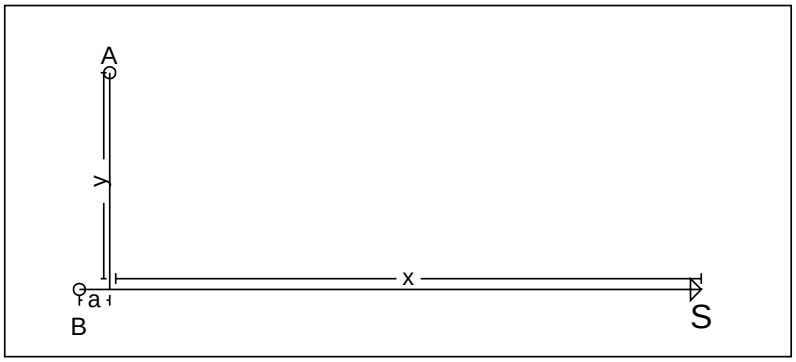


Figure 4: Judge 'S' and two objects 'A' and 'B'.

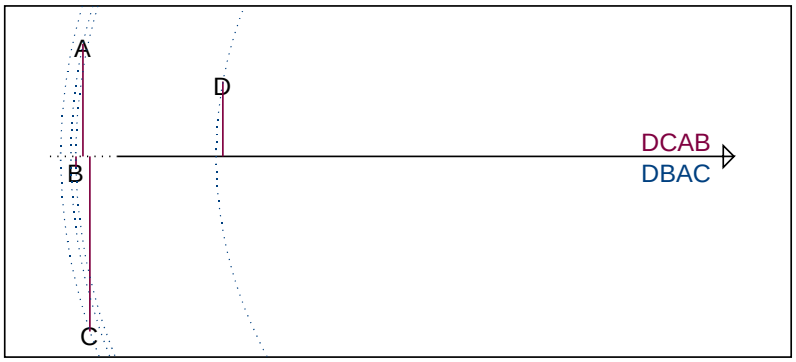


Figure 5: Distant ideal point for which the reproduced preference orders under the distance model do not equal those reproduced under the vector model.

a small ($a \to 0$) or the objects lie almost on a line orthogonal to the vector,

$$\forall\ x > 1\ \exists\ y, a > 0 : y^2 > a^2 + ax \tag{28}$$

where the condition is fulfilled for example by taking $a = x^{-1}$, $y = 2$ and this even for very large x. Note that this is exactly what happens with the Roskam data: in Figure 3b 'HuRe' and 'JCONS' lie almost on a line orthogonal to the vector labelled with '3' (or, a is very small). In practice, this observation means that some ideal points might never be represented by vector points, no matter how distant they are. This aspect is further illustrated in Figure 5, where the projection lines and isopreference contours are drawn for a distant subject in a situation where the projection lines fall close together for three objects though these objects are separated from each other.

How can we understand the difference between a very distant ideal point and a vector representation from a substantive point of view? Let us consider a hypothetical example of two companies that select their employees on the basis of their performance on two intelligence tests, one testing verbal intelligence and the other performance intelligence (see Figure 6). Furthermore, both companies occupy the same position in a preference space made up by a performance IQ dimension and a verbal IQ dimension but company I uses the vector model and company II the distance model to rank the candidates. The position these companies occupy is characterized by a very high and equal value on both dimensions with the consequence that both rank candidates with respect to their overall intelligence: the higher the overall intelligence of a candidate, the more he is preferred by both companies. Suppose now that there are six candidates 'A', 'B', 'C', 'D', 'E', and 'F' as depicted in Figure 6. Most candidates are ranked the same by both companies, except for the two candidates with the highest overall IQ: these are candidates 'A' and 'B' who differ little with respect

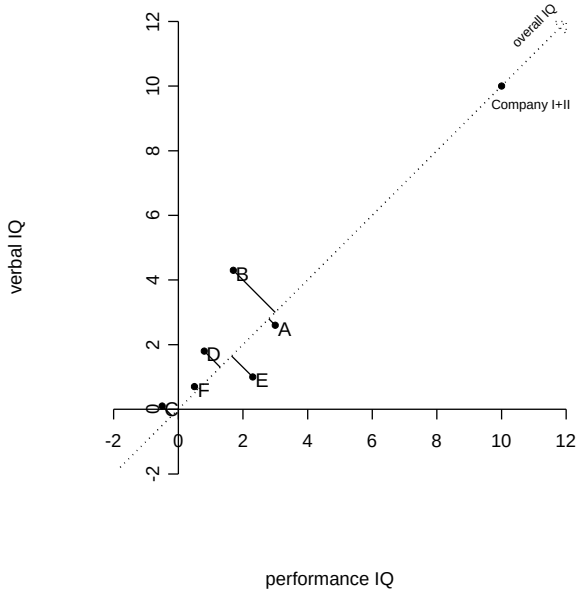


Figure 6: Illustration of the substantive difference between a distant ideal point and a vector representation for a hypothetical example.

to their overall intelligence. Company I selects the candidate with the highest overall IQ (this is, candidate 'B') while company II selects, among candidates with an almost equal overall score, the candidate who scores more or less the same on both dimensions (this is, candidate 'A') .

