

Notes on the C Version of Smacof

Jan de Leeuw - University of California Los Angeles

Started October 10 2023, Version of December 02, 2023

Abstract

TBD

Contents

1	Introduction	2
2	Example	2
3	Unweighted, full matrix	2
3.1	Unnormalized	2
3.2	Normalized	2
4	Unweighting	3
5	Thoughts on ALS	3
5.1	The Single-Step approach	4
5.2	The Two-step Approach	4
6	smacofStructure	5
7	Appendix: Code	5
8	smacofMaximumSum	6
9	smacofElegant	6
10	smacofAdjustDiagonal	6
11	smacofImpute	6
12	smacofHildreth	6
13	smacofDykstra	7

14 smacofJacobi	7
15 smacofIndividualDifferenceModels.c	7
16 smacofSort	8
17 Code	8
17.1 smacofSort.R	8
17.2 smacofSort.c	8
References	8

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/del-eeuw/mdsStruct>.

1 Introduction

The loss function in (metric, least squares, Euclidean, symmetric) Multidimensional Scaling (MDS) is

$$\sigma(X) := \frac{1}{2} \sum_{1 \leq j < i \leq n} w_{ij} (\delta_{ij} - d_{ij}(X))^2.$$

This assumes symmetry and it uses all elements below the diagonal of both W , Δ , and $D(X)$. For missing data we set $w_{ij} = 0$.

2 Example

Here is a small input example.

3 Unweighted, full matrix

3.1 Unnormalized

$W = E - I$ and $V = (I - E) + (n - 1)I = nI - E = nJ$. Thus $V^+ = n^{-1}J$ and $V^+B = n^{-1}B$.

3.2 Normalized

After normalization $W = \frac{1}{n(n-1)}(E - I)$ and $\sum \sum w_{ij} \delta_{ij}^2 = 1$ or $\sum \sum \delta_{ij}^2 = n(n-1)$. Also $V = \frac{1}{n(n-1)}(nI - E) = \frac{1}{(n-1)}J$, which means $V^+ = (n-1)J$. For

$$B(X) = \sum_{i < j} w_{ij} \frac{\delta_{ij}}{d_{ij}(X)} A_{ij} = \frac{1}{n(n-1)} \sum_{i < j} \frac{\delta_{ij}}{d_{ij}(X)} A_{ij}$$

and thus

$$V^+ B(X) = \frac{1}{n} \sum_{i < j} \frac{\delta_{ij}}{d_{ij}(X)} A_{ij}$$

4 Unweighting

$$\sigma(d) = \sum_{k=1}^K w_k (\delta_k - d_k)^2 \tag{1}$$

$$= \sum_{k=1}^K w_k (\delta_k - \bar{d}_k)^2 - 2 \sum_{k=1}^K w_k (\delta_k - \bar{d}_k)(d_k - \bar{d}_k) + \sum_{k=1}^K w_k (d_k - \bar{d}_k)^2. \tag{2}$$

Now suppose $w_k \leq w_\star$ and define

$$r_k = \frac{w_k}{w_\star} (\delta_k - \bar{d}_k)$$

$$\begin{aligned} \sigma(d) &\leq \sigma(\bar{d}) + w_\star \left\{ \sum_{k=1}^K (d_k - \bar{d}_k)^2 - \sum_{k=1}^K r_k (d_k - \bar{d}_k) \right\} \\ &= \sigma(\bar{d}) + w_\star \left\{ \sum_{k=1}^K (d_k - (\bar{d}_k + r_k))^2 - \sum_{k=1}^K r_k^2 \right\}. \end{aligned}$$

Note

$$\delta_k(X) := \frac{w_k}{w_\star} \delta_k + \left(1 - \frac{w_k}{w_\star}\right) d_k(X)$$

So given X computed the adjusted dissimilarities $\Delta(X)$ and perform one or more unweighted smacof steps to compute the update X^+ . Of course if all w_k are the same then $\Delta(X) = \Delta$ and we compute a regular unweighted smacof. It is of some interest to study how many smacof steps to make before computing a new $\Delta(X)$.

5 Thoughts on ALS

In nonmetric MDS we minimize

$$\sigma(\Delta, X) = \sum_{k=1}^K w_k (\delta_k - d_k(X))^2$$

over the configurations X and the transformed dissimilarities Δ , where we assume $\Delta \in \mathfrak{D}$, with \mathfrak{D} the intersection of a convex cone and a sphere.

5.1 The Single-Step approach

Kruskal (1964a), Kruskal (1964b) defines

$$\sigma_*(X) := \min_{\Delta \in \mathfrak{D}} \sigma(\Delta, X)$$

and

$$\Delta(X) = \operatorname{argmin}_{\Delta \in \mathfrak{D}} \sigma(\Delta, X)$$

Thus

$$\sigma_*(X) = \sigma(\Delta(X), X)$$

which is now a function of X only. Under some conditions, which are usually true in MDS,

$$\mathcal{D}\sigma_*(X) = \mathcal{D}_2\sigma(\Delta(X), X)$$

where $\mathcal{D}_2\sigma(\Delta(X), X)$ are the partials of σ with respect to X . Thus the partials of σ_* can be computed by evaluating the partials of σ with respect to X at $(X, \Delta(X))$. This has created much confusion in the past. We can now solve the problem of minimizing σ_* , which is a function of X alone.

