Smacof at 50: A Manual
Part 4: Smacof for Cartwheel Data

Jan de Leeuw - University of California Los Angeles

Started May 19 2024, Version of May 19, 2024

Contents

Note: This is a working manuscript which will be expanded/updated frequently. All suggestions for improvement are welcome. All Rmd, tex, html, pdf, R, and C files are in the public domain. Attribution will be appreciated, but is not required. The files can be found at https://github.com/deleeuw in the smacofPC directory of the repositories smacofCode, smacofManual, and smacofExamples.

1 Loss Function

We start with a general non-metric MDS problem. The formulas for the general case are simpler than those for the various special cases implemented in smacofNM.

The loss function is

$$\sigma(X, \Delta_1, \cdots, \Delta_s) := \sum_{r=1}^{R} \sigma_r(X, \Delta_r), \tag{1}$$

with

$$\sigma_r(X, \Delta_r) := \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ijr}(\delta_{ijr} - d_{ij}(X))^2. \tag{2}$$

As usual, the symbol $:=$ is used for definitions.

The *weights* $W_r = \{w_{ijr}\}$ are known non-negative numbers and $D(X) := \{d_{ij}(X)\}$ is a matrix of Euclidean distances between the rows of the $n \times p$ *configuration* $X = \{x_{is}\}$, which are interpreted as n points in $\mathbb{R}^p$.

Loss function (1) must be minimized over *configurations* X and over the K matrices of *disparities* $\Delta_r = \{\delta_{ijr}\}$. The minimization problem has some constraints, both on X and on the Δ_r. We require that $X \in \mathcal{X} \subseteq \mathbb{R}^{n \times p}$. Here $\mathcal{X}$ is the set of *column-centered* (columns add up to zero) $n \times p$ matrices that are *normalized* by

$$\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^{\star} d_{ij}^2(X)) = 1, \tag{3}$$

where

$$w_{ij}^{\star} := \sum_{r=1}^{R} w_{ijr}. \tag{4}$$

The *disparities* $\Delta_r = \{\delta_{ijr}\}$ are required to satisfy $\Delta_r \in \mathcal{C}_r$. The $\mathcal{C}_r$ are polyhedral convex cones, which are subcones of the cone of non-negative matrices. Each of the cones is defined by partial orders $\leq_r$ over the elements of the Δ_r. In general, neither the (known) weight matrices W_r nor the (unknown) disparity matrices Δ_r need to be symmetric and/or hollow (i.e. have zero diagonal).

The *data* of the MDS problem are the weights W_r and the cones $\mathcal{C}_r$. Each pair $(W_r, \mathcal{C}_r)$ is called a *slice* of the data. The *unknowns* or *parameters* of the problem are $X \in \mathcal{X}$ and the $\Delta_r \in \mathcal{C}_r$.

2 The Data

2.1 Data Generation

Utility

3 Algorithm

3.1 Initial Configuration

In metric and non-linear MDS the default initial configuration is the classical Torgerson metric MDS solution. That is not available for smacofPC, because there are no numerical dissimilarities. In De Leeuw (1970) (section 5.1) and De Leeuw (1974) (chapter 4) a different eigenvalue-eigenvector based initial solution is proposed. We discuss a somewhat modernized version here. It is sometime known as the *maximum sum method*.

The data are an order over a number of pairs of pairs.

$$\omega(X) := \sum_{(i,j)\prec(k,l)}\sum (d_{ij}^2(X) - d_{kl}^2(X)), \tag{5}$$

where $(i,j) \prec (k,l)$ means that we want $d_{ij}(X) \leq d_{kl}(X)$. We want all terms in $\omega(X)$ to be positive, but for purposes of the initial configuration we relax this to wanting their sum to be large. This explains the "maximum sum" name.

The sum (5) can be written as the quadratic form $\omega(X) = \text{tr } X'A^\star X$, with

$$A^\star := \left\{\sum_{(i,j)}\sum_{(k,l)}(A_{ij} - A_{kl})\right\} \tag{6}$$

Note that $A^\star$ is symmetric and doubly-centered. Moreover its trace is zero, and consequently it has one zero, some negative, and some positive eigenvalues.

Because X is column centered we have $\omega(X) = \text{tr } X'(A^\star + \theta J)X$ where J is the centering matrix $I - n^{-1}ee'$ and θ is arbitrary. For the non-zero eigenvalues we have $\lambda_s(A^\star + \theta J)) = \lambda_s(A^\star) + \theta$, and thus for $\theta \geq -\lambda_{\text{min}}(A^\star)$ the matrix $A^\star + \theta J$ is positive semi-definite.

