Smacof at 50: A Manual
Part 5: Unfolding in Smacof

Jan de Leeuw - University of California Los Angeles

Started December 12 2022, Version of December 22, 2024

Abstract

TBD

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 repositories smacofCode, smacofManual, and smacofExamples.

1 Introduction

In Multidimensional Unfolding (MDU) the objects of an MDS problem are partitioned into two sets. There is a set of n row-objects and a set of m column-objects, and a corresponding $n \times p$ row-configuration X and $m \times p$ column-configuration Y. We minimize stress defined as

$$\sigma(X, Y) := \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(\delta_{ij} - d(x_i, y_j)^2. \tag{1}$$

over both X and Y. Here

$$d(x_i, y_j) := \sqrt{(x_i - y_j)'(x_i - y_j)} \tag{2}$$

Thus the within-set dissimilarities are missing, or ignored even if they are available, and only the between-set dissimilarities are fitted by between-set distances.

If we define Z as

$$Z := \tag{3}$$

and U as

$$U := \tag{4}$$

then we can also write

$$\sigma(Z) = \sum_{i=1}^{n+m} \sum_{j=1}^{n+m} u_{ij}(\delta_{ij} - d_{ij}(Z))^2. \tag{5}$$

Data Preferences Type A and Type B Conditional

The unfolding model for preference judgments is often attributed to Clyde H. Coombs (1950), with further developments by Coombs and his co-workers reviewed in C. H. Coombs (1964). After this path-breaking work the digital computer took over, and minimization of loss function (1) and its variations was started by Roskam (1968) and Kruskal and Carroll (1969).

In this manual we are not interested in MDU as a psychological theory, as a model for preference judgments. We merely are interested in mapping off-diagonal dissimilarity relations into low-dimensional Euclidean space, i.e. in making a picture of the data. In some cases (distance completion, distances with errors, spatial basis)

Sixty years ago Joseph B. Kruskal published his basic non-metric multidimensional scaling papers (Kruskal (1964a), Kruskal (1964b)). They provided the necessary tools to analyze what later became known as symmetric one-mode data, i.e. symmetric dissimilarities between n objects. It did not take long before Roskam (1968) and Kruskal and Carroll (1969) generalized to rectangular two-mode data, i.e. to non-metric multidimensional unfolding. The canonical example of such data is a number of individuals ranking their preferences (interpreted as similarities) for a number of objects. The ordinal information in unfolding data is much smaller than that in symmetric one mode data. Not only are there no dissimilarities between the individuals and between the objects, but in addition the rankings are conditional, which means that preferences can only be compared within individuals. It is safe to say that as a consequence of this paucity of information the unmodified Kruskal approach to non-metric unfolding did not work.

If one applies a Kruskal-type non-metric scaling program to preference rank orders, then one invariable finds what has become known as a trivial solution. To understand this, let us consider what non-metric multidimensional unfolding want to accomplish. Suppose there are n individuals, and each individual i provides us with a partial order $\prec_i$ on the m objects. We want to represent the individuals as n points x_i and the objects as m points y_j in a low-dimensional Euclidean space, in such a way that $j \prec_i \ell$ in the data corresponds with $d(x_i, y_j) < d(x_i, y_\ell)$ in the representation. Here $d(x, y)$ is Euclidean distance.

The Kruskal approach quantifies this objective using a least squares loss function called stress. The stress for individual i is

$$\sigma_i(X, Y) := \min_{\delta \in K_i} \sum_{j=1}^{m} w_{ij}(\delta_j - d(x_i, y_j))^2.$$

Here K_i is the set of all vectors δ_i with m elements that satisfy $\delta_{ij} \leq \delta_{i\ell}$ for all (i, j, ℓ) for which $j \prec_i \ell$. The total stress $\sigma(X, Y)$ is the sum of the $\sigma_i(X, Y)$ stresses over the n individuals.

Thus we do not require directly that the distances satify the ordinal constraints, we require that the distances are numerically as close as possible to a vector δ that does satisfy the ordinal constraints. Note, however, that the constraints on δ use $\leq$, *less than or equal to*, instead if $<$, *less than*, which means we are OK with $\delta_{ij} = \delta_{i\ell}$ for some, or even all, $j \prec_i \ell$. And that is where the trivial solutions come in.

The most obvious trivial solution collapses all x_i and all y_j into a single point. This makes all distances zero, and thus $\sigma(X, Y) = 0$. We have achieved our objective, which was minimizing stress, but the solution is completely independent of the data. A slightly more elaborate, but equally trivial, solution puts all x_i in one point and all y_j in another. All distances are non-zero, but equal. Even more elaborately, we can put all x_i in the origin, and all y_j on a circle with the origin as center.