What the hypothetical example illustrates, is that the difference between vectors and distant ideal points is connected with the relative importance of the dimensions that make up the preference space: the relative amount in which each of the dimensions contribute determines the preference scores more for distant ideal points then for vectors. In the latter case, a low value on one or more dimension can be compensated for by extremely high values on other dimensions. Note that this difference between vectors and distant ideal points is only important when the orthogonal projections fall close together. However, given the number of objects and subjects that make up a preference data set one will often encounter close projections.

5 Adding constraints

To avoid degenerate solutions in the VIPSCAL model, we will build in some constraints. A first constraint is to set $\|\mathbf{x}_i\|^2 \leq 1$ or, equivalently, $\sum_p x_{ip}^2 \leq 1$. In this way, the maximal distance from a subject to the center is one and this should avoid that ideal points run away. However, the same type of solution can still be obtained if the object points cluster together in the center. Therefore, we add the constraint $\mathbf{Y}'\mathbf{Y} = m(2p)^{-1}\mathbf{I}$. The factor $m(2p)^{-1}$ is chosen in order to have the $\mathbf{x}_i$ and $\mathbf{y}_j$ points in the same range of values: the part mp^{-1} accounts for the fact that on average, we want each $\mathbf{y}_j$ point to have a length one. The factor 2^{-1} was found by experience and accounts for the fact that not all $\mathbf{x}_i$ have a length one. In this way, the average distance of the item and subject points to the origin should be approximately equal.

As explained by Kiers (2002), the equality constraint $\|\mathbf{x}_i\|^2 = 1$ can be imposed by adapting Steps 4 and 5 (see section 3 on page 6) as follows. Equation (15) can be rewritten as a linear function under the constraint $\mathbf{x}_i'\mathbf{x}_i = 1$ as $k_x + \text{tr}\mathbf{F}_x\mathbf{x}_i + \sum_{q=1}^{3} \lambda_q \text{tr}\mathbf{C}_{qx}$. This function can be minimized by maximizing $\text{tr}-\mathbf{F}_x\mathbf{x}_i$ so that the update is given by $\mathbf{x}^+ = \mathbf{V}_x\mathbf{U}_x'$ where $\mathbf{V}_x$ and $\mathbf{U}_x$ come from the singular value decomposition of $-\mathbf{F}_x = \mathbf{U}_x\mathbf{\Sigma}_x\mathbf{V}_x'$. To impose the inequality constraint $\|\mathbf{x}_i\|^2 \leq 1$, first calculate the unconstrained update. If the resulting $\mathbf{x}_i$ has a length greater then 1, compute the constrained update. In the same way, the constrained update for $\mathbf{Y}$ is given by $\mathbf{Y}^+ = m^{1/2}(2p)^{-1/2}\mathbf{V}_y\mathbf{U}_y'$. In case of constraints, the relaxed update is no longer of use as convergence cannot be guaranteed anymore. Note that a model with all judges on a circle can be obtained by calculating the constrained update for all i. Such a model has compensatory characteristics as the sum of squared coordinates is equal to one, implying that large (absolute) values on one dimension result in small values on the remaining dimensions. It is also possible to constrain the norm of the $\mathbf{x}_i$ to another length then one. Taking a small value will yield an object points degeneracy which can be interpreted by using a signed compensatory

Table 2: Models that can be analyzed with the algorithm proposed here. The length constraints on $\mathbf{x}_i$ are always combined with the orthogonality constraint $\mathbf{Y}'\mathbf{Y} = m(2p)^{-1}\mathbf{I}$. VIP stands for a mixed vector ideal point configuration and MDU for a full ideal point or multidimensional unfolding configuration.

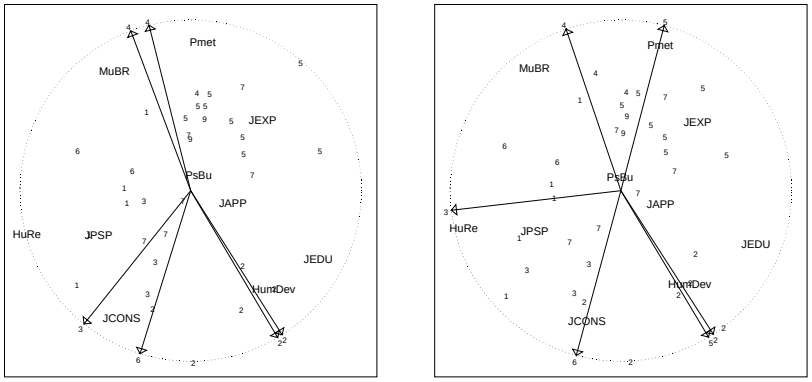
	Representation of the judges	Length Constraint	Model
1	Mixed	No	Vector Ideal Point
2	Mixed	$\|\mathbf{x}_i\| \leq 1$	Constrained VIP
3	Mixed	$\|\mathbf{x}_i\| = 1$	Circular VIP
4	Mixed	$\|\mathbf{x}_i\| \leq 0.01$	Compensatory VIP
5	All Ideal Points	No	Unfolding
6	All Ideal Points	$\|\mathbf{x}_i\| \leq 1$	Constrained MDU
7	All Ideal Points	$\|\mathbf{x}_i\| = 1$	Circular MDU
8	All Ideal Points	$\|\mathbf{x}_i\| \leq 0.01$	Compensatory MDU
9	All Vectors	Length is of minor importance	Vector

distance model (see Van Deun et al., in press).