Of course maximizing ω over all X does not make sense, because by making X larger we make ω larger. Thus the supremum over all X is $+\infty$ and the maximum is not attained. We need some kind of normalization. The obvious choice is $\text{tr } X'X = 1$, but unfortunately that does not work. It gives a rank-one solution with all columns of X equal to the eigenvector corresponding with the dominant eigenvalue of $A^\star$. Instead we use $\text{tr } (X'X)^2 = 1$. This gives the solution $X = K\Lambda^{\frac{1}{2}}$ with Λ the largest p eigenvalues of $A^\star$ (assumed to be non-negative) and K the corresponding normalized eigenvectors. This is our version of the Torgerson initial solution for non-metric MDS.

We have been deliberately vague about what to do if the number of positive eigenvalues of $A^\star$ is less than p, which is of course a problem the Torgerson metric MDS solution has as well. In the program we simply choose θ equal to $-\lambda_p(A^\star)$ if $\lambda_p(A^\star) < 0$. We expect the problem to be rare, and the actual choice of θ to be fairly inconsequential.

3.2 The Iterations

As in other smacof implementations our minimization method is based on the *alternating least squares* principle. Start iteration $\nu = 0$ with the initial normalized configuration $X^{(0)}$. We do

not need an initial $(\Delta_1^{(0)}, \cdots, \Delta_1^{(0)})$. Then in each iteration ν we improve the configuration and the disparities by solving two subproblems. They are

$$\text{Step } \nu.1: \quad \text{For } k = 1, \cdots, K \text{ set } \Delta_r^{(\nu+1)} = \underset{\Delta_r \in \mathcal{C}_r}{\operatorname{argmin}} \sigma_r(X^{(\nu)}, \Delta_r), \tag{7}$$

$$\text{Step } \nu.2: \quad \text{select } X^{(\nu+1)} \text{ from } \underset{X \in \mathcal{X}}{\operatorname{argmin}} \sigma(X, \Delta_1^{(\nu+1)}, \cdots, \Delta_s^{(k+1)}). \tag{8}$$

3.3 First Subproblem

In formulating the first subproblem (7) we have used the fact that the projections on the cones $\mathcal{C}_r$ can be carried out separately for each k. We also know that the projections exist and are unique.

3.4 Second Subproblem

To minimize loss over $X \in \mathcal{X}$ for the current best value of the Δ_r. This subproblem is simplified by using the least squares partitioning

$$\sigma(X, \Delta_1, \cdots, \Delta_s) = \sum_{r=1}^{R} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ijr}(\delta_{ijr} - \delta_{ij}^\star)^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^\star(\delta_{ij}^\star - d_{ij}(X))^2, \tag{9}$$

where

$$\delta_{ij}^\star = \frac{\sum_{r=1}^{R} w_{ijr}\delta_{ijr}}{\sum_{r=1}^{R} w_{ijr}}. \tag{10}$$

Minimizing over X can be done by minimizing the second term on the right in (9), which is a standard metric smacof problem. But unfortunately there is no closed form solution for this problem, and the minimizer must be computed by an infinite iterative process. Of course we do not want to have an infinite iterative process with the infinite alternating least square process, and thus we truncate the minimizations in the second subproblem. Although the resulting algorithm is not strictly of the form (7)-(8) any more, it does decrease the loss function in each iteration and consequently produces a stable and convergent algorithm. Since the smacof iterations use majorization (or MM), our overall algorithm combines alternating least squares and majorization.

3.5 Slice-Independence

We will assume throughout that $w_{ijr} = w_{ij}\epsilon_{ijr}$, where $\epsilon_{ijr} = 1$ is either zero or one. If ϵ_{ijr} is one, we say that pair (i, j) *participates* in slice r. Thus $\epsilon_{ijr} = 0$ for all pairs that do not particpate. We refer to the assumption $w_{ijr} = w_{ij}\epsilon_{ijr}$ on the weights as the *slice-independent* case.

