Symbolic Multidimensional Scaling

P.J.F. Groenen* Y. Terada †

May 27, 2015

Econometric Institute Report EI 2015-15

Abstract

Multidimensional scaling (MDS) is a technique that visualizes dissimilarities between pairs of objects as distances between points in a low dimensional space. In symbolic MDS, a dissimilarity is not just a value but can represent an interval or even a histogram. Here, we present an overview of developments for symbolic MDS. We discuss how interval dissimilarities they can be represented by (concentric) circles or rectangles, how replications can be represented by a three-way MDS version, and show how nested intervals of distances can be obtained for representing histogram dissimilarities. The various models are illustrated by empirical examples.

*Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands (e-mail: groenen@few.eur.nl)

†HHS office 3, The Center for Information and Neural Networks (CiNet), 1-4 Yamadaoka, Suita City, Osaka, Japan, 565-0871 (e-mail: terada@nict.go.jp (teradayoshikazu@gmail.com))

1 Introduction

Standard multidimensional scaling (MDS) is a common visualizing technique to construct a point configuration in a low-dimensional space that preserves given dissimilarities among a set of objects as distances between points. Here, we present an overview of MDS-type models for dissimilarities that are not a single value, but are a range (interval) or even a histogram. In case the intervals or histograms collapse to a single value for all dissimilarities, the methods discussed here simplify into regular MDS. To model the interval of a dissimilarity, one needs a visualization of the pairs of objects that allows to obtain a minimum and maximum of the distance as measured between the two representations of the objects in a low dimensional space. For interval dissimilarities, two shapes for representing objects were proposed by Denœux and Masson (2000): the circle and the rectangle in two dimensions, or the hypersphere and hyperbox in higher dimensionalities. The minimal and maximal distances are simply defined by the smallest and largest Euclidean distances between the pairs of shapes.

The main purpose of this report is to provide a current overview for symbolic MDS. We discuss the definitions, the loss functions that are optimized, and present some illustrative examples.

The remainder of this report is organized as follows. In Section 2, we introduce interval-valued dissimilarity data and describe two important model, called the hypersphere model and the hyperbox model, for MDS of interval-valued dissimilarities. We also discuss the extension of interval MDS for three-way interval-valued MDS in Section 3. Next, we deal with histogram-valued (or percentile-valued) dissimilarity data and introduce some natural extensions of the hypersphere model and the hyperbox model for histogram-valued dissimilarity data. Finally, we conclude this report and discuss about some future works.

2 Interval MDS

The main characteristic of interval MDS is that there is not a single observed value of the dissimilarities but that they are represented by an interval, that is,

$$\Delta_{\text{Int}} := \begin{bmatrix} - & [\delta_{12}^{(L)}, \delta_{12}^{(U)}] & \cdots & [\delta_{1n}^{(L)}, \delta_{1n}^{(U)}] \\ [\delta_{21}^{(L)}, \delta_{21}^{(U)}] & - & \cdots & [\delta_{2n}^{(L)}, \delta_{2n}^{(U)}] \\ \vdots & \vdots & \ddots & \vdots \\ [\delta_{n1}^{(L)}, \delta_{n1}^{(U)}] & [\delta_{n2}^{(L)}, \delta_{n2}^{(U)}] & \cdots & - \end{bmatrix}, \tag{1}$$

where the upper and lower bounds of the interval are given by $\delta_{ij}^{(U)}$ and $\delta_{ij}^{(L)}$ with $\delta_{ij}^{(U)} \geq \delta_{ij}^{(L)} \geq 0$, $\delta_{ij}^{(L)} = \delta_{ji}^{(L)}$, and intervals are assumed to be symmetric, that is, $\delta_{ij}^{(U)} = \delta_{ji}^{(U)}$ $(i, j = 1, \ldots, n)$ with n the number of objects. Denœux and Masson (2000) proposed to find regions $R_i \subset \mathbb{R}^q$ $(i = 1, \ldots, n)$ of a given q dimensional space in such a way that the differences between given dissimilarity intervals $[\delta_{ij}^{(L)}, \delta_{ij}^{(U)}]$ and reconstructed distance intervals $[d_{ij}^{(L)}(\mathcal{R}), d_{ij}^{(U)}(\mathcal{R})]$ are minimized. Therefore, the largest and smallest distances between any two regions are defined by

$$
\begin{aligned}
d_{ij}^{(L)}(\mathcal{R}) &:= \min_{\mathbf{x}_i \in R_i,\, \mathbf{x}_j \in R_j} \|\mathbf{x}_i - \mathbf{x}_j\|, \text{ and} \\
d_{ij}^{(U)}(\mathcal{R}) &:= \max_{\mathbf{x}_i \in R_i,\, \mathbf{x}_j \in R_j} \|\mathbf{x}_i - \mathbf{x}_j\|.
\end{aligned}
$$

Figure 1 gives an example of two irregularly shaped regions R_i and the corresponding maximum distance $d_{ij}^{(U)}(\mathcal{R})$ and minimum distance $d_{ij}^{(L)}(\mathcal{R})$.

Denœux and Masson (2000) defined interval MDS as the minimization of the Stress function

$$
\begin{aligned}
\sigma_{\text{Int}}^2(\mathcal{R}) \;:=\; & \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij} \left(\delta_{ij}^{(L)} - d_{ij}^{(L)}(\mathcal{R})\right)^2 + \\
& \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij} \left(\delta_{ij}^{(U)} - d_{ij}^{(U)}(\mathcal{R})\right)^2, \qquad (2)
\end{aligned}
$$

where w_{ij} is a given nonnegative weight. We shall assume that the matrix with w_{ij} is irreducible, that is, it is not possible to partition the objects in subsets such that all w_{ij} between objects in different subsets are zero. If so, then the problem can be split into separate interval MDS problems. Note that $\sigma_{\text{Int}}^2(\mathcal{R})$ is sometimes referred to as the least-squares approach to interval MDS. Unfortunately, it is difficult to optimize Stress function $\sigma_{\text{Int}}(\mathcal{R})$ (or the Stress function of the possibility approach) over all sets of n connected closed region since the possible shapes of the regions R_i are potentially very irregular and do not allow for a simple parametrization for which an explicit formula for $d_{ij}^{(L)}(\mathcal{R})$ and $d_{ij}^{(U)}(\mathcal{R})$ is available. Therefore, some assumptions on the admissible shapes of the regions R_i are needed to be able to optimize the Stress function and obtain a configuration that is more easily interpretable. To this extent, two main models were proposed by Denœux and Masson (2000): (a) the circle (hypersphere) model that represents the regions by spheres with different radii and (b) the rectangle (hyperbox) model representing each object by a hyperbox of varying size.

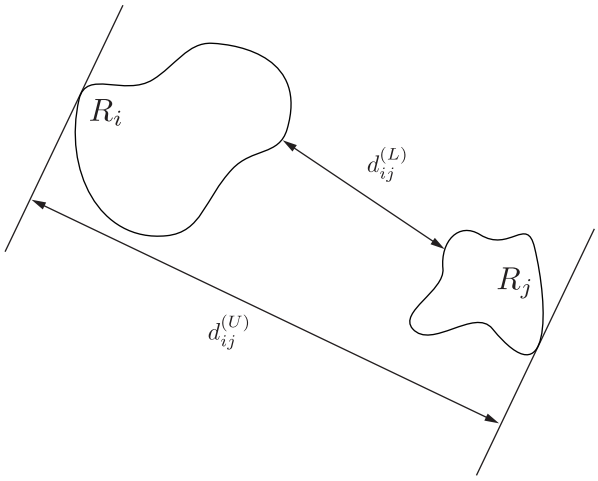


Figure 1: An illustration of the minimal and maximal distances in dimensionality $q = 2$ between the regions R_i and R_j as used in he general framework of interval MDS.

