Singularities and Zero Distances in Multidimensional Scaling

Jan de Leeuw, Patrick Groenen, Patrick Mair

Version 01, Februari 15, 2016

Contents

Note: This is a working paper which will be expanded/updated frequently. The directory deleeuwpdx.net/pubfolders/zeroes has a pdf copy of this article, the complete Rmd file with all code chunks, the bib file, and the R source code.

1 Problem

Multidimensional scaling (MDS) in this paper is minimization of the loss function *stress*, introduced by Kruskal (1964a), Kruskal (1964b). Stress is defined as

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

and must be minimized over all $n \times p$ *configurations* X. Here $W = \{w_{ij}\}$ and $\Delta = \{\delta_{ij}\}$ are matrices of, respectively, *weights* and *dissimilarities*. Both W and Δ are symmetric, non-negative, and hollow (zero diagonal). The matrix $D(X) = \{d_{ij}(X)\}$ has *Euclidean distances* between the rows of X, defined as

$$d_{ij}^2(X) := (x_i - x_j)'(x_i - x_j) = \mathbf{tr}\ X'A_{ij}X, \tag{2}$$

where, using unit vectors e_i and e_j,

$$A_{ij} := (e_i - e_j)(e_i - e_j)'. \tag{3}$$

Thus dissimilarities between objects are represented as points in low-dimensional Euclidean space.

Two matrices, introduced in MDS by De Leeuw (1977), play an important role in this article. They are both symmetric, doubly-centered, and diagonally dominant. The first matrix V is

$$v_{ij} := -w_{ij},$$
$$v_{ii} := -\sum_{1 \leq j \neq i \leq n} v_{ij},$$

and second matrix, more precisely the matrix valued function, $B(X)$ is

$$b_{ij}(X) := \begin{cases} -w_{ij}\frac{\delta_{ij}}{d_{ij}(X)} & \text{if } d_{ij}(X) > 0, \\ 0 & \text{if } d_{ij}(X) = 0, \end{cases}$$
$$b_{ii}(X) := -\sum_{1 \leq j \neq i \leq n} b_{ij}(X)$$

An important result on zero distances at a local minimum was proved by De Leeuw (1984). It is based on a formula for the Dini directional derivative (see also De Leeuw, Groenen, and Mair (2016)).

Theorem 1: [Directional Derivative]

$$\lim_{\epsilon \downarrow 0} \frac{\sigma(X + \epsilon Y) - \sigma(X)}{\epsilon} = \mathbf{tr}\ Y'(V - B(X))X - \sum\{w_{ij}\delta_{ij}d_{ij}(Y) \mid d_{ij}(X) = 0\}$$

Proof: Use

$$d_{ij}(X + \epsilon Y) = d_{ij}(X) + \epsilon \begin{cases} \frac{1}{d_{ij}(X)}\mathbf{tr}\ X'A_{ij}Y & \text{if } d_{ij}(X) > 0, \\ d_{ij}(Y) & \text{if } d_{ij}(X) = 0. \end{cases}$$

to get the required result. **QED**

Theorem 2: [Necessary Condition] If σ has a local minimum at X then $(V - B(X))X = 0$ and $d_{ij}(X) > 0$ for all $i \neq j$ such that $w_{ij}\delta_{ij} > 0$.

Proof: Direct from theorem 1. **QED**

If σ has a local minimum at X and is differentiable at X, then the derivative vanishes. In the usual MDS problems we have $w_{ij}\delta_{ij} > 0$ for all $i \neq j$, and theorem 2 says that stress is differentiable at local minima.

Points where the derivative exists and vanishes are stationary points, which are not necessarily local minima. Because of the possibility of zero distances we need a more general notion of a stationary point, which we now define as points where $0 \in \partial\sigma(X)$, with $\partial\sigma(X)$ the

2

subdifferential of σ at X. We know (De Leeuw (1977)) that σ is the difference of a convex quadratic $\eta := \mathbf{tr}\ X'VX$ and a convex function $\rho(X) : +\mathbf{tr}\ X'B(X)X$, and thus the subdifferential is simply a translation of the convex subdifferential, or $\partial\sigma(X) = 2(VX - \partial\rho(X))$.

Theorem 3: **[Stationary Points]** We have $0 \in \partial\sigma(X)$ if and only if for all $i < j$ with $d_{ij}(X) = 0$ there exist u_{ij} with $u'_{ij}u_{ij} \leq 1$ such that

$$(V - B(X))X = \sum\sum_{1 \leq i < j \leq n} \{w_{ij}\delta_{ij}(e_i - e_j)u'_{ij} \mid d_{ij}(X) = 0\}$$

Proof: This follows from the formula for the subdifferential of the distance function

$$\partial d_{ij}(X) = \begin{cases} \left\{\frac{1}{d_{ij}(X)}A_{ij}X\right\} & \text{if } d_{ij}(X) > 0, \\ \{(e_i - e_j)u' \mid u'u \leq 1\} & \text{if } d_{ij}(X) = 0. \end{cases}$$

QED

Note that theorem 3 implies that always $(V - B(X))X \in \partial\sigma(X)$, whether there are zero distances or not.

