Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2006) 360–378
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

I-Scal: Multidimensional scaling of interval dissimilarities

P.J.F. Groenen[a,*], S. Winsberg[b], O. Rodríguez[c], E. Diday[d]

[a] *Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands*
[b] *Predisoft, 18 Rue Visconti, Paris 75006, France*
[c] *CIMPA-PIMAD, Escuela de Matemática, Universidad de Costa Rica, 2060 San José, Costa Rica, Brazil*
[d] *CEREMADE, Université Paris Dauphine, F-75775 Paris Cédex 16, France*

Available online 2 May 2006

Abstract

Multidimensional scaling aims at reconstructing dissimilarities between pairs of objects by distances in a low-dimensional space. However, in some cases the dissimilarity itself is unknown, but the range of the dissimilarity is given. Such fuzzy data give rise to a data matrix in which each dissimilarity is an interval of values. These interval dissimilarities are modelled by the ranges of the distances defined as the minimum and maximum distance between two rectangles representing the objects. Previously, two approaches for such data have been proposed and one of them is investigated. A new algorithm called I-Scal is developed. Because I-Scal is based on iterative majorization it has the advantage that each iteration is guaranteed to improve the solution until no improvement is possible. In addition, a rational start configuration is proposed that is helpful in locating a good quality local minima. In a simulation study, the quality of this algorithm is investigated and I-Scal is compared with one previously proposed algorithm. Finally, I-Scal is applied on an empirical example of dissimilarity intervals of sounds.

Keywords: Multidimensional scaling; Interval-type data; Iterative majorization; I-Scal

1. Introduction

Classical multidimensional scaling (MDS) represents the dissimilarities among a set of objects as distances between points in a low-dimensional space. The aim of these MDS methods is to reveal relationships among the objects and uncover the dimensions giving rise to the space. Many of the applications of MDS revolve around the analysis of proximity data collected in studies related to the social sciences or to fields like product marketing and development (e.g., the objects studied may be acoustical sounds, consumer products, etc.). The goal in these studies is to visualize the objects and the distances among them and to discover the dimensions underlying the dissimilarity ratings.

Sometimes the proximity data are collected for n objects yielding a single dissimilarity matrix with the entry for the ith row and the jth column being the dissimilarity between the ith and jth object (with $i = 1, \ldots, n$ and $j = 1, \ldots, n$). Techniques for analyzing this form of data (two-way one mode) have been developed by Kruskal (1964a, b), see also Winsberg and Carroll (1989) or Borg and Groenen (2005). Sometimes the proximity data are collected from K sources such as a panel of K judges or under K different conditions, yielding three-way two mode data and an $n \times n \times K$ array.

* Corresponding author.
E-mail addresses: groenen@few.eur.nl (P.J.F. Groenen), suzanne.winsberg@predisoft.com (S. Winsberg), oldemar.rodriguez@predisoft.com (O. Rodríguez), diday@ceremade.dauphine.fr (E. Diday).

0167-9473/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.04.003

Techniques have been developed to deal with this form of data permitting the study of individual or group differences underlying the dissimilarity ratings (see, for example, Carroll, 1972; Carroll and Winsberg, 1995; Winsberg and De Soete, 1993, 1997).

All of these MDS techniques require that each entry of the dissimilarity matrix be a single numerical value. However, it may be interesting to collect dissimilarity data where the dissimilarity between object i and object j is fuzzy. Then the fuzzy dissimilarity might be represented by an interval of values rather than a single value; consequently the ijth entry of the $n \times n$ dissimilarity matrix is an interval of values, $\left[\delta_{ij}^{(\mathrm{L})}, \delta_{ij}^{(\mathrm{U})}\right]$, where $\delta_{ij}^{(\mathrm{U})}$ is the upper bound of the interval, and $\delta_{ij}^{(\mathrm{L})}$ is the lower bound. Thus, in this paper we consider data to be fuzzy in the sense of interval data. For example, it may be that a judge prefers to indicate a range of values of dissimilarity to express the difference between object i and object j. The extent of the range may vary from object pair to object pair, because he or she wishes to capture the relative precision of his individual judgment. Note that this feature cannot be reflected by a single value of dissimilarity. Or it may be that the objects in the set under consideration are of such a complex nature that the dissimilarity between each pair of them is better represented by a range or an interval of values, rather than a single value. In this latter case, the dissimilarity between object i and object j simply cannot be represented by a single value so it must be represented by an interval of values.

More and more we are conducting statistical analyses of huge data sets. In the MDS case, for example, one may wish to deal with a very large number of objects. One might wish to collect dissimilarity ratings on 100 musical extracts, each one lasting 20 seconds. Here, since one has 100 stimuli or objects, that would normally require each judge to make 4950 judgments of dissimilarity resulting in a task that would be much too long for any judge to endure. To make the task feasible approximately 100 judgments could be collected from each of say 1000 judges. Then it would be necessary to preprocess the data prior to an MDS analysis. Of course, one way to preprocess the data is to calculate the average value of dissimilarity for each object pair. Then the resulting data matrix would consist of single-valued data, say a dissimilarity rating of 27 for object pair (1, 2) and a classical MDS such as MDSCAL (Kruskal, 1964a, b) could be performed. However, this last approach to preprocessing of the data, results in the loss of the valuable information contained in the variation of the dissimilarity ratings for each object pair. It would thus be much more interesting to use the interval of dissimilarity ratings observed for each object pair. Then, the resulting data matrix, after the preprocessing will contain interval values, that is, it will be multivalued, say the interval [23,30] for object pair (1, 2). When the observed variable(s) are represented by an interval of values, then this data can be considered as symbolic data. See Bock and Diday (2000) for a detailed presentation of symbolic data analysis. We note that Denœux and Masson (2000) also point out that interval-type data may be considered as symbolic data. We remark that for any multivalued data, it may be inappropriate to use existing data analytic techniques developed for single-valued data.

In all of these circumstances, the resulting entry of the $n \times n$ dissimilarity matrix would be an interval of values $[a, b]$ corresponding to $\left[\delta_{ij}^{(\mathrm{L})}, \delta_{ij}^{(\mathrm{U})}\right]$ rather than a single value. Of course, if a given entry of the matrix was single-valued, it could be represented in interval form as $[a, a]$.

Denœux and Masson (2000) and Masson and Denœux (2002) have developed two MDS techniques that treat dissimilarity matrices composed of interval data. One technique (see Denœux and Masson, 2000) yields a representation in which each object is represented by a hypersphere in a low-dimensional space. In addition, they have used a hyperbox (hypercube) representation as well. Although the distance between (hyper)spheres has some interesting properties, in this paper, we focus on a technique which yields a representation of the objects as hyperboxes in a low-dimensional Euclidean space rather than hyperspheres because the hyperbox representation is reflected as a conjunction of p properties where p is the dimensionality of the space. This representation as a conjunction is appealing most of all from a linguistic point of view. In everyday language, if we have objects consisting of repeated sound bursts differing with respect to loudness and the number of bursts per second, a given sound or a group of sounds might be referred to as having a loudness lying between 2 and 3 dbSPR and a repetition rate corresponding to between 300 and 400 ms between bursts, that is, as a conjunction of two properties. We would not refer to a sound as a hypersphere with a loudness and repetition rate centered at 2.5 dbSPR and 350 ms and a radius of a given value to be expressed in just what units. We note that perceptually a sound might not have a precise loudness or repetition rate to a listener. Second, since one of the principal aims of MDS is to reveal relationships among the objects in terms of the underlying dimensions, it is most useful for this type of data to express the location of each object in terms of a range of each of these underlying attributes or dimensions. Therefore, in the remainder of this paper, we only focus on the hyperbox representation of MDS on interval data.

The results of MDS analyses with hyperboxes can be represented in two ways. The first is a plot for each pair of dimensions displaying each object as a rectangle. The second is a graph for each underlying dimension displaying the location and range for each object on that dimension. The latter approach may facilitate the dimension wise interpretation over the former approach.

Denœux and Masson (2000) developed a technique which minimizes a loss function for the hyperbox representation using a gradient descent algorithm. Their loss function is closely related to the I-Stress (for interval stress) defined in the next section. Our I-Stress function is somewhat more general as it allows us to weight each object pair differently. One important application of differential weighting would be to weight the larger dissimilarities less than the smaller ones. It is well known that judges use less discrimination when comparing two large dissimilarities than they do when comparing two small dissimilarities. In addition, zero-one weights can be used to accommodate missing data. Another difference of the present paper with that of Denœux and Masson (2000) lies in the minimization method used: here we derive a new majorization algorithm and they use a gradient descent algorithm. The majorization algorithm not only guarantees an improvement at each step until convergence (for a gradient method this depends on the step-size procedure), but as shown by De Leeuw and Heiser (1980) it is easy to add constraints in the majorization approach and thus it is superior in this respect as well. Moreover, Denœux and Masson (2000) recognize that I-Stress can have local minima that are not global and that their gradient method only reaches a local minimum. However, they have not studied how severe the local minimum problem is and how well the gradient method can reach a (candidate) global minimum. Below, we study the local minima issue and compare the majorization algorithm in this respect to the gradient algorithm. To do so, we use a multistart strategy together with a rational start that we have developed (see Appendix B) and which is a useful tool to deal with the local minima problem (see Section 3).

The other technique developed by Masson and Denœux (2002) which they call the possibility model, fits the data from above; that is the model value for the interval of the distance between object i and object j is constrained to be greater than the data dissimilarity value between object i and object j for all pairs. This model or technique is clearly different from our approach, in which the model value for the distance interval can be greater than, equal to, or less than the data value for the dissimilarity interval. Moreover, they apply this second approach only to the hypersphere representation.

In this paper, we focus on interval-type data; that is, on fuzzy dissimilarity data represented by an interval of values. The remainder of this paper is organized as follows. In the next section, we propose the I-Scal algorithm for MDS of interval dissimilarities based on iterative majorization. Then, we investigate the quality of this algorithm by looking at the local minimum problem and perform a small simulation study. We also discuss an empirical data set on interval dissimilarities of sounds. We end the paper with a discussion of the results and conclusions.

2. MDS of interval dissimilarities

To develop MDS for interval dissimilarities, the ranges of dissimilarities must be represented by ranges of distances. Here, we choose to represent the objects by rectangles and approximate the upper bound of the dissimilarity by the maximum distance between the rectangles and the lower bound by the minimum distance between the rectangles. Fig. 1 shows an example of rectangle representation and how the minimum and maximum distance between two rectangles is defined.

