

Differentiability of Stress at Local Minima

Jan de Leeuw

Version May 24, 2018

Abstract

Earlier papers have shown that stress is differentiable at local minima if certain conditions on the weights and dissimilarities are satisfied. In this note we show the result remains true without these additional conditions.

Contents

1	Introduction	1
2	Differentiation	2
3	Final Result	3
	References	3

Note: This is a working paper which will be expanded/updated frequently. All suggestions for improvement are welcome. The directory gifi.stat.ucla.edu/zero has a pdf version, the bib file, and the complete Rmd file.

1 Introduction

The multidimensional scaling *stress* loss function is defined as

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

where W and Δ are non-negative symmetric and hollow matrixes of *weights* and *dissimilarities*, where X is the $n \times p$ *configuration*, and where $d_{ij}(X)$ is the Euclidean distance between rows i and j of X . Thus

$$d_{ij}^2(X) = (e_i - e_j)' X X' (e_i - e_j) = \mathbf{tr} X' A_{ij} X,$$

where e_i and e_j are unit vectors (columns of the identity matrix), and

$$A_{ij} \triangleq (e_i - e_j)(e_i - e_j)'$$

Multidimensional scaling is minimization of stress over configurations.

2 Differentiation

The directional derivative of σ at X in direction Y is defined as

$$D\sigma(X, Y) \triangleq \lim_{\epsilon \downarrow 0} \frac{\sigma(X + \epsilon Y) - \sigma(X)}{\epsilon}.$$

Stress is not differentiable at configurations for which some $d_{ij}(X)$ are zero, but it has a finite directional derivatives everywhere. We will show this by actually giving the formula, which was first given by De Leeuw (1984). See also De Leeuw, Groenen, and Mair (2016). The directional derivative is interesting because clearly a necessary condition for σ to have a local minimum at X is that $D\sigma(X, Y) \geq 0$ for all Y .

In order to derive a convenient expression for $D\sigma(X, Y)$ we give some definitions. First some indicators for zero distances.

$$\alpha_{ij}(X) \triangleq \begin{cases} 1 & \text{if } d_{ij}(X) = 0, \\ 0 & \text{if } d_{ij}(X) > 0. \end{cases}$$

$$\beta_{ij}(X) \triangleq \begin{cases} 0 & \text{if } d_{ij}(X) = 0, \\ \frac{1}{d_{ij}(X)} & \text{if } d_{ij}(X) > 0. \end{cases}$$

Then some matrices.

$$V \triangleq \sum_{1 \leq i < j \leq n} w_{ij} A_{ij},$$

$$B(X) \triangleq \sum_{1 \leq i < j \leq n} w_{ij} \delta_{ij} \beta_{ij}(X) A_{ij}$$

And finally

$$\theta(X, Y) \triangleq \sum_{1 \leq i < j \leq n} w_{ij} \delta_{ij} \alpha_{ij}(X) d_{ij}(Y).$$

Theorem