The resulting algorithm incorporates nine different models, determined by two factors: the first factor pertains to the representation of the judges (mixture of vectors and ideal points, all ideal points or all vectors), the second to the length constraint (no constraint, a maximum length of one, an exact length of one or a maximal small length). Note that the length factor is of no importance for the interpretation of a full vector model. Constraining the length of the vectors, however, has an influence on the estimation of the coordinates. Therefore, with the same starting configuration, the constrained configurations slightly differ from the unconstrained configurations. Using different lengths in the fully constrained model makes no difference as the loss function is not influenced by linear transformations (a change in length is compensated for by a change in the value of the regression parameter b_i). The models with only ideal points incorporate the classical unfolding model (no length constraint) that is known to degenerate in the majority of cases, but also three different constrained unfolding models. To our knowledge, VIPSCAL is the first unfolding algorithm that allows for length restrictions on the configuration.

6 Applications of the constrained models

Some of the constrained analyses are illustrated here. As a first application, we reconsider the mixed model for the journal data(Roskam, 1968) but now with the inequality constraint $\|\mathbf{x}_i\|^2 \leq 1$. Then, we consider several ways of modelling the breakfast data (Green & Rao, 1972). In all cases, the same stopping criteria as previously are used. Here, we will show two solutions for each model: one based on a multistart procedure with 50 random starts and one based on rational starting values. For each solution, the loss value and the average proportion of recovered preference orders are reported. The latter measure is also known as index $C1$ (see Kim et al., 1999) and is calculated in the following way: for each



(a) Rational start. $Loss = 0.2880$, $C1 = 0.8450$

(b) Best of 50 random starts. $Loss = 0.3171$, $C1 = 0.8251$

Figure 7: Roskam data, constrained VIP model with interval transformations. In the left panel, the rational based solution is given and in the right panel the best of 50 random based solutions. For each solution, the average squared correlation and the average proportion of recovered preference orders is reported.

subject, the number of pairs of objects for which the configuration preserves the order relation is calculated and this number is divided by the total number of pairs, this is $m(m-1)/2$. Then, the average is taken over all subjects.

To have an illustration of metric analyses, the journal data were analysed using interval transformations (see also the end of this section). The solutions for the constrained VIP analysis are given in Figure 7 where the solution in the left panel is based on rational starting values while the solution in the right panel is the best of those based on random values. Both solutions are very similar with the rational one having a lower loss value and a better fit. In these configurations, the preference of each judge can be easily derived: for example, the group of people who consider the department of experimental psychology most important for their work (labelled by '5') often consult the Journal of Experimental Psychology, Psychological Bulletin and Psychometrika while they least consult the Journal of Consulting, Human Relations, and Human Development. The judges labelled by '2' (educational and developmental psychology), on the other hand, often consult Human Development, the Journal of Education, and the Journal of Consulting while they least consult Multivariate Behavioral Research and Psychometrika. Note the central position of Psychological Bulletin.

In a second application, we illustrate some more of the possibilities the algorithm offers. The data we use here, are the breakfast data on preference for 15 breakfast items expressed by 21 MBA students and their wives (Green & Rao, 1972). The different constrained models that we will consider are the mixed con-

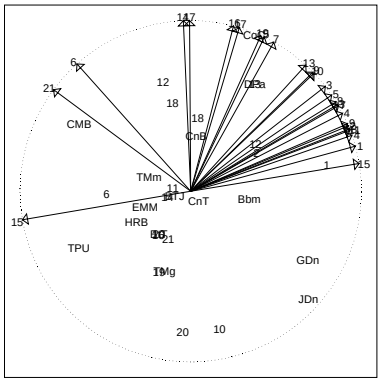
Table 3: Value of the loss function and proportion of recovered preference orders for different models applied to the breakfast data.

	Multistart		Rational	
	Loss	$C1$	Loss	$C1$
Mixed, $\|\mathbf{x}_i\| \leq 1$	0.2983	0.6927	0.1880	0.7612
All vectors, $\|\mathbf{x}_i\| \leq 1$	0.2020	0.6884	0.2025	0.6839
All ideal points, $\|\mathbf{x}_i\| \leq 1$	0.2610	0.6753	0.2314	0.7512
All ideal points, $\|\mathbf{x}_i\| = 1$	0.1876	0.6524	0.2311	0.7624

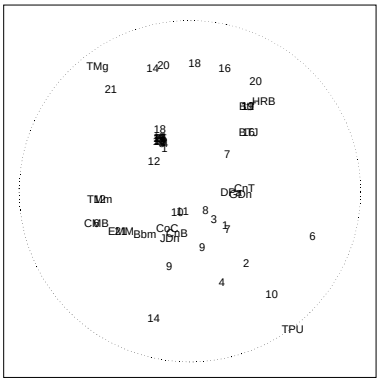
strained model, the constrained unfolding model, the constrained vector model and the circular unfolding model.

