Multidimensional Scaling Using Majorization: SMACOF in **R**

Jan de Leeuw
University of California, Los Angeles

Patrick Mair
WU 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 (R Development Core Team 2009) 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

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

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

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

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 $\rho(X) \geq \textbf{tr } X'B(Y)Y$. Thus we minorize the convex function $\rho(X)$ with a linear function. This gives us a majorization of *stress*

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

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)

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

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 $\Lambda_k = \mathrm{diag}(\lambda_{ij})$. 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 ID-IOSCAL, 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 representation of the cities makes perfect sense˜(Cox and Cox 1991). Furthermore, it may be appropriate for these nonlinear configurations to measure distance as the shortest geodesic on the non-linear manifold. Again, using the earth as an example, the Euclidean distance, which goes through the earth, may not be as relevant as the geodesic distance measured over the surface of the earth. In addition, since a sphere in three dimensions is really two-dimensional, we could look for ways to portray non-linear structures in higher dimensions

locally faithfully in fewer dimensions. This is basically the classical problem of cartography, in which we compute a suitable projection of the surface of the earth on the plane.

Much more recently, the problem of finding the best MDS representation with points on a sphere came up in computer vision. Ron Kimmel and his group at the Technion in Haifa have published a number of papers in which approximate geodetic distances along arbitrary surfaces are used in MDS with great-circle distances along a sphere. The two-dimensional spherical coordinates are then used to flatten the sphere. A representative paper is~Elad, Keller, and Kimmel (2005). Nonlinear methods for dimensionality reduction are described in Roweis and Saul (2000) and Tenenbaum, De Silva, and Langford (2000).

5.1. Basic formulations for MDS-Q

As mentioned above, in this section we are interested in the case in which the points in the configuration are constrained to lie on a quadratic surface in $\mathbb{R}^p$. In $\mathbb{R}^2$, important special cases are a circle, ellipse, hyperbola, and parabola; in $\mathbb{R}^3$, corresponding special cases are a sphere and an ellipsoid.

We call the technique of placing the points on the MDS with quadratic constraints MDS-Q. Borg and Groenen (2005) call this type of MDS *weakly constrained MDS* since the external quadratic restrictions are not necessarily forced. In the most general form of MDS-Q the vector of configurations $\mathbf{x}_i$, each of length p, must satisfy

$$\mathbf{x}_i'\Lambda\mathbf{x}_i + 2\mathbf{x}_i'\boldsymbol{\beta} + \gamma = 0, \tag{32}$$

for some $p \times p$ matrix Λ, some p-element vector $\boldsymbol{\beta}$, and some constant γ. Because of the invariance of the distance function under translations we can put the center of the surface in the origin. And because distance is invariant under rotation, we can also require, without loss of generality, that Λ is diagonal. This covers conics (ellipse, hyperbola, parabola) in $\mathbb{R}^2$, and the various kinds of ellipsoids, hyperboloids, paraboloids, and cylinders in $\mathbb{R}^3$. In the case of ellipsoids and hyperboloids we can choose $\beta = 0$ and $\gamma = -1$, such that the constraints become $\mathbf{x}_i'\Lambda\mathbf{x}_i = 1$. For ellipsoids, the matrix Λ is positive semi-definite which means that we can also write

$$\mathbf{x}_i = \Lambda^{1/2}\mathbf{z}_i, \text{ where } \|\mathbf{z}_i\| = 1 \text{ for all } i. \tag{33}$$

And, of course, spheres are ellipses in which the matrix Λ is scalar, i.e. $\Lambda = \lambda I$.

5.2. Geodesic distances on quadratic surfaces

Analogously to simple MDS in Section 3.1, we want to find configurations X such that $d_{ij}(X) \approx \delta_{ij}$. Having a quadratic surface we can use the Euclidean distance $d_{ij}(X)$. Alternatively, we can also think of using *geodesic distances* which leads to geodesic MDS-Q. In this case, again, the points are required to lie on a quadratic surface, but now we define distance to be the length of the shortest geodesic along the surface.

In the simple case of a sphere with radius λ these are commonly referred to as *great-circle distances*. The metric embedding problem for spherical space is discussed extensively by~Blumenthal (1953, Chapter VII). Define

$$\breve{d}_{ij}(X) = \lambda \arccos\left(\frac{\mathbf{x}_i'\mathbf{x}_j}{\lambda^2}\right) \tag{34}$$

with $\breve{d}_{ij}(X)$ as the great-circle distance between the points measured along the sphere. Spherical distance is monotonic with Euclidean distance. The two distance scales are quite different at the higher end, because Euclidean distance between two points on the sphere is bounded above by 2λ, while spherical distance is bounded by $2\lambda\pi$. If points are close together the two distances are, of course, approximately equal.

Unfortunately, matters become more complicated if we go from the sphere to the ellipsoid. In $\mathbb{R}^2$ computing the length of an elliptical arc means evaluating an incomplete elliptic integral of the second kind. Suppose the ellipse is $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. The length of arc from point $(0,1)$ to point $(a\sin\theta, b\cos\theta)$, with $0 \leq \theta \leq \frac{1}{2}\pi$, is

$$\breve{d}(\theta) = \int_0^\theta \sqrt{a^2 \sin^2 t + b^2 \cos^2 t}\, dt. \tag{35}$$

Arc length distances between arbitrary points on the ellipse can be computed by adding and subtracting integral terms of this form. Distance along the ellipse is still monotonic with Euclidean distance, but the relationship is no longer simple.

For a parabola $y = \frac{x^2}{a}$ the situation is somewhat simpler because we can measure along the horizontal axis. We find

$$\breve{d}(x_0, x_1) = \frac{1}{2a} \int_{2x_0}^{2x_1} \sqrt{a^2 + u^2}\, du =$$

$$= \frac{u}{4a}\sqrt{a^2 + u^2} + \frac{1}{4}a\log(u + \sqrt{a^2 + u^2})\Big|_{2x_0}^{2x_1},$$

and thus the integral can be evaluated in closed form.

Computing geodesics on the ellipsoid in $\mathbb{R}^3$ is even more complicated than the ellipse. Since the work of Jacobi and Weierstrass we know what the geodesics on ellipsoids look like, and we know how to compute them, but the analytic expressions are complicated and difficult to work with in an MDS context˜(Knörrer 1980; Tabanov 1996; Perelomov 2000). In geodesy there are many programs easily available to do the computations, but for now we have found no easy way to fit geodesics on ellipsoids with MDS. The same thing is true for higher dimensions, and for hyperboloids and paraboloids.

Summarizing, it looks like MDS-Q may be feasible for any dimension and for any quadratic surface. Geodesic MDS-Q, however, seems limited for now to spheres in any dimension, with the possible exception of ellipses and parabolas in $\mathbb{R}^2$.

5.3. Primal methods for MDS-Q

There are several strategies for fitting MDS-Q. The package **smacof** allows for the following approaches: *Primal methods*, in which the constraints are incorporated in parametric form directly into the loss function, and *dual methods*, where constraints are imposed at convergence by using penalty or Lagrangian terms. In this section we will focus on the primal method.

