Multidimensional Scaling Using Majorization: SMACOF in R

Jan de Leeuw
University of California, Los Angeles

Patrick Mair
Wirtschaftsuniversität Wien

Abstract

In this paper we present the methodology of multidimensional scaling problems (MDS) solved by means of the majorization algorithm. The objective function to be minimized is known as stress and functions which majorize stress are elaborated. This strategy to solve MDS problems is called SMACOF and it is implemented in an **R** package of the same name which is presented in this article. We extend the basic SMACOF theory in terms of configuration constraints, three-way data, unfolding models, and projection of the resulting configurations onto spheres and other quadratic surfaces. Various examples are presented to show the possibilities of the SMACOF approach offered by the corresponding package.

Keywords: smacof, multidimensional scaling, majorization, R.

1. Introduction

From a general point of view, *multidimensional scaling* (MDS) is a set of methods for discovering "hidden" structures in multidimensional data. Based on a proximity matrix derived from variables measured on objects as input entity, these distances are mapped on a lower dimensional (typically two or three dimensions) spatial representation. A classical example concerns airline distances between US cities in miles as symmetric input matrix. Applying MDS, it results in a two-dimensional graphical representation reflecting the US map (see Kruskal and Wish 1978). Depending on the nature of the original data various proximity/dissimilarity measures can be taken into account. For an overview see Cox and Cox (2001, Chapter 1) and an implementation of numerous proximity measures in R is given by Meyer and Buchta (2007). Typical application areas for MDS are, among others, social and behavioral sciences, marketing, biometrics, and ecology.

Mardia, Kent, and Bibby (1979, Chapter 14) provide a MDS chapter within the general framework of multivariate analysis. An early historical outline of MDS is given in Shepard (1980). For introductory MDS reading we refer to Kruskal and Wish (1978) and more advanced topics can be found in Borg and Groenen (2005) and Cox and Cox (2001).

The traditional way of performing MDS is referred to as *classical scaling* (Torgerson 1958)

which is based on the assumption that the dissimilarities are precisely Euclidean distances without any additional transformation. With further developments over the years, MDS techniques are commonly embedded into the following taxonomies (see e.g. Cox and Cox 2001):

- 1-way vs. multi-way MDS: In K-way MDS each pair of objects has K dissimilarity measures from different "replications" (e.g., repeated measures, multiple raters etc.).
- 1-mode vs. multi-mode MDS: Similar to the former distinction but the K dissimilarities are qualitatively different (e.g., experimental conditions, subjects, stimuli etc.)

For each MDS version we provide metric and non-metric variants. Non-metric MDS will be described in a separate section since, within each majorization iteration, it includes an additional optimization step (see Section 4). However, for both approaches, the particular objective function (or loss function) we use in this paper is a sum of squares, commonly called *stress*. We use majorization to minimize stress and this MDS solving strategy is known as SMACOF (Scaling by MAjorizing a COmplicated Function).

Furthermore we will provide several extensions of the basic SMACOF approach in terms of constraints on the configuration, individual differences (i.e., three-way data structures), rectangular matrices, and quadratic surfaces. From early on in the history of MDS its inventors, notably Torgerson (1958) and Shepard (1962a,b), discovered that points in MDS solutions often fell on, or close to, quadratic manifolds such as circles, ellipses, or parabolas. Fitting quadratics with SMACOF will be one of the focal points of this paper (see Section 5). Some of the first examples analyzed with the new techniques were the color similarity data of Ekman (1954) and the color naming data of Fillenbaum and Rapaport (1971). Another early application involved triadic comparisons of musical intervals (Levelt, Van De Geer, and Plomp 1966), where the points appeared to fall on a parabola. A critical discussion can be found in Shepard (1974, p. 386-388). Around the same time, the triadic comparisons of Dutch political parties (De Gruijter 1967) were carried out, which showed a curved left-right dimension, with parties ordered along an ellipse.