Not only the distances are represented by ranges, the coordinates themselves are also ranges. Let the rows of the $n \times p$ matrix $\mathbf{X}$ contain the coordinates of the center of the rectangles, where n is the number of objects and p the dimensionality. The distance from the center of rectangle i along axis s, denoted the spread, is represented by r_{is}. Note that $r_{is} \geqslant 0$. The maximum Euclidean distance between rectangles i and j is given by

$$d_{ij}^{(\mathrm{U})}(\mathbf{X}, \mathbf{R}) = \left(\sum_{s=1}^{p} \left[|x_{is} - x_{js}| + (r_{is} + r_{js}) \right]^2 \right)^{1/2} \tag{1}$$

and the minimum Euclidean distance by

$$d_{ij}^{(\mathrm{L})}(\mathbf{X}, \mathbf{R}) = \left(\sum_{s=1}^{p} \max\left[0, |x_{is} - x_{js}| - (r_{is} + r_{js})\right]^2 \right)^{1/2}. \tag{2}$$

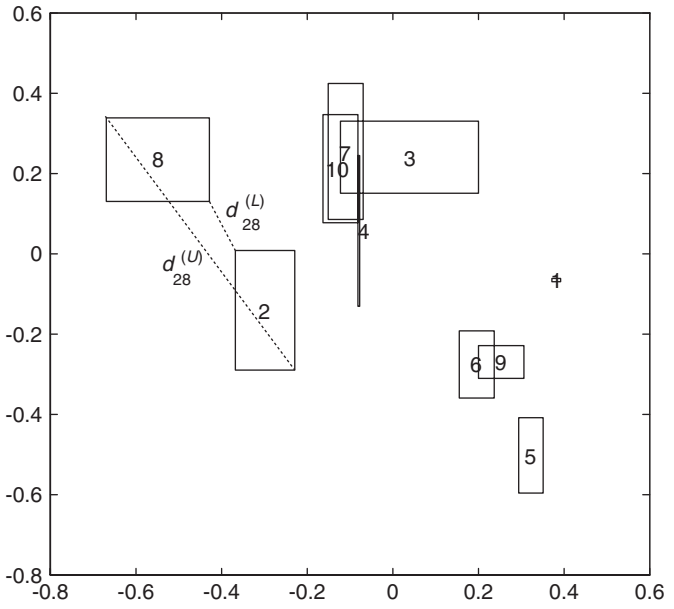

Fig. 1. Example of distances in MDS for interval dissimilarities where the objects are represented by rectangles.

The definition of $d_{ij}^{(\mathrm{L})}(\mathbf{X}, \mathbf{R})$ ensures that if two hyperboxes overlap, their minimum distances is zero. Even though Euclidean distances are used between the hyperboxes, the lower and upper distances change when the solution is rotated. The reason is that the hyperboxes are defined with respect to the axes. For a dimensional interpretation, the property of rotational uniqueness can be seen as an advantage of our interval MDS. Of course, if $\mathbf{R} = \mathbf{0}$ all hyperboxes shrink to points and then interval MDS simplifies into ordinary MDS, which can be freely rotated.

Note that another natural distance measure between rectangles would be the (symmetric or asymmetric) Hausdorff metric (for a definition, see, for example, Bock and Diday, 2000, p. 302). However, it yields a single value for the distance between two (hyper)polygons. Therefore, it is not applicable here where the data consist of a range or interval of dissimilarities for each object pair and we model the range of dissimilarity by a range of distances, not a single value.

The objective of MDS for interval dissimilarities is to represent the lower and upper bounds of the dissimilarities by minimum and maximum distances between rectangles as well as possible in least-squares sense. The I-Stress loss function that models this objective and needs to be minimized over $\mathbf{X}$ and $\mathbf{R}$ is given by

$$\sigma_{\mathrm{I}}^2(\mathbf{X}, \mathbf{R}) = \sum_{i<j}^{n} w_{ij}\left[\delta_{ij}^{(\mathrm{U})} - d_{ij}^{(\mathrm{U})}(\mathbf{X}, \mathbf{R})\right]^2 + \sum_{i<j}^{n} w_{ij}\left[\delta_{ij}^{(\mathrm{L})} - d_{ij}^{(\mathrm{L})}(\mathbf{X}, \mathbf{R})\right]^2, \tag{3}$$

where $\delta_{ij}^{(\mathrm{U})}$ is the upper bound of the dissimilarity of objects i and j, $\delta_{ij}^{(\mathrm{L})}$ is the lower bound, and w_{ij} is a given nonnegative weight. In our experiments, we always report a normalized version of I-Stress, that is,

$$\frac{\sigma_{\mathrm{I}}^2(\mathbf{X}, \mathbf{R})}{\sum_{i<j}^{n} w_{ij}\left[\delta_{ij}^{(\mathrm{U})}\right]^2 + \sum_{i<j}^{n} w_{ij}\left[\delta_{ij}^{(\mathrm{L})}\right]^2} \tag{4}$$

that has the advantage that at a local minimum, (4) is always between zero and one independent of the size of the values of $\delta_{ij}^{(\mathrm{U})}$ and $\delta_{ij}^{(\mathrm{L})}$ (see Borg and Groenen, 2005, Section 11.1). Clearly, division by the constant in (4) does not change the position of local minima.

Below, we derive a majorization algorithm called I-Scal to minimize I-Stress for two reasons. First, iterative majorization is guaranteed to reduce I-Stress in each iteration from any starting configuration until a stationary point is obtained that, in practice, almost always coincides with a local minimum. Second, as in each iteration the algorithm

operates on a quadratic function in $\mathbf{X}$ and $\mathbf{R}$ it is easy to impose constraints that have well-known solutions for quadratic functions. This property can be useful for extensions of I-Scal that require constraints.

2.1. A majorization algorithm

In this section, we develop a majorization algorithm to minimize (3) over $\mathbf{X}$ and $\mathbf{R}$. The basic idea of iterative majorization is that the original loss function is replaced in each iteration by an auxiliary function that is easier to handle. The auxiliary function, the so-called majorizing function, needs to satisfy two requirements: (i) the majorizing function is equal to the original function at the current estimate, and (ii) the majorizing function is always larger than or equal to the original function. Usually, the majorizing function is linear or quadratic so that the minimum of the majorizing function can be calculated easily. From the requirements it can be derived that (a) the loss of the majorizing function and the original loss function is equal at the current estimate, (b) at the update the majorizing function is smaller than at the current estimate, so that (c) the original loss function is smaller at the update since the original loss function is never larger than the majorizing function. This reasoning proves that if the conditions (i) and (ii) are satisfied, the iterative majorization algorithm yields a series of nonincreasing function values. For more details on iterative majorization, we refer to De Leeuw (1994), Heiser (1995), Kiers (2002) or, for an introduction, to Hunter and Lange (2004) and Borg and Groenen (2005, Chapter 8).

In what follows, we shall derive inequalities to find a majorizing function of $\sigma_{\mathrm{I}}^2(\mathbf{X},\mathbf{R})$. To do this, we expand (3) as

$$\begin{aligned}\sigma_{\mathrm{I}}^2(\mathbf{X},\mathbf{R}) &= \sum_{i<j}^{n} w_{ij}\left[\delta_{ij}^{(\mathrm{U})}\right]^2 + \sum_{i<j}^{n} w_{ij}\left[d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R})\right]^2 - 2\sum_{i<j}^{n} w_{ij}\delta_{ij}^{(\mathrm{U})} d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R}) \\ &\quad + \sum_{i<j}^{n} w_{ij}\left[\delta_{ij}^{(\mathrm{L})}\right]^2 + \sum_{i<j}^{n} w_{ij}\left[d_{ij}^{(\mathrm{L})}(\mathbf{X},\mathbf{R})\right]^2 - 2\sum_{i<j}^{n} w_{ij}\delta_{ij}^{(\mathrm{L})} d_{ij}^{(\mathrm{L})}(\mathbf{X},\mathbf{R}). \end{aligned} \tag{5}$$

Because w_{ij}, $\delta_{ij}^{(\mathrm{U})}$, and $\delta_{ij}^{(\mathrm{L})}$ are nonnegative, (5) can be considered as a weighted sum of $\left[d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R})\right]^2$, $-d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R})$, $\left[d_{ij}^{(\mathrm{L})}(\mathbf{X},\mathbf{R})\right]^2$, and $-d_{ij}^{(\mathrm{L})}(\mathbf{X},\mathbf{R})$. Thus, to find a majorizing function for (5) it suffices to find a majorizing function for each of the four distance terms. In Appendix A, we derive for each of these four terms a quadratic or linear majorizing function in the parameters. Once we have obtained the four majorizing functions, these majorizing functions are substituted in (5) which gives the overall majorizing function of $\sigma_{\mathrm{I}}^2(\mathbf{X},\mathbf{R})$ that is quadratic in the parameters $\mathbf{X}$ and $\mathbf{R}$. We shall derive majorizing functions of the form $\alpha\left(x_{is}-x_{js}\right)^2 - 2\beta\left(x_{is}-x_{js}\right)\left(y_{is}-y_{js}\right)+\gamma$ for the center coordinates $\mathbf{X}$ and $\alpha r_{is}^2 - 2\beta r_{is} + \gamma$ for the width parameters $\mathbf{R}$. Note that the parameters α, β, and γ are only functions of $\mathbf{Y}$ and $\mathbf{Q}$, the known current estimates of $\mathbf{X}$ and $\mathbf{R}$. We shall prove that the majorization algorithm automatically finds values of $\mathbf{R}$ that are nonnegative.

2.2. Combining the majorization results

In Appendix A, we have derived majorizing functions for each of the terms in (5). To obtain an overall majorizing function, we substitute each of the respective terms in (5) by (A.6), (A.11), (A.15), and (A.19). This substitution gives the overall majorizing inequality

$$\begin{aligned}\sigma_{\mathrm{I}}^2(\mathbf{X},\mathbf{R}) \leqslant & \sum_{s=1}^{p}\sum_{i<j}\left(\alpha_{ijs}^{(1)}+\alpha_{ij}^{(3)}\right)\left(x_{is}-x_{js}\right)^2 - 2\sum_{s=1}^{p}\sum_{i<j}\left(\beta_{ijs}^{(1)}+\beta_{ijs}^{(3)}+\beta_{ijs}^{(5)}\right)\left(x_{is}-x_{js}\right)\left(y_{is}-y_{js}\right) \\ & + \sum_{s=1}^{p}\sum_{i<j}\left(\alpha_{ijs}^{(2)}+\alpha_{ijs}^{(4)}+\alpha_{ijs}^{(5)}\right) r_{is}^2 + \sum_{s=1}^{p}\sum_{i<j}\left(\alpha_{jis}^{(2)}+\alpha_{jis}^{(4)}+\alpha_{jis}^{(5)}\right) r_{js}^2 \\ & - 2\sum_{s=1}^{p}\sum_{i<j}\left(\beta_{ijs}^{(2)}+\beta_{ijs}^{(4)}\right)\left(r_{is}+r_{js}\right) + \sum_{s=1}^{p}\sum_{i<j}\left(\gamma_{ijs}^{(1)}+\gamma_{ijs}^{(2)}\right). \end{aligned} \tag{6}$$

We first consider the terms that are linear and quadratic in $\mathbf{X}$. Let $\mathbf{A}_s^{(1)}$ be a matrix with elements

$$a_{ijs}^{(1)} = \begin{cases} -\left(\alpha_{ijs}^{(1)} + \alpha_{ij}^{(3)}\right) & \text{if } i \neq j, \\ -\sum_{k \neq i} a_{iks}^{(1)} & \text{if } i = j. \end{cases} \tag{7}$$

Note that $\alpha_{ijs}^{(1)}$ and $\alpha_{ij}^{(3)}$ are only dependent on the known current estimates $\mathbf{Y}$ and $\mathbf{Q}$. It may be verified that

$$\sum_{s=1}^{p} \sum_{i<j} \left(\alpha_{ijs}^{(1)} + \alpha_{ij}^{(3)}\right) (x_{is} - x_{js})^2 = \sum_{s=1}^{p} \mathbf{x}_s' \mathbf{A}_s^{(1)} \mathbf{x}_s.$$