I think Guttman calls this the *single step approach*. A variation of Kruskal's single-step approach defines

$$\sigma_G(X) = \sum_{k=1}^K w_k (\delta_k^\#(X) - d_k(X))^2$$

where the $\delta_k^\#(X)$ are *Guttman's rank images*, i.e. the permutation of the $d_k(X)$ that makes it monotone with the δ_k (Guttman (1968)). Or, alternatively, we define

$$\sigma_S(X) := \sum_{k=1}^K w_k (\delta_k^\circ(X) - d_k(X))^2$$

where the $\delta_k^\circ(X)$ are *Shepard's rank images*, i.e. the permutation of the δ_k that makes it monotone with the $d_k(X)$ (Shepard (1962a), Shepard (1962b)).

The Shepard and Guttman alternatives are computationally more intricate and more complicated than the Kruskal *monotone regression* approach, mostly because of problems with uniqueness and differentiation, but they are obviously both single step approaches.

5.2 The Two-step Approach

The *two-step approach* or *alternating least squares* approach alternates minimization of $\sigma(\Delta, X)$ over X for our current best estimate of Δ with minimization of $\sigma(\Delta, X)$ over $\Delta \in \mathfrak{D}$ for our current best value of X . Thus an update looks like

$$\Delta^{(k)} = \operatorname{argmin}_{\Delta} \sigma(\Delta, X^{(k)}), \tag{3}$$

$$X^{(k+1)} = \operatorname{argmin}_X \sigma(\Delta^{(k)}, X). \tag{4}$$

This approach to MDS was in the air since the early (unsuccessful) attempts around 1968 of Young and De Leeuw to combine Torgerson’s classic metric MDS method with Kruskal’s monotone regression transformation.

As formulated, however, there are some problems with the ALS algorithm. Step (??) is easy to carry out, using monotone regression. Step (??) means solving a metric scaling problem, which is an iterative process that requires an infinite number of iterations. Thus, what is usually implemented, is to combine step (??) with one of more iterations of a convergent iterative procedure for metric MDS, such as *smacof*. If we take only one of these *inner iterations* the algorithm becomes indistinguishable from Kruskal’s single step method. This has also created much confusion in the past.

It is somewhat worrisome that in the ALS approach we solve the first subproblem (??) exactly, while we take only a single step towards the solution for given Δ in the second subproblem (??). If we have an infinite iterative procedure to compute the optimal $\Delta \in \mathfrak{D}$ for given X , then a more balanced approach is to take several inner iterations in the first step and several inner iterations in the second step. How many of each, nobody knows.

6 *smacofStructure*

We use this example as input for the R version of *smacofSort()*, which results in the *smacofStructure*

The *dist* element in the data frame is zero, because we have not computed distances yet. Given a configuration X the *row* and *col* elements in the data frame allow for straightforward computation of distances. In the case of nonmetric MDS, or more generally in MDS with transformed dissimilarities, the *dhat* column will be filled as well. The fact that preprocessing with *smacofSort()* gives the ordered dissimilarities, as well as the tie blocks, is especially useful in the ordinal case, both with monotone polynomials and monotone splines. In the metric (ratio) case there is no need for the *dhat* column.

It is obvious how this *smacofStructure* can be adapted if there are multiway data. If we have row-conditional or matrix-conditional data, then we use one of these *smacofStructures* for each row or each matrix.

7 Appendix: Code

We use *qsort* to sort the rows of the input data frame by increasing delta. The sorting is done in C, the R version is a wrapper around the compiled C code. The C code also contains a main which analyzes the same small example as we have used in the text. Compile with “clang -o runner -O2 smacofSort.c” and then start “runner” in the shell. The C code uses the *.C()* interface in R, which can be improved using *.Call()*, but probably in this case with little gain.

8 smacofMaximumSum

Maximize

$$\sum_{1 \leq j < i \leq n} \sum w_{ij} \delta_{ij}^2 d_{ij}^2(X)$$

over X with $\text{tr}(X'X)^2 = 1$. This gives $BX = X\Lambda$, with

$$B = \sum_{1 \leq j < i \leq n} \sum w_{ij} \delta_{ij}^2 A_{ij}$$

There is also a non-metric version. Maximize over $\text{tr}(X'X)^2$

$$\sum_{1 \leq j < i \leq n} \sum_{1 \leq k < l \leq n} w_{ij,kl} \text{sign}(\delta_{ij} - \delta_{kl}) (d_{ij}^2(X) - d_{kl}^2(X))$$

which simplifies to

$$2 \sum_{1 \leq j < i \leq n} \sum d_{ij}^2(X) \sum_{1 \leq k < l \leq n} w_{ij,kl} \text{sign}(\delta_{ij} - \delta_{kl})$$