The smacof package has many options, and makes it possible to represent a given data matrix in many different ways. In some contexts, such as molecular conformation or geographic map making, we have prior knowledge what the dimensionality is, or what the appropriate constraints are. But the model selection issue is obviously an important problem in "exploratory" MDS. If there is no such prior knowledge, we have to rely on pseudo-inferential techniques such as Ramsay (1982) or De Leeuw and Meulman (1986).

2. Basic majorization theory

Before describing details about MDS and SMACOF we give a brief overview on the general concept of *majorization* which optimizes a particular objective function; in our application referred to as *stress*. More details about the particular stress functions and their surrogates for various SMACOF extensions will be elaborated below.

In a strict sense, majorization is not an algorithm but rather a prescription for constructing optimization algorithms. The principle of majorization is to construct a surrogate function which majorizes a particular objective function. For MDS, majorization was introduced by

De Leeuw (1977a) and further elaborated in De Leeuw and Heiser (1977) and De Leeuw and Heiser (1980). One way to think about this approach for optimizing objective functions is as a generalization of the EM-algorithm (Dempster, Laird, and Rubin 1977). In fact, Lange (2004) uses the term *MM-algorithm* which stands for either *majorize/minimize* or *minorize/maximize*. De Leeuw (1994) puts it (together with EM, ALS and others) into the framework of block-relaxation methods.

From a formal point of view majorization requires the following definitions. Let us assume we have a function $f(x)$ to be minimized. Finding an analytical solution for complicated $f(x)$ can be rather cumbersome. Thus, the majorization principle suggests to find a simpler, more manageable surrogate function $g(x, y)$ which majorizes $f(x)$, i.e. for all x

$$g(x, y) \geq f(x) \tag{1}$$

where y is some fixed value called the *supporting point*. The surrogate function should touch the surface at y, i.e., $f(y) = g(y, y)$, which, at the minimizer $x^\star$ of $g(x, y)$ over x, leads to the inequality chain

$$f(x^\star) \leq g(x^\star, y) \leq g(y, y) = f(y) \tag{2}$$

called the *sandwich inequality*.

Majorization is an iterative procedure which consists of the following steps:

1. Choose initial starting value $y := y_0$.
2. Find update $x^{(t)}$ such that $g(x^{(t)}, y) \leq g(y, y)$.
3. Stop if $f(y) - f(x^{(t)}) < \epsilon$, else $y := x^{(t)}$ and proceed with step 2.

This procedure can be extended to multidimensional spaces and as long as the sandwich inequality in 2 holds, it can be used to minimize the corresponding objective function. In MDS the objective function called *stress* is a multivariate function of the distances between objects. We will use majorization for *stress* minimization for various SMACOF variants as described in the following sections. Detailed elaborations of majorization in MDS can be found in Groenen (1993); Borg and Groenen (2005).

3. Basic SMACOF methodology

3.1. Simple SMACOF for symmetric dissimilarity matrices

MDS input data are typically a $n \times n$ matrix Δ of *dissimilarities* based on observed data. Δ is symmetric, non-negative, and hollow (i.e. has zero diagonal). The problem we solve is to locate $i, j = 1, \ldots, n$ points in low-dimensional Euclidean space in such a way that the distances between the points approximate the given dissimilarities δ_{ij}. Thus we want to find an $n \times p$ matrix X such that $d_{ij}(X) \approx \delta_{ij}$, where

$$d_{ij}(X) = \sqrt{\sum_{s=1}^{p} (x_{is} - x_{js})^2}. \tag{3}$$

The index $s = 1, \ldots, p$ denotes the number of dimensions in the Euclidean space. The elements of X are called *configurations* of the objects. Thus, each object is scaled in a p-dimensional space such that the distances between the points in the space match as well as possible the observed dissimilarities. By representing the results graphically, the configurations represent the coordinates in the configuration plot.