For an explanation of this expression, see, for example, Groenen et al. (1996) or Section 8.6 of Borg and Groenen (2005). In a similar way, the linear term can be easily written in matrix algebra by defining matrix $\mathbf{B}_s^{(1)}$ with elements

$$b_{ijs}^{(1)} = \begin{cases} -\left(\beta_{ijs}^{(1)} + \beta_{ijs}^{(3)} + \beta_{ijs}^{(5)}\right) & \text{if } i \neq j, \\ -\sum_{k \neq i} b_{iks} & \text{if } i = j, \end{cases} \tag{8}$$

so that

$$\sum_{s=1}^{p} \sum_{i<j} \left(\beta_{ijs}^{(1)} + \beta_{ijs}^{(3)} + \beta_{ijs}^{(5)}\right) (x_{is} - x_{js}) (y_{is} - y_{js}) = \sum_{s=1}^{p} \mathbf{x}_s' \mathbf{B}_s^{(1)} \mathbf{y}_s.$$

We now turn to a compact expression for the terms with r_{is}^2 and r_{is}. The first thing to notice is that

$$\sum_{s=1}^{p} \sum_{i<j} \left(\alpha_{ijs}^{(2)} + \alpha_{ijs}^{(4)} + \alpha_{ijs}^{(5)}\right) r_{is}^2 + \sum_{s=1}^{p} \sum_{i<j} \left(\alpha_{jis}^{(2)} + \alpha_{jis}^{(4)} + \alpha_{jis}^{(5)}\right) r_{js}^2 = \sum_{s=1}^{p} \sum_{i=1}^{n} r_{is}^2 \sum_{j \neq i} \left(\alpha_{ijs}^{(2)} + \alpha_{ijs}^{(4)} + \alpha_{ijs}^{(5)}\right).$$

Let $\mathbf{A}_s^{(2)}$ be a diagonal matrix with diagonal elements $a_{iis}^{(2)} = \sum_{j \neq i} \left(\alpha_{ijs}^{(2)} + \alpha_{ijs}^{(4)} + \alpha_{ijs}^{(5)}\right)$, so that

$$\sum_{s=1}^{p} \sum_{i=1}^{n} r_{is}^2 \sum_{j \neq i} \left(\alpha_{ijs}^{(2)} + \alpha_{ijs}^{(4)} + \alpha_{ijs}^{(5)}\right) = \sum_{s=1}^{p} \mathbf{r}_s' \mathbf{A}_s^{(2)} \mathbf{r}_s,$$

where $\mathbf{r}_s$ is column s of $\mathbf{R}$. For the terms linear in r_{is}, we note that

$$\sum_{s=1}^{p} \sum_{i<j} \left(\beta_{ijs}^{(2)} + \beta_{ijs}^{(4)}\right) (r_{is} + r_{js}) = \sum_{s=1}^{p} \sum_{i=1}^{s} r_{is} \sum_{j \neq i} \left(\beta_{ijs}^{(2)} + \beta_{ijs}^{(4)}\right).$$

Define $\mathbf{b}_s^{(2)}$ as a vector with elements $\sum_{j \neq i} \left(\beta_{ijs}^{(2)} + \beta_{ijs}^{(4)}\right)$ so that we may write

$$\sum_{s=1}^{p} \sum_{i=1}^{n} r_{is} \sum_{j \neq i} \left(\beta_{ijs}^{(2)} + \beta_{ijs}^{(4)}\right) = \sum_{s=1}^{p} \mathbf{r}_s' \mathbf{b}_s^{(2)}.$$

The reformulation above allows us to write the right-hand side of (6) as

$$\sigma_{\mathrm{I}}^2(\mathbf{X}, \mathbf{R}) \leqslant \sum_{s=1}^{p} \left(\mathbf{x}_s' \mathbf{A}_s^{(1)} \mathbf{x}_s - 2\mathbf{x}_s' \mathbf{B}_s^{(1)} \mathbf{y}_s\right) + \sum_{s=1}^{p} \left(\mathbf{r}_s' \mathbf{A}_s^{(2)} \mathbf{r}_s - 2\mathbf{r}_s' \mathbf{b}_s^{(2)}\right) + \sum_{s=1}^{p} \sum_{i<j} \left(\gamma_{ijs}^{(1)} + \gamma_{ijs}^{(2)}\right). \tag{9}$$

From (9), a dimensionwise update for $\mathbf{X}$ and $\mathbf{R}$ can be derived. Since the right-hand side of (9) is quadratic in $\mathbf{X}$, a minimum is obtained by equating the gradient to zero, i.e.,

$$2\mathbf{A}_s^{(1)} \mathbf{x}_s - 2\mathbf{B}_s^{(1)} \mathbf{y}_s = \mathbf{0}, \quad \text{or,} \quad \text{equivalently } \mathbf{A}_s^{(1)} \mathbf{x}_s = \mathbf{B}_s^{(1)} \mathbf{y}_s. \tag{10}$$

To solve the linear system (10) we need a generalized inverse of $\mathbf{A}_s^{(1)}$, since it is not of full rank. We use the Moore–Penrose inverse $\mathbf{A}_s^{(1)+}$ which equals here $\mathbf{A}_s^{(1)+} = \left(\mathbf{A}_s^{(1)} + n^{-1}\mathbf{1}\mathbf{1}'\right)^{-1} - n^{-1}\mathbf{1}\mathbf{1}'$ with $\mathbf{1}$ a vector of ones of appropriate length (see, for example Groenen et al., 1995, 1999). The update for $\mathbf{x}_s$ is defined by

$$\mathbf{x}_s = \mathbf{A}_s^{(1)+}\mathbf{B}_s^{(1)}\mathbf{y}_s. \tag{11}$$

The update for $\mathbf{r}_s$ is found in a similar fashion as

$$\mathbf{r}_s = \mathbf{A}_s^{(2)-1}\mathbf{b}_s^{(2)}, \quad \text{or,} \quad \text{equivalently } r_{is} = \frac{b_{is}^{(2)}}{a_{iis}^{(2)}}. \tag{12}$$

The restriction on r_{is} is that it is nonnegative. It may be verified that all terms that make up $a_{iis}^{(2)}$ and $b_{is}^{(2)}$ are nonnegative, so that (12) automatically yields an update that is also nonnegative thereby satisfying the restrictions.

The I-SCAL algorithm based on iterative majorization can be summarized in Algorithm 1.

Algorithm 1. The I-Scal algorithm.

Set $\mathbf{X}_0$ to some initial matrix for the coordinate centers.
Set $\mathbf{R}_0$ to some matrix of nonnegative values for the width.
Set iteration counter $k := 0$.
Set $\mathbf{X}_{-1} := \mathbf{X}_0$ and $\mathbf{R}_{-1} := \mathbf{R}_0$.
Set the convergence criterion ε to a small positive value, e.g., 10^{-6}.
while $k = 0$ or $\sigma_{\mathrm{I}}^2(\mathbf{X}_{k-1}, \mathbf{R}_{k-1}) - \sigma_{\mathrm{I}}^2(\mathbf{X}_k, \mathbf{R}_k) \leqslant \varepsilon$ **do**
  $k := k + 1$
  Set $\mathbf{Y} := \mathbf{X}_{k-1}$ and $\mathbf{Q} := \mathbf{R}_{k-1}$.
  for $s = 1$ to p **do**
    Compute $\mathbf{A}_s^{(1)}$ by (7) and $\mathbf{B}_s^{(1)}$ by (8).
    Compute the update of $\mathbf{x}_s$ by (11).
    Compute $\mathbf{A}_s^{(2)}$ and $\mathbf{b}_s^{(2)}$.
    Compute the update of $\mathbf{r}_s$ by (12).
  end for
  Set $\mathbf{X}_k := \mathbf{X}$ and $\mathbf{R}_k := \mathbf{R}$.
end while

3. Investigating the quality of the majorization algorithm

To validate the I-Scal algorithm, we have analyzed several artificial data sets. First, we investigated the local minimum problem. The algorithm permits considering both the rational start described in the Appendix B and many random starts and then chooses the best global solution. We generated artificial interval-type data with random values from the uniform distribution for $\mathbf{X}$ from 0 to 1 for each coordinate, and random values from the uniform distribution for $\mathbf{R}$ from 0 to 0.2 for each coordinate. Since we have generated both values of $\mathbf{X}$ and $\mathbf{R}$, we obtain intervals from the coordinates. Then interval values, that is upper and lower value of the distance, were generated as follows: upper and lower values of the distance were calculated from (1) and (2) and these values were used as the upper bound of the dissimilarities $\delta_{ij}^{(\mathrm{U})}$ and the lower bound $\delta_{ij}^{(\mathrm{L})}$. If there are no local minima and only a single global minimum, then the majorization algorithm should be able to find zero σ_{I}^2 to recover these data perfectly. In this study, we chose $n = 20$ and $p = 2$.

To see how bad the local minimum problem is, we performed multiple random starts on these perfect data and additionally the rational start from Appendix B. We use this algorithm for the rational start because it gives us starting values for all the parameters of the model. That would not be achieved by an ordinary MDS on the centers. We applied both the I-Scal algorithm and used the gradient approach given by Denœux and Masson (2000). As they only present the gradient, we used these gradients in the `fminunc` function from MatLab's 7.0 optimization toolbox. We used the same random starts for both methods (except the rational start from Appendix B that was used only in combination with I-Scal).

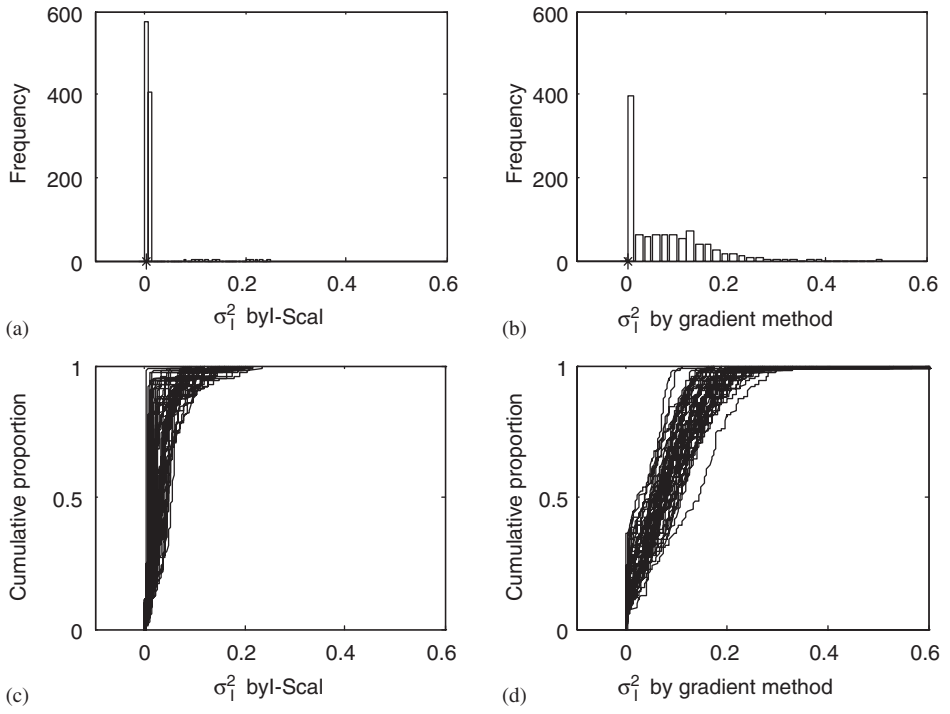

Fig. 2. Distribution of σ_I^2 for perfect data obtained by 1000 random starts followed by I-Scal (Panel a) or by the gradient method of Denœux and Masson (2000) (Panel b). The '*' indicates the position of σ_I^2 obtained by the rational start discussed in Appendix B followed by either of the two algorithms. The cumulative distributions in Panels c and d are based on 100 random starts for 50 replication.