Ever since Roskam (1968) and Kruskal and Carroll (1969) trivial solutions have been acknowledged and ways to avoid them in non-metric multidimensional unfolding have been proposed, since 1980 mostly by Willem Heiser and his students. There are excellent discussions of these proposed solutions in the recent dissertations of Van Deun (2005) and Busing (2010). None of them, however, solves the version of the non-metric multidimensional unfolding problem where we require that $d(x_i, y_j) \leq d(x_i, y_\ell)$ if $j \prec_i \ell$. In that version we do not need a computer to find a perfect solution of the relevant inequalities, we just choose one of the trivial solutions.

In this paper we study one early attempt to salvage the Kruskal-Roskam approach, using row-wise normalization of stress. The trivial solutions have in common that the $d(x_i, y_j) = d(x_i, y_\ell)$ for all (i, j, ℓ), and we can prevent them from happening by redefining stress as

$$\sigma(X, Y) = \sum_{i=1}^{n} \frac{\min_{\delta \in K_i} \sum_{j=1}^{m} w_{ij}(\delta_j - d(x_i, y_j))^2}{\min_{\delta \in L} \sum_{j=1}^{m} w_{ij}(\delta_j - d(x_i, y_j))^2},$$

{#eq-rosstress} where L is the set of all vectors for which all m elements are the same. The optimal δ_j in denominator i are all equal to the weighted mean of the $d(x_i, y_j)$. Thus, at a trivial solution, both nominators and denominators are zero for all i, and since $0/0$ is undefined, stress is undefined at trivial solutions. Thus the algorithm cannot converge to a trivial solution.

In practice, however, the Kruskal-Roskam approach does not work well. All to often solutions still look like trivial solutions, or are partly trivial. An early attempt to explain why there are these failures was De Leeuw (1983) (reissued as De Leeuw (2006)). That paper studies how the loss function from (**eq-rosstress?**) behaves near trivial solutions. Since the 1983 paper is somewhat tentative, we present a more complete, more general, and more rigorous version in this paper. We look again at the idea that since we do not like $0/0$ the iterative algorithm must not like it too. # Loss function

1.1 Metric

1.2 Non-linear

1.3 Non-metric

1.4 Constraints

2 smacofUF

2.1 Initial Configuration

$$\sigma(C)=\sigma(\tilde{C}+(C-\tilde{C}))=\sum_{i=1}^n\sum_{j=1}^m((\delta_{ij}^2-\text{tr }A_{ij}\tilde{C})-\text{tr }A_{ij}(C-\tilde{C}))^2$$

$$\sigma(C)=\sigma(\tilde{C})-2\sum_{i=1}^n\sum_{j=1}^m(\delta_{ij}^2-\text{tr }A_{ij}\tilde{C})\text{tr }A_{ij}(C-\tilde{C})+\sum_{i=1}^n\sum_{j=1}^m\{\text{tr }A_{ij}(C-\tilde{C})\}^2$$

From De Leeuw, Groenen, and Pietersz (2006)

$$\sum_{i=1}^n\sum_{j=1}^m\{\text{tr }A_{ij}(C-\tilde{C})\}^2\leq(n+m+2)\text{tr }(C-\tilde{C})^2$$

Define

$$B(\tilde{C}):=\frac{1}{n+m+2}\sum_{i=1}^n\sum_{j=1}^m(\delta_{ij}^2-\text{tr }A_{ij}\tilde{C})A_{ij}$$

So we minimize

$$-2\text{ tr }B(\tilde{C})C+\text{tr }C^2-2\text{tr }C\tilde{C}=\text{tr }(C-\{\tilde{C}+B(\tilde{C})\})^2$$

2.2 Constraints

Basic smacof theory ((**deleeuw_heiser_C_81?**)) tells us that having constraints on X and Y

Let

$$\overline{Z}:= \tag{6}$$

be the Guttman transform of $Z^{(k)}$. In constrained smacof we have to minimize, in each iteration,

$$\text{tr }(X-\overline{X})'V_{11}(X-\overline{X})+2\text{ tr }(X-\overline{X})'V_{12}(Y-\overline{Y})+\text{tr }(Y-\overline{Y})'V_{22}(Y-\overline{Y})$$

over X and Y satisfying the constraints. That defines $Z^{(k+1)}$. As in other block relaxation cases it is not necessary to actually minimize …, it suffices to decrease it in some systematic way.

If there are no constraints we have $Z^{(k+1)}=\Gamma(Z^{(k)})$.