To see the consequences of slice-independence for our equations we define

$$\mathcal{J}_r := \{(i, j) \mid \epsilon_{ijr} = 1\} \tag{11}$$

so that

$$\sigma_r(X, \Delta_1, \cdots, \Delta_s) = \sum_{(i,j) \in \mathcal{J}_r} w_{ij}(\delta_{ijr} - d_{ij}(X))^2 \tag{12}$$

Equation (4) gives $w_{ij}^\star = w_{ij}\epsilon_{ij}^\star$, where $\epsilon_{ij}^\star$ is the number of times pair (i,j) occurs in the R slices. A set of slices is *balanced* if all $\epsilon_{ij}^\star$ are equal. Also, from (10),

$$\delta_{ij}^\star = \frac{\sum_{r=1}^R \epsilon_{ijr}\delta_{ijr}}{\sum_{r=1}^R \epsilon_{ijr}} \tag{13}$$

which does not depend on the w_{ij}.

It follows that in our computations we have to deal with various different sets of weights. There are the w_{ijr}, the $w_{ij}^\star$, the w_{ij}, and the $\epsilon_{ij}^\star$. In the first ALS subproblem where we minimize over the Δ_r for fixed X we use equation (12), i.e. we use the w_{ij}. If we minimize over X for fixed Δ_r we use (9), which means we use $w_{ij}^\star = w_{ij}\epsilon_{ij}^\star$ and $\delta_{ij}^\star$ given by (13). Of course this all simplifies dramatically if $w_{ij} = 1$ for all pairs (i,j) (the *unweighted* case).

3.6 Paired Comparisons

THe paired comparison method of data collection is the simplest and the most basic one of the cartwheel methods.

Positive Orthant / Absolute Value / Pairwise

De Leeuw (1970) De Leeuw (2018) Hartmann (1979) Guttman (1969) Johnson (1973)

Suppose datum r says that that $(i,j) \prec (k,l)$. In the slice independent case $w_{ijr} = w_{ij}$ and $w_{klr} = w_{kl}$ can be non-zero and all other elements of W_r are zero.

Use $w_{(i,j)_r}$ If $(i,j) \prec (k,l)$

$$\sigma_r(X,\Delta_r) = w_{ij}(\delta_{ijr} - d_{ij})^2 + w_{kl}(\delta_{klr} - d_{kl})^2$$

Must be minimized over $\delta_{ijr} \leq \delta_{klr}$. If $d_{ij} \leq d_{kl}$ then $\hat{d}_{ijr} = d_{ij}$ and $\hat{d}_{klr} = d_{kl}$, and otherwise

$$\hat{d}_{ijr} = \hat{d}_{klr} = \frac{w_{ij}d_{ij} + w_{kl}d_{kl}}{w_{ij} + w_{kl}}$$

Thus

$$w_{ij}(\hat{d}_{ijr} - d_{ij})^2 + w_{kl}(\hat{d}_{klr} - d_{kl})^2$$

is zero if the order of d_{ij} and d_{kl} is the same as the order in the data and

$$\frac{w_{ij}w_{kl}}{w_{ij} + w_{kl}}(d_{ij} - d_{kl})^2$$

Also

$$w_{ij}(\hat{d}_{ijr} - d_{ij})^2 + w_{kl}(\hat{d}_{klr} - d_{kl})^2 =$$

$$w_{ij}\hat{d}^2_{ijr} + w_{kl}\hat{d}^2_{klr} + \ -2w_{ij}\hat{d}_{ijr}d_{ij} - 2w_{kl}\hat{d}_{klr}d_{kl} +$$

So far we have only considered the forced-choice situation in which the subject has to choose one of the two pairs. If we allow for the alternative that (i, j) and (k, l) are equally similar then we can choose between two different approaches. In the *primary approach* we incur no loss for this pair, no matter what $d_{ij}(X)$ and $d_{kl}(X)$ are. In the *secondary approach* we require that $\delta_{ijr} = \delta_{klr}$ and consequently we use … and add to the loss if $d_{ij}X) \neq d_{kl}(X)$.

De Leeuw, J. 1970. "The Positive Orthant Method for Nonmetric Multidimensional Scaling." Research Report 001-70. Leiden, The Netherlands: Department of Data Theory FSW/RUL.

———. 1974. *Canonical Analysis of Categorical Data*. Leiden, The Netherlands: Psychological Institute, Leiden University.

———. 2018. "The Positive Orthant Method ." 2018.

Guttman, L. 1969. "Smallest Space Analysis by the Absolute Value Principle." In *Proceedings of the XIX International Congress of Psychology, London*.

Hartmann, W. 1979. *Geometrische Modelle zur Analyse empirischer Data*. Akademie Verlag.

Johnson, R. M. 1973. "Pairwise Nonmetric Multidimensional Scaling." *Psychometrika* 38 (12–18).