The results of this comparison study for 1000 random starts on a single perfect data set is shown in the histograms of Fig. 2a and b. The distributions show that both algorithms are capable in locating the global minimum of zero. We also see that a lot of different local minima were found. Apparently, the loss function of σ_I^2 has a severe local minimum problem. As the distribution of the gradient method of Denœux and Masson (2000) in Fig. 2b is flatter than the I-Scal distribution in Fig. 2a, we know that at least for this data set, I-Scal obtains better quality local minima and is capable to locate the global optimum in the vast majority of the cases.

To see if the result is data specific we repeated this experiment 50 times each time with a new data set. However, in this part of the experiment we used 100 random starts for each of the 50 data sets. Within each replication (i.e., data set) both methods under comparison used the same random perfect data set and the same random start configuration in the multiple random start runs. Fig. 2c shows for I-Scal the cumulative distributions of the σ_I^2 values for each of the 50 replications and Fig. 2d for the gradient method. In many of the replications, I-Scal reaches low or zero I-Stress for many of the random starts. However, the gradient method of Denœux and Masson (2000) performs worse as the largest part of the cumulative σ_I^2 values lies below those of I-Scal. This finding suggests that the majorization algorithm I-Scal is better in finding find good quality local minima than Denœux and Masson's gradient method. On average, the cpu-time required for I-Scal was about 2.9 s per run and 3.6 s for the gradient method. This result indicates that the computational effort is about the same for both methods with a slight gain of about 25% for I-Scal.

The effect of the rational start that we have developed in Appendix B is quite strong: using it to start either of the two algorithms yielded I-Stress values close to the global minimum of zero. Therefore, in the sequel, we apply I-Scal always in conjunction with the multistart strategy (that is the rational start combined with random starts) to avoid bad fitting local minima. Preliminary experimentation suggested that at least 50 random starts are needed to circumvent local minima problems, when using the multistart strategy. Note that although in this comparison we have used our rational start in conjunction with the gradient algorithm (and indeed it helps) Denœux and Masson (2000) and Masson and Denœux (2002) do not use any rational start and do not examine the local minima problem.

In the second study, we investigate how well I-Scal performs for a variety of artificial data sets. Because the previous study showed that I-Scal yielded good quality solutions much more consistently, we limit this study to I-Scal only. These data sets were generated artificially for two values of n ($n = 10, 20$), for three error levels (no error, error variance 5% of the total variance, and error variance 15% of the total variance), and for two values of p ($p = 2, 3$). All of these factors were crossed in our study and 10 data sets were generated for each combination using the following procedure: random values of $\mathbf{X}$ and $\mathbf{R}$ were generated as described above, upper and lower values for the distance were calculated using (1) and (2), and then normally distributed error was added to these distance values to obtain the dissimilarity matrix. For example, in the case of 15% error, the variance of the error was set to 15% of the variance of the true upper and lower distances. To ensure that the lower and upper bounds of the dissimilarities were nonnegative and that the lower bound is not larger than the upper bounds, we enforced the restrictions $0 \leqslant \delta_{ij}^{(\mathrm{L})} \leqslant \delta_{ij}^{(\mathrm{U})}$ for all ij.

To measure the ability of I-Scal to recover the true object configuration, we computed the root mean square deviation between the true and recovered coordinates. This measure called DEL, for delta, originally defined by Weinberg and Menil (1993), is defined for the centers of the objects, by

$$\mathrm{DELX} = \left[(np)^{-1} \sum_{i=1}^{n} \sum_{s=1}^{p} \left(\hat{x}_{is} - x_{is} \right)^2 \right]^{1/2}, \tag{13}$$

where $\hat{x}_{is}$ is the true value of the center for object i for the sth coordinate, and x_{is} is the recovered value and p is the number of dimensions. We also define a measure DEL for the spreads of the objects by

$$\mathrm{DELR} = \left[(np)^{-1} \sum_{i=1}^{n} \sum_{s=1}^{p} \left(\hat{r}_{is} - r_{is} \right)^2 \right]^{1/2}. \tag{14}$$

In addition to seeing how well we recover the location of the objects it is interesting to see how well we recover the size of the objects, which in some experimental situations is a measure of the how well defined, or how well known are the objects to the judges. In such situations, it is this lack of precise knowledge of the objects which gives rise to the range of dissimilarities, that is the fuzziness of the dissimilarities. So, finally we define a measure DEL for the recovery of size of the objects by

$$\mathrm{DELSIZE} = \left[n^{-1} \sum_{i=1}^{n} \left[\prod_{s=1}^{p} \left(2\hat{r}_{is} \right) - \prod_{s=1}^{p} \left(2r_{is} \right) \right]^2 \right]^{1/2}. \tag{15}$$

A DEL value of 0 is clearly the ideal and indicates perfect recovery.

To measure how well the method is able to reconstruct the true upper and lower bounds of the dissimilarities $\left(\hat{d}_{ij}^{(\mathrm{U})} \text{ and } \hat{d}_{ij}^{(\mathrm{L})} \right)$, we calculate Tucker's coefficient of congruence, $\lambda^{(\mathrm{U})}$ and $\lambda^{(\mathrm{L})}$ of the true and reconstructed vectors of upper and lower bounds. For the upper bounds, Tucker's coefficient of congruence (Tucker, 1951), $\lambda^{(\mathrm{U})}$ is defined by

$$\lambda^{(\mathrm{U})} = \frac{\sum_{i<j} \hat{d}_{ij}^{(\mathrm{U})} d_{ij}^{(\mathrm{U})}}{\left[\sum_{i<j} \left(\hat{d}_{ij}^{(\mathrm{U})} \right)^2 \sum_{i<j} \left(d_{ij}^{(\mathrm{U})} \right)^2 \right]^{1/2}}. \tag{16}$$

Note that Tucker's coefficient of congruence can be viewed as a correlation coefficient without subtracting the mean. Geometrically, this coefficient is equal to the cosine of the angle formed by two vectors. Therefore, this measure is always between -1 and 1, but since distances are positive by definition, $\lambda^{(\mathrm{U})}$ and $\lambda^{(\mathrm{L})}$ will be between 0 and 1.

The results of the simulation study are presented in Table 1. Each value in the table is the average of 10 simulations for that case, followed by in parentheses the standard deviation obtained. We note that for DELSIZE one cannot compare values obtained for two dimensions with those obtained for three dimensions without realizing that errors in the r_{is}'s are multiplied when determining the size of an object. The errors are multiplied in DELSIZE because parameters recovered relate to the dimensions, and the SIZE and thus DELSIZE relate to the volume so the errors in recovering the volume are the product of the errors in recovering the dimensions.

Table 1
Recovery values for the simulations

Error	n	p	$\lambda^{(L)}$	$\lambda^{(U)}$	DELX	DELR	DELSIZE $\times 10^{-7}$
0.00	10	2	0.9987 (0.0020)	0.9998 (0.0002)	0.0062 (0.0086)	0.0013 (0.0016)	101.2 (92.9)
0.00	10	3	0.9988 (0.0008)	0.9998 (0.0001)	0.0183 (0.0141)	0.0034 (0.0020)	8.2 (9.9)
0.00	20	2	0.9998 (0.0001)	0.9999 (0.0000)	0.0003 (0.0003)	0.0001 (0.0001)	7.6 (13.8)
0.00	20	3	0.9985 (0.0015)	0.9998 (0.0002)	0.0101 (0.0104)	0.0013 (0.0013)	3.2 (3.4)
0.05	10	2	0.9949 (0.0034)	0.9989 (0.0005)	0.0047 (0.0050)	0.0019 (0.0015)	319.1 (325.5)
0.05	10	3	0.9956 (0.0022)	0.9987 (0.0002)	0.0194 (0.0121)	0.0050 (0.0018)	9.4 (7.2)
0.05	20	2	0.9987 (0.0003)	0.9996 (0.0001)	0.0003 (0.0001)	0.0003 (0.0001)	49.3 (45.7)
0.05	20	3	0.9972 (0.0012)	0.9993 (0.0002)	0.0085 (0.0075)	0.0015 (0.0007)	3.4 (3.4)
0.15	10	2	0.9903 (0.0046)	0.9964 (0.0009)	0.0118 (0.0092)	0.0037 (0.0018)	554.0 (491.5)
0.15	10	3	0.9911 (0.0038)	0.9963 (0.0009)	0.0272 (0.0173)	0.0093 (0.0034)	16.0 (9.6)
0.15	20	2	0.9967 (0.0008)	0.9987 (0.0003)	0.0009 (0.0006)	0.0007 (0.0002)	103.0 (43.8)
0.15	20	3	0.9944 (0.0019)	0.9983 (0.0004)	0.0107 (0.0069)	0.0032 (0.0010)	7.2 (4.2)

Each cell has 10 replications. For these replications, the mean and the standard deviation (within parentheses) are reported.

The mean I-Stress over the 40 replications in the zero error case is 0.000631 (stand. dev. 0.000607), for the 0.05 error condition 0.003647 (stand. dev. 0.000947), and for the 0.15 error condition 0.009229 (stand. dev. 0.002507).

The values in the table indicate excellent recovery of the true values using our algorithm. We note that for example Tucker's coefficient of congruence decreases slightly when error is increased as would be expected. Also the lower bounds appear to be recovered better than the upper bounds. DELX and DELR and consequently DELSIZE also increase slightly with increasing error. All measures of recovery improve with increasing n. DELX and DELR are both recovered better for the two-dimensional case than for the three-dimensional case; this is because there are more degrees of freedom in the three-dimensional case.

4. Interval MDS for empirical data

We now consider two real data sets where the entries in the dissimilarity matrix are an interval-valued measures. The objects in the study are 10 sounds differing with respect to only two physical parameters: the spectral center of gravity and log attack time. Previous studies of musical timbre have shown that these two physical parameters are highly correlated with the perceptual axes found when dissimilarity judgments are collected for sounds from different musical instruments playing the same note at the same loudness for the same duration. Until about 35 years ago timbre was considered to be a perceptual parameter of sound that was complex and multidimensional, defined primarily by what it was not, that is, what distinguishes two sounds presented in a similar manner equal in pitch, subjective duration, and loudness (see Plomp, 1970). MDS studies have shown that these two attributes of sound, namely spectral center of gravity and log attack time explain the factors we use to distinguish, say, middle C on the piano from middle C on some other instrument (see, e.g., McAdams et al., 1995; McAdams and Winsberg, 1999). So when middle C is sounded on the piano the sound has some unidimensional attributes such as pitch, corresponding to the frequency of the fundamental, loudness, and duration. In addition, it is characterized by its timbre, that is, it is a note from a piano not some other instrument. This last attribute is perceptually multidimensional with two important underlying dimensions related to spectral center of gravity and log attack time. The spectral center of gravity is the weighted average of the harmonics generated when the note is sounded averaged over the duration of the tone with a running time window of, say, 12 ms and is higher say for the harpsichord than it is for the piano. The log attack time is the logarithm of the rise time measured from the time the amplitude envelope reaches a threshold of 2% of the maximum amplitude to the time it takes to reach the maximum amplitude and is longer for a wind instrument like the trumpet than it is for a string instrument like the harp. The sounds in this study were artificially generated to represent the range of values found in natural instruments according to the design in Fig. 3. The data represents dissimilarity judgments from the same expert listener taken on two occasions, see Table 2. On each occasion the expert listened to each pair of sounds and indicated a range of dissimilarity for each pair on a slider scale going from very similar to very different.