Now we make the optimization problem more precise by defining *stress* $\sigma(X)$ by

$$\sigma(X) = \sum_{i<j} w_{ij}(\delta_{ij} - d_{ij}(X))^2. \tag{4}$$

Here, W is a known $n \times n$ matrix of *weights* w_{ij}, also assumed to be symmetric, non-negative, and hollow. We assume, without loss of generality, that

$$\sum_{i<j} w_{ij}\delta_{ij}^2 = n(n-1)/2 \tag{5}$$

and that W is irreducible (De Leeuw 1977a), so that the minimization problem does not separate into a number of independent smaller problems. W can for instance be used for imposing missing value structures: $w_{ij} = 1$ if δ_{ij} is known and $w_{ij} = 0$ if δ_{ij} is missing. However, other kinds of weighting structures are allowed along with the restriction $w_{ij} \geq 0$.

Following De Leeuw (1977a), stress, as given in (4), can be decomposed as

$$\begin{aligned}\sigma(X) &= \sum_{i<j} w_{ij}\delta_{ij}^2 + \sum_{i<j} w_{ij}d_{ij}^2(X) - 2\sum_{i<j} w_{ij}\delta_{ij}d_{ij}(X) = \\ &= \eta_\delta^2 + \eta^2(X) - 2\rho(X).\end{aligned}$$

From restriction (5) it follows that the first component $\eta_\delta^2 = n(n-1)/2$. The second component $\eta^2(X)$ is a weighted sum of the squared distances $d_{ij}^2(X)$, and thus a convex quadratic. The third one, i.e. $-2\rho(X)$, is the negative of a weighted sum of the $d_{ij}(X)$, and is consequently concave.

The third component is the crucial term for majorization. Let us define the matrix $A_{ij} = (e_i - e_j)(e_i - e_j)'$ whose elements equal 1 at $a_{ii} = a_{jj} = 1$, -1 at $a_{ij} = a_{ji}$, and 0 elsewhere. Furthermore, we define

$$V = \sum_{i<j} w_{ij}A_{ij} \tag{6}$$

as the weighted sum of row and column centered matrices A_{ij}. Hence, we can rewrite

$$\eta^2(X) = \mathbf{tr}\ X'VX. \tag{7}$$

For a similar representation of $\rho(X)$ we define the matrix

$$B(X) = \sum_{i<j} w_{ij}s_{ij}(X)A_{ij} \tag{8}$$

where

$$s_{ij}(X) = \begin{cases}\delta_{ij}/d_{ij}(X) & \text{if } d_{ij}(X) > 0, \\ 0 & \text{if } d_{ij}(X) = 0.\end{cases}$$

Using $B(X)$ we can rewrite $\rho(X)$ as

$$\rho(X) = \mathbf{tr}\ X'B(X)X \tag{9}$$

and, consequently, the stress decomposition becomes

$$\sigma(X) = 1 + \mathbf{tr}\ X'VX - 2\mathbf{tr}\ X'B(X)X. \tag{10}$$

At this point it is straightforward to find the majorizing function of $\sigma(X)$. Let us denote the supporting point by Y which, in the case of MDS, is a $n \times p$ matrix of configurations. Similar to (8) we define

$$B(Y) = \sum_{i<j} w_{ij} s_{ij}(Y) A_{ij} \tag{11}$$

with

$$s_{ij}(Y) = \begin{cases} \delta_{ij}/d_{ij}(Y) & \text{if } d_{ij}(Y) > 0, \\ 0 & \text{if } d_{ij}(Y) = 0. \end{cases}$$

The Cauchy-Schwartz inequality implies that for all pairs of configurations X and Y, we have $\rho(X) \geq \mathbf{tr}\ X'B(Y)Y$. Thus we minorize the convex function $\rho(X)$ with a linear function. This gives us a majorization of *stress*

$$\begin{aligned} \sigma(X) &= 1 + \mathbf{tr}\ X'VX - 2\mathbf{tr}X'B(X)X \\ &\leq 1 + \mathbf{tr}\ X'VX - 2\mathbf{tr}\ X'B(Y)Y = \tau(X, Y). \end{aligned} \tag{12}$$

Obviously, $\tau(X, Y)$ is a (simple) quadratic function in X which majorizes stress. Finding its minimum analytically involves

$$\frac{\partial \tau(X, Y)}{\partial X} = 2VX - 2B(Y)Y = \mathbf{0}. \tag{13}$$

To solve this equation system we use the Moore-Penrose inverse $V^+ = (V + n^{-1}\mathbf{11}')^{-1} - n^{-1}\mathbf{11}'$ which leads to

$$\overline{X} = V^+ B(Y)Y. \tag{14}$$

This is known as the Guttman transform (Guttman 1968) of a configuration. Note that if $w_{ij} = 1$ for all $i \neq j$ we have $V = 2n(I - n^{-1}\mathbf{11}')$ and the Guttman transform simply becomes $\overline{X} = n^{-1}B(Y)Y$.