We start with primal gradient projection and how this can be solved using majorization. The constraints˜(33) lead to the problem of minimizing $\mathbf{tr}\ (\lambda Z - Y)'V(\lambda Z - Y)$ over all scalars λ and over all Z with $\mathbf{diag}\ ZZ' = I$. The optimum λ for given Z is

$$\hat{\lambda} = \frac{\mathbf{tr}\ Y'VZ}{\mathbf{tr}\ Z'VZ}, \tag{36}$$

and the problem we need to solve is equivalent to the maximization of

$$\rho(Z) = \frac{[\mathbf{tr}\ Y'VZ]^2}{\mathbf{tr}\ Z'VZ}. \tag{37}$$

In order to maximize the function $\rho(Z)$ we use the fractional programming technique of˜Dinkelbach (1967). Suppose $\tilde{Z}$ is our current best configuration. Define

$$\eta(Z, \tilde{Z}) = [\mathbf{tr}\ Y'VZ]^2 - \rho(\tilde{Z})\mathbf{tr}\ Z'VZ. \tag{38}$$

If we find Z^+ such that $\eta(Z^+, \tilde{Z}) > \eta(\tilde{Z}, \tilde{Z}) = 0$, then $\rho(Z^+) > \rho(\tilde{Z})$. Thus for global convergence it is sufficient to increase $\eta(Z, \tilde{Z})$.

We increase $\eta(Z, \tilde{Z})$ by *block relaxation*˜(De Leeuw 1994), i.e. we cycle through all $\mathbf{z}_i$, optimizing over each of the $\mathbf{z}_i$ in turn, while keeping the others fixed at their current best values. Various different strategies are possible. We can perform a number of cycles updating Z, while keeping $\rho(\tilde{Z})$ fixed at its current value. Or we could compute a new $\rho(\tilde{Z})$ after computing the update of each single row $\mathbf{z}_i$ of Z. It is unclear which strategy is best, and some numerical experimentation will be useful. Expanding we find

$$\eta(Z, \tilde{Z}) = \mathbf{z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{z}_i + 2\mathbf{z}_i'\mathbf{h}_i + \text{ terms not depending on } \mathbf{z}_i, \tag{39}$$

where $U = VY$ and

$$\mathbf{h}_i = \mathbf{u}_i\left(\sum_{j \neq i}^n \mathbf{u}_j'\mathbf{z}_j\right) - \rho(\tilde{Z})\sum_{j \neq i}^n v_{ij}\mathbf{z}_j. \tag{40}$$

Differentiate and introduce a Lagrange multiplier θ for the side condition $\mathbf{z}_i'\mathbf{z}_i = 1$. The stationary equations are

$$(u_i'z_i)u_i + h_i = \theta z_i. \tag{41}$$

Premultiplying both sides by $\mathbf{z}_i'$ shows that if there are multiple solutions, we want the one with the largest value of θ.

The stationary equations have the solution $\mathbf{z}_i = -(\mathbf{u}_i\mathbf{u}_i' - \theta I)^{-1}\mathbf{h}_i$. Thus if we define

$$\mathcal{F}(\theta) = h_i'(u_iu_i' - \theta I)^{-2}h_i \tag{42}$$

we can find θ by solving the equation $\mathcal{F}(\theta) = 1$. Such equations, often called *secular equations*, have been studied systematically in great detail in numerical mathematics, possibly starting with˜Forsythe and Golub (1965); Spjøtvoll (1972). There are excellent reviews of secular equation theory in˜Tao and An (1995) and of solvers in˜Conn, Gould, and Toint (2000, Chapter 7).

Since we are dealing with a simple special case, we can actually solve the secular equation quite simply. Define $\tau_i = \mathbf{u}_i'\mathbf{u}_i$, the projector $P_i = \mathbf{u}_i\mathbf{u}_i'/\tau_i$ and its orthogonal complement $Q_i = I - P_i$. The equation becomes

$$\mathcal{F}(\theta) = \frac{p_i}{(\tau_i - \theta)^2} + \frac{q_i}{\theta^2} = 1, \tag{43}$$

where $p_i = \mathbf{h}_i'P_i\mathbf{h}_i$ and $q_i = \mathbf{h}_i'Q_i\mathbf{h}_i$. We can find the solutions by solving the quartic equation

$$\theta^2 p_i + (\tau_i - \theta)^2 q_i - \theta^2(\tau_i - \theta)^2 = -\theta^4 + 2\tau_i\theta^3 + (p_i + q_i - \tau_i^2)\theta^2 - 2\tau_i q_i\theta + \tau_i^2 q_i = 0 \tag{44}$$
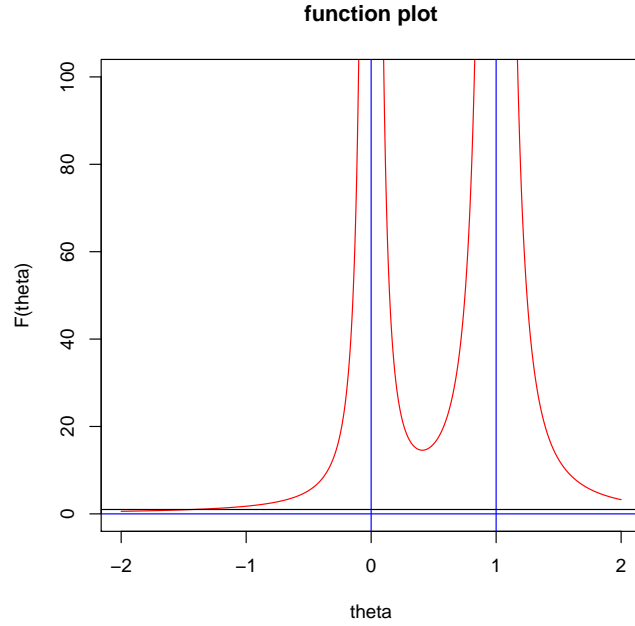
function plot

Figure 1: Root location

for its largest real root.

This makes it possible to describe the behavior of the function $\mathcal{F}$ and to show where the relevant roots are located. The plot in Figure 1 is typical. It has $\tau_i = 1$. The function is always positive, it has a horizontal asymptote at zero and two vertical asymptotes, one at zero and the other at τ_i. Between 0 and τ_i it has its unique stationary value, a local minimum, at

$$\hat{\theta} = \frac{\sqrt[3]{q_i}}{\sqrt[3]{q_i} + \sqrt[3]{p_i}} \tau_i, \tag{45}$$

equal to

$$\mathcal{F}(\hat{\theta}) = \frac{(\sqrt[3]{p_i} + \sqrt[3]{q_i})^3}{\tau_i^2}. \tag{46}$$

Thus the equation $\mathcal{F}(\theta) = 1$ has one negative root and one root larger than τ_i. There may or may not be two additional roots, depending on whether the value of the function at the local minimum is smaller than or larger than one. Correspondingly, the quartic~(44) has either two real roots (and two complex conjugates) or four real roots (of which two could be equal).

5.4. Dual methods: CMDA

The dual method which is used in **smacof** is known as CMDA (Constrained/Confirmatory Monotone Distance Analysis). It was proposed by~Borg and Lingoes (1979, 1980) and further discussed in Borg and Groenen (2005, Section 10.4). The idea is to impose the restrictions directly on the distances and not on the configurations. This makes the method more specific to MDS.