An overview of the loss and fit values is given in Table 3, the graphical representations for the random solutions can be found in Figure 8 and for the rational solutions in Figure 9. A comparison of the random and rational configurations shows that the latter are clearly more interpretable, except for the constrained vector model. For example, most of the breakfast items are clustered together in the mixed configuration of Figure 8a, while the rational solution in Figure 9a clearly shows a cluster of the hard items (with the exception of toast pop-up, 'TPU'), of the doughnuts and, of most of the soft items. Similarly, taking a look at the configuration in Figure 8b, we see that a lot of breakfast items are clustered. Note also that 14 subjects (the large dot in the configuration) are clustered together in the center of most of the breakfast items. The solution for the same model but based on the rational starting configuration is again clearly interpretable (see Figure 9b). And finally, if we consider the circular ideal point solutions (see Figures 8d and 9d), we see that the breakfast items and a lot of judges (in fact, 15, this is the large dot at the right side of the configuration) are clustered together in the center of almost all breakfast items while the rational solution is again interpretable. The vector model (see Figures 8c and 9c), however, holds very similar configurations that are both not interpretable: most vectors lie in a direction orthogonal to the breakfast items such that the projections of the items on the vectors fall close together. A comparison of the columns labelled 'Loss' in Table 3, shows that the rational solutions have lower values for the models with constraint $\mathbf{x}_i \leq 1$ in two cases (mixed and ideal points). For the circular ideal point model and for the vector model, the solutions from the multistart procedure have a lower loss value. An indication of the fit of the configurations to the data can be found in the columns labelled $C1$ in Table 3: with the exception of the vector model, the rational solutions clearly fit better.

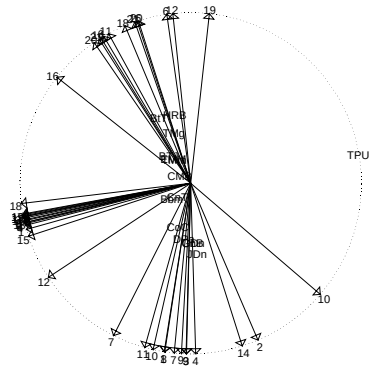
We further evaluated the constrained models for several other empirical data sets (being the data on preference for family structures of Delbeke, 1968, data on rankings of persons tested on their English speech, and the MBA and aspirin data which are described in Kim et al., 1999). The breakfast (Green & Rao, 1972) and journal data (Roskam, 1968) were included too, so in total we



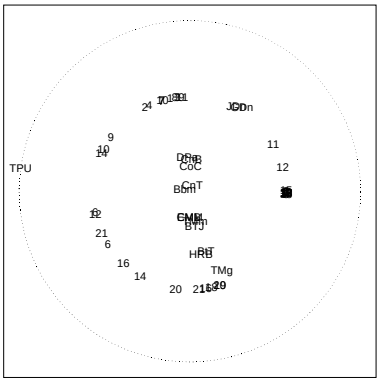
(a) Constrained VIP.



(b) Constrained MDU.

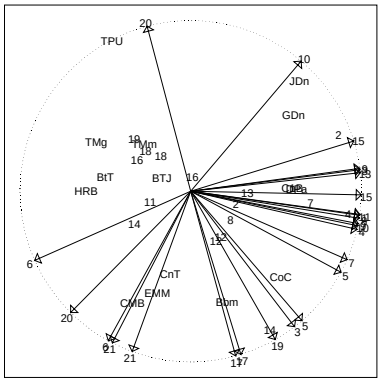


(c) Constrained vector model.

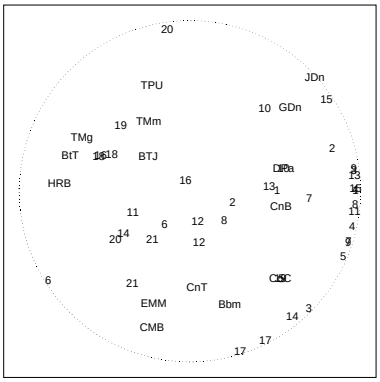


(d) Circular MDU.

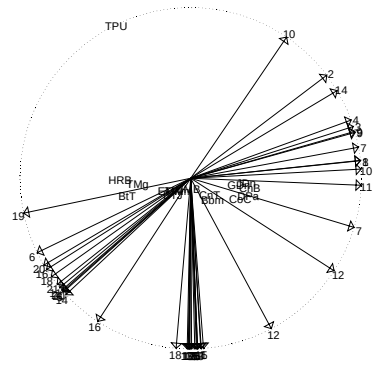
Figure 8: Breakfast data, random start. The breakfast items are labelled by the three letter codes and the 21 couples who ranked the items by the numbers.