2.1 Circle Model

Perhaps the most simple region is the circle (for $q = 2$) and its generalization of a hypersphere in higher dimensionality. Indeed, a hypersphere can be simply parametrized by its center coordinates $\mathbf{x} \in \mathbb{R}^q$ and its radius $r > 0$. For an example of minimal and maximal distances between circles, see Figure 2. The circle model has a total of $n(q+1)$ parameters to be estimated, nq for center coordinates in the $n \times q$ matrix $\mathbf{X}$ and n for the n vector $\mathbf{r}$ of radii. The lower and upper distances between hypersphere i and j are given respectively by

$$d_{ij}^{(L)}(\mathbf{X}, \mathbf{r}) := \max\left(0, d_{ij}(\mathbf{X}) - r_i - r_j\right) \text{ and} \tag{3}$$

$$d_{ij}^{(U)}(\mathbf{X}, \mathbf{r}) := d_{ij}(\mathbf{X}) + r_i + r_j, \tag{4}$$

where $d_{ij}(\mathbf{X}) := \|\mathbf{x}_i - \mathbf{x}_j\|$.

There are two approaches, called the least-squares approach and the possibility approach, to obtain a hypersphere representation of objects. In the least-squares approach, for given interval-valued dissimilarity data Δ, the Interval Stress function of the circle model is simply obtained by substituting (3) and (4) for $d_{ij}^{(L)}(\mathcal{R})$ and $d_{ij}^{(U)}(\mathcal{R})$ in the interval MDS Stress of (2). To find a hypersphere representation, the function $\sigma_{\text{Int}}(\mathbf{X}, \mathbf{r})$ is minimized over $\mathbf{X}$ and $\mathbf{r}$ with the constraints $r_i \geq 0$ $(i = 1, \ldots, n)$.

The possibility model provides an exact representation in some sense whereas the least-squares approach provides an approximate representation

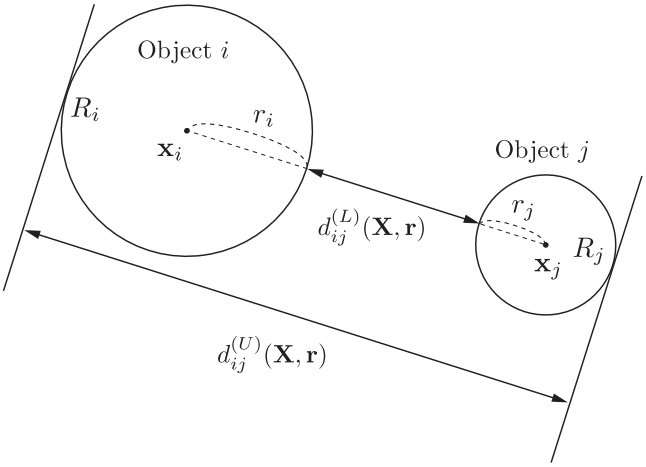


Figure 2: Example of minimal and maximal distances as defined by the circle (hypersphere) model in dimensionality $q = 2$.

in which lower and upper distances approximate as well as possible the given lower and upper dissimilarities, respectively. By analogy with the possibilistic fuzzy regression model Tanaka, Uejima, and Asai (1982), the goal of the possibility approach is to find the smallest hypersphere representation satisfying the following constraints:

$$[\delta_{ij}^{(L)}, \delta_{ij}^{(U)}] \subseteq [d_{ij}^{(L)}(\mathbf{X},\mathbf{r}), d_{ij}^{(U)}(\mathbf{X},\mathbf{r})] \quad (i,j = 1,\ldots,n).$$

To do so, we assume that the center coordinate matrix has already been obtained, say $\mathbf{X}^*$. For example, in Masson and Denœux (2002), $\mathbf{X}^*$ is determined by the standard least-squares MDS on the centers of the interval-valued dissimilarities, that is, by minimizing the following objective function over $\mathbf{X}$:

$$\sigma(\mathbf{X}) := \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[\delta_{ij} - d_{ij}(\mathbf{X})\right]^2,$$

with $\delta_{ij} := (\delta_{ij}^{(L)} + \delta_{ij}^{(U)})/2$. Then, the possibility approach of the circle model is defined by the linear program

$$\begin{aligned} &\min_{\mathbf{r}\in\mathbb{R}^n} \sum_{i=1}^{n} r_i \\ &\text{subject to } \delta_{ij}^{(L)} \geq d_{ij}^{(L)}(\mathbf{X}^*,\mathbf{r}), \\ &\qquad\qquad \delta_{ij}^{(U)} \leq d_{ij}^{(U)}(\mathbf{X}^*,\mathbf{r}), \text{and} \\ &\qquad\qquad r_i \geq 0 \ (i,j = 1,\ldots,n). \end{aligned}$$

Table 1: Eight musical instruments and their abbreviation from the timbre data.

Instrument	Abbrev.	Instrument	Abbrev.
1. Guitar	Gu	5. Oboe	Ob
2. Harp	Hr	6. Clarinet	Cl
3. Violin pizzicato	Vp	7. Horn	Ho
4. Bowed Violin	Vl	8. Trumpet	Tr

Although the two-step approach of the possibility approach is attractive in that the linear programming part that fits the radii has a global solution, it has the disadvantage that the $\mathbf{X}$ and $\mathbf{r}$ are not simultaneously optimal. Therefore, in the remainder of this report we shall focus on the interval MDS approach.

As an example of the interval MDS method using circle model, the interval-valued dissimilarity data about timbre (Marozeau, de Cheveigné, McAdams, & Winsberg, 2003) that we shall use through out this report (see `http://recherche.ircam.fr/pcm/archive/timbref0/`). First, we briefly describe the details of these data. 24 subjects judged timbre dissimilarity between pairs of stimuli produced by a set of twelve musical instruments with equal fundamental frequency, duration, and loudness. There were three sessions, each at a different fundamental frequency (B3, 247Hz; C#4, 277Hz; Bb4, 466Hz). For more details about these data, see Marozeau et al. (2003). Here, we aggregate this data at fundamental frequency C#4 (277Hz) to an interval-valued dissimilarity data by lower and upper 10%-percentiles of these empirical dissimilarity distributions. Table 1 shows the eight instruments that we consider.

The results of the circle model for the timbre data are given in Figure 3. For most of the instruments, the radii are similar, except for the horn (7. Ho) and clarinet (6. Cl) which are larger, and the trumpet (8. Tr) which is smaller. Therefore, the largest interval is obtained for horn and clarinet and in general these two instruments will have larger intervals with the other instruments, for example with the bowed violin (4. Vl) and oboe (5. Ob). The most similar instruments are the guitar (1. Gu) and (2. Hr) and being perceived only slightly different from the violin pizzicato (3. Vp), all three being string instruments. Another cluster of instruments is formed by the bowed violin (4. Vl) and oboe (5. Ob). Finally, the trumpet (8. Tr) is considered both similar to the group of bowed violin (4. Vl) and oboe (5. Ob) and the group of horn (7. Ho) and clarinet (6. Cl).