The data in Table 2 were analyzed by I-Scal for both occasions separately using 1000 random starts. The resulting solutions are given in Fig. 4. Visual examination of the two solutions reveals the following. The horizontal axis represents

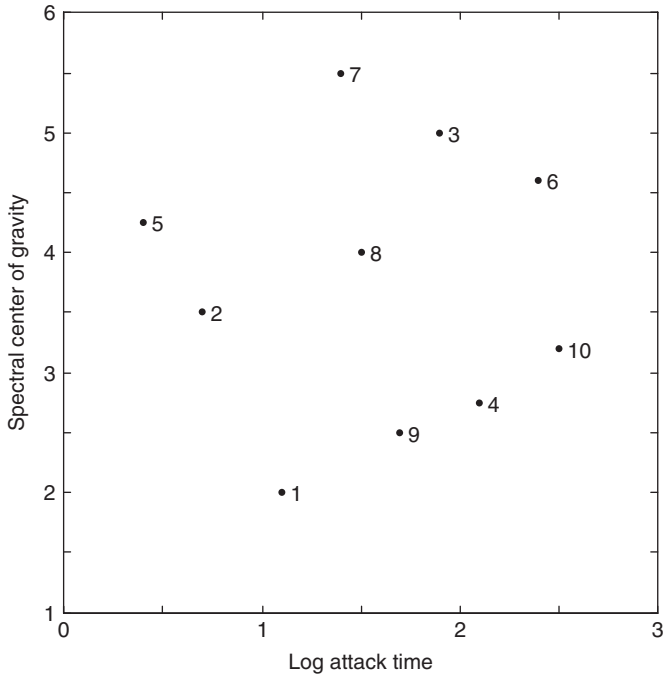

Fig. 3. Design of the 10 sounds according to spectral center of gravity (vertical axis) and log attack time (horizontal axes).

Table 2
Interval dissimilarities of 10 sounds judged by an expert at occasions 1 (lower triangle) and 2 (upper triangle)

Tone	1	2	3	4	5	6	7	8	9	10
1	[–, –]	[62, 81]	[82, 95]	[6, 22]	[62, 87]	[58, 87]	[67, 81]	[64, 77]	[0, 13]	[48, 61]
2	[73, 88]	[–, –]	[39, 59]	[53, 68]	[3, 16]	[68, 92]	[17, 41]	[45, 57]	[40, 66]	[92, 98]
3	[93, 100]	[6, 21]	[–, –]	[48, 74]	[51, 78]	[0, 8]	[0, 11]	[23, 56]	[46, 69]	[33, 61]
4	[7, 25]	[46, 66]	[60, 72]	[–, –]	[51, 68]	[17, 41]	[72, 92]	[44, 55]	[0, 20]	[31, 42]
5	[95, 100]	[4, 36]	[38, 58]	[63, 74]	[–, –]	[34, 54]	[0, 5]	[3, 23]	[68, 79]	[45, 70]
6	[73, 90]	[37, 63]	[16, 22]	[33, 46]	[1, 8]	[–, –]	[9, 26]	[8, 37]	[42, 54]	[22, 54]
7	[90, 100]	[49, 71]	[4, 13]	[87, 98]	[10, 21]	[28, 46]	[–, –]	[21, 42]	[47, 77]	[77, 91]
8	[64, 79]	[7, 36]	[10, 26]	[36, 54]	[26, 45]	[28, 50]	[32, 60]	[–, –]	[54, 79]	[18, 33]
9	[0, 8]	[37, 63]	[58, 78]	[8, 19]	[66, 81]	[71, 84]	[76, 90]	[29, 46]	[–, –]	[3, 18]
10	[35, 44]	[78, 88]	[49, 75]	[0, 7]	[69, 82]	[65, 81]	[75, 91]	[75, 88]	[20, 53]	[–, –]

log attack time and the vertical axis the spectral center of gravity. Without imposing any restriction, I-Scal seems to be able to reconstruct the physical space. The results for the second occasion in Fig. 4b reflect the physical space quite well. Notice the groupings 10, 9, 4, 1 and 2, 5, 7 and 3, 6, 8 reflect well how these stimuli are grouped in the physical space. Moreover, the relation of these groups to one another approximates their disposition in the physical space reasonably well. However, the solution from the first occasion shows some deviations from the physical space: 8, 3, 6 are too far to the left, 3 is too low, 7 is too far to the left, and 1 is too far to the right. It is interesting to note that these differences from one occasion to another are greater than the range of uncertainty reflected in the solutions. The improved results on the second occasion indicate that the task is better performed with some practice and with greater familiarity with the group of sounds. It also appears from the figures that sounds with long attack times are more difficult to localize. This last remark seems to be the case in general. As one may note from the figure on the right, however, object 10 seems to be well located along this dimension even though it has a long attack time. This type of data with the I-Scal

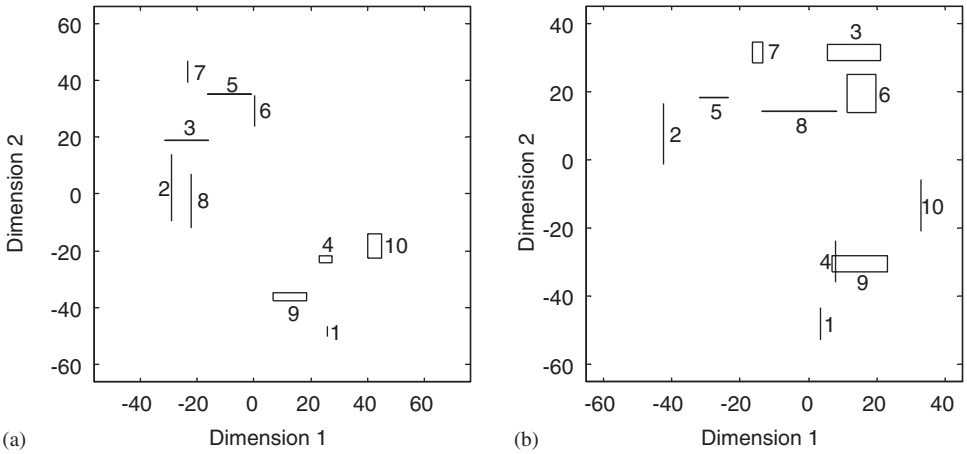

Fig. 4. I-Scal solutions for the sound data in Table 2. Panel (a) gives the results for Occasion 1 with I-Stress 0.02861128 and Panel (b) for Occasion 2 with I-Stress 0.04893295.

solution might possibly be used to establish norms which could then be used to detect people with specific types of hearing impairment. In this section, dealing with a real data example we have focussed on the interpretation of the results, in contrast to the previous section where we focussed on the statistical quality of the results. We note, however, that collecting interval dissimilarities as opposed to single-valued dissimilarities provides in this real data case a better framework with which to compare the hearing disabled with normal hearing. Moreover, we can see from the results that the I-Scal algorithm provided a solution which is consistent with all we know about these stimuli.

5. Discussion and conclusions

We have presented an MDS technique for data that deals with fuzzy dissimilarities consisting of an interval of values observed for each pair of objects. In this technique, each object is represented as a hyperbox, that is, a hypercube in a p-dimensional space. By representing the objects as hypercubes, we are able to convey information contained when the dissimilarity between the objects or for any object pair needs to be expressed as a range of values not a single value. It may be so, moreover, that the precision inherent in the dissimilarities is such that the precision in one recovered dimension is worse than that for the other dimensions. Our technique is able to tease out and highlight this kind of information.

We proposed the I-Scal algorithm for doing MDS of interval dissimilarities. This algorithm is based on iterative majorization. The advantage is that each iteration yields better I-Stress until no improvement is possible. Simulation studies have shown that I-Scal combined with multiple random start and a rational start yields good quality solutions.

Denœux and Masson (2000) discuss an extension that allows the upper and lower bounds to be transformed. Although it is technically feasible to do so in our case, we do not believe that transformations are generally useful for MDS with interval data. The reason is that by having the available information of a given interval for each dissimilarity, it seems unnatural to destroy this information. The only case making sense to us for applying ordinal transformations is when a subject has given the intervals of rank orders of the pairs of objects. However, we also believe this to be a difficult task for a judge and thus doubt whether it is practical. Therefore, in general we recommend to apply interval MDS without any transformation and perform it directly on the upper and lower bounds.

Moreover, we are easily able to extend our method to deal with the case in which the dissimilarity between object i and object j is an empirical distribution of values or, equivalently, a histogram. For example, we may have enough detailed information so that we can represent the empirical distribution of the dissimilarity as a histogram, for example by 0.10[0, 1], 0.30[1, 2], 0.40[2, 3], 0.20[3, 4], where the first number indicates the relative frequency and values between the brackets define the bin. Then, each object may be represented in the MDS plane by a series of embedded rectangles, one for each bin. For spherical representations, a loss function was provided by Masson and Denœux (2002) without providing an algorithm.

As the I-Scal algorithm is based on iterative majorization, each majorizing function is quadratic in the parameters. Therefore, restrictions such as symbolic MDS for histogram data or the extension of interval MDS to three-way data (by, for example, the weighted Euclidean model) can be easily derived combined with the I-Scal algorithm. We intend to pursue these extensions in future publications.

Acknowledgment

We would like to thank Thierry Denœux and three anonymous reviewers for their comments and suggestions that have improved the quality of this paper.

Appendix A. Majorizing the terms in I-Stress

A.1. Majorizing $\left[d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R})\right]^2$

The square of the upper bound of the distance $\left[d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R})\right]^2$ can be written as

$$\left[d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R})\right]^2=\sum_{s=1}^{p}\left[\left(x_{is}-x_{js}\right)^2+\left(r_{is}+r_{js}\right)^2+2\left|x_{is}-x_{js}\right|\left(r_{is}+r_{js}\right)\right]. \tag{A.1}$$

The term $\left(x_{is}-x_{js}\right)^2$ is standard in MDS and is quadratic in $\mathbf{X}$. The product $\left|x_{is}-x_{js}\right|\left(r_{is}+r_{js}\right)$ can be seen as the product of two functions, i.e., a_1a_2. Consider the following inequality:

$$\left(\frac{a_1}{b_1}-\frac{a_2}{b_2}\right)^2\geqslant 0 \tag{A.2}$$

with strict equality if $a_1=b_1$ and $a_2=b_2$. Expanding (A.2) gives

$$\begin{aligned}0&\leqslant\frac{a_1^2}{b_1^2}+\frac{a_2^2}{b_2^2}-2\frac{a_1a_2}{b_1b_2},\\ 2\,\frac{a_1a_2}{b_1b_2}&\leqslant\frac{a_1^2}{b_1^2}+\frac{a_2^2}{b_2^2},\\ 2a_1a_2&\leqslant\frac{b_2}{b_1}a_1^2+\frac{b_1}{b_2}a_2^2.\end{aligned} \tag{A.3}$$

Note that (A.3) only holds when $b_1>0$ and $b_2>0$. If $b_1=0$ (or $b_2=0$) then we shall replace it by a small ε. Note that this adaptation violates restriction (i) of the requirements of a majorizing function, but by making ε small enough, it should not distort the convergence properties of the majorizing algorithm. In the sequel, we shall use this adaptation implicitly whenever necessary.