For Euclidean MDS, with points constrained on a circle or a sphere, Borg and Lingoes (1979) introduce an extra point x_0 into the MDS problem, and define the family of penalized loss functions

$$\sigma_\kappa(X) = \min_{\Delta \in \mathcal{D}_L} \sigma_L(X, \Delta) + \kappa \min_{\Delta \in \mathcal{D}_C} \sigma_C(X, \Delta) \tag{47}$$

The whole set $\mathcal{D} = \mathcal{D}_L \cup \mathcal{D}_C$ consists of all non-negative and hollow symmetric matrices that satisfy the constraints. $\Delta \in \mathcal{D}_L$ is the observed dissimilarity matrix and $\Delta \in \mathcal{D}_L$ expresses the side constraints. In the case of the sphere, the side constraints are that x_0 has equal distance to all other points, while the rest of the dissimilarities in $\mathcal{D}_L$ are missing (i.e. have zero w_{ij}). Correspondingly, we have two stress functions $\sigma_L(X, \Delta)$ and $\sigma_C(X, \Delta)$. The non-negative quantity κ is a penalty parameter. If $\kappa \to \infty$ the second term is forced to zero, and we minimize the first term under the conditions that the second term is zero, i.e. that the x_i are on a sphere with center at x_0 and with radius λ.

The CMDA approach has the advantage that it can be implemented quite simply by using the standard Euclidean MDS majorization method. It has the usual disadvantage that we have to select a penalty parameter, or a sequence of penalty parameters, in some way or another. Moreover the Hessian of *stress* will become increasingly ill-conditioned for large penalties, and convergence can consequently be quite slow.

The majorization algorithm for the penalty function~(47) uses the iterations

$$X^{(k+1)} = (V + \kappa I)^{-1}(V\overline{X}^{(k)} + \kappa \tilde{X}^{(k)}), \tag{48}$$

where $\tilde{x}_i^{(k)} = \lambda x_i^{(k)}/\|x_i^{(k)}\|$. For large κ this means

$$X^{(k+1)} \approx \tilde{X}^{(k)} + \frac{1}{\kappa}V(\overline{X}^{(k)} - \tilde{X}^{(k)}), \tag{49}$$

which indicates that convergence will tend to be slow.

The stress functions for additional quadratic surfaces are the following. We know that an ellipse can be defined as the locus of all points that have equal sum of distances from two focal points. Focal points can be chosen on the horizontal axis, at equal distances from the origin. This means that we can fit an ellipse with CMDA by introducing two additional points x_0 and $-x_0$ on the horizontal axis into our MDS problem, and minimize

$$\sigma_\kappa(X, \lambda) = \sigma(X) + \kappa \sum_{i=1}^{n}(d(x_i, x_0) + d(x_i, -x_0) - \lambda)^2. \tag{50}$$

The hyperbola is the locus of all points that have equal difference of distances from two focal points. Thus, analogous to the case of the ellipse, we minimize

$$\sigma_\kappa(X, \lambda) = \sigma(X) + \kappa \sum_{i=1}^{n}(|d(x_i, x_0) - d(x_i, -x_0)| - \lambda)^2. \tag{51}$$

The parabola is the locus of all points that have equal distance to a line (the *directrix*) and a point (the *focal point*). The directrix can be chosen to be the horizontal axis. More generally, for any conic section, the distance to the focal point must be a constant multiple of the distance to the directrix. This multiple, the *eccentricity*, is equal to one for the parabola, large than one for the hyperbola, and less than one for the ellipse.

Suppose x_0 is the focal point, and $\overline{x}_i$ is the projection of x_i on the horizontal axis. Then we must minimize

$$\sigma_\kappa(X, \lambda) = \sigma(X) + \kappa \sum_{i=1}^{n} (d(x_i, x_0) - \lambda d(x_i, \overline{x}_i))^2. \tag{52}$$

6. The R package smacof

So far there does not exist a comprehensive R package for MDS. There are functions in **MASS** to perform classical MDS, Sammon mapping, and non-metric MDS. For the latter variant there are also some functionalities in **vegan** (Oksanen, Kindt, Legendre, O'Hara, Henry, and Stevens 2007), **labdsv** (Roberts 2006), **ecodist** (Goslee and Urban 2007), and **xgobi** (Maechler 2005). Individual differences MDS is implemented in the **SensoMineR** (Husson and Le 2007) package.

In the current version of **smacof** ellipsoids, parabolas, hyperbolas as well as geodesic distances are not implemented. The main functions are `smacofSym()` for basic (symmetric) SMACOF, `smacofIndDiff()` for three-way data, `smacofConstraint()` for SMACOF with external constraints, and, finally, `smacofSphere.primal()` and `smacofSphere.dual()` for sphere projections. Print, summary, and residual S3 methods are provided. The two-dimensional plot options are the configuration plot, Shepard diagram, residual plot, and stress decomposition diagram. The configuration plot is provided for three-dimensional solutions as well (static and dynamic).

6.1. SMACOF for Ekman's color data

Ekman (1954) presents similarities for 14 colors (wavelengths from 434 to 674 nm). The similarities are based on a rating by 31 subjects where each pair of colors was rated on a 5-point scale (0 = no similarity up to 4 = identical). After averaging, the similarities were divided by 4 such that they are within the unit interval.

We perform a two-dimensional basic SMACOF solution using `smacofSym()` which uses the classical scaling solution (Torgerson 1958) as default starting configuration. Note that all SMACOF functions allow for user-specified starting configurations as well.

We start with the transformation of similarities into dissimilarities by subtracting them from 1. Then we perform a two-dimensional basic SMACOF and a quadratic SMACOF, projecting the configurations on a sphere. Both computations are performed in a non-metric manner.

```
> require("smacof")
> data("ekman")
> ekman.d <- sim2diss(ekman, method = 1)
> res.basic <- smacofSym(ekman.d, metric = FALSE)


> res.sphere <- smacofSphere.primal(ekman.d, metric = FALSE)


> plot(res.basic, main = "Configurations Basic SMACOF", asp = 1)
> plot(res.sphere, main = "Configurations Sphere SMACOF", asp = 1)
```
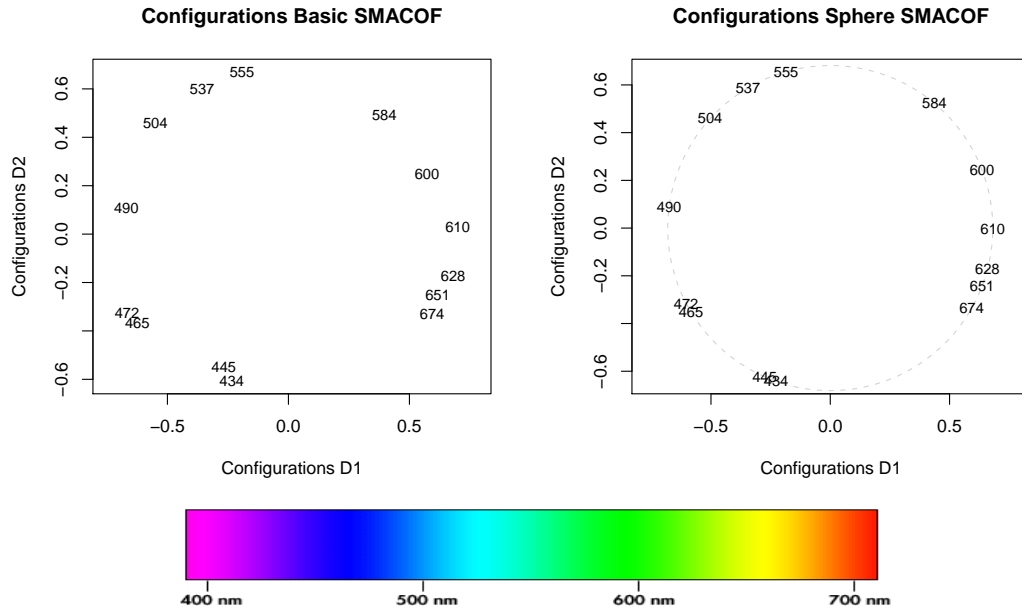
Figure 2: Configuration plots and color wavelengths.