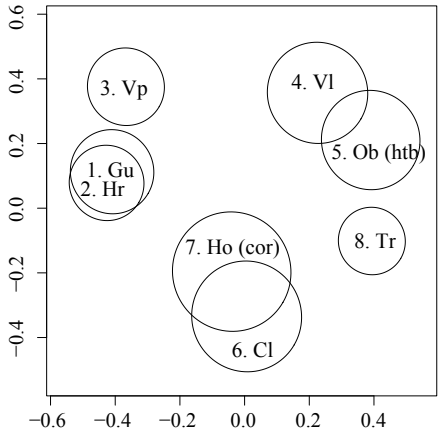


Figure 3: The result of the circle model with $q = 2$ for the timbre dissimilarity data ($\sigma_{\mathrm{C}}^2 = 0.47$).

2.2 Rectangle Model

In the circle model, the radii of each hypersphere represent variabilities of each object. However, from a circle representation, we cannot interpret the variability of the underlying dimensions. In addition, it is difficult interpret the dimensions due to the freedom of rotation of the circle model. As noted by Groenen, Winsberg, Rodríguez, and Diday (2006), one of the important aims of MDS is to discover the relationships among the objects in terms of the underlying dimensions. Thus, it is most useful for interval-valued dissimilarity data to express the location and the variability of each object in terms of its underlying dimensions.

From this this point of view, the rectangle representation seems to be appropriate since the sides of the rectangles correspond directly to the dimensions. The parametrization of a rectangle is slightly more complicated than that of a circle. A rectangle can be parametrized by a pair of center coordinate $\mathbf{x} \in \mathbb{R}^q$ and lengths of each dimension $\mathbf{r} \in \mathbb{R}_+^q$. Denote $\mathbf{X}$ as the matrix with rows $\mathbf{x}_i$ $(i = 1, \ldots, n)$ representing the center coordinates of the hyperbox for each object and $\mathbf{R}$ be a matrix whose rows $\mathbf{r}_i$ $(i = 1, \ldots, n)$ represent the side length vectors of the hyperbox with each object. In the rectangle model, the smallest distance between two rectangles representing objects i and j is defined by

$$d_{ij}^{(L)}(\mathbf{X}, \mathbf{R}) = \left(\sum_{s=1}^{q} \max[0, |x_{is} - x_{js}| - (r_{is} + r_{js})]^2 \right)^{1/2}, \tag{5}$$

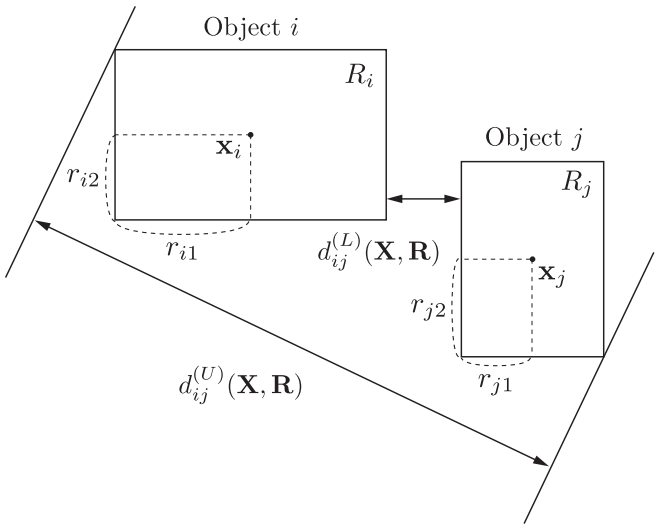


Figure 4: The rectangle model with $q = 2$.

and the maximum distance by

$$d_{ij}^{(U)}(\mathbf{X},\mathbf{R}) = \left(\sum_{s=1}^{q}[|x_{is}-x_{js}| + (r_{is}+r_{js})]^2\right)^{1/2}. \tag{6}$$

Figure 4 shows an example of the minimal and maximal distances the rectangle model with $q = 2$. In the least-squares approach, the Stress function is obtained by replacing $d_{ij}^{(L)}(\mathcal{R})$ and $d_{ij}^{(L)}(\mathcal{R})$ in (2) by $d_{ij}^{(L)}(\mathbf{X},\mathbf{R})$ and $d_{ij}^{(U)}(\mathbf{X},\mathbf{R})$ as defined in (5) and (6).

A solution of the rectangle model applied to the timbre data is presented in Figure 5. There, the variation the range of distances along the first dimension is caused by the guitar (1. Gu), harp (2. Hr) and to a lesser extent by the bowed violin (4. Vl) and horn (7. Ho). Along the second dimension, there are large ranges for the violin pizzicato (3. Vp), the hobo (5. Ob), horn (7. Ho), clarinet (6. Cl), and trumpet (8. Tr). Note that the maximal Euclidean distances for, for example, the violin pizzicato (3. Vp) and the clarinet (6. Cl) are obtained diagonally from the upper end of the line representing the violin pizzicato (3. Vp) and the lower end of the line representing clarinet (6. Cl). It seems that the rectangle solution spreads the ranges more over the dimensions than the circle model does. It is fair to say that most of the interval information of the dissimilarities is mostly shown on the second dimension, perhaps with the exception of objects pairs that involve the guitar (1. Gu), harp (2. Hr).

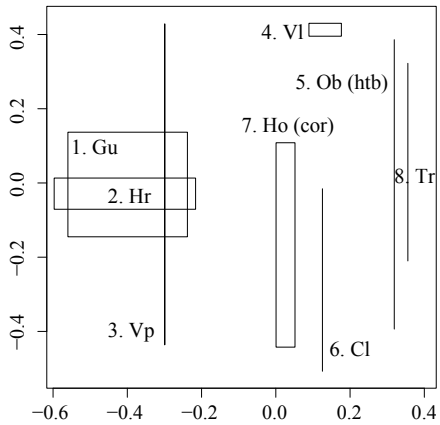


Figure 5: The result of the rectangle model with $q = 2$ for the timbre dissimilarity data ($\sigma^2_{\mathrm{R}} = 0.160$).

3 Three-way Interval MDS

The basic interval MDS models are the circle and rectangle models. Both can be easily extended to two-mode three-way interval-valued dissimilarity data. In this section, we describe the some extensions of the circle model and the rectangle model for the following data:

$$\mathcal{D}_L := (\Delta_1, \ldots, \Delta_L),$$

where each Δ_l ($l = 1, \ldots, L$) is an dissimilarity matrix of intervals as defined in (1). The present dissimilarity matrix can be thought of as a three-way-two-mode data matrix.

