Smacof at 50: A Manual
Part 4: Smacof for Pairs of Pairs

Jan de Leeuw - University of California Los Angeles

Started May 19 2024, Version of May 22, 2024

Contents

Note: This is a working paper 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 Introduction

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) (republished as Gttman (1979)) Johnson (1973)

2 Loss Function

We start with a general non-metric MDS problem with loss function

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

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 $\Delta_r \in \mathcal{C}_r$.

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

$$\sigma(X, \Delta_1, \cdots, \Delta_s) = \sum_{r=1}^{R} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ijr}(\delta_{ijr} - \delta_{ij}^\star)^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^\star(\delta_{ij}^\star - d_{ij}(X))^2, \tag{5}$$

where

$$\delta_{ij}^\star = \frac{\sum_{r=1}^{R} w_{ijr}\delta_{ijr}}{\sum_{r=1}^{R} w_{ijr}}. \tag{6}$$

2.1 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

$$\mathcal{J}_r := \{(i, j) \mid \epsilon_{ijr} = 1\} \tag{7}$$

so that

$$\sigma_r(X, \Delta_1, \cdots, \Delta_s) = \sum_{r=1}^{s} \sum_{(i,j)\in\mathcal{J}_r} w_{ij}(\delta_{ijr} - d_{ij}(X))^2 \tag{8}$$

Equation (4) gives $w^\star_{ij} = w_{ij}\epsilon^\star_{ij}$, where $\epsilon^\star_{ij}$ is the number of times pair (i, j) occurs in the R slices. A set of slices is *balanced* if all $\epsilon^\star_{ij}$ are equal. Also, from (6),

$$\delta^\star_{ij} = \frac{\sum_{r=1}^{R} \epsilon_{ijr}\delta_{ijr}}{\sum_{r=1}^{R} \epsilon_{ijr}} \tag{9}$$

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 $w^\star_{ij}$, the w_{ij}, and the $\epsilon^\star_{ij}$. In the first ALS subproblem where we minimize over the Δ_r for fixed X we use equation (8), i.e. we use the w_{ij}. If we minimize over X for fixed Δ_r we use (5), which means we use $w^\star_{ij} = w_{ij}\epsilon^\star_{ij}$ and $\delta^\star_{ij}$ given by (9). Of course this all simplifies dramatically if $w_{ij} = 1$ for all pairs (i, j) (the *unweighted* case).

2.2 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 (1973) (republished as De Leeuw (1984)), 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.

$$\omega(X) := \sum_{(i,j)\prec(k,l)} \sum (d^2_{ij}(X) - d^2_{kl}(X)), \tag{10}$$

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

$$A^\star := \left\{\sum_{(i,j)} \sum_{(k,l)} (A_{ij} - A_{kl})\right\} \tag{11}$$

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 $\theta \geq -\lambda_{\min}(A^\star)$ 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 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 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.

2.3 Pairwise Monotone Regression

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 $d_{ij}X) \neq d_{kl}(X)$.

3 Program

3.1 Parameters

The smacofPC function in R has the following parameters (with default values).

```
smacofPC <- function(data,
                     nobj = max(data),
                     ndim = 2,
                     wmat = NULL,
                     xold = NULL,
                     labels = NULL,
                     width = 15,
                     precision = 10,
                     itmax = 1000,
                     eps = 1e-10,
                     verbose = TRUE,
                     kitmax = 5,
                     keps = 1e-10,
                     kverbose = 0,
                     init = 1,
                     ties = 0)
```

- If xold is non-null then an initial configuration matrix must be provided.
- If labels is non-null then a character vector of plot labels must be provided.
- width and precision are relevant for the format of (optional) major iteration output.
- itmax and eps determine when the major iterations stop.
- If verbose = TRUE itel and stress for each major iteration are written to stdout.
- kitmax and keps determine the number of inner Guttman transform iterations between two monotone regressions.
- If kverbose = TRUE then itel and stress for each inner iteration are written to stdout.
- If init = 1 the maximum sum initial configuration is computed, if init = 2 a random initial configuration is used.
- Ties is either 0, 1, or 2. If ties = 0 the data are forced choice, no ties are allowed. If ties = 1 or ties = 2 the the primary or secondary approach to ties is used.