Since majorization is an iterative procedure, in step $t = 0$ we set $Y := X^{(0)}$ where $X^{(0)}$ is a start configuration. Within each iteration t we compute $\overline{X}^{(t)}$ which, for simple SMACOF, gives us the update $X^{(t)}$. Now the stress $\sigma(X^{(t)})$ can be calculated and we stop iterating if $\sigma(X^{(t)}) - \sigma(X^{(t-1)}) < \epsilon$ or a certain iteration limit is reached. Majorization guarantees a series of non-increasing stress values with a linear convergence rate (De Leeuw 1988). As shown in Groenen and Heiser (1996) local minima are more likely to occur in low-dimensional solutions (especially unidimensional scaling). For high-dimensional solutions local minima are rather unlikely whereas fulldimensional scaling has no local minimum problem.

3.2. SMACOF with restrictions on the configurations

De Leeuw and Heiser (1980) introduced a SMACOF version with restrictions on the configuration matrix X which Borg and Groenen (2005, Chapter 10) call *confirmatory MDS with*

external constraints. The basic idea behind this approach is that the researcher has some substantive underlying theory regarding a decomposition of the dissimilarities. We start with the simplest restriction in terms of a linear combination, show the majorization solution and then present some additional possibilities for constraints. The linear restriction in its basic form is

$$X = ZC \tag{15}$$

where Z is a known predictor matrix of dimension $n \times q$ ($q \geq p$). The predictors can be numeric in terms of external covariates or one can specify an ANOVA-like design matrix. C is a $q \times p$ matrix of regression weights to be estimated.

The optimization problem is basically the same as in the former section: We minimize stress as given in (4) with respect to C. The expressions for V and B as well as $\tau(X, Y)$ can be derived analogous to simple SMACOF and, correspondingly, the Guttman transform is $\overline{X} = V^{+}B(Y)Y$. It follows that Equation 12 can be rewritten as (see De Leeuw and Heiser 1980, Theorem 1)

$$\begin{aligned}\tau(X, Y) &= 1 + \mathbf{tr} X'VX - 2\mathbf{tr} X'V\overline{X} \\ &= 1 + \mathbf{tr}(X - \overline{X})'V(X - \overline{X}) - \mathbf{tr}\overline{X}'V\overline{X}\end{aligned} \tag{16}$$

where the second term denotes the *lack of confirmation fit* and becomes zero if no restrictions are imposed. Thus, in each iteration t of the majorization algorithm we first compute the Guttman transform $\overline{X}^{(t)}$ of our current best configuration, and then solve the *configuration projection problem* of the form

$$\min_{X \in \mathcal{S}} \mathbf{tr} \left(X - \overline{X}^{(t)}\right)' V \left(X - \overline{X}^{(t)}\right). \tag{17}$$

In other words, we project $\overline{X}^{(t)}$ on the manifold of constrained configurations. With linear restrictions this projection gives us the update

$$X^{(t)} = ZC^{(t)} = Z(Z'VZ)^{-1}Z'V\overline{X}^{(t)}. \tag{18}$$

with $\sigma\left(X^{(t+1)}\right) < \sigma\left(X^{(t)}\right)$.