In standard three-way MDS, each individual dissimilarity matrix obtains a so called individual space consisting of configuration $\mathbf{X}_l \in \mathbb{R}^{n\times q}$ such that the distances between the points in the individual space match the corresponding individual dissimilarities. However, in three-way MDS, these individual spaces are constrained to be a simple transformation of a common underlying configuration $\mathbf{X} \in \mathbb{R}^{n\times q}$. There are three major MDS models for three-way single-valued dissimilarity data: (a) the dilation model $f_l(\mathbf{X}) := v_l\mathbf{X}$ ($v_l \geq 0$), (b) the weighted Euclidean model $f_l(\mathbf{X}) := \mathbf{X}\mathbf{V}_l$ ($\mathbf{V}_l$: a $q \times q$ positive diagonal matrix), and (c) the generalized Euclidean model $f_l(\mathbf{X}) := \mathbf{X}\mathbf{T}_l$ ($\mathbf{T}_l \in \mathbb{R}^{q\times q}$). The dilation model is hardly used in standard three-way MDS because the dilation only reflects the overall fit of the common space to the individual dissimilarity matrix. The weighted Euclidean model is the most commonly used three-way MDS model: the individual space is obtained by stretching or shrinking the dimensions of the common

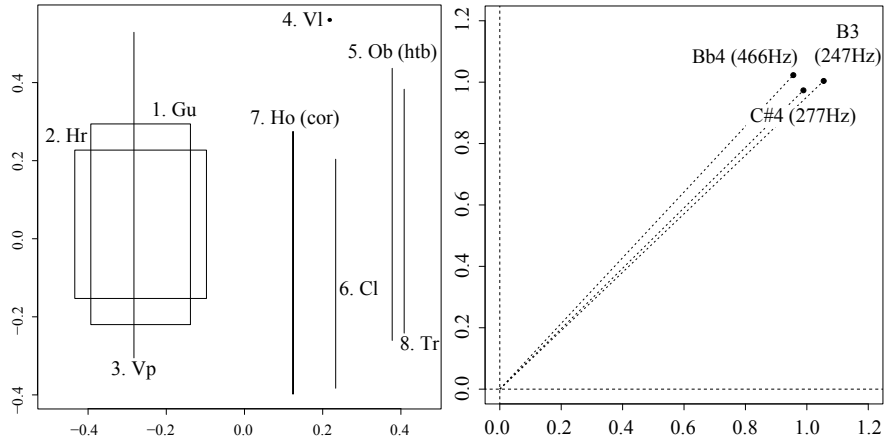
space. The generalized Euclidean model allows for a rotation before doing the stretching and shrinking.

To apply three-way models to interval dissimilarities, we have to consider the type of representations (circle and rectangle) and the possibility of properly parameterizing the individual spaces. Using the weighted Euclidean model with the circle model would lead to representing the circles by ellipses. However, we also need analytic expressions for the upper distances between two ellipses but these are not available. Therefore, it will be hard to optimize the Stress function of the weighted Euclidean model for the circle representation. For generalized Euclidean model, a similar reasoning holds. The dilation model, however, can be considered for the circle model as the circles do not change in shape.

On the other hand, the rectangle model is good to fit the weighted Euclidean model. We assume that the common space has a rectangle representation of objects and the objects in the individual space are also represented by rectangles. More precisely, if $R_i \subset \mathbb{R}^q$ is a hyperbox with the center coordinate $\mathbf{x}_i$ and the side lengths $\mathbf{r}_i$, a region $R_{il} := \{\mathbf{V}_l\mathbf{y} \in \mathbb{R}^q \mid \mathbf{y} \in R_i\}$ in each individual space is also a hyperbox with the center coordinate $\mathbf{x}_{il} := \mathbf{V}_l\mathbf{x}_i$ and the side lengths $\mathbf{r}_{il} := \mathbf{V}_l\mathbf{r}_i$. According to the least-squares approach, the Stress function of the weighted Euclidean model for the rectangle model is defined by

$$\begin{aligned}\sigma^2_{\text{3Way}}(\mathbf{X}, \mathbf{R}, \mathbf{V}_1, \ldots, \mathbf{V}_L) = &\sum_{l=1}^{L}\sum_{i<j}^{n} w_{ij}\left[\delta^{(U)}_{ijl} - d^{(U)}_{ij}(\mathbf{X}\mathbf{V}_l, \mathbf{R}\mathbf{V}_l)\right]^2 \\ &+ \sum_{l=1}^{L}\sum_{i<j}^{n} w_{ij}\left[\delta^{(L)}_{ijl} - d^{(L)}_{ij}(\mathbf{X}\mathbf{V}_l, \mathbf{R}\mathbf{V}_l)\right]^2. \quad (7)\end{aligned}$$

As described above, there are the three different frequency conditions on the timbre data and thus three interval-valued dissimilarity matrices were constructed as in Section 2.1. A solution of the weighted Euclidean model based on the hyperbox model applied to this three-way timbre data is presented in Figure 6. In the first dimension, we see a salient contrast between impulsive and sustained instruments. Therefore, we interpret the first dimension to represent the sustainability of the sounds. The sounds of bowed violin (Vl) and oboe (Ob) have several overtones which are perceived as bright sounds. On the other hand, the sounds of clarinet (Cl) and horn (Ho) have only a few overtones and are perceived as dark sounds. Consequently, the second dimension represents the brightness of sounds and those with a large absolute value in the second dimension are dark sounds (see `http://recherche.ircam.fr/pcm/archive/timbref0/`). Focusing on the



(a) (b)

Figure 6: The result of the weighted Euclidean model for the hyperbox model. Panel (a) shows the common configuration and Panel (b) the weights for each dimension.

weights of the first dimension in Figure 6b, we can see that the differences in the weights are small. There is some evidence that the weights increase with smaller of frequency. This would mean that the difference between impulsive and sustained sounds become more clear with decreasing frequency. Under the low frequency condition (B3, 247Hz), the impulsive sounds disappear faster than under the higher frequency conditions. For the second dimension, the weight of the middle frequency (C#4, 277Hz) is the smallest and in that condition the variability of the perception of the brightness of sounds is smaller than under extreme frequency conditions.

4 Histogram MDS

In most cases, interval-valued dissimilarity data consists of maximum and minimum values. However, these values are susceptible to the effect of outliers. The lower and upper α-percentiles $\alpha^{(L)}$ and $\alpha^{(U)}$ of a dissimilarity distribution are more appropriate to construct interval-valued dissimilarity data. Consider the set A with elements A_k containing the ranges of the percentile values $[\alpha_k^{(L)}, \alpha_k^{(U)}]$ such that the ranges become increasingly larger and a smaller one contains the larger one, that is, $A_k \subset A_l$ if $k < l$. A typical example is $A_1 = [.40, 60]$, $A_2 = [.30, .70]$, and $A_3 = [.20, .80]$. We have the following type dissimilarity data, called histogram-valued or percentile-valued dissimilarity data:

$$\Delta_{\text{Hist}} := (\Delta_{A_1}, \ldots, \Delta_{A_K}),$$

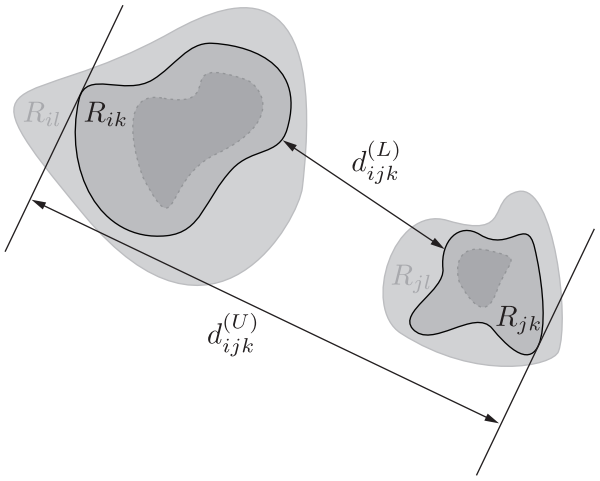


Figure 7: Histogram representation for general regions where regions R_{ik} is contained in R_{il} if percentile range $A_k \subset A_l$.