The resulting configuration plots are given in Figure 2. For the nonmetric basic SMACOF solution we see that the wavelength configurations are approximately circularly arranged. Starting with the lowest wavelength we see that the pairs 434-445 and 465-472 correspond to bluish, 490 to turqoise, the set of 504-555 to greenish, 584-610 to yellow-orange, and finally 628-674 to reddish. Since the color palette is commonly represented as a circle, it is somewhat natural to compute a two-dimensional (nonmetric) spherical SMACOF solution which can be found on the right-hand side of Figure 2. Obviously, the configurations lie perfectly on a circle with a radius $\lambda = .0714$.

Now we try to evaluate the basic solution by comparing the SMACOF configuration to that obtained from other data. We can use the CIE Luv space (see e.g. Fairchild 2005) which is extensively used for applications such as computer graphics which deal with colored lights. It is based on human average color matching data and wavelength discrimination data and, thus, it is quite different than the Ekman judgments based on large color differences. Let us first obtain the Luv coordinates in R. First, we read-in the CIE XYZ 31 tristimulus values for an equal energy spectrum at the wavelengths used in the Ekman dataset from an online site. The last line in the code below adjusts these to equiluminance (a constant luminance plane).

```
> wl <- c(434, 445, 465, 472, 490, 504, 537, 555, 584, 600, 610, 628, 651, 674)
> CIE31 <- read.table("http://cvision.ucsd.edu/database/data/cmfs/ciexyz31_1.txt", header
> names(CIE31) <- c("WL", "xbar", "ybar", "zbar")
> ekmanWL <- subset(CIE31, WL %in% wl)
> ekmanWL[, -1] <- ekmanWL[, -1]/ekmanWL[, "ybar"]
```

In the next step for any given wavelength we extract the distance to the next higher one from the final MDS configuration. We do the same thing for the Luv data:

Ekman MDS vs. Luv

Figure 3: Ekman configurations vs. Luv distances.
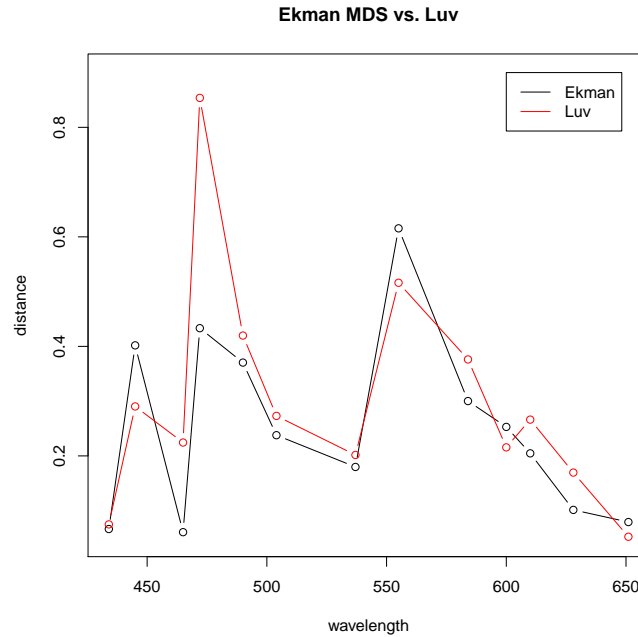
```
> ekman.mds.wld <- res.basic$confdiss[cumsum(c(1, 13:2))]
> require(grDevices)
> ekmanLuv <- convertColor(ekmanWL[, -1], "XYZ", "Luv")
> Luv.dist <- dist(ekmanLuv)
> Luv.wld <- Luv.dist[cumsum(c(1, 13:2))]
```

On the x-axis we plot the wavelengths and on the y-axis the distances to the subsequent one.

```
> plot(wl[-14], ekman.mds.wld, type = "b", ylim = c(0.05, 0.9), xlab = "wavelength", ylab
> lines(wl[-14], Luv.wld/400, col = "red", type = "b")
> legend(600, 0.9, c("Ekman", "Luv"), lty = 1, col = c("black", "red"))
```

Note that the division of the Luv distances by 400 is arbitrary. The result is given in Figure 3. It shows striking similarities in the distances across different wavelengths.

6.2. Breakfast rating for rectangular SMACOF

As a metric unfolding example we use the breakfast dataset from Green and Rao (1972) which is also analyzed in Borg and Groenen (2005, Chapter 14). 42 individuals were asked to order 15 breakfast items due to their preference. These items are: toast = toast pop-up, butoast = buttered toast, engmuff = English muffin and margarine, jdonut = jelly donut, cintoast = cinnamon toast, bluemuff = blueberry muffin and margarine, hrolls = hard rolls and butter, toastmarm = toast and marmalade, butoastj = buttered toast and jelly, toastmarg = toast and margarine, cinbun = cinnamon bun, danpastry = Danish pastry,
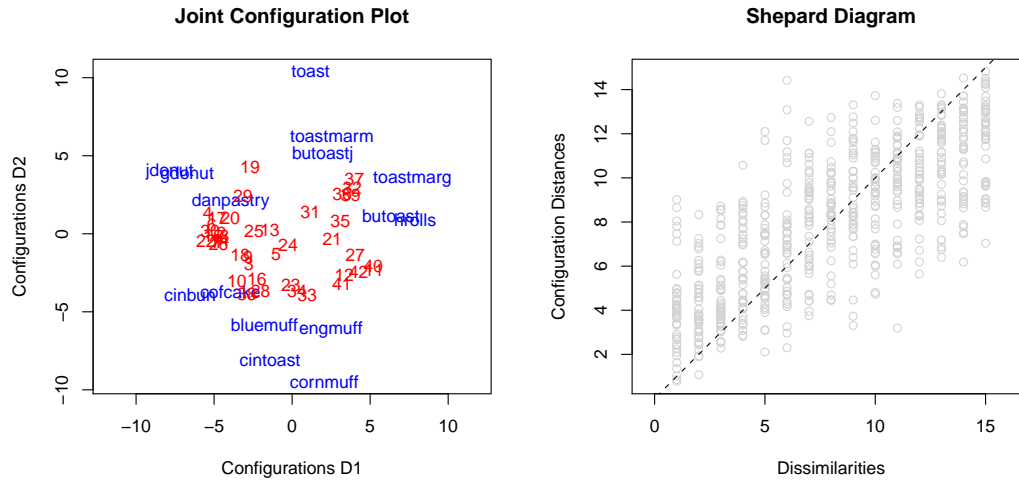
Figure 4: Joint configuration plot/Shepard diagram for breakfast data.

gdonut = glazed donut, cofcake = coffee cake, and cornmuff = corn muffin and butter. For this 42×15 matrix we compute a rectangular SMACOF solution.

```
> data("breakfast")
> res.rect <- smacofRect(breakfast, itmax = 1000)


> plot(res.rect, joint = TRUE, xlim = c(-10, 10), asp = 1)
> plot(res.rect, plot.type = "Shepard", asp = 1)
```

The configuration plot on the left hand side in Figure 4 generated by the method plot.smacofR() represents the coordinates of the breakfast types and the rater jointly. First, let us focus on the breakfast configurations. At the top and top-right we see a large toast cluster or, since hard rolls are there too, we could characterize it also as bread cluster. At the bottom we can identify a muffin cluster including the cinnamon toast. Moving more to the right the cinnamon bun and the coffee cake are grouped together. Finally, at the top-right corner the jelly and glazed donut form a donut cluster with the danish pastry close to them.