Let $\mathbf{Y}$ and $\mathbf{Q}$ be the current known estimates of $\mathbf{X}$ and $\mathbf{R}$. Then, substituting $\left|x_{is}-x_{js}\right|$ for a_1, $\left(r_{is}+r_{js}\right)$ for a_2, $\left|y_{is}-y_{js}\right|$ for b_1, $\left(q_{is}+q_{js}\right)$ for b_2 into (A.3) gives

$$2\left|x_{is}-x_{js}\right|\left(r_{is}+r_{js}\right)\leqslant\frac{q_{is}+q_{js}}{\left|y_{is}-y_{js}\right|}\left(x_{is}-x_{js}\right)^2+\frac{\left|y_{is}-y_{js}\right|}{\left(q_{is}+q_{js}\right)}\left(r_{is}+r_{js}\right)^2. \tag{A.4}$$

To get rid of the crossproduct term $2r_{is}r_{js}$ in $\left(r_{is}+r_{js}\right)^2$, we apply (A.3) again:

$$2r_{is}r_{js}\leqslant\frac{q_{js}}{q_{is}}r_{is}^2+\frac{q_{is}}{q_{js}}r_{js}^2. \tag{A.5}$$

Combining (A.1), (A.4), and (A.5), and multiplying by w_{ij} gives the majorizing inequality

$$w_{ij}\left[d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R})\right]^2 \leqslant \sum_{s=1}^{p}\left[\alpha_{ijs}^{(1)}\left(x_{is}-x_{js}\right)^2+\alpha_{ijs}^{(2)}r_{is}^2+\alpha_{jis}^{(2)}r_{js}^2\right] \tag{A.6}$$

with $\alpha_{ijs}^{(1)}=w_{ij}\left[1+\left(q_{is}+q_{js}\right)/\left|y_{is}-y_{js}\right|\right]$ and $\alpha_{ijs}^{(2)}=w_{ij}\left[\left|y_{is}-y_{js}\right|+\left(q_{is}+q_{js}\right)\right]/q_{is}$.

A.2. Majorizing $-d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R})$

Minus the upper bound of the distance, $-d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R})$, can be written as

$$-d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R})=-\left(\sum_{s=1}^{p}\left[\left|x_{is}-x_{js}\right|+\left(r_{is}+r_{js}\right)\right]^2\right)^{1/2}. \tag{A.7}$$

This term is concave in the parameters $\mathbf{X}$ and $\mathbf{R}$ and hence can be majorized by a linear function in $\mathbf{X}$ and $\mathbf{R}$. Consider the Cauchy–Schwarz inequality $\|\mathbf{a}\|\|\mathbf{b}\|\geqslant\mathbf{a}'\mathbf{b}$. Dividing both sides of the inequality by $-\|\mathbf{b}\|$ (assuming that $\|\mathbf{b}\|>0$) gives

$$-\|\mathbf{a}\|\leqslant\begin{cases}-\mathbf{a}'\mathbf{b}/\|\mathbf{b}\| & \text{if } \|\mathbf{b}\|>0,\\ 0 & \text{if } \|\mathbf{b}\|=0.\end{cases} \tag{A.8}$$

Note that the inequality $-\|\mathbf{a}\|\leqslant 0$ if $\|\mathbf{b}\|=0$ does not violate any of the majorization requirements. Applying (A.8) to $-d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R})$ gives

$$-d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R})\leqslant\begin{cases}-\dfrac{\sum_{s=1}^{p}\left[\left|x_{is}-x_{js}\right|+\left(r_{is}+r_{js}\right)\right]\left[\left|y_{is}-y_{js}\right|+\left(q_{is}+q_{js}\right)\right]}{d_{ij}^{(\mathrm{U})}(\mathbf{Y},\mathbf{Q})} & \text{if } d_{ij}^{(\mathrm{U})}(\mathbf{Y},\mathbf{Q})>0,\\ 0 & \text{if } d_{ij}^{(\mathrm{U})}(\mathbf{Y},\mathbf{Q})=0.\end{cases} \tag{A.9}$$

Applying (A.8) to the single term $-\left|x_{is}-x_{js}\right|$ gives

$$-\left|x_{is}-x_{js}\right|\leqslant\begin{cases}-\dfrac{\left(x_{is}-x_{js}\right)\left(y_{is}-y_{js}\right)}{\left|y_{is}-y_{js}\right|} & \text{if } \left|y_{is}-y_{js}\right|>0,\\ 0 & \text{if } \left|y_{is}-y_{js}\right|=0\end{cases} \tag{A.10}$$

(Kiers and Groenen, 1996). Combining the results of (A.9) and (A.10) and multiplying by $w_{ij}\delta_{ij}^{(\mathrm{U})}$ gives the majorizing inequality

$$-w_{ij}\delta_{ij}^{(\mathrm{U})}d_{ij}^{(\mathrm{U})}(\mathbf{X},\mathbf{R})\leqslant-\sum_{s=1}^{p}\left[\beta_{ijs}^{(1)}\left(x_{is}-x_{js}\right)\left(y_{is}-y_{js}\right)+\beta_{ijs}^{(2)}\left(r_{is}+r_{js}\right)\right] \tag{A.11}$$

with

$$\beta_{ijs}^{(1)}=\begin{cases}\dfrac{w_{ij}\delta_{ij}^{(\mathrm{U})}\left[\left|y_{is}-y_{js}\right|+\left(q_{is}+q_{js}\right)\right]}{\left|y_{is}-y_{js}\right|d_{ij}^{(\mathrm{U})}(\mathbf{Y},\mathbf{Q})} & \text{if } \left|y_{is}-y_{js}\right|>0 \quad\text{and}\quad d_{ij}^{(\mathrm{U})}(\mathbf{Y},\mathbf{Q})>0,\\ 0 & \text{if } \left|y_{is}-y_{js}\right|=0 \quad\text{or}\quad d_{ij}^{(\mathrm{U})}(\mathbf{Y},\mathbf{Q})=0,\end{cases}$$

$$\beta_{ijs}^{(2)}=\begin{cases}\dfrac{w_{ij}\delta_{ij}^{(\mathrm{U})}\left[\left|y_{is}-y_{js}\right|+\left(q_{is}+q_{js}\right)\right]}{d_{ij}^{(\mathrm{U})}(\mathbf{Y},\mathbf{Q})} & \text{if } d_{ij}^{(\mathrm{U})}(\mathbf{Y},\mathbf{Q})>0,\\ 0 & \text{if } d_{ij}^{(\mathrm{U})}(\mathbf{Y},\mathbf{Q})=0.\end{cases}$$

A.3. Majorizing $[d_{ij}^{(\mathrm{L})}(\mathbf{X},\mathbf{R})]^2$

To majorize $\left[d_{ij}^{(\mathrm{L})}(\mathbf{X},\mathbf{R})\right]^2$, we start by considering a majorizing function for $\max\left[0,\left|x_{is}-x_{js}\right|-\left(r_{is}+r_{js}\right)\right]^2$. Let $a_1=\left|x_{is}-x_{js}\right|$, $a_2=r_{is}+r_{js}$, $b_1=\left|y_{is}-y_{js}\right|$, and $b_2=q_{is}+q_{js}$. Then $\max\left[0,\left|x_{is}-x_{js}\right|-\left(r_{is}+r_{js}\right)\right]^2=\max\left[0,a_1-a_2\right]^2$. To simplify notation even further, we use $a=a_1-a_2$ and $b=b_1-b_2$ so that $\max\left[0,a_1-a_2\right]^2=\max[0,a]^2$. Thus, the function $\max[0,a]^2$ is the zero line when $a<0$ and is a^2 for $a\geqslant 0$. This function can be quadratically majorized by two different functions, depending on the current estimate b. For $b\geqslant 0$ the function $\max[0,a]^2=a^2$, so that $\max[0,a]^2$ can be majorized by a^2. For positive a, the majorizing function coincides with the original function $\max[0,a]^2$ and for negative a, $\max[0,a]^2<a^2$ as should be the case in majorization. For $b<0$, the function $\min[0,a]^2$ can be majorized by the function $(a-b)^2$. It can be verified that in this case, too, the majorizing function is never smaller than the original function. Summarizing these majorization results and resubstituting gives

$$\max\left[0,a_1-a_2\right]^2\leqslant\begin{cases}(a_1-a_2)^2 & \text{if } b_1-b_2\geqslant 0,\\ \left[(a_1-a_2)-(b_1-b_2)\right]^2 & \text{if } b_1-b_2<0.\end{cases}\tag{A.12}$$

It can be verified that (A.12) satisfies the requirements of a majorizing function. Working out both conditions of (A.12) yield terms a_1^2, a_2^2, a_1, a_2, and $-2a_1a_2$. The only complicating term is the product $-2a_1a_2$ for which we derive a majorizing function. Consider the following inequalities

$$\begin{aligned}&0\leqslant\left[(a_1+a_2)-(b_1+b_2)\right]^2,\\ &0\leqslant(a_1+a_2)^2+(b_1+b_2)^2-2\left(a_1+a_2\right)\left(b_1+b_2\right),\\ &0\leqslant a_1^2+a_2^2+2a_1a_2+(b_1+b_2)^2-2\left(a_1+a_2\right)\left(b_1+b_2\right),\\ &-2a_1a_2\leqslant a_1^2+a_2^2+(b_1+b_2)^2-2\left(a_1+a_2\right)\left(b_1+b_2\right).\end{aligned}\tag{A.13}$$

Thus, (A.13) is the majorization inequality for majorizing minus the product of two functions. Combining (A.12) and (A.13) gives

$$\max\left[0,a_1-a_2\right]^2\leqslant\begin{cases}2\left(a_1^2+a_2^2\right)-2\left(a_1+a_2\right)\left(b_1+b_2\right)+\left(b_1+b_2\right)^2 & \text{if } b_1\geqslant b_2,\\ 2\left(a_1^2+a_2^2\right)-4a_1b_1-4a_2b_2+2\left(b_1^2+b_2^2\right) & \text{if } b_1<b_2.\end{cases}\tag{A.14}$$