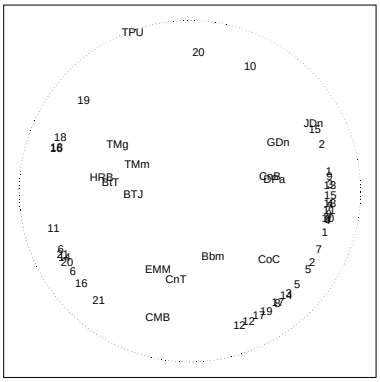
(a) Constrained VIP.



(b) Constrained MDU.



(c) Constrained vector model.



(d) Circular MDU.

Figure 9: Breakfast data, rational start. The breakfast items are labelled by the three letter codes and the 21 couples who ranked the items by the numbers.

Table 4: Table of the number of the six data sets analyzed, that yielded an interpretable configuration. The columns refer to six different restricted models and the rows to a combination of the measurement level and starting conditions.

	mixed		ideal points		vectors	
	$\mathbf{x}_i \leq 1$	$\mathbf{x}_i = 1$	$\mathbf{x}_i \leq 1$	$\mathbf{x}_i = 1$	$\mathbf{x}_i \leq 1$	$\mathbf{x}_i = 1$
Interval: Best	5	6	6	5	6	6
Interval: Rational	6	6	6	6	6	6
Ordinal: Best	2	2	4	2	0	0
Ordinal: Rational	5	4	6	4	0	0

considered six data sets. We are mainly interested in the interpretability of the configurations. For this reason, the model with the $\|\mathbf{x}_i\|$ restricted to a small value is not included here.

Table 4 reports the number of data sets that yielded an interpretable configuration for each of the six restricted models. The different rows of the table refer to a combination of the measurement level and the starting conditions: 'Best' reports the results for the solution with the lowest loss of 51, with 50 solutions based on random starts and one on a rational start. A comparison of the first two lines of numbers in Table 4 with the last two lines, shows that almost all metric solutions are interpretable while this is clearly not the case for the nonmetric solutions. To avoid the occurrence of degenerate solutions, the restrictions used here are clearly not sufficient for ordinal analyses. This problem can be overcome in some cases by using a rational starting configuration. With respect to the different models, two observations can be made: first, none of the pure vector configurations analyzed at an ordinal level was interpretable and second, the best results were obtained for the restricted ideal point model. Concerning the fit of the solution to the data, as measured by index $C1$ (not reported here), we observed mainly that higher values were found for the metric analyses. There were also differences between the different models: the circular and vector models fitted less well and this effect was somewhat more pronounced for the metric case.

7 Discussion

We proposed a loss function that combines the ideal point model of unfolding and the vector model. An alternating least squares and iterative majorization algorithm was developed, named VIPSCAL, that can be used to minimize this loss. Contrary to our initial expectations, degenerate solutions were obtained. Inspection of these degeneracies showed that, contrary to what is usually believed, representing a (very) distant ideal point by a vector does not necessarily hold the same preference ranking: for objects that have almost equal values on the line connecting the point with the origin (the *direction*), the vector model only takes the value on the direction into account while the distance model takes the relative contributions of each of the dimensions into account.

To further prevent the occurrence of distant ideal points, restrictions on the configuration were added and it was shown how these could easily be incorporated in the algorithm. As a result, VIPSCAL envelopes nine different models including the classical ideal point and vector model but also, for example, a constrained ideal point model and an ideal point model with all ideal points on a circle. A small study of the performance of the constrained VIPSCAL algorithm on six empirical data sets, showed that it seems to perform satisfactorily in the metric case: well fitting, nondegenerate solutions were obtained for all data sets when using a rational start and for almost all cases when using a multistart procedure. In the nonmetric case, the restrictions proposed here are not always enough to avoid degenerate solutions. Here, it seems important to use a rational start, but even then a degenerate solution might appear. Only for the ideal point model with the restriction $\|\mathbf{x}_i\| \leq 1$ a nice solution was obtained for all data sets. For the vector model, all nonmetric solutions were fitting badly and not interpretable. Summarized, VIPSCAL seems to give nice solutions for interval data and, with a rational start, in the majority of cases for ordinal analyses and especially for the constrained ideal point model. Further evaluation of the proposed algorithm is needed, with special attention to the issue of rational starting configurations. The importance of good starting values in order to avoid degeneracies in multidimensional unfolding, was already stressed out by Roskam (1977) but little research into this topic has been done so far. It is possible that better ways of obtaining a rational start exist then the one proposed here.

Extensions of the proposed loss function are possible. For example, the compensatory distance model could be included in the loss function and the tools of alternating least squares and iterative majorization can be used to develop an accompanying algorithm. We do not believe, however, that such a model would solve the problem of degenerate solutions and it has the disadvantage of complicating the interpretation of the configuration. More interesting is the use of VIPSCAL to study the relation between the vector and ideal point models. For example, with small adaptations of VIPSCAL an algorithm can be built that allows for a latent class analysis of a combined vector ideal point model that yields two different object configurations, one for the ideal points and one for the vectors. In fact, the merits of VIPSCAL are to be sought more in the possibilities it offers to get a better understanding of the relation between the vector and ideal point model then in it's capability of avoiding degenerate solutions.