3 Normalization Restrictions

$X'V_{11}X=I$ or tr $X'V_{11}X=1$.

3.0.1 Centroid Restriction

$$Z=\begin{bmatrix}X\\Y\end{bmatrix}=\begin{bmatrix}I\\D^{-1}G'\end{bmatrix}X=HX$$

Minimize tr $(\overline{Z}-HX)'V(\overline{Z}-HX)$. If there are no further restrictions on X the minimum is attained at $\hat{X}=(H'VH)^{+}H'V\overline{Z}$. Otherwise write $X=\hat{X}+(X-\hat{X})$. Then

$$\text{tr }(\overline{Z}-H\hat{X}-H(X-\hat{X}))'V(\overline{Z}-H\hat{X}-H(X-\hat{X}))=\text{tr }(\overline{Z}-H\hat{X})'V(\overline{Z}-H\hat{X})+\text{tr }(X-\hat{X})'H'VH(X-\hat{X})$$

and we must minimize tr $(X-\hat{X})'H'VH(X-\hat{X})$ for example over $X'H'VHX=I$. That is maximizing $H'VH\hat{X}=H'VHXM$ with M a symmetric matrix of Lagrange multipliers. Thus $M^2=\hat{X}'H'VH\hat{X}$ and $X=\hat{X}(\hat{X}'H'VH\hat{X})^{-\frac{1}{2}}$.

Over tr $X'H'VHX=1$ we get $H'VH(X-\hat{X})=\lambda H'VHX$ or $X=(1-\lambda)H'VHX=H'VH\hat{X}$, wich means X is proportional to $\hat{X}$ and we just have to normalize $\hat{X}$ to find X.

The constraint $X'V_{11}X=I$ is more difficult to deal with

$$\text{tr }(X-\hat{X})'H'VH(X-\hat{X})=$$

Oblique Procrustus

3.0.2 Linearly Restricted Unfolding

3.0.2.1 External Unfolding

$$\begin{bmatrix}X\\Y\end{bmatrix}=\begin{bmatrix}R&0\\0&S\end{bmatrix}\begin{bmatrix}P\\Q\end{bmatrix}$$

Normalization Restrictions

$X'V_{11}X=I$.

3.0.3 Rank-one Restriction

$Y = za'$

Minimize

$$2 \text{ tr } (za' - \overline{Y})' V_{21}(X - \overline{X}) + \text{tr } (za' - \overline{Y})' V_{22}(za' - \overline{Y})$$

Removing irrelevant terms

$$2\, z'\{V_{21}(X - \overline{X}) - V_{22}\overline{Y}\}a + a'a.z'V_{22}z$$

Let $H = -V_{21}(X - \overline{X}) - V_{22}\overline{Y}$. Minimize using $z'V_{22}z = 1$.

$$a = H'z = \{\overline{Y}' V_{22} - (X - \overline{X})' V_{12})\}z$$

Only You

Each row object coincides with the first choice column object

4 Examples

4.1 Roskam

4.2 Breakfast

4.3 Gold

4.4 Indicator matrix / Matrices

References

Busing, Frank M. T. A. 2010. "Advances in Multidimensional Unfolding." PhD thesis, Leiden University.

Coombs, C. H. 1964. *A Theory of Data*. Wiley.

Coombs, Clyde H. 1950. "Psychological Scaling Without a Unit of Measurement." *Psychological Review* 57: 148–58.

De Leeuw, J. 1983. "On Degenerate Multidimensional Unfolding Solutions." Research Report. Leiden, The Netherlands: Department of Data Theory FSW/RUL.

———. 2006. "On Degenerate Nonmetric Unfolding Solutions." Preprint Series 507. Los Angeles, CA: UCLA Department of Statistics. https://jansweb.netlify.app/publication/deleeuw-r-06-a/deleeuw-r-06-a.pdf.

De Leeuw, J., P. J. F. Groenen, and R. Pietersz. 2006. "Optimizing Functions of Squared Distances." UCLA Department of Statistics. https://jansweb.netlify.app/publication/deleeuw-groenen-pietersz-u-06/deleeuw-groenen-pietersz-u-06.pdf.

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.

Kruskal, J. B., and J. D. Carroll. 1969. “Geometrical Models and Badness of Fit Functions.” In *Multivariate Analysis, Volume II*, edited by P. R. Krishnaiah, 639–71. North Holland Publishing Company.

Roskam, E. E. 1968. “Metric Analysis of Ordinal Data in Psychology.” PhD thesis, University of Leiden.

Van Deun, K. 2005. “Degeneracies in Multidimensional Scaling.” PhD thesis, KU Leuven.