Note that (A.14) has terms with $-a_1$ that after resubstitution yields a term with $-\left|x_{is}-x_{js}\right|$ that can be majorized by (A.10). Also, a_2^2 equals after resubstitution $r_{is}^2+r_{js}^2+2r_{is}r_{js}$. The crossproduct term $2r_{is}r_{js}$ is majorized by (A.5). Combining all results and multiplying by w_{ij} gives the following majorization inequality:

$$\begin{aligned}w_{ij}\left[d_{ij}^{(\mathrm{L})}(\mathbf{X},\mathbf{R})\right]^2\leqslant\sum_{s=1}^{p}\Big[&\alpha_{ij}^{(3)}\left(x_{is}-x_{js}\right)^2+\alpha_{ijs}^{(4)}r_{is}^2+\alpha_{jis}^{(4)}r_{js}^2\\ &-2\beta_{ijs}^{(3)}\left(x_{is}-x_{js}\right)\left(y_{is}-y_{js}\right)-2\beta_{ijs}^{(4)}\left(r_{is}+r_{js}\right)+\gamma_{ijs}^{(1)}\Big]\end{aligned}\tag{A.15}$$

with

$$\begin{aligned}\alpha_{ij}^{(3)}&=2w_{ij},\\ \alpha_{ijs}^{(4)}&=2w_{ij}\left(1+q_{js}/q_{is}\right),\\ \beta_{ijs}^{(3)}&=\begin{cases}\dfrac{w_{ij}\left[\left|y_{is}-y_{js}\right|+\left(q_{is}+q_{js}\right)\right]}{\left|y_{is}-y_{js}\right|} & \text{if } \left|y_{is}-y_{js}\right|\geqslant q_{is}+q_{js} \quad\text{and}\quad \left|y_{is}-y_{js}\right|>0,\\ 2w_{ij} & \text{if } \left|y_{is}-y_{js}\right|<q_{is}+q_{js} \quad\text{and}\quad \left|y_{is}-y_{js}\right|>0,\\ 0 & \text{if } \left|y_{is}-y_{js}\right|=0,\end{cases}\\ \beta_{ijs}^{(4)}&=\begin{cases}w_{ij}\left[\left|y_{is}-y_{js}\right|+\left(q_{is}+q_{js}\right)\right] & \text{if } \left|y_{is}-y_{js}\right|\geqslant q_{is}+q_{js},\\ 2w_{ij}\left(q_{is}+q_{js}\right) & \text{if } \left|y_{is}-y_{js}\right|<q_{is}+q_{js},\end{cases}\\ \gamma_{ijs}^{(1)}&=\begin{cases}w_{ij}\left[\left|y_{is}-y_{js}\right|+\left(q_{is}+q_{js}\right)\right]^2 & \text{if } \left|y_{is}-y_{js}\right|\geqslant q_{is}+q_{js},\\ 2w_{ij}\left[\left|y_{is}-y_{js}\right|^2+\left(q_{is}+q_{js}\right)^2\right] & \text{if } \left|y_{is}-y_{js}\right|<q_{is}+q_{js}.\end{cases}\end{aligned}$$

A.4. Majorizing $-d_{ij}^{(\mathrm{L})}(\mathbf{X},\mathbf{R})$

To find a majorizing function for minus the lower bound of the distance, $-d_{ij}^{(\mathrm{L})}(\mathbf{X},\mathbf{R})$, we can make use of the majorization inequality (A.8) based on Cauchy–Schwarz. This allows us to write

$$-d_{ij}^{(\mathrm{L})}(\mathbf{X},\mathbf{R}) \leqslant \begin{cases} \sum_{s=1}^{p} - \frac{\max[0,|x_{is}-x_{js}|-(r_{is}+r_{js})]\max[0,|y_{is}-y_{js}|-(q_{is}+q_{js})]}{d_{ij}^{(\mathrm{L})}(\mathbf{Y},\mathbf{Q})} & \text{if } d_{ij}^{(\mathrm{L})}(\mathbf{Y},\mathbf{Q})>0, \\ 0 & \text{if } d_{ij}^{(\mathrm{L})}(\mathbf{Y},\mathbf{Q})=0. \end{cases} \tag{A.16}$$

Thus, (A.16) yields a function of the sum of $-\max\left[0,\left|x_{is}-x_{js}\right|-\left(r_{is}+r_{js}\right)\right]$. Using the same notational simplification as in the previous subsection, this can be rewritten as $-\max[0,a_1-a_2]$. This function can be majorized as follows:

$$-\max[0,a_1-a_2] \leqslant \begin{cases} -(a_1-a_2) & \text{if } b_1 \geqslant b_2, \\ 0 & \text{if } b_1 < b_2. \end{cases} \tag{A.17}$$

Again we majorize the terms $-a_1$ (which is equal to $-\left|x_{is}-x_{js}\right|$ after resubstitution) by (A.10). The term $+a_2$ leads after resubstitution to a term $\left(r_{is}+r_{js}\right)$. For algorithmic reasons it is better to only have linear terms that are negative. For this reason, we apply another majorization step to $+r_{is}$ (and r_{js}), i.e.,

$$r_{is} \leqslant \frac{1}{2}\frac{r_{is}^2}{q_{is}} + \frac{1}{2}q_{is}. \tag{A.18}$$

Combining these results and multiplication by $w_{ij}\delta_{ij}^{(\mathrm{L})}$ yields the majorizing function

$$-w_{ij}\delta_{ij}^{(\mathrm{L})}d_{ij}^{(\mathrm{L})}(\mathbf{X},\mathbf{R}) \leqslant \sum_{s=1}^{p}\left[\frac{1}{2}\alpha_{ijs}^{(5)}r_{is}^2 + \frac{1}{2}\alpha_{jis}^{(5)}r_{js}^2 - \beta_{ijs}^{(5)}(x_{is}-x_{js})\left(y_{is}-y_{js}\right) + \frac{1}{2}\gamma_{ijs}^{(2)}\right] \tag{A.19}$$

with

$$\alpha_{ijs}^{(5)} = \begin{cases} \dfrac{w_{ij}\delta_{ij}^{(\mathrm{L})}\max\left[0,\left|y_{is}-y_{js}\right|-\left(q_{is}+q_{js}\right)\right]}{q_{is}d_{ij}^{(\mathrm{L})}(\mathbf{Y},\mathbf{Q})} & \text{if } \left|y_{is}-y_{js}\right| \geqslant q_{is}+q_{js}, \\ 0 & \text{if } \left|y_{is}-y_{js}\right| < q_{is}+q_{js}, \end{cases}$$

$$\beta_{ijs}^{(5)} = \begin{cases} \dfrac{w_{ij}\delta_{ij}^{(\mathrm{L})}\max\left[0,\left|y_{is}-y_{js}\right|-\left(q_{is}+q_{js}\right)\right]}{\left|y_{is}-y_{js}\right|d_{ij}^{(\mathrm{L})}(\mathbf{Y},\mathbf{Q})} & \text{if } \left|y_{is}-y_{js}\right| \geqslant q_{is}+q_{js} \text{ and } \left|y_{is}-y_{js}\right|>0, \\ 0 & \text{if } \left|y_{is}-y_{js}\right| < q_{is}+q_{js} \text{ or } \left|y_{is}-y_{js}\right|=0, \end{cases}$$

$$\gamma_{ijs}^{(2)} = \begin{cases} \dfrac{w_{ij}\delta_{ij}^{(\mathrm{L})}\left(q_{is}+q_{js}\right)\max\left[0,\left|y_{is}-y_{js}\right|-\left(q_{is}+q_{js}\right)\right]}{d_{ij}^{(\mathrm{L})}(\mathbf{Y},\mathbf{Q})} & \text{if } \left|y_{is}-y_{js}\right| \geqslant q_{is}+q_{js}, \\ 0 & \text{if } \left|y_{is}-y_{js}\right| < q_{is}+q_{js}. \end{cases}$$

Appendix B. Rational start for MDS of interval data

It is well known that classical scaling (Torgerson, 1958; Gower, 1966) has a duality property with standard or classical principal components analysis, PCA, when the dissimilarities are Euclidean distances. To avoid local minimum problems we propose to use a rational start, using InterScal (Rodríguez, 2000), an algorithm for MDS of interval data that produces results similar to the vertices method for PCA of interval data, developed by Chouakria et al. (2000, pp. 200–212). The interval data for n objects on m variables can be described by two matrices: the $n \times m$ matrices $\mathbf{H}^{(\mathrm{U})}$ and $\mathbf{H}^{(\mathrm{L})}$ with the upper and lower bounds. The trick in the vertices method is that each object can be thought of as being somewhere in a hyperbox defined by all the intervals on the m variables. Therefore, the vertices method for PCA

represents each object by all its vertices in the $2^m \times m$ matrix $\mathbf{M}_i$. An example of $\mathbf{M}_i$ with $m = 3$ variables is given by

$$\mathbf{M}_i = \begin{bmatrix} h_{i1}^{(\mathrm{L})} & h_{i2}^{(\mathrm{L})} & h_{i3}^{(\mathrm{L})} \\ h_{i1}^{(\mathrm{L})} & h_{i2}^{(\mathrm{L})} & h_{i3}^{(\mathrm{U})} \\ h_{i1}^{(\mathrm{L})} & h_{i2}^{(\mathrm{U})} & h_{i3}^{(\mathrm{L})} \\ h_{i1}^{(\mathrm{L})} & h_{i2}^{(\mathrm{U})} & h_{i3}^{(\mathrm{U})} \\ h_{i1}^{(\mathrm{U})} & h_{i2}^{(\mathrm{L})} & h_{i3}^{(\mathrm{L})} \\ h_{i1}^{(\mathrm{U})} & h_{i2}^{(\mathrm{L})} & h_{i3}^{(\mathrm{U})} \\ h_{i1}^{(\mathrm{U})} & h_{i2}^{(\mathrm{U})} & h_{i3}^{(\mathrm{L})} \\ h_{i1}^{(\mathrm{U})} & h_{i2}^{(\mathrm{U})} & h_{i3}^{(\mathrm{U})} \end{bmatrix}. \tag{B.1}$$

The vertices method for PCA proceeds by applying standard PCA to the $2^m n \times m$ matrix $\mathbf{M}$ that has all matrices $\mathbf{M}_i$ stacked underneath each other, that is,

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \mathbf{M}_3 \\ \vdots \\ \mathbf{M}_n \end{bmatrix}.$$

Let the singular value decomposition of $\mathbf{M}$ be given by $\mathbf{M} = \mathbf{P}\mathbf{\Phi}\mathbf{Q}'$ with $\mathbf{P}'\mathbf{P} = \mathbf{Q}'\mathbf{Q} = \mathbf{I}$ and $\mathbf{\Phi}$ a diagonal matrix with nonnegative singular values ordered from large to small. Then, the rank p approximation of $\mathbf{H}$ in PCA solution is given by $\mathbf{P}_p\mathbf{\Phi}_p\mathbf{Q}_p' = \mathbf{X}\mathbf{Q}_p'$, where the subscript p denotes that the first p columns are used for $\mathbf{P}$ and $\mathbf{Q}$ and the first p rows and columns of $\mathbf{\Phi}$.

To see the equivalence between classical MDS and PCA, we need a compact expression of the matrix of squared distances $\mathbf{D}$ of squared distances between the rows of $\mathbf{M}$, that is,

$$\mathbf{D} = \mathbf{a}\mathbf{1}' + \mathbf{1}\mathbf{a}' - 2\mathbf{M}\mathbf{M}', \tag{B.2}$$

where $\mathbf{a}$ is the vector with the diagonal elements of $\mathbf{M}\mathbf{M}'$. Let $\mathbf{J} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}'$ be the centering matrix so that $\mathbf{J}\mathbf{1} = \mathbf{0}$ and $\mathbf{1}'\mathbf{J} = \mathbf{0}'$. Pre- and post-multiplying (B.2) by $\mathbf{J}$ and also multiplying by the factor $-\frac{1}{2}$ gives

$$\mathbf{M}\mathbf{M}' = -\tfrac{1}{2}\mathbf{J}\mathbf{D}\mathbf{J}. \tag{B.3}$$

Classical MDS minimizes $\|\mathbf{X}\mathbf{X}' - (-\frac{1}{2}\mathbf{J}\mathbf{D}\mathbf{J})\|^2$ by computing the eigendecomposition $-\frac{1}{2}\mathbf{J}\mathbf{D}\mathbf{J} = \mathbf{P}\mathbf{\Phi}^2\mathbf{P}'$ and choosing $\mathbf{X} = \mathbf{P}_p\mathbf{\Phi}_p$. The above proves that classical MDS on the matrix of squared Euclidean distances between the rows of $\mathbf{M}$ is the same as PCA on $\mathbf{M}$ directly. Classical MDS can also be used directly on a dissimilarity matrix $\mathbf{\Delta}$. In this case, the eigendecomposition of $-\frac{1}{2}\mathbf{J}\mathbf{\Delta}^{(2)}\mathbf{J}$ is taken, where $\mathbf{\Delta}^{(2)}$ denotes the matrix of squared dissimilarities.