A Majorizing $g_i a_i^2 \mathbf{d}_i' \mathbf{J} \mathbf{d}_i$

Here, we show how the first term of (11) involving distances can be majorized by a function of the coordinates that is of the form (8). A good reference for finding majorizing inequalities is Groenen (2002). The centering operator $\mathbf{J}$ equals $\mathbf{I} - m^{-1}\mathbf{1}\mathbf{1}'$, so that

$$g_i a_i^2 \mathbf{d}_i' \mathbf{J} \mathbf{d}_i = g_i a_i^2 \mathbf{d}_i' \mathbf{d}_i - \frac{g_i a_i^2}{m} \mathbf{d}_i' \mathbf{1}\mathbf{1}' \mathbf{d}_i, \tag{29}$$

The first term in the right hand side of (29) can be rewritten as

$$\begin{aligned} g_i a_i^2 \mathbf{d}_i' \mathbf{d}_i &= g_i a_i^2 \mathrm{tr} (\mathbf{1}\mathbf{x}_i' - \mathbf{Y})' (\mathbf{1}\mathbf{x}_i' - \mathbf{Y}) \\ &= c_{1i} \mathrm{tr} \mathbf{x}_i \mathbf{1}'\mathbf{1} \mathbf{x}_i' - 2 c_{1i} \mathrm{tr} \mathbf{x}_i \mathbf{1}' \mathbf{Y} + c_{1i} \mathrm{tr} \mathbf{Y}' \mathbf{Y} \end{aligned} \tag{30}$$

with $c_{1i} = g_i a_i^2$. Rewriting (30) in function of $\mathbf{x}_i$ gives,

$$g_i a_i^2 \mathbf{d}_i' \mathbf{d}_i = k_{2x} + \mathrm{tr} \mathbf{B}_{2x} \mathbf{x}_i \mathbf{C}_{2x} \mathbf{x}_i' - 2 \mathrm{tr} \mathbf{A}_{2x} \mathbf{x}_i, \tag{31}$$

with $k_{2x} = c_{1i} \mathrm{tr} \mathbf{Y}'\mathbf{Y}$, $\mathbf{B}_{2x} = c_{1i}$, $\mathbf{C}_{2x} = \mathbf{1}'\mathbf{1}$, and $\mathbf{A}_{2x} = c_{1i}\mathbf{1}'\mathbf{Y}$. In function of $\mathbf{Y}$ we get

$$\sum_i g_i a_i^2 \mathbf{d}_i' \mathbf{d}_i = k_{2y} + \mathrm{tr} \mathbf{B}_{2y} \mathbf{Y} \mathbf{C}_{2y} \mathbf{Y}' - 2 \mathrm{tr} \mathbf{A}_{2y} \mathbf{Y}, \tag{32}$$

with $k_{2y} = \sum_i c_{1i} \mathrm{tr} \mathbf{x}_i \mathbf{1}'\mathbf{1} \mathbf{x}'$, $\mathbf{B}_{2y} = \sum_i c_{1i}$, $\mathbf{C}_{2y} = \mathbf{I}$, and $\mathbf{A}_{2y} = \sum_i c_{1i} \mathbf{x}_i \mathbf{1}'$. Consider the second term in the right hand side of (29): $-\mathbf{d}_i' \mathbf{1}\mathbf{1}' \mathbf{d}_i$ is minus the squared sum of d_{ij}'s which can be majorized using the inequality $-cs^2 \leq cs_0^2 - 2css_0$, with c a positive constant. Take $c = m^{-1} g_i a_i^2$ and $s = \mathbf{d}_i' \mathbf{1}$, we then obtain

$$\begin{aligned} -\frac{g_i a_i^2}{m} \mathbf{d}_i' \mathbf{1}\mathbf{1}' \mathbf{d}_i &\leq \frac{g_i a_i^2}{m} \mathbf{d}_{i0}' \mathbf{1}\mathbf{1}' \mathbf{d}_{i0} - \frac{2 g_i a_i^2}{m} \mathbf{d}_i' \mathbf{1}\mathbf{1}' \mathbf{d}_{i0} \\ &= \frac{g_i a_i^2}{m} \mathbf{d}_{i0}' \mathbf{1}\mathbf{1}' \mathbf{d}_{i0} - \frac{2 g_i a_i^2}{m} (\mathbf{1}'\mathbf{d}_{i0}) \mathbf{d}_i' \mathbf{1}. \end{aligned} \tag{33}$$