3.2 Input

The data are either a four column (if ties = 0) or a five column (if ties = 1 or ties = 2) matrix (or data frame). Here is an example of a data matrix with ties = 2 from the hoogeveen example (see below). The first four columns are indices. Row one, for example, tells us that $\delta_{23} \leq \delta_{34}$. Row two has an entry in the fifth column and tells us that $\delta_{23} = \delta_{45}$, and that we are supposed to use the second approach to ties, i.e. require $\hat{d}_{23} = \hat{d}_{45}$.

```
##    V1 V2 V3 V4 V5
## 1   2  3  3  4  0
```

```
## 2   2  3  4  5  2
## 3   2  3  3  5  0
## 4   2  3  2  4  0
## 5   4  5  1  2  0
## 6   2  5  1  2  2
## 7   3  5  4  5  0
## 8   1  3  1  2  0
## 9   1  2  2  4  0
## 10  2  5  2  4  0
## 11  3  5  3  4  0
## 12  1  3  2  3  0
## 13  1  2  2  5  0
## 14  3  5  2  5  0
## 15  1  5  1  4  0
## 16  1  3  3  4  0
## 17  3  5  1  2  0
## 18  1  5  1  4  0
## 19  3  5  2  3  2
## 20  2  5  2  4  0
```

- The data can have replications of some or all comparisons.
- It is not necessary that the data are consistent with any partial order, i.e. there can be intransivities and irreflexivities.
- It is possible to use the secondary approach to ties for some comparisons and the primary approach for others.
- For more information on data generation see section 4 on utilities.

3.3 Algorithm

3.4 Output

```
h <- list(
  nobj = nobj,
  ndim = ndim,
  snew = snew,
  itel = itel,
  xnew = xnew,
  dhat = dhat,
  dmat = dmat,
  wvec = wmat,
  labels = labels
)
```

4 Utilities

4.1 Data Generation and Collection

Let us first address the elephant in the room. Even for moderate n there are a lot of pairs of pairs, and it quickly becomes impossible to present all of them to a subject, even if that subject is paid or is an undergraduate psychology student.

If we only consider distinct pairings of distinct pairs there are already $\binom{\binom{n}{2}}{2}$ pairs of pairs, which is of the order $\frac{1}{4}n^4$. Here is a little table.

```
##     3          3
##     4         15
##     5         45
##     6        105
##     7        210
##     8        378
##     9        630
##    10        990
##    11       1485
##    12       2145
##    13       3003
##    14       4095
##    15       5460
##    16       7140
##    17       9180
##    18      11628
##    19      14535
##    20      17955
```

There are a number of ways to deal with this fundamental problem. If we take the psychophysical point of view that subjects are merely replications then we can use multiple subjects, each handling a subset of the pairs of pairs. Alternatively, we can select a random subset of the total set of pairs of pairs. Or we could use a design to select a preferably balanced subset.

Take all $\binom{n}{2}$ pairs. Random permutation for the first pair, random permutation for the second pair. Repeat m times. Note: this may generate comparisons of (i,j) with (i,j).

4.2 Plots

5 Examples

5.1 Hoogeveen

5.2 Parties

5.3 Ekman (1954)

References

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. https://jansweb.netlify.app/publication/deleeuw-r-70-a/deleeuw-r-70-a.pdf.

———. 1973. "Canonical Analysis of Categorical Data." PhD thesis, University of Leiden, The Netherlands. https://jansweb.netlify.app/publication/deleeuw-b-73/deleeuw-b-73.pdf.

———. 1984. *Canonical Analysis of Categorical Data*. Leiden, The Netherlands: DSWO Press. https://jansweb.netlify.app/publication/deleeuw-b-84/deleeuw-b-84.pdf.

———. 2018. "The Positive Orthant Method ." 2018. https://jansweb.netlify.app/publication/deleeuw-e-18-d/deleeuw-e-18-d.pdf.

Ekman, G. 1954. "Dimensions of Color Vision." *Journal of Psychology* 38: 467–74.

Gttman, L. 1979. "Smallest Space Analysis by the Absolute Value Principle." In *Geometric Representations of Relational Data*, edited by J. C. Lingoes, E. E. Roskam, and I. Borg, 707–12. Mathesis Press.

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).