where

$$\Delta_k := \begin{bmatrix} - & [\delta_{12k}^{(L)}, \delta_{12k}^{(U)}] & \cdots & [\delta_{1nk}^{(L)}, \delta_{1nk}^{(U)}] \\ [\delta_{21k}^{(L)}, \delta_{21k}^{(U)}] & - & \cdots & [\delta_{2nk}^{(L)}, \delta_{2nk}^{(U)}] \\ \vdots & \vdots & \ddots & \vdots \\ [\delta_{n1k}^{(L)}, \delta_{n1k}^{(U)}] & [\delta_{n2k}^{(L)}, \delta_{n2k}^{(U)}] & \cdots & - \end{bmatrix},$$

with $[\delta_{ijk}^{(L)}, \delta_{ijk}^{(U)}] \subseteq [\delta_{ijl}^{(L)}, \delta_{ijl}^{(U)}]$ for $k \leq l$.

For the histogram-valued dissimilarity data, we will use a constrained version of the standard interval MDS applied to each Δ_{A_k}. Let R_{ki} be a region of object i corresponding to percentile range A_k. It seems natural to expect that regions R_{ki} $(k = 1, \ldots, K)$ are satisfying $R_{ki} \subset R_{li}$ for $A_k \subset A_l$. Figure 7 shows this idea for general regions R_{ki}.

Thus, we consider the following MDS stress function for the histogram-valued dissimilarity data:

$$\begin{aligned} \min \sigma_{\text{Hist}}^2(\mathcal{R}_1, \ldots, \mathcal{R}_K) := & \sum_{k=1}^{K} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ijk} \left(\delta_{ijk}^{(L)} - d_{ij}^{(L)}(\mathcal{R}_k) \right)^2 \\ & + \sum_{k=1}^{K} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ijk} \left(\delta_{ijk}^{(U)} - d_{ij}^{(U)}(\mathcal{R}_k) \right)^2 \qquad (8) \\ & \text{subject to } R_{ik} \subset R_{il} \text{ for } A_k \subset A_l. \end{aligned}$$

Next, four models for the histogram MDS can be thought of as shown Figure 8. They vary by the shape (circle model versus rectangle model)

and by sharing a common center (concentric) or not (non-concentric). In all cases, the restriction that a region representing a smaller range of α values is contained in the region of a larger α range, thus $R_{ik} \subset R_{il}$ for $A_k \subset A_l$.

We first provide explicit descriptions for the concentric circle and rectangle models. As a simple extension of the circle model for interval-valued dissimilarity data, Masson and Denœux (2002) proposed the concentric circle model for the histogram-valued dissimilarity data. In the concentric circle model, the regions R_{ik} that represents object i form nested circles that have the same center point. Therefore, the regions R_{ik} can be described by a common center $\mathbf{x}_i$ (with $\mathbf{x}_i$ being row i of $\mathbf{X}$) and the radii r_{ki} being ordered, that is,

$$0 \leq r_{i1} \leq r_{i2} \leq \cdots \leq r_{iK}.$$

The upper left panel of Figure 8 shows the regions of this model. Replacing $d_{ij}^{(L)}(\mathcal{R}_k)$ and $d_{ij}^{(U)}(\mathcal{R}_k)$ in the histogram MDS stress of (8) by $d_{ij}^{(L)}(\mathbf{X}, \mathbf{r}_k)$ and $d_{ij}^{(U)}(\mathbf{X}, \mathbf{r}_k)$ gives the stress function of the concentric circle model. Similarly to Groenen and Winsberg (2006), an algorithm can be constructed that combines majorization minimization with monotone regression for imposing the order constraints on the r_{ik}s. Terada and Yadohisa (2011a) used a different approach for imposing the order constraints by defining $r_{ik} = \sum_{l=1}^{k} a_{il}$ with $a_{ik} \geq 0$. In their algorithm, the positivity restrictions on a_{il} are automatically satisfied.

Groenen and Winsberg (2006) proposed the concentric rectangle model albeit not under that name. It has a common center $\mathbf{x}_i$ for the rectangles representing object i and a nested series of rectangles. The lower left panel of Figure 8 shows the concentric rectangle model. The stress function of the concentric rectangle model is obtained by replacing $d_{ij}^{(L)}(\mathcal{R}_k)$ and $d_{ij}^{(U)}(\mathcal{R}_k)$ in (8) by the lower and upper distances $d_{ij}^{(L)}(\mathbf{X}, \mathbf{R}_k)$ and $d_{ij}^{(U)}(\mathbf{X}, \mathbf{R}_k)$ defined in (5) and (6). To ensure nestedness, the constraints

$$0 \leq r_{is1} \leq r_{is2} \leq \cdots \leq r_{isK},$$

need to be imposed. Groenen and Winsberg (2006) proposed to use a of majorization and monotone regression for imposing the order constraints and Terada and Yadohisa (2011a) developed an algorithm that automatically satisfies these constraints by a similar alternative parametrization as they did for the concentric circle model.

The common center constraint of the concentric models is quite strict condition and may not natural in some cases. The nonconcentric models relieve this constraint at the cost of being somewhat more complicated. We first consider the non-concentric circle model in which each object is represented by nested circles that do not necessarily have the same center point.