Simplifies more $w_{ij,kl} = w_{ij}$ or $w_{ij,kl} = w_{kl}$

Also note

$$\rho(X) = \sum_{1 \leq j < i \leq n} \sum w_{ij} \delta_{ij} d_{ij}(X) = \sum_{1 \leq j < i \leq n} \sum \frac{w_{ij}}{\delta_{ij} d_{ij}(X)} \delta_{ij}^2 d_{ij}^2(X) \approx \sum_{1 \leq j < i \leq n} \sum \frac{w_{ij}}{\delta_{ij}^2} \delta_{ij}^2 d_{ij}^2(X) = \eta^2(X)$$

9 smacofElegant

10 smacofAdjustDiagonal

11 smacofImpute

12 smacofHildreth

Hildreth (1957)

Consider the QP problem of minimizing $f(x) = \frac{1}{2}(x - y)'W(x - y)$ over all $x \in \mathbb{R}^n$ satisfying $Ax \geq 0$, where A is $m \times n$. Wlg we can assume $a_j'Wa_j = 1$. The Lagrangian is

$$\mathcal{L}(x, \lambda) = \frac{1}{2}(x - y)'W(x - y) - \lambda'Ax.$$

$$\max_{\lambda \geq 0} \mathcal{L}(x, \lambda) = \begin{cases} \frac{1}{2}(x - y)'W(x - y) & \text{if } Ax \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

and thus

$$\min_{x \in \mathbb{R}^n} \max_{\lambda \geq 0} \mathcal{L}(x, \lambda) = \min_{Ax \geq 0} \frac{1}{2}(x - y)'W(x - y).$$

By duality

$$\min_{x \in \mathbb{R}^n} \max_{\lambda \geq 0} \mathcal{L}(x, \lambda) = \max_{\lambda \geq 0} \min_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda).$$

The inner minimum over x is attained for

$$x = y + W^{-1} A' \lambda,$$

and is equal to

$$\min_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda) = -\frac{1}{2} \lambda' A W^{-1} A' \lambda - \lambda' A y.$$

Thus

$$\max_{\lambda \geq 0} \min_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda) = -\frac{1}{2} \min_{\lambda \geq 0} \{ (W y + A' \lambda)' W^{-1} (W y + A' \lambda) - y' W y \}$$

We minimize $h(\lambda) = (W y + A' \lambda)' W^{-1} (W y + A' \lambda)$ with coordinate descent. Let $\lambda_j(\epsilon) = \lambda + \epsilon e_j$. Then

$$h(\lambda_j(\epsilon)) = (W y + A' \lambda + \epsilon a_j)' W^{-1} (W y + A' \lambda + \epsilon a_j) = \epsilon^2 a_j' W^{-1} a_j + 2\epsilon a_j' W^{-1} (y + W^{-1} A' \lambda) +$$

which must be minimized over $\epsilon \geq -\lambda_j$. So the minimum is attained at

$$\epsilon = -\frac{a_j' W^{-1} x}{a_j' W^{-1} a_j}$$

with $x = y + W^{-1} A' \lambda$ (cf ...), provided ... satisfies .. Otherwise $\epsilon = -\lambda_j$. Now update both λ and x , and go to the next j .

13 smacofDykstra

14 smacofJacobi

taken from De Leeuw (2017)

partial jacobi

15 smacofIndividualDifferenceModels.c

$$X_k = X, \tag{5}$$

$$X_k = X \Lambda_k \text{ with } \Lambda_k \text{ diagonal}, \tag{6}$$

$$X_k = X C_k, \tag{7}$$

$$X_k = X \Lambda_k Y' \text{ with } \Lambda_k \text{ diagonal}, \tag{8}$$

$$X_k = X C_k Y', \tag{9}$$

$$X_k = X \Lambda_k Y'_k \text{ with } \Lambda_k \text{ diagonal} \tag{10}$$

$$X_k = X C_k Y' \text{ with } C_k = \sum_{s=1}^r z_{ks} H_s. \tag{11}$$

16 smacofSort

17 Code

17.1 smacofSort.R

17.2 smacofSort.c

References

- De Leeuw, J. 2017. “Jacobi Eigen in R/C with Lower Triangular Column-wise Compact Storage.” 2017. <https://jansweb.netlify.app/publication/deleeuw-e-17-o/deleeuw-e-17-o.pdf>.
- Guttman, L. 1968. “A General Nonmetric Technique for Fitting the Smallest Coordinate Space for a Configuration of Points.” *Psychometrika* 33: 469–506.
- Hildreth, C. 1957. “A Quadratic Programming Procedure.” *Naval Research Logistic Quarterly* 14 (79–85).
- Kruskal, J. B. 1964a. “Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis.” *Psychometrika* 29: 1–27.
- . 1964b. “Nonmetric Multidimensional Scaling: a Numerical Method.” *Psychometrika* 29: 115–29.
- Shepard, R. N. 1962a. “The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function. I.” *Psychometrika* 27: 125–40.
- . 1962b. “The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function. II.” *Psychometrika* 27: 219–46.