The last term of function (33) still contains distances that depend on $\mathbf{x}_i$ and $\mathbf{Y}$. The Cauchy Schwarz inequality can be used to obtain a majorizing function of the coordinates as $-d = -\|\mathbf{x} - \mathbf{y}\| \leq -d_0^{-1} (\mathbf{x} - \mathbf{y})' (\mathbf{x}_0 - \mathbf{y}_0)$. Set $c_{2i} = m^{-1} g_i a_i^2 (\mathbf{1}'\mathbf{d}_{i0})$ so that

$$\begin{aligned} -2 c_{2i} \mathbf{d}_i' \mathbf{1} &\leq -2 c_{2i} \mathrm{tr} (\mathbf{1}\mathbf{x}_i' - \mathbf{Y})' \mathbf{D}_{i0}^{-1} (\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0) \\ &= -2 c_{2i} \mathrm{tr} \mathbf{x}_i \mathbf{1}' \mathbf{D}_{i0}^{-1} (\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0) + 2 c_{2i} \mathrm{tr} \mathbf{Y}' \mathbf{D}_{i0}^{-1} (\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0), \end{aligned} \tag{34}$$

with $\mathbf{D}_{i0}^{-1}$ the inverse diagonal matrix of $\mathbf{d}_{i0}$'s. When $d_{i0} = 0$, we set the corresponding majorizing term equal to zero. In function of the $\mathbf{x}_i$ we obtain

$$-\frac{g_i a_i^2}{m} \mathbf{d}_i' \mathbf{1}\mathbf{1}' \mathbf{d}_i \leq k_{3x} - 2 \mathrm{tr} \mathbf{A}_{3x} \mathbf{x}_i, \tag{35}$$

with $k_{3x} = c_{2i}[\mathbf{1}'\mathbf{d}_{i0} + 2\text{tr}\mathbf{Y}'\mathbf{D}_{i0}^{-1}(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)]$ and $\mathbf{A}_{3x} = c_{2i}\mathbf{1}'\mathbf{D}_{i0}^{-1}(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)$. In function of $\mathbf{Y}$

$$-\sum_i \frac{g_i a_i^2}{m}\mathbf{d}_i'\mathbf{1}\mathbf{1}'\mathbf{d}_i \leq k_{3y} + 2\text{tr}\mathbf{A}_{3y}\mathbf{Y}, \tag{36}$$

with $k_{3y} = \sum_i c_{2i}[\mathbf{1}'\mathbf{d}_{i0} - 2\text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{D}_{i0}^{-1}(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)]$ and $\mathbf{A}_{3y} = \sum_i c_{2i}\mathbf{D}_{i0}^{-1}(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)$.

B Majorizing $-2g_i a_i \mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i$

First we replace $\mathbf{J}$ by it's full expression,

$$-2g_i a_i \mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i = -2g_i a_i \mathbf{d}_i'\boldsymbol{\gamma}_i + \frac{2g_i a_i}{m}\mathbf{d}_i'\mathbf{1}\mathbf{1}'\boldsymbol{\gamma}_i. \tag{37}$$

Both terms in the right hand side can be majorized. For the first term, which is again minus a linear function of the distances provided that the γ_{ij} are non-negative, we get by the Cauchy Schwarz inequality

$$\begin{aligned} -2g_i a_i \mathbf{d}_i'\boldsymbol{\gamma}_i &\leq -2g_i a_i \text{tr}(\mathbf{1}\mathbf{x}_i' - \mathbf{Y})'\mathbf{D}_{i0}^{-1}\boldsymbol{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0) \\ &= -2g_i a_i \text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{D}_{i0}^{-1}\boldsymbol{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0) + \\ &\quad 2g_i a_i \text{tr}\mathbf{Y}'\mathbf{D}_{i0}^{-1}\boldsymbol{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0), \end{aligned} \tag{38}$$

where $\boldsymbol{\Gamma}_i$ is a diagonal matrix of $\boldsymbol{\gamma}_i$ and all $\boldsymbol{\gamma}_{ij} \geq 0$. In function of the $\mathbf{x}_i$ this is

$$-2g_i a_i \mathbf{d}_i'\boldsymbol{\gamma}_i \leq k_{4x} - 2\text{tr}\mathbf{A}_{4x}\mathbf{x}_i, \tag{39}$$

with $k_{4x} = 2g_i a_i \text{tr}\mathbf{Y}'\mathbf{D}_{i0}^{-1}\boldsymbol{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)$ and $\mathbf{A}_{4x} = g_i a_i \mathbf{1}'\mathbf{D}_{i0}^{-1}\boldsymbol{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)$ while in function of $\mathbf{Y}$

$$-2\sum_i g_i a_i \mathbf{d}_i'\boldsymbol{\gamma}_i \leq k_{4y} + 2\text{tr}\mathbf{A}_{4y}\mathbf{Y}. \tag{40}$$

Here, $k_{4y} = -2\sum_i g_i a_i \text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{D}_{i0}^{-1}\boldsymbol{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)$ and $\mathbf{A}_{4y} = \sum_i g_i a_i \mathbf{D}_{i0}^{-1}\boldsymbol{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)$. The second term in 37, which contains positive sums of distances, can be majorized by the inequality $cs \leq c(2s_o)^{-1}s^2 + c2^{-1}|s_0|$ with c a positive constant. Take $c_{3i} = m^{-1}a_i g_i(\mathbf{1}'\boldsymbol{\gamma}_i)$

$$\begin{aligned} c_{3i}\mathbf{d}_i'\mathbf{1} &\leq c_{3i}\text{tr}(\mathbf{1}\mathbf{x}_i' - \mathbf{Y})'\mathbf{D}_{i0}^{-1}(\mathbf{1}\mathbf{x}_i' - \mathbf{Y}) + c_{3i}\mathbf{d}_{i0} \\ &= c_{3i}\text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{D}_{i0}^{-1}\mathbf{1}\mathbf{x}_i' + c_{3i}\text{tr}\mathbf{Y}'\mathbf{D}_{i0}^{-1}\mathbf{Y} - 2c_{3i}\text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{D}_{i0}^{-1}\mathbf{Y} + c_{3i}\mathbf{d}_{i0} \end{aligned} \tag{41}$$