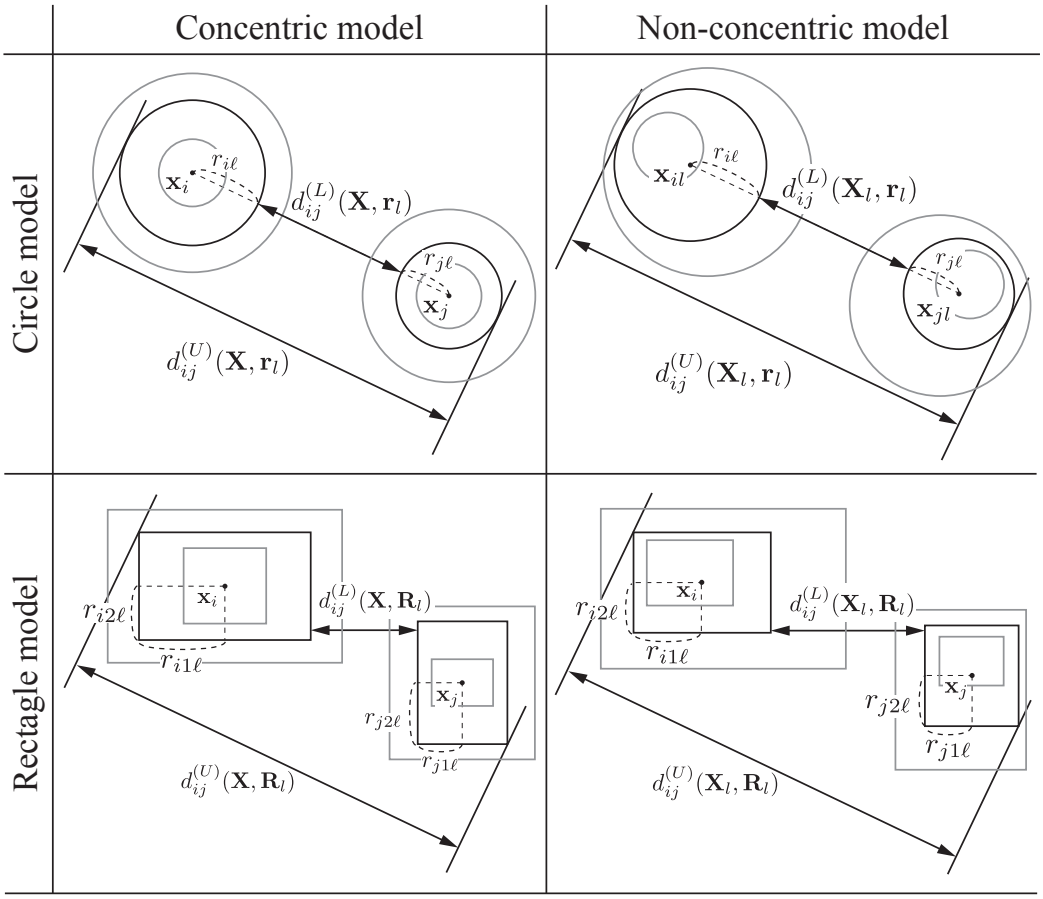


Figure 8: Four models for representing the histogram dissimilarities. The upper panels represent circle models and the lower panels the rectangle models. In the left two panels, the regions representing an object have the same center and the right two panels they have not.

The top right panel of Figure 8 shows this mode. The nested constraint $R_{i1} \subset \cdots \subset R_{iK}$ can be rewritten by

$$\|\mathbf{x}_{ik} - \mathbf{x}_{i(k+1)}\| + r_{ik} \leq r_{i(k+1)} \quad (k = 1, \ldots, K)$$

where $\mathbf{x}_{ik} := (x_{i1k}, \ldots, x_{iqk})^T$ and r_{ik} are the center point and the radius of the circle corresponding to α range k of object i, respectively. Similar to the concentric model, the stress function σ^2_{NCC} for the nonconcentric circle model is obtained by substituting $d_{ij}^{(L)}(\mathbf{X}_k, \mathbf{r}_k)$ and $d_{ij}^{(U)}(\mathbf{X}_k, \mathbf{r}_k)$ for $d_{ij}^{(L)}(\mathcal{R}_k)$ and $d_{ij}^{(U)}(\mathcal{R}_k)$ in the histogram MDS stress of (8). Unfortunately, it is difficult to directly optimize this complex constrained optimization problem. Terada and Yadohisa (2011b) used unconstrained optimization with a penalty that penalizes the deviations from the constraint. The penalty term is defined by

$$\begin{aligned} &g^2_{\mathrm{NCC}}(\mathbf{X}_1, \ldots, \mathbf{X}_K, \mathbf{r}_1, \ldots, \mathbf{r}_K) \\ &:= \sum_{i=1}^{n} \sum_{k=1}^{K-1} \max\left[0, \|\mathbf{x}_{ik} - \mathbf{x}_{i(k+1)}\| + r_{ki} - r_{i(k+1)}\right]^2 \end{aligned}$$

By minimizing the extended stress function $\tilde{\sigma}^2_{\mathrm{NCC}} := \sigma^2_{\mathrm{NCC}} + \lambda g^2_{\mathrm{NCC}}$ for a sufficiently large $\lambda > 0$ a solution is obtained that satisfies the nestedness constraints.

In the nonconcentric rectangle model, each object is represented by nested rectangles that do not necessarily have the same center point. The lower right panel of Figure 8 presents the nonconcentric rectangle model. The stress function can be obtained by substituting $d_{ij}^{(L)}(\mathcal{R}_k)$ and $d_{ij}^{(U)}(\mathcal{R}_k)$ with $d_{ij}^{(L)}(\mathbf{X}_k, \mathbf{R}_k)$ and $d_{ij}^{(U)}(\mathbf{X}_k, \mathbf{R}_k)$ in the general histogram MDS stress (8). Using center coordinates and side length values, we rewrite the nestedness constraints of the stress function by the inequalities

$$r_{isk} + |x_{isk} - x_{is(k+1)}| \leq r_{is(k+1)} \tag{9}$$

for $(i = 1, \ldots, n, k = 1, \ldots, K-1,$ and $s = 1, \ldots, p)$.

Surprisingly, there exists an efficient parametrization of the nonconcentric hyperbox model in which the nested constraints are eliminated. Following Terada and Yadohisa (2011a), new parameters $a_{isk}^{(L)}$ and $a_{isk}^{(U)}$ are introduced so that

$$x_{isk} = x_{is1} + \frac{1}{2} \sum_{l=2}^{k} \left(a_{isl}^{(U)} - a_{isl}^{(L)} \right)$$

and

$$r_{isk} = r_{is1} + \frac{1}{2} \sum_{l=2}^{k} \left(a_{isl}^{(U)} + a_{isl}^{(L)} \right),$$

instead of using x_{isk} and r_{isk} $(k = 2, \ldots, K)$. Using this parametrization, the constraints in (9) can be replacing by the following nonnegativity constraints

$$r_{is1} \geq 0,\ a_{isk}^{(L)} \geq 0, \text{ and } a_{isk}^{(U)} \geq 0.$$

Introducing new variables $\rho_{0is},\ \alpha_{isk}^{(L)}$ and $\alpha_{isk}^{(U)}$ $(i = 1, \ldots, n;\ k = 2, \ldots, K,\ s = 1, \ldots, p)$ so that $r_{is1} = \rho_{is1}^2,\ a_{isk}^{(L)} = {\alpha_{isk}^{(L)}}^2$, and $a_{isk}^{(U)} = {\alpha_{isk}^{(U)}}^2$, the nonnegativity condition can be ensured. Therefore, the nonconcentric hyperbox model MDS can be directly solved by unconstrained optimization.

4.1 Example: Timbre Data

A solution of the concentric circle model applied to the timbre data with $A_1 = (10\%, 90\%), A_2 = (5\%, 95\%)$, and $A_3 = (0\%, 100\%)$ is presented in the top left panel of Figure 9. The radii of the nested circles of bowed violin (Vl) evenly reduce with increasing the α. This means that the variability of feeling for bowed violin (Vl) is reduced with increasing α. On the other hand, there is only small difference between the radii of clarinet (Cl) corresponding with A_1 and A_2. Thus, we can interpret that the variability of feeling for clarinet (Cl) is less sensitive to the increase of α.

The bottom left panel of Figure 9 is the resulting configuration of the concentric rectangle model for the timbre data. Compared to the concentric circle model the variability of the different α ranges is mostly related to the first dimension for all objects. Since the brightness of sounds highly depends on the associated perceived feelings, people can more easily distinguish impulsive sounds from sustained sounds. The variability shown on the first dimension reflects this fact.

The top right panel of Figure 9 shows the result of the nonconcentric circle model applied to the timbre data. With decreasing α ranges, the centers of violin pizzicato (Vp) and oboe (Ob) shift to the right and the left, respectively. This means that for a decreasing α range the violin pizzicato (Vp) is perceived more as an impulsive sound whereas the oboe (Ob) more as a sustained sound. The concentric circle model cannot reflect this kind of information.