2   Zero Distances

Theorem 2 tells us that at local minima $w_{ij}\delta_{ij} = 0$ if $d_{ij}(X) = 0$. But this leaves open some interesting questions about stationary points and local minima that are not covered by this result.

Can there be local minima with zero distances ? If we allow for zero dissimilarities the answer is clearly yes. Take Δ equal to $D(X)$ where the configuration X has at least one zero distance. Then X gives the global minimum of stress, equal to zero. And a global minimum is, of course, also a local minimum. If we allow for zero weights, then the answer is again yes, even if all dissimilarties are positive. Suppose we have four objects with all dissimilarities 1. We are fitting two-dimensional configurations. If all weights are one, the global minimum is attained for four points in the corners of a square. If we set $w_{12} = 0$, however, we get perfect fit by putting points in the corners of an equilateral triangle, with objects 1 and 2 together in one corner. Since this makes stress zero, it is the global minimum. If we set, in addition, $w_{34} = 0$ then we get the global minimum by having two points on the line, one with objects 1 and 2, and one with objects 3 and 4.

A related question is if there can there be solutions to $(V - B(X))X$ with zero distances ? We answer this more systematically. Suppose X satisfies the stationary equations $(V - B(X))X = 0$ and X is of block form, with all rows within a block the same, and rows in different blocks different. If there are m blocks, then we can write $X = G\overline{X}$, where G is an $n \times m$ binary indicator matrix, indicating block membership. Suppose I_j are the indices of the objects in block j.

It follows that $(V - B(X))G\overline{X} = 0$ and thus $(\overline{V} - \overline{B}(X))\overline{X} = 0$, where $\overline{V} := G'VG$ and $\overline{B}(X) := G'B(X)X$. We can write, somewhat suggestively, $\overline{B}(X) = B(\overline{X})$. Use the fact that all between-set distances in $B_{j\ell}(X)$ are equal to $d_{j\ell}(\overline{X})$. Thus if, for $j \neq \ell$

$$\overline{w}_{j\ell} := -\sum_{i \in I_j} \sum_{k \in I_\ell} w_{ik}, \tag{4}$$

$$\overline{b}_{j\ell}(X) := -\frac{1}{d_{j\ell}(\overline{X})} \sum_{i \in I_j} \sum_{k \in I_\ell} w_{ik}\delta_{ik}, \tag{5}$$

and the diagonal elements are filled in the make the matrices doubly-centered.

So the stationary equations for an MDS problem with zero distances have a *reduced* or *clustered* set of stationary equations, of the order of the number of *clusters* of points, which do not involve zero distance. This reduced MDS problem can also be interpreted as the original MDS problem, but with the constraints that there is some fixed clustering of points.

Conversely, if we have stationary equations for an MDS problem with nono-zero weights and distances, then we can always expand them to stationary equations for a problem with zero distances, simply by finding W and Δ that satisfy (4) and (5).

3 Singularities

Suppose $(V - B(X))X = 0$ and the $n \times p$ matrix X has rank $r < p$. Define $\overline{X} = K\Lambda$, using the left singular vectors K corresponding to the r non-zero singular values in Λ. Then $D(X) = D(\overline{X})$ and consequently $(V - B(\overline{X}))\overline{X} = 0$. Thus we can reduced the problem to a non-singular one.

Conversely, if $(V - B(X))X = 0$ we can define $\overline{X} = (X \mid 0)L$, with L square orthonormal. Again $D(X) = D(\overline{X})$ and thus again $(V - B(\overline{X}))\overline{X} = 0$. Thus we can expand the problem to a singular one.

4 NEWS

001

- First upload

References

De Leeuw, J. 1977. "Applications of Convex Analysis to Multidimensional Scaling." In *Recent Developments in Statistics*, edited by J.R. Barra, F. Brodeau, G. Romier, and B.

Van Cutsem, 133–45. Amsterdam, The Netherlands: North Holland Publishing Company. http://deleeuwpdx.net/janspubs/1977/chapters/deleeuw_C_77.pdf.

———. 1984. "Differentiability of Kruskal's Stress at a Local Minimum." *Psychometrika* 49: 111–13. http://deleeuwpdx.net/janspubs/1984/articles/deleeuw_A_84f.pdf.

De Leeuw, J., P. Groenen, and P. Mair. 2016. "Differentiability of rStress at a Local Minimum." doi:10.13140/RG.2.1.1249.8968.

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.