We rewrite the majorizing function (41) in function of the $\mathbf{x}_i$,

$$c_{3i}\mathbf{d}_i'\mathbf{1} \leq k_{5x} + \text{tr}\mathbf{B}_{3x}\mathbf{x}_i\mathbf{C}_{3x}\mathbf{x}_i' - 2\text{tr}\mathbf{A}_{5x}\mathbf{x}_i, \tag{42}$$

with $k_{5x} = c_{3i}(\text{tr}\mathbf{Y}'\mathbf{D}_{i0}^{-1}\mathbf{Y} + \mathbf{d}_{i0})$, $\mathbf{B}_{3x} = c_{3i}$, $\mathbf{C}_{3x} = \mathbf{1}'\mathbf{D}_{i0}^{-1}\mathbf{1}$, and $\mathbf{A}_{5x} = c_{3i}\mathbf{1}'\mathbf{D}_{i0}^{-1}\mathbf{Y}$. In function of $\mathbf{Y}$,

$$\sum_i c_{3i}\mathbf{d}_i'\mathbf{1} \leq k_{5y} + \text{tr}\mathbf{B}_{3y}\mathbf{Y}\mathbf{C}_{3y}\mathbf{Y}' - 2\text{tr}\mathbf{A}_{5y}\mathbf{Y}, \tag{43}$$

with $k_{5y} = \sum_i c_{3i}(\text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{D}_0^{-1}\mathbf{1}\mathbf{x}_i' + \mathbf{d}_{i0})$, $\mathbf{B}_{3y} = \sum_i c_{3i}\mathbf{D}_{i0}^{-1}$, $\mathbf{C}_{3y} = \mathbf{I}$, and $\mathbf{A}_{5y} = \sum_i c_{3i}\mathbf{x}_i\mathbf{1}'\mathbf{D}_{i0}^{-1}$.

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. (1975). A note on the relation between the vector model and the unfolding model for preferences. *Psychometrika*, *40*, 115-116.

Delbeke, L. (1968). *Construction of preference spaces: an investigation into the applicability of multidimensional scaling models.* Leuven: Leuvense universitaire uitgaven.

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* (p. 308-325). Berlin: Springer-Verlag.

De Leeuw, J., & Heiser, W. J. (1980). Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (Ed.), *Multivariate analysis V* (pp. 501–522). Amsterdam: North Holland.

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

Gifi, A. (1990). *Nonlinear multivariate analysis.* Chicester, England: Wiley.

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. (2002). *Iterative majorization algorithms for minimizing loss functions in classification.* (Working paper presented at the 8th conference of the IFCS, Krakow, Poland)

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.

Heiser, W. J. (1995). Convergent computation by iterative majorization: theory and applications in multidimensional data analysis. In W. J. Krzanowski (Ed.), *Recent advances in descriptive multivariate analysis* (pp. 157–189). Oxford: Oxford University Press.

Kiers, H. A. L. (2002). Setting up alternating least squares and iterative majorization algorithms for solving various matrix optimization problems. *Computational Statistics and Data Analysis*, *41*, 157–170.

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., & Carroll, J. D. (1969). Geometrical models and badness-of-fit functions. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (Vol. 2, p. 639-671). New York: Academic Press.

Lange, K., Hunter, D. R., & Yang, I. (2000). Optimization transfer using surrogate objective functions. *Journal of computational and graphical statistics*, *9*, 1–20.

Meulman, J., Heiser, W. J., & Carroll, J. D. (1986). *PREFMAP-3 user's guide.* Bell Laboratories, Murray Hill NJ. H16. (Unpublished)

Roskam, E. E. C. I. (1968). *Metric analysis of ordinal data in psychology.* Voorshoten: VAM.

Roskam, E. E. C. I. (1977). A survey of the Michigan-Israel-Netherlands-Integrated series. In J. C. Lingoes (Ed.), *Geometric representations of relational data* (pp. 289–347). Ann Arbor, MI: Mathesis Press.

Van Deun, K., Groenen, P. J. F., Heiser, W. J., Busing, F. M. T. A., & Delbeke, L. (in press). Interpreting degenerate solutions in unfolding by use of the vector model and the compensatory distance model. *Psychometrika*.

Van Deun, K., Heiser, W. J., & Delbeke, L. (2004). Multidimensional unfolding by nonmetric multidimensional scaling of spearman distances in the extended permutation polytope. (Manuscript submitted for publication)

Young, F. W., Takane, Y., & De Leeuw, J. (1978). The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika*, *43*, 279–281.