A similar effect is found for the nonconcentric rectangle model shown in the bottom right panel of Figure 9, that is, the solution is similar to the concentric rectangle model except that the rectangles are pushed to the outside from the origin.

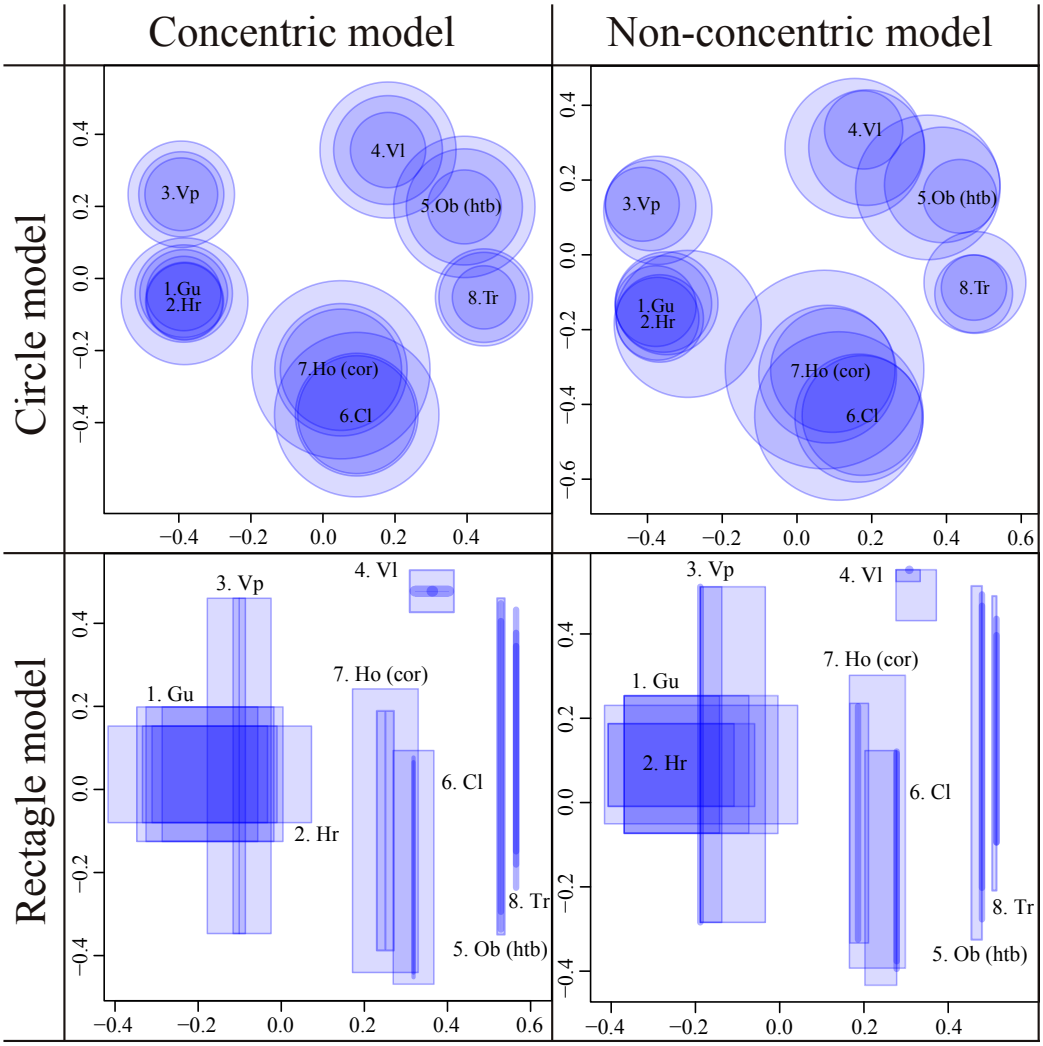


Figure 9: The results of the four models for representing the histogram timbre dissimilarities.

4.2 Histogram MDS and Fuzzy Theory

The present formulation of histogram MDS can be considered as a similar problem for general fuzzy dissimilarities in the context of fuzzy theory as done by Masson and Denœux (2002). They assume that each object is represented by a p-dimensional fuzzy number f_i in $\mathbb{R}^p$. Let f_i be the membership function of the p-dimensional fuzzy number of object i and $R_i(\alpha)$ be the α-cut of f_i for $\alpha \in [0, 1]$. In this setting, Masson and Denœux (2002) mentioned that the fuzzy distance between two fuzzy numbers f_i and f_j can be defined by the following membership function:

$$d_{ij}(c) := \sup_{\mathbf{x},\mathbf{y}\in\mathbb{R}^p : c=\|\mathbf{x}-\mathbf{y}\|} \min[f_i(\mathbf{x}), f_j(\mathbf{y})].$$

Moreover, Masson and Denœux (2002) show that each α-cut of d_{ij} is a closed interval $\xi_{ij}(\alpha) := [d_{ij}^{(L)}(\alpha), d_{ij}^{(U)}(\alpha)]$, where $d_{ij}^{(L)}(\alpha)$ and $d_{ij}^{(U)}(\alpha)$ are respectively the lower and upper distances between the α-cuts $R_i(\alpha)$ and $R_j(\alpha)$. For $\alpha_1 > \alpha_2$, we have $R_i(\alpha_1) \subset R_i(\alpha_2)$ $(i = 1, \ldots, n)$. Histogram MDS can then be considered as the problem to estimate the hidden α-cuts $R_i(\alpha_1), \ldots, R_i(\alpha_K)$ of hidden fuzzy number f_i $(i = 1, \ldots, n)$ for given α-cuts $\xi_{ij}(\alpha_k) := \left[\xi_{kij}^{(L)}, \xi_{kij}^{(U)}\right]$ $(i, j = 1, \ldots, n;\ k = 1, \ldots, K)$ of fuzzy dissimilarities in least-squares sense. In particular, percentile intervals of a dissimilarity distribution can be considered as the α-cuts of fuzzy dissimilarities.

5 Local Minima

The stress function discussed in this report need to be minimized by iterative algorithms as no analytic solution is available. For MDS and symbolic MDS two types of algorithms are available. First, general purpose minimization algorithms such as BFGS can be applied. The latter is a steepest descent approach that needs the gradient of the stress function. If the gradient is not available, then it can estimate the gradient by multiple function evaluations near the present estimates of the parameters. For interval MDS problems with large n (say, $n > 30$), numerical gradients most likely could lead to a serious slowdown of the algorithm. A second type of algorithm is the minimization by majorization (MM) algorithm that is guaranteed to reduce the Stress function in each iteration (Groenen et al., 2006; Terada & Yadohisa, 2011a, 2011b). MM algorithms can be seen as steepest descent algorithms with a constant stepsize. Here, we use three versions. The first one is the plain majorization algorithm. The second one allows in each iteration to dilate (by a single scalar) the estimates obtained from a majorization update. This scales all parameters to their size such that it minimizes stress.

The third one uses a trick that doubles the step length in each iteration, the so-called relaxed update. The idea is that an ordinary MM step reduces the stress and that a step twice as long cannot increase stress. Often, the MM with step doubling decreases the number of iterations by a half. Here, we use the step doubling in combination with the optimal dilation.

The Stress function for MDS of interval dissimilarities (2) does not have nice properties such as convexity. Consequently, it cannot be guaranteed *a priori* that an algorithm minimizing (2) yields a global minimum. In fact, many local minima can occur. Here, we investigate the attraction to a (candidate) global minimum of five algorithms: (a) the majorization algorithm (MM), (b) MM with optimal dilation, (c) MM with acceleration and dilation, (d) the standard BFGS algorithm for minimizing functions in R using explicit gradients, and (e) the BFGS algorithm that uses numerically estimated gradients. We compare their performance (final Stress divided by the best Stress found over all runs by any of the algorithms) over 1000 random starts for the circle and rectangle models of interval MDS.

Figure 10a shows the results for the circle model. It presents the cumulative distribution of the performance (stress divided by overall best stress). We find that for the circle mode BFGS with numeric gradients is approximately in 30% of the runs attracted to the (candidate) global minimum whereas the three MM varieties about 5%. However, the MM algorithms and particularly MM with dilation are best able to obtain an almost optimal performance in 50 to 60% of the runs. The hyperbox models seems to be much more prone to local minima because Figure 10b shows only a very small vertical line in the lower left corner of about a 40% attraction to the (candidate) global minimum by the BFGS algorithm with numerical gradients. There is a small region of attraction (at most 10 %) of the MM methods to solutions close to the candidate global minimum. We conclude from Figure 10 that at least for the timbre data, the local minimum problem for σ^2_{Int} seems less severe for the circle model than for the hyperbox model. For the circle model, the majorization approaches seem to be beneficial to increase the probability of finding the best local minimum and thus the candidate global minimum. For the hyperbox model, one tend to find many local minima and the BFGS method with numerical gradients seems to be better able to find good local minima.

Table 2 shows the computational time of these experiments for each method. For the circle model, three MM varieties are much faster than the BFGS methods where the MM algorithm with dilation seems to be best choice from both the computational cost and the local minimum problem perspectives. For the rectangle model we found that the BFGS method with numerical gradients is best able to find the best local minimum. However,

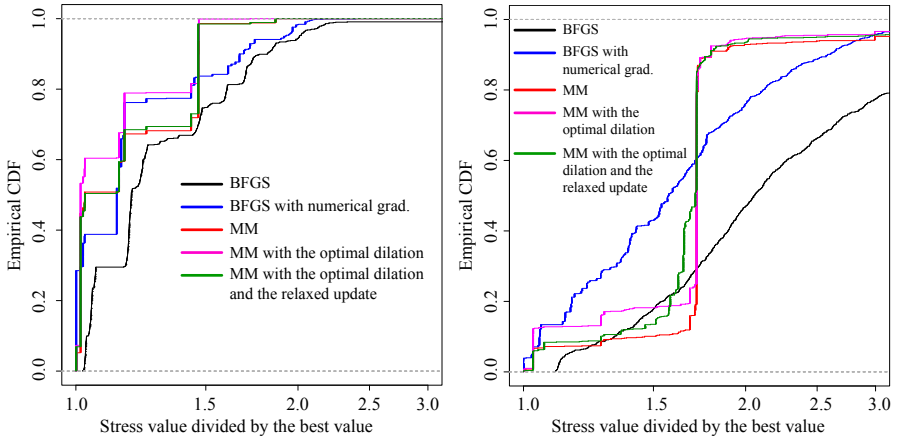


Figure 10: Empirical cumulative distributions of Stress values on the timbre data for 1000 random starts using four different minimization methods. Panel (a) contains the distributions for the circle model and Panel (b) for the rectangle model.

the three MM methods are approximately three times faster than the BFGS method with numerical gradients. Therefore, we can try 3000 random starts by using the MM methods at the same computational costs of 1000 random starts of the BFGS method with numerical gradients. Thus, considering both the computational costs and the quality of the local minimum, the MM methods (especially MM with dilation) seems to be a good choice even for the rectangle model.

6 Conclusions and Discussion

In this report, we have provided an overview of the present state of the art of symbolic MDS. We discussed interval MDS where the dissimilarities are intervals and the objects can be represented by circles or rectangles. The dissimilarity range between any pair of objects in then modeled by the minimum and maximum Euclidean distances between pairs of circles or rectangles. We discussed an extension for three-way interval MDS when replications of interval MDS dissimilarities are available. The third type of symbolic MDS data discussed here are histogram dissimilarities. The histogram MDS model is based on a nested sets of ranges representing ranges of percentiles of the distribution of the dissimilarity. These ranges are represented by nested circles or rectangles that can be either concentric or not concentric. The minimum and maximum Euclidean distances of a certain percentile range between the

Table 2: Computational time of this experiments for each method.

Circle model	time (s)
1. BFGS	35.79
2. BFGS with numerical grad.	364.37
3. MM	6.76
4. MM with the optimal dilation	6.35
5. MM with the optimal dilation and the relaxed update	5.27
Rectangle model	time (s)
1. BFGS	20.90
2. BFGS with numerical grad.	297.05
3. MM	103.57
4. MM with the optimal dilation	111.10
5. MM with the optimal dilation and the relaxed update	70.88

circles (or rectangles) of objects i and j models the same range of percentiles of the distribution of the dissimilarity for i and j.

It is common practice in MDS to allow the dissimilarities to be optimally transformed, for example, by an ordinal transformation. We believe that not good to do so in the context of symbolic MDS. The reason is that the very fact that interval MDS is based on a lower and upper bound of the dissimilarities implies that these values are measured on at least an interval scale. Therefore, any model that represents such intervals should be on exactly the same scale and, thus, nonlinear transformations do not make sense.

For interval MDS, the R package **smds** (Terada & Groenen, 2015) is available. With this package, the circle and rectangle models for interval-valued dissimilarity data can be applied. In addition, it allows for choosing the type of optimization algorithm used (e.g., the MM algorithm). The core of this package is written in C++ which ensures fast computations of the solutions.

References

Denœux, T., & Masson, M. (2000). Multidimensional scaling of interval-valued dissimilarity data. *Pattern Recognition Letters*, *21*, 83–92.

Groenen, P. J. F., & Winsberg, S. (2006). Multidimensional scaling of histogram dissimilarities. In V. Bategelj, H. H. Bock, A. Ferligoj, & A. Žiberna (Eds.), *Data science and classification* (p. 581-588). Springer-Verlag.

Groenen, P. J. F., Winsberg, S., Rodríguez, O., & Diday, E. (2006). I-scal: Multidimensional scaling of interval dissimilarities. *Computational Statistics & Data Analysis*, *51*, 360–378.

Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *The Journal of the Acoustical Society of America*, *114*, 2946–2957.

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

Tanaka, H., Uejima, S., & Asai, K. (1982). Linear regression analysis with fuzzy model. *IEEE Transactions on Systems, Man and Cybernetics*, *12*, 903–907.

Terada, Y., & Groenen, P. J. F. (2015). **smds***: Symbolic multidimensional scaling.* http://cran.r-project.org/web/packages/smds/index.html.

Terada, Y., & Yadohisa, H. (2011a). Multidimensional scaling with hyperbox model for percentile dissimilarities. In J. Watada, G. Phillips-Wren, L. C. Jain, & R. J. Howlett (Eds.), *Intelligent decision technologies* (pp. 779–788). Springer-Verlag.

Terada, Y., & Yadohisa, H. (2011b). Multidimensional scaling with the nested hypersphere model for percentile dissimilarities. *Procedia Computer Science*, *6*, 364–369.