Examining the raters we see, from a general perspective, that the individuals lie pretty much in the center of the configuration plots. This is reasonable since each rater was supposed to rate each breakfast type on an ordinal level. An interesting group of individuals (raters 32, 36, 37, 39) appears in the toast cluster. Looking closer at their ratings in the data file, their preferences in general are very similar, and, of course, the toasts are among their first preferences.

The Shepard diagram on the right-hand side in Figure 4 represents the observed distances on the abscissae and the resulting configuration distances based on the final SMACOF solution on the ordinate. An isotonic regression is fitted through the corresponding pairs of observations. The Shepard diagram is useful for the goodness-of-fit examination of the results.

6.3. Three-way SMACOF based on bread ratings

The data set we provide for three-way SMACOF is described in Bro (1998). The raw data consist of ratings of 10 breads on 11 different attributes carried out by 8 raters. Note that the bread samples are pairwise replications: Each of the 5 different breads, which have a different salt content, was presented twice for rating. The attributes are bread odor, yeast odor, off-flavor, color, moisture, dough, salt taste, sweet taste, yeast taste, other taste, and total. First we fit an unconstrained solution followed by a model with identity restriction.

```
> data("bread")
> res.uc <- smacofIndDiff(bread)
> res.uc

Call: smacofIndDiff(delta = bread)

Model: Three-way SMACOF
Number of objects: 10

Metric stress: 0.5218725
Number of iterations: 385

> res.id <- smacofIndDiff(bread, constraint = "identity")
> res.id

Call: smacofIndDiff(delta = bread, constraint = "identity")

Model: Three-way SMACOF
Number of objects: 10

Metric stress: 0.5343962
Number of iterations: 98

> plot(res.uc, main = "Group Configurations Unconstrained", xlim = c(-1.2,1), ylim = c(-1.
> plot(res.id, main = "Group Configurations Identity", xlim = c(-1.2,1), ylim = c(-1.2,1),
> plot(res.uc, plot.type = "resplot", main = "Residuals Unconstrained", xlim = c(2,14), yl
> plot(res.id, plot.type = "resplot", main = "Residuals Indentity", xlim = c(2,14), ylim =
```

The identity restriction leads to the same configurations across the raters. The stress value for the identity solution is considerably higher than for the unconstrained solution. The unconstrained solution reflects nicely the bread replication pairs. Their configurations at the top of Figure 5 are very close to each other. The residual plots on the bottom show a descending trend for both solutions: small distances are systematically overestimated, large distances are underestimated. Of course, the residuals for the unconstrained solution are in general smaller than those for the identity constrained.

6.4. Constrained SMACOF on kinship data

As an example for SMACOF with external constraints we use a data set collected by Rosenberg and Kim (1975). Based on their similarity, students sorted the following 15 kinship terms into
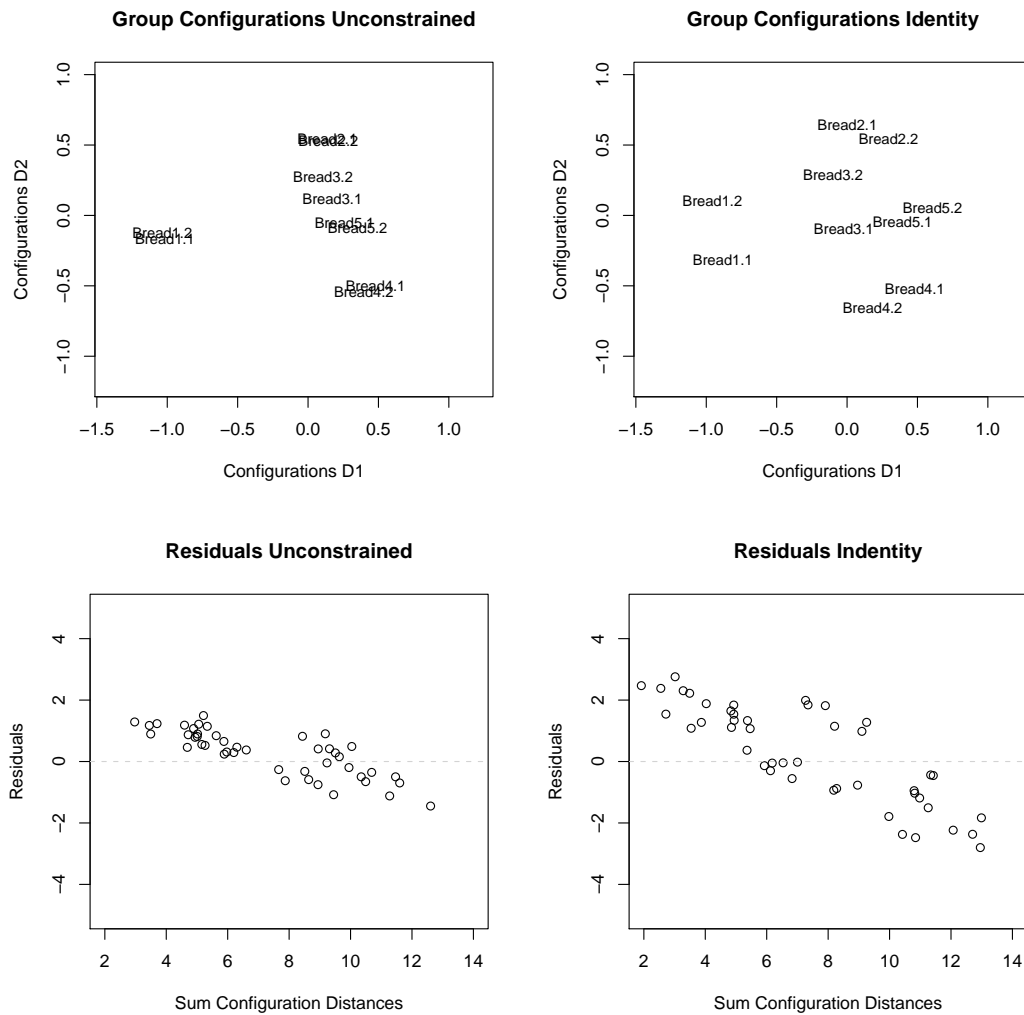
Figure 5: Group configuration and residual plots for bread data.

groups: aunt, brother, cousin, daughter, father, granddaughter, grandfather, grandmother, grandson, mother, nephew, niece, sister, son, and uncle. Thus, the data set has some obvious "pairs" such as sister-brother, daughter-son, mother-father etc.