To get a symbolic MDS method that has a *duality property* with vertices PCA, when dissimilarity is modelled by an Euclidean distance, we need as input the dissimilarities between all $2^m n$ rows of the matrix $\mathbf{M}$ defined above, because vertices PCA effects a classical PCA of the matrix $\mathbf{M}$. Thus, we would need as input a matrix $\mathbf{\Delta}$ of size $2^m n \times 2^m n$ but it is clearly impossible to construct a matrix of this size, because we only have two dissimilarities, that is the maximum and the minimum, for each pair of objects. So it seems impossible to find a symbolic MDS method that has a *duality property* with vertices PCA. Therefore, we propose to find an approximate solution.

The idea, then, is to carry out an MDS of the distance matrix $\widetilde{\mathbf{\Delta}}$ defined below. For each hypercube (thus for each object i), the matrix $\widetilde{\mathbf{\Delta}}$ has two rows. In the first row, we use the minimum dissimilarity and the maximum dissimilarity among a hypercube and itself, whereas we use the dissimilarity minimum and the average dissimilarity among each different couple of hypercubes, that is to say, we use $2m$ dissimilarities. In the second row of the matrix $\widetilde{\mathbf{\Delta}}$, we use the maximum dissimilarity and the minimum dissimilarity among a hypercube and itself, and we use the average dissimilarity and the maximum dissimilarity among each different couple of hypercubes, in this row we also use $2n$

dissimilarities, but as the average dissimilarities were already employed we really use n dissimilarities, therefore for each hypercube we use $3n$ dissimilarities. Then, since $d(x, y) = d(y, x)$, in total we use $(3/2)n(n+1) > n(n+1)$ dissimilarities. Note that $\widetilde{\mathbf{\Delta}}$ is a symmetric matrix and its size is $2n \times 2n$. Since for each hyperbox we have two rows, we can compute a coordinate minimum and maximum, that is, coordinates of interval type. The matrix $\widetilde{\mathbf{\Delta}}$ is given by

$$\widetilde{\mathbf{\Delta}} = \begin{bmatrix} 0 & 0 & \delta_{12}^{(\mathrm{L})} & \overline{\delta}_{12} & \delta_{13}^{(\mathrm{L})} & \overline{\delta}_{13} & \cdots & \delta_{1m}^{(\mathrm{L})} & \overline{\delta}_{1m} \\ 0 & 0 & \overline{\delta}_{12} & \delta_{12}^{(\mathrm{U})} & \overline{\delta}_{13} & \delta_{13}^{(\mathrm{U})} & \cdots & \overline{\delta}_{1m} & \delta_{1m}^{(\mathrm{U})} \\ \delta_{21}^{(\mathrm{L})} & \overline{\delta}_{21} & 0 & 0 & \delta_{23}^{(\mathrm{L})} & \overline{\delta}_{23} & \cdots & \delta_{2m}^{(\mathrm{L})} & \overline{\delta}_{2m} \\ \overline{\delta}_{21} & \delta_{21}^{(\mathrm{U})} & 0 & 0 & \overline{\delta}_{23} & \delta_{23}^{(\mathrm{U})} & \cdots & \overline{\delta}_{2m} & \delta_{2m}^{(\mathrm{U})} \\ \delta_{31}^{(\mathrm{L})} & \overline{\delta}_{31} & \delta_{32}^{(\mathrm{L})} & \overline{\delta}_{32} & 0 & 0 & \cdots & \delta_{3m}^{(\mathrm{L})} & \overline{\delta}_{3m} \\ \overline{\delta}_{31} & \delta_{31}^{(\mathrm{U})} & \overline{\delta}_{32} & \delta_{32}^{(\mathrm{U})} & 0 & 0 & \cdots & \overline{\delta}_{3m} & \delta_{3m}^{(\mathrm{U})} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \delta_{m1}^{(\mathrm{L})} & \overline{\delta}_{m1} & \delta_{m2}^{(\mathrm{L})} & \overline{\delta}_{m2} & \delta_{m3}^{(\mathrm{L})} & \overline{\delta}_{m3} & \cdots & 0 & 0 \\ \overline{\delta}_{m1} & \delta_{m1}^{(\mathrm{U})} & \overline{\delta}_{m2} & \delta_{m2}^{(\mathrm{U})} & \overline{\delta}_{m3} & \delta_{m3}^{(\mathrm{U})} & \cdots & 0 & 0 \end{bmatrix}, \tag{B.4}$$

where $\overline{\delta}_{ij} = \left(\delta_{ij}^{(\mathrm{L})} + \delta_{ij}^{(\mathrm{U})}\right) \Big/ 2$.

The InterScal algorithm for rational start for MDS of interval-value dissimilarity data is performed as follows:

1. Obtain the dissimilarities $\delta_{ij}^{(\mathrm{L})}$ and $\delta_{ij}^{(\mathrm{U})}$ for all $i, j = 1, 2, \ldots, n$.
2. Compute the matrix $\widetilde{\mathbf{\Delta}}$ according to (B.4).
3. Find the matrix $\mathbf{B} = -\frac{1}{2}\mathbf{J}\widetilde{\mathbf{\Delta}}^{(2)}\mathbf{J}$ with $\mathbf{J}$ the centering matrix.
4. Find the eigenvalues $\mathbf{\Phi}^2$ and eigenvectors $\mathbf{P}$ of $\mathbf{B}$.
5. Compute the coordinates of the $2n$ points in p dimensions using the formula

$$y_{is} = p_{is}\phi_{ss} \quad \text{for } i = 1, 2, \ldots, 2n \quad \text{and} \quad s = 1, 2, \ldots, p.$$

6. Construct the center coordinates $\mathbf{X}$ and the spread $\mathbf{R}$ of the hypercube for object i by for each dimension s

$$x_{is} = \left(y_{2i,s} + y_{2i-1,s}\right) \Big/ 2,$$
$$r_{is} = \left|y_{2i,s} - y_{2i-1,s}\right| \Big/ 2.$$

Note that the solution for $\mathbf{X}$ is not unique as $\mathbf{B} = \mathbf{P}\mathbf{\Phi}^2\mathbf{P}' = \mathbf{Y}\mathbf{T}\mathbf{T}'\mathbf{Y}'$ for any $\mathbf{T}\mathbf{T}' = \mathbf{I}$. Any rigid rotation is an example of matrix of type $\mathbf{T}$. We choose the solution corresponding to principal axes. The first axis maximizes the variance of the vertices of the hypercube. However, since any rotation is also a solution, one may wish to rotate the principal axes solution in order to obtain axes which are more interpretable.

Let us consider the special case when all the intervals of the dissimilarities are zero, that is, $\delta_{ij}^{(\mathrm{L})} = \delta_{ij}^{(\mathrm{U})} = \delta_{ij}$. In this case, $\overline{\delta}_{ij}$ also equals δ_{ij}. Then $\widetilde{\mathbf{\Delta}}$ has blocks of 2×2 matrices for all combination ij with all four elements equal to δ_{ij}. Then, INTERSCAL is exactly equal to classical scaling on the $n \times n$ matrix of dissimilarities with elements δ_{ij}, except that each object appears twice, hence the intervals collapse to points.

References

Bock, H.-H., Diday, E., 2000. Analysis of Symbolic Data. Springer, Berlin.

Borg, I., Groenen, P.J.F., 2005. Modern Multidimensional Scaling: Theory and Applications. second ed. Springer, New York.

Carroll, J.D., 1972. Individual differences and multidimensional scaling. In: Theory and Applications in the Behavioral Sciences, vol. I, Theory. Seminar Press, New York.

Carroll, J.D., Winsberg, S., 1995. Fitting an extended INDSCAL model to 3-way proximity data. J. Classification 12, 57–71.

Chouakria, A., Cazes, P., Diday, E., 2000. Symbolic Principal Components Analysis. Springer, Berlin, pp. 200–212.

De Leeuw, J., 1994. Block relaxation algorithms in statistics. In: Bock, H.-H., Lenski, W., Richter, M.M. (Eds.), Information Systems and Data Analysis. Springer, Berlin, pp. 308–324.

De Leeuw, J., Heiser, W.J., 1980. Multidimensional scaling with restrictions on the configuration. In: Krishnaiah, P.R. (Ed.), Multivariate Analysis, vol. V. North-Holland Publishing Company, Amsterdam, The Netherlands, pp. 501–522.

Denœux, T., Masson, M., 2000. Multidimensional scaling of interval-valued dissimilarity data. Pattern Recognition Lett. 21, 83–92.

Gower, J.C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53, 325–338.

Groenen, P.J.F., Mathar, R., Heiser, W.J., 1995. The majorization approach to multidimensional scaling for Minkowski distances. J. Classification 12, 3–19.

Groenen, P.J.F., DeLeeuw, J., Mathar, R., 1996. Least squares multidimensional scaling with transformed distances. In: Gaul, W., Pfeifer, D. (Eds.), Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, pp. 177–185.

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

Heiser, W.J., 1995. Convergent Computation by Iterative Majorization: Theory and Applications in Multidimensional Data Analysis. Oxford University Press, Oxford, pp. 157–189.

Hunter, D.R., Lange, K., 2004. A autorial on MM algorithms. Amer. Statist. 39, 30–37.

Kiers, H.A.L., 2002. Setting up alternating least squares and iterative majorization algorithms for solving various matrix optimization problems. Comput. Statist. Data Anal. 41, 157–170.

Kiers, H.A.L., Groenen, P.J.F., 1996. A monotonically convergent algorithm for orthogonal congruence rotation. Psychometrika 61, 375–389.

Kruskal, J.B., 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29, 1–27.

Kruskal, J.B., 1964b. Nonmetric multidimensional scaling: a numerical method. Psychometrika 29, 115–129.

Masson, M., Denœux, T., 2002. Multidimensional scaling of fuzzy dissimilarity data. Fuzzy Sets and Systems 128, 339–352.

McAdams, S., Winsberg, S., 1999. Multidimensional scaling of musical timbre constrained by physical parameters. J. Acoust. Soc. Amer. 105, 1273.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., Krimphoff, J., 1995. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. Psychological Res. 58, 177–192.

Plomp, R., 1970. Timbre as a Multidimensional Attribute of Complex Tones. Sijthoff, Leiden, pp. 397–414 (Chapter 4).

Rodríguez, O., 2000. Classification et modèles linéaires en analyse des données symboliques. Ph.D. Thesis, Université Paris IX Dauphine, Paris.

Torgerson, W.S., 1958. Theory and Methods of Scaling. Wiley, New York.

Tucker, L.R., 1951. A method for synthesis of factor analysis studies. Personnel Research Section Report 984, Department of the Army, Washington, DC.

Weinberg, S.L., Menil, V.C., 1993. The recovery of structure in linear and ordinal data: INDSCAL versus ALSCAL. Multivariate Behavioral Res. 28, 215–233.

Winsberg, S., Carroll, J.D., 1989. A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. Psychometrika 54, 217–219.

Winsberg, S., De Soete, G., 1993. A latent class approach to fitting the weighted euclidean model, CLASCAL. Psychometrika 58, 315–330.

Winsberg, S., De Soete, G., 1997. Multidimensional scaling with constrained dimensions: CONSCAL. British J. Math. Statist. Psych. 50, 55–72.