Basically, the **smacof** package allows the user to implement arbitrary configuration restrictions by specifying a corresponding update function for X as given in (18). Nevertheless, we provide additional restriction possibilities which are commonly used. Besides the classical linear restriction described above, for the special case of number of predictors equal number of dimensions, i.e. $q = p$, the square matrix C can be restricted to be diagonal: $C = \mathbf{diag}(c_{11}, c_{22}, \ldots, c_{ss}, \ldots, c_{qq})$. For each column of Z we suppose $\mathbf{z}_s'V\mathbf{z}_s = 1$. For majorization, let $\overline{\mathbf{x}}_s^{(t)}$ be the s-th column of the Guttman transformed matrix $\overline{X}^{(t)}$ in iteration t, the corresponding C diagonal update is given by

$$c_{ss}^{(t)} = \mathbf{z}_s'V\overline{\mathbf{x}}_s^{(t)}. \tag{19}$$

Combining unrestricted, linearly restricted and the diagonally restricted models leads to a framework of a partially restricted X. De Leeuw and Heiser (1980) use the block notation

$$X = \begin{bmatrix} X_1 & ZC_1 & C_2 \end{bmatrix} \tag{20}$$

in which X_1 is the unrestricted part and of dimension $n \times q_1$. ZC_1 is the linearly restricted part of dimension $n \times q_2$ and C_2 is a diagonal matrix of order n which can be either present or absent. The corresponding models are commonly coded as triples (q_1, q_2, q_3) denoting the number of dimensions contributed by each component: q_1 is the number of unrestricted dimensions, q_2 the number of linearly restricted dimensions, and q_3 is either zero or one, depending on presence or absence of the diagonal matrix C_2. An important special case and the one which is implemented also in **smacof** is $(q, 0, 1)$ which is a q-dimensional MDS model with uniquenesses (Bentler and Weeks 1978).

Further specifications of this partially restricted framework can be found in De Leeuw and Heiser (1980) and additional elaborations in Borg and Groenen (2005, Chapter 10).

3.3. SMACOF for individual differences

The last version of our basic SMACOF routines is SMACOF for individual differences (also known as three-way SMACOF). It is a somewhat natural extension of the classical MDS setting from Section 3.1 in terms of $k = 1, \ldots, K$ separate $n \times n$ symmetric dissimilarty matrices Δ_k. A typical situation is, e.g., that we have K judges and each of them produces a dissimilarity matrix or that we have K replications on some MDS data. The very classical approach for MDS computation on such structures is INDSCAL (INdividual Differences SCaling; Carroll and Chang 1970). An elaborated overview of additional algorithms is given in Borg and Groenen (2005, Chapter 22).

We will focus on the majorization solution and collect the Δ_k matrices in a block-diagonal structure

$$\Delta^* = \begin{bmatrix} \Delta_1 & & & \\ & \Delta_2 & & \\ & & \ddots & \\ & & & \Delta_K \end{bmatrix}.$$

The corresponding observed distances are denoted by $\delta_{ij,k}$. Similarly, we merge the resulting configurations X_k into the configuration supermatrix

$$X^* = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{bmatrix}.$$

Correspondingly, V^* is block diagonal and one block consists of the submatrix V_k based on the configuration weights W_k and is computed according to Equation 6. Based on these weight matrices W_k with elements $w_{ij,k}$, the total stress to be minimized, consisting of the single $\sigma(X_k)$'s, can be written as

$$\sigma(X^*) = \sum_{k=1}^{K} \sum_{i<j} w_{ij,k}(\delta_{ij,k} - d_{ij}(X_k))^2. \tag{21}$$

In individual difference models there is an additional issue regarding the distance computations. We compute a configuration matrix X_k for each individual, but we constrain the X_k

by only allowing differential weighting of each dimension by each individual. If we think of a linear decomposition of X_k, as described in the former section, we have

$$X_k = ZC_k \tag{22}$$

with the C_k diagonal matrices of order p. The weighted Euclidean distance can be expressed as

$$d_{ij}(ZC_k) = \sqrt{\sum_{s=1}^{p}(c_{ss,k}z_{is} - c_{ss,k}z_{js})^2} = \sqrt{\sum_{s=1}^{p} c_{ss,k}^2(z_{is} - z_{js})^2}. \tag{23}$$

Z is the $n \times p$ matrix of coordinates of the so called *group stimulus space* or *common space*. If $C_k = I$ for all k we get the so called *identity model*.

Regarding majorization, we have to minimize stress in (21) with respect to Z and C_k. The optimization procedure is basically the same as in the former section. Within each iteration t we compute the Guttman transform $\overline{X}^{*(t)}$. Analogous to Equation 17 we have to solve the configuration projection problem

$$\min_{X \in \mathcal{S}} \mathbf{tr} \left(X - \overline{X}^{*(t)}\right)' V^* \left(X - \overline{X}^{*(t)}\right). \tag{24}$$

over X to obtain the update $X^{(t)}$.

In brief, we present three extensions of the classical INDSCAL approach above. Carroll and Chang (1970) extend differential weighting by means of the *generalized Euclidean distance*, allowing the C_k in (22) to be general, and not necessarily diagonal. This means

$$d_{ij}(X_k) = \sqrt{\sum_{s=1}^{p}\sum_{s'=1}^{p}(x_{is} - x_{js})h_{ss',k}(x_{is'} - x_{js'})}, \tag{25}$$

with $H_k = C_k C_k'$. This is known as the IDIOSCAL (Individual DIfferences in Orientation SCALing) model. For identification purposes H_k can be decomposed in various ways. The spectral decomposition (*Carroll-Chang decomposition*) leads to

$$H_k = U_k \Lambda U_k' \tag{26}$$

where $U_k U_k' = I$ and $\Lambda_k = \text{diag}(\lambda_{ij})$. The *Tucker-Harshman decomposition* implies

$$H_k = D_k R_k D_k \tag{27}$$

where D_k is a diagonal matrix of standard deviations and R_i a correlation matrix. This is often combined with the normalization

$$\frac{1}{K}\sum_{k=1}^{K} H_k = I \tag{28}$$

proposed by Schönemann (1972). The models currently implemented in **smacof** are IDIOSCAL, INDSCAL with C_k restricted to be diagonal, and the identity model with $C_k = I$. Additional models can be found in Cox and Cox (2001, Chapter 10).

3.4. SMACOF for rectangular dissimilarity matrices

The prototypical case for rectangular MDS input matrices is that we have n_1 individuals or judges which rate n_2 objects or stimuli. Therefore, MDS becomes a model for preferential choice which is commonly referred to as an *unfolding model*. The basic idea is that the ratings and the judges are represented on the same scale and for each judge, the corresponding line can be folded together at the judge's point and his original rankings are observed (Cox and Cox 2001, p.165). This principle of scaling is sometimes denoted as *Coombs scaling* (Coombs 1950). Detailed explanations on various unfolding techniques can be found in Borg and Groenen (2005, Chapters 14-16). We will limit our explanations to the SMACOF version of metric unfolding.

Let us assume an observed dissimilarity (preference) matrix Δ of dimension $n_1 \times n_2$ with elements δ_{ij}. For rectangular SMACOF the resulting configuration matrix X is partitioned into two matrices: X_1 of dimension $n_1 \times p$ as the individual's or judge's configuration, and X_2 of dimension $n_2 \times p$ as the object's configuration matrix. Consequently, stress can be represented as

$$\sigma(X_1, X_2) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} w_{ij}(\delta_{ij} - d_{ij}(X_1, X_2))^2 \tag{29}$$

with

$$d_{ij}(X_1, X_2) = \sqrt{\sum_{s=1}^{p}(x_{1is} - x_{2js})^2}. \tag{30}$$

Let X be the $(n_1 + n_2) \times p$ joint matrix of configurations. It follows that we can accomplish the same simple SMACOF representation as in (10). The weights w_{ij} are collected into the $n_1 \times n_2$ matrix W_{12} of weights having the same properties as in the former section. The reason we use W_{12} is that, due to the decomposition of X, W has the following block structure:

$$W = \begin{bmatrix} W_{11} & W_{12} \\ W_{12}' & W_{22} \end{bmatrix} = \begin{bmatrix} 0 & W_{12} \\ W_{12}' & 0 \end{bmatrix}$$

The input data structure X does not allow for within-sets proximities. Therefore, W_{11} and W_{22} have 0 entries.

V is computed following Equation 6 and $B(X)$ following Equation 8. Based on the decomposition of X, V can be partitioned into

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{12}' & V_{22} \end{bmatrix},$$

and $B(X)$ into

$$B(X) = \begin{bmatrix} B_{11}(X) & B_{12}(X) \\ B_{12}(X)' & B_{22}(X) \end{bmatrix}.$$

$B_{11}(X)$ is a $n_1 \times n_1$ diagonal matrix with minus the row sums of $B_{12}(X)$ in the diagonal. Correspondingly, B_{22} is $n_2 \times n_2$ with minus the column sums of $B_{12}(X)$ on the diagonal.

To apply majorization we need to define the supporting matrix Y which for rectangular SMACOF consists of the two blocks Y_1 and Y_2. In analogy to $B(X)$ the block structure

$$B(Y) = \begin{bmatrix} B_{11}(Y) & B_{12}(Y) \\ B_{12}(Y)' & B_{22}(Y) \end{bmatrix}$$

results. The sandwich inequality is the same as in (12). To optimize the majorizing function we compute the Moore-Penrose inverse V^+ and the updating formula (Guttman transform) within each iteration t is again $\overline{X}^{(t)} = V^+B(Y)Y$. It should be mentioned that because of the special structure of W the computation of the Moore-Penrose inverse simplifies, especially in the case where the (off-diagonal) weights are all equal

4. Nonmetric SMACOF variants

Looking at various loss functions as for instance given in (4), (21), and (29), we see that we do not have any transformation on the dissimilarities δ_{ij}. If the dissimilarities are on an ordinal scale, we can think of transformations that preserve this rank order. If such a transformation f obeys only the monotonicity constraint $\delta_{ij} < \delta_{i'j'} \Rightarrow f(\delta_{ij}) < f(\delta_{i'j'})$, it is referred to as *nonmetric*.

The resulting $\hat{d}_{ij} = f(\delta_{ij})$ are commonly denoted as *disparities* which have to be chosen in an optimal way. Straightforwardly, the stress function (for the symmetric case) becomes

$$\sigma(X, \widehat{D}) = \sum_{i<j} w_{ij}(\hat{d}_{ij} - d_{ij}(X))^2 \tag{31}$$

which we have to minimize with respect to the configurations X and, simultaneously, with respect to the disparity matrix $\widehat{D}$. Regarding majorization, there is one additional step after each Guttman transform in iteration t: The computation of optimal $\hat{d}_{ij}^{(t)}$ (with subsequent normalization) such that the monotonicity constraint is fulfilled. If the order of $d_{ij}(X^{(t)})$ is the same as the order of $\hat{d}_{ij}^{(t)}$, the optimal update is clearly $\hat{d}_{ij}^{(t)} = d_{ij}(X^{(t)})$. If the orders differ, the optimal update is found by *monotone regression* which we will discuss in brief below.

Before that, we have to consider the case of ties in the observed ordinal dissimilarity matrix Δ, i.e., the case of $\delta_{ij} = \delta_{i'j'}$. Having this case, we distinguish between three approaches: the *primary approach* does not necessarily require that $\hat{d}_{ij} = \hat{d}_{i'j'}$, whereas the (more restrictive) *secondary approach* does. An even less restrictive version is the tertiary approach from De Leeuw (1977b), in which we merely requite that the means of the tie-blocks are in the correct order. More details can be found in Cox and Cox (2001).

When solving the monotone (or isotonic) regression problem in step t, one particular tie approach has to be taken into account. In MDS literature this problem is referred to as *primary monotone least squares regression* and **smacof** solves it by means of the *pooled-adjacent-violators algorithm* (PAVA, Ayer, Brunk, Ewing, Reid, and Silverman 1955; Barlow, Bartholomew, Bremner, and Brunk 1972). This package performs monotone regression using weighted means (i.e., weighted with elements w_{ij}).

5. Extended SMACOF: Quadratic surfaces

The fact that quadratic surfaces frequently show up empirically leads to some interesting technical and methodological problems. In some cases it may be appropriate to require that the points computed by MDS are indeed located exactly on some parametric surface. If we are measuring distances between cities on earth, for example, an exact spherical or elliptical