Based on this grouping they created a dissimilarity matrix according to the following rule: 1 if the objects were sorted in different groups, 0 for the same group. Based on these values we can compute the percentages of how often terms were not grouped together over all students (see Borg and Groenen 2005, p. 83).

As external scales we have gender (male, female, missing), generation (two back, one back, same generation, one ahead, two ahead), and degree (first, second, etc.) of the kinship term. Using these external scales we fit a two-dimensional SMACOF solution with linear constraints and an ordinary SMACOF solution without covariates.

```
> data("kinshipdelta")
```

```
> data("kinshipscales")
> res.sym <- smacofSym(kinshipdelta)
> res.lin <- smacofConstraint(kinshipdelta, constraint = "linear", external = kinshipscale

> plot(res.sym, main = "Configuration Plot SMACOF", xlim = c(-1, 1), ylim = c(-1, 1.5), as
> plot(res.lin, main = "Configuration Plot SMACOF Constraint",xlim = c(-1, 1), ylim = c(-1
> plot(res.sym, plot.type = "stressplot", main = "Stress Decomposition SMACOF", xlim = c(0
> plot(res.lin, plot.type = "stressplot", main = "Stress Decomposition SMACOF Constraint",
```

In Figure 6 we see clearly the effect of the external scale on the SMACOF solution. The left
hand side, i.e. the ordinary SMACOF solution, shows the typical circular alignment of the
objects. Especially in negative x-direction the natural pairs such as sister-brother, daughter-
son, and mother-father are quite distant from each other. The stress decomposition charts
show the proportions of each object to total stress.

The plot for the constrained solution shows a far more consistent picture. All these natural
pairs have almost the same x-coordinates. The distance in y-direction reflects a gender effect
(females are always above males). Cousin is somewhat apart from the other star-like con-
figurations. This makes sense since cousin, which is genderless (covariate gender is missing),
consequently does not have any natural pair.

6.5. 3D spherical SMACOF

Cox and Cox (2001) took data from the New Geographical Digest in 1986 on which countries
traded with other countries. For 20 countries the main trading partners are dichotomously
scored (1 = trade performed, 0 = trade not performed). Based on this dichotomous matrix the
dissimilarities are computed using the Jaccard coefficient. The most intuitive MDS approach
is to project the resulting distances to a sphere which gives us a "trading globe". In this
example we use the dual algorithm for spherical projection.

```
> data("trading")
> res.sphere <- smacofSphere.dual(trading, ndim = 3)
> plot3d(res.sphere, plot.sphere = FALSE)
```

Note that if we apply the dual algorithm, the first row in the configuration object `res.sphere$conf`
corresponds to the center of the sphere. However, on the left circumference of the globe we
have a cluster of Commonwealth Nations (New Zealand, Australia, Canada, India). China is
somewhat separated which reflects its situation in 1986. For this time, the cluster of the com-
munist governed countries USSR, Czechoslovakia, and Eastern Germany is also reasonable.
Around the "south pole" we have Japan, USA, West Germany, and UK.

7. Discussion

In this final section we will discuss some relations to other models as well as some future
extensions of the package. The **homals** package (De Leeuw and Mair 2009) allows for the
computation of various Gifi models (Gifi 1990). Meulman (1992) incorporates the MDS
approach into Gifi's optimal scaling model family. Takane, Young, and De Leeuw (1977)

Configuration Plot SMACOF

Configuration Plot SMACOF Constraint

Stress Decomposition SMACOF

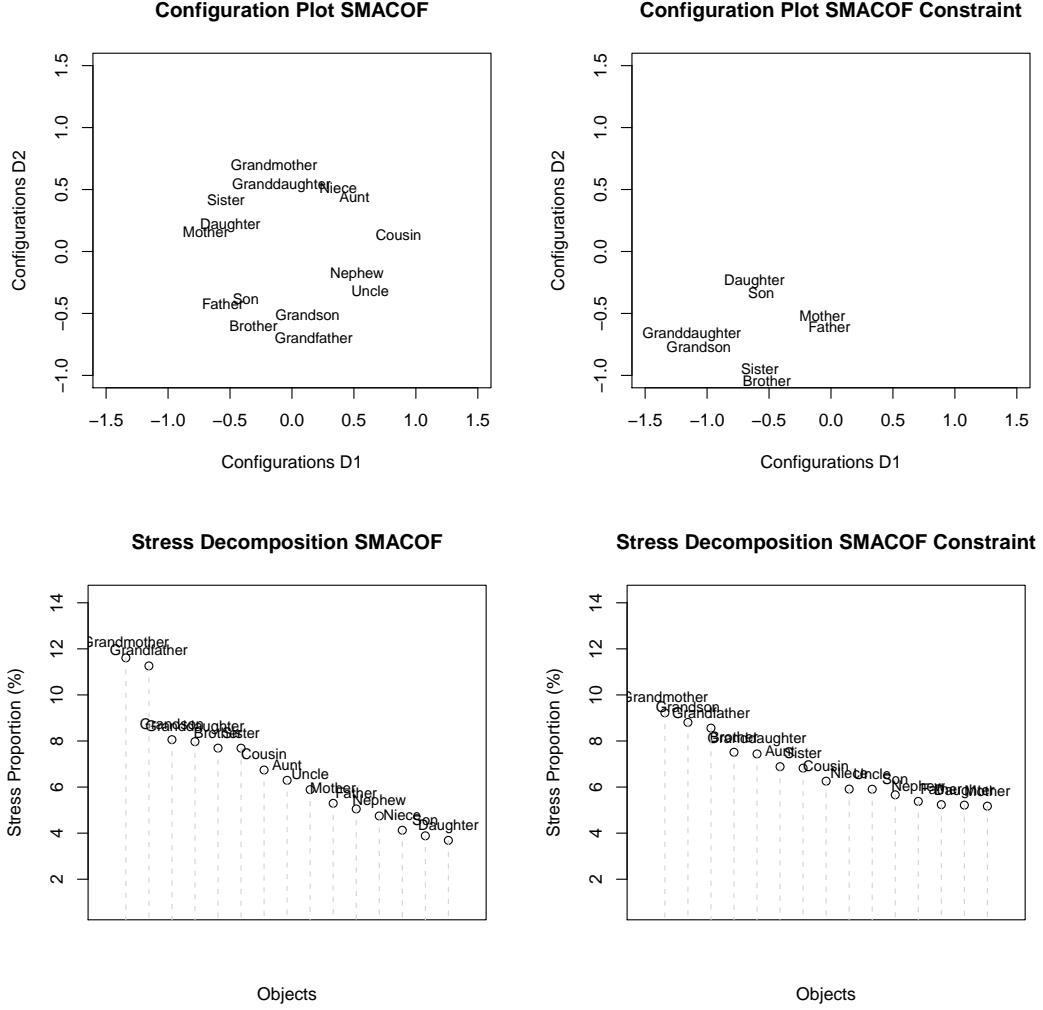**Stress Decomposition SMACOF Constraint**

Figure 6: Configuration plots without/with external scales.

developed ALSCAL (Alternating Least Squares SCALing) for stress minimization can be regarded as an alternative to majorization. Note that Gifi mainly uses ALS (alternating least squares) to minimize the corresponding loss functions. Especially, by looking at the restricted models in Section 3.2 the similarity between SMACOF and the Gifi approach becomes obvious (see also Borg and Groenen 2005, p. 233). A general discussion can be found also in Cox and Cox (2001, Chapter 11).

MDS is also related to correspondence analysis (CA) which, within the context of this special issue, is provided by the package **anacor**. CA is basically a two-mode technique, displaying row and column objects. From this point of view CA is related to the unfolding MDS, or, in our framework, to rectangular SMACOF from Section 3.4.

Similar to CA, so far the **smacof** package can only handle positive dissimilarities. In subsequent versions we will allow also for negative dissimilarities. Furthermore, projections onto
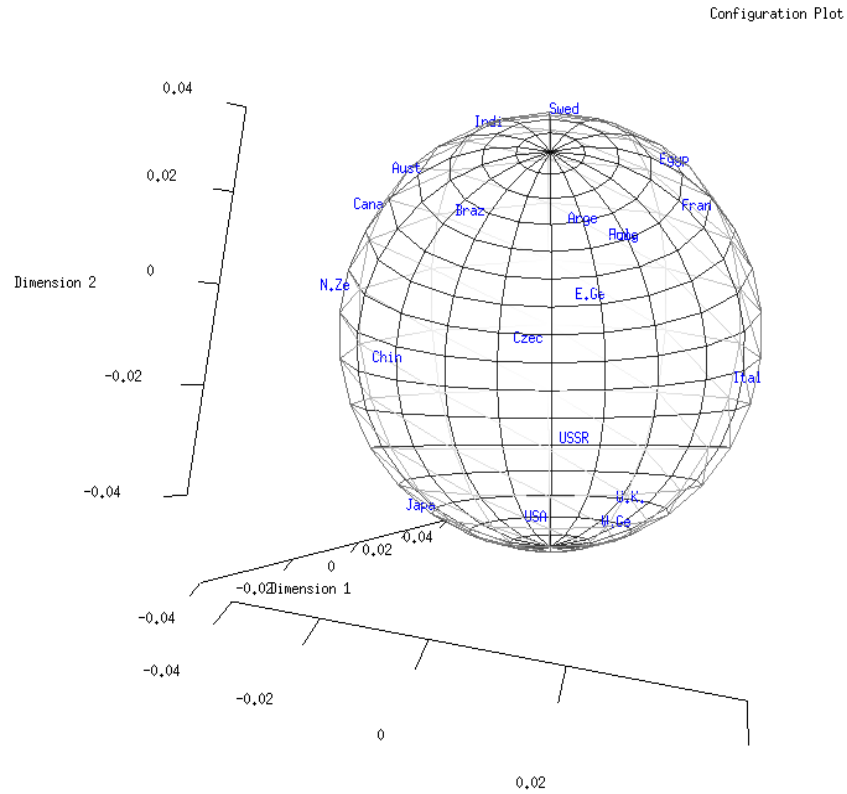
Configuration Plot

Figure 7: 3D sphere projection of trading data.

ellipsoids, hyperbolas and parabolas will be possible. The formal groundwork was given in Section 5.4. This also includes the implementation of geodesic distances from Section 5.2. Furthermore, we can think of implementing additional plot types as given in Chen and Chen (2000).

Another issue is the speed of convergence of the majorization approach. Especially for large nonmetric problems the computational time can become quite extensive. Rosman, Bronstein, Bronstein, Sidi, and Kimmel (2008) and De Leeuw (2008b,a) give details on how to accelerate the majorization algorithm for MDS. The R code included in the reports by De Leeuw is developed for unrestricted metric scaling, but it can be applied to restricted and non-metric scaling as well. Boosting the algorithms with various of these accelerators would represent another improvement of the package.

References

Ayer M, Brunk HD, Ewing GM, Reid WT, Silverman E (1955). "An Empirical Distribution Function for Sampling with Incomplete Information." *The Annals of Mathematical Statistics*, **26**, 641–647.

Barlow RE, Bartholomew RJ, Bremner JM, Brunk HD (1972). *Statistical Inference Under Order Restrictions.* John Wiley & Sons, New York.

Bentler PM, Weeks DG (1978). "Restricted Multidimensional Scaling Models." *Journal of Mathematical Psychology*, **17**, 138–151.

Blumenthal LM (1953). *Theory and Applications of Distance Geometry.* Clarendon Press, Oxford, England.

Borg I, Groenen PJF (2005). *Modern Multidimensional Scaling: Theory and Applications.* 2nd edition. Springer, New York.

Borg I, Lingoes JC (1979). "Multidimensional Scaling with Side Constraints on the Distances." In JC Lingoes and EE Roskam and I Borg (ed.), *Geometric Representations of Relational Data.* Mathesis Press, Ann Arbor, Michigan.

Borg I, Lingoes JC (1980). "A Model and Algorithm for Multidimensional Scaling with External Constraints on the Distances." *Psychometrika*, **45**, 25–38.

Bro R (1998). *Multi-Way Analysis in the Food Industry: Models, Algorithms, and Applications.* Ph.D. thesis, University of Amsterdam (NL) & Royal Veterinary and Agricultural University (DK).

Carroll J, Chang J (1970). "Analysis of Individual Differences in Multidimensional Scaling Via an N-way Generalization of Eckart-Young Decomposition." *Psychometrika*, pp. 283–319.

Chen CH, Chen JA (2000). "Interactive Diagnostic Plots for Multidimensional Scaling with Applications in Psychosis Disorder Data Analysis." *Statistica Sinica*, **10**, 665–691.

Conn A, Gould N, Toint P (2000). *Trust-Region Methods.* MPS-SIAM Series on Optimization. SIAM, Philadelphia, PA.

Coombs CH (1950). "Psychological Scaling Without a Unit of Measurement." *Psychological Review*, **57**, 145–158.

Cox TF, Cox MAA (1991). "Multidimensional Scaling on a Sphere." *Communications in Statistics: Theory and Methods*, **20**, 2943–2953.

Cox TF, Cox MAA (2001). *Multidimensional Scaling.* 2nd edition. Chapman & Hall/CRC, Boca Raton, FL.

De Gruijter DNM (1967). "The Cognitive Structure of Dutch Political Parties in 1966." *Report E019-67*, Psychological Institute, University of Leiden.

De Leeuw J (1977a). "Applications of Convex Analysis to Multidimensional Scaling." In J˜Barra, F˜Brodeau, G˜Romier, B˜van Cutsem (eds.), *Recent Developments in Statistics*, pp. 133–145. North Holland Publishing Company, Amsterdam, The Netherlands.

De Leeuw J (1977b). "Correctness of Kruskal's Algorithms for Monotone Regression with Ties." *Psychometrika*, **42**, 141–144.

De Leeuw J (1988). "Convergence of the Majorization Method for Multidimensional Scaling." *Journal of Classification*, **5**, 163–180.

De Leeuw J (1994). "Block Relaxation Algorithms in Statistics." In HH˜Bock, W˜Lenski, MM˜Richter (eds.), *Information Systems and Data Analysis*, pp. 308–324. Springer.

De Leeuw J (2008a). "Accelerated Least Squares Multidimensional Scaling." *Preprint series*, UCLA Department of Statistics. URL http://preprints.stat.ucla.edu/493/rate.pdf.

De Leeuw J (2008b). "Polynomial Extrapolation to Accelerate Fixed Point Iterations." *Preprint series*, UCLA Department of Statistics. URL http://preprints.stat.ucla.edu/542/linear.pdf.

De Leeuw J, Heiser WJ (1977). "Convergence of Correction Matrix Algorithms for Multidimensional Scaling." In J˜Lingoes (ed.), *Geometric Representations of Relational Data*, pp. 735–752. Mathesis Press, Ann Arbor, Michigan.

De Leeuw J, Heiser WJ (1980). "Multidimensional Scaling with Restrictions on the Configuration." In P˜Krishnaiah (ed.), *Multivariate Analysis, Volume V*, pp. 501–522. North Holland Publishing Company, Amsterdam, The Netherlands.

De Leeuw J, Mair P (2009). "Gifi Methods for Optimal Scaling in R: The Package **homals**." *Journal of Statistical Software*, **forthcoming**, 1–30.

De Leeuw J, Meulman J (1986). "A Special Jackknife for Multidimensional Scaling." *Journal of Classification*, **3**, 97–112.

Dempster AP, Laird NM, Rubin DB (1977). "Maximum Likelihood for Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society*, **B39**, 1–38.

Dinkelbach W (1967). "On Nonlinear Fractional Programming." *Management Science*, **13**, 492–498.

Ekman G (1954). "Dimensions of Color Vision." *Journal of Psychology*, **38**, 467–474.

Elad A, Keller Y, Kimmel R (2005). "Texture Mapping via Spherical Multi-dimensional Scaling." In R˜Kimmel, N˜Sochen, J˜Weikert (eds.), *Scale-Space 2005, LNCS 3459*, pp. 443–455. Springer-Verlag, Berlin, Heidelberg, New York.

Fairchild MD (2005). *Color Appearance Models*. 2nd edition. John Wiley & Sons, Chichester.

Fillenbaum S, Rapaport A (1971). *Structure in the Subjective Lexicon*. Academic Press, New York, New York.

Forsythe G, Golub G (1965). "On the Stationary Values of a Second Degree Polynomial on the Unit Sphere." *Journal of the Society for Industrial and Applied Mathematics*, **13**, 1050–1068.

Gifi A (1990). *Nonlinear Multivariate Analysis*. John Wiley & Sons, Chichester, England.

Goslee S, Urban D (2007). *ecodist: Dissimilarity-based functions for ecological analysis*. R package version 1.1.3, URL http://CRAN.R-project.org/package=ecodist.

Green P, Rao V (1972). *Applied Multidimensional Scaling*. Holt, Rinehart & Winston.

Groenen PJF (1993). *The Majorization Approach to Multidimensional Scaling: Some Problems and Extensions*. Ph.D. thesis, University of Leiden.

Groenen PJF, Heiser WJ (1996). "The Tunneling Method for Global Optimization in Multidimensional Scaling." *Psychometrika*, **61**, 529–550.

Guttman L (1968). "A General Nonmetric Technique for Fitting the Smallest Coordinate Space for a Configuration of Points." *Psychometrika*, **33**, 469–506.

Husson F, Le S (2007). *SensoMineR: Sensory Data Analysis with R*. R package version 1.07, URL http://CRAN.R-project.org/package=SensoMineR.

Knörrer H (1980). "Geodesics on the Ellipsoid." *Inventiones Mathematicae*, **59**, 119–143.

Kruskal JB, Wish M (1978). *Multidimensional Scaling (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-011)*. Sage Publications, Newbury Park, CA.

Lange K (2004). *Optimization*. Springer, New York.

Levelt WJM, Van De Geer JP, Plomp R (1966). "Triadic Comparisons of Musical Intervals." *British Journal of Mathematical and Statistical Psychology*, **19**, 163–179.

Maechler M (2005). *xgobi: Interface to the XGobi and XGvis Programs for Graphical Data Analysis*. R package version 1.2-13, URL http://CRAN.R-project.org/package=xgobi.

Mardia KV, Kent JT, Bibby JM (1979). *Multivariate Analysis*. Academic Press, New York.

Meulman JJ (1992). "The Integration of Multidimensional Scaling and Multivariate Analysis With Optimal Transformations." *Psychometrika*, **57**, 539–565.

Meyer D, Buchta C (2007). *proxy: Distance and Similarity Measures*. R package version 0.3, URL http://CRAN.R-project.org/package=proxy.

Oksanen J, Kindt R, Legendre P, O'Hara B, Henry M, Stevens H (2007). *vegan: Community Ecology Package*. R package version 1.8-8, URL http://CRAN.R-project.org/package=vegan.

Perelomov AM (2000). "A Note on Geodesics on an Ellipsoid." *Regular and Chaotic Dynamics*, **5**, 89–94.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ramsay JO (1982). "Some Statistical Approaches to Multidimensional Scaling Data." *Journal of the Royal Statistical Society, Series A*, **145**, 285–312.

Roberts DW (2006). **labdsv**: *Laboratory for Dynamic Synthetic Vegephenomenology*. R package version 1.2-2, URL http://CRAN.R-project.org/package=labdsv.

Rosenberg S, Kim MP (1975). "The Method of Sorting as a Data Gathering Procedure in Multivariate Research." *Multivariate Behavioral Research*, **10**, 489–502.

Rosman G, Bronstein AM, Bronstein MM, Sidi A, Kimmel R (2008). "Fast Multidimensional Scaling using Vector Extrapolation." *Technical report CIS-2008-01*, Computer Science Department, Technion, Haifa. URL http://www.cs.technion.ac.il/users/wwwb/cgi-bin/tr-info.cgi/2008/CIS/CIS-2008-01.

Roweis S, Saul LK (2000). "Nonlinear Dimensionality Reduction by Locally Linear Embedding." *Science*, **290**, 2323–2326.

Schönemann PH (1972). "An Algebraic Solution for a Class of Subjective Metrics Models." *Psychometrika*, **37**, 441–451.

Shepard R (1962a). "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function (Part I)." *Psychometrika*, **27**, 125–140.

Shepard R (1962b). "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function (Part II)." *Psychometrika*, **27**, 219–246.

Shepard RN (1974). "Representation of Structure in Similarity Data: Problems and Prospects." *Psychometrika*, **39**, 373–421.

Shepard RN (1980). "Multidimensional Scaling, Tree-Titting, and Clustering." *Science*, **210**, 390–398.

Spjøtvoll E (1972). "A Note on a Theorem of Forsythe and Golub." *SIAM Journal on Applied Mathematics*, **23**, 307–311.

Tabanov MB (1996). "New Ellipsoidal Confocal Coordinates and Geodesics on an Eliipsoid." *Journal of Mathematical Sciences*, **82**, 3851–3858.

Takane Y, Young FW, De Leeuw J (1977). "Nonmetric Individual Differences in Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features." *Psychometrika*, **42**, 7–67.

Tao PD, An LTH (1995). "Lagrangian Stability and Global Optimality in Nonconvex Quadratic Minimization Over Euclidean Balls and Spheres." *Journal of Convex Analysis*, **2**, 263–276.

Tenenbaum JB, De Silva V, Langford JC (2000). "A Global Geometric Framework for Nonlinear Dimensionality Reduction." *Science*, **290**, 2319–2323.

Torgerson W (1958). *Theory and Methods of Scaling*. John Wiley & Sons, New York.

Affiliation:

Jan de Leeuw
Department of Statistics

University of California, Los Angeles
E-mail: deleeuw@stat.ucla.edu
URL: http://www.stat.ucla.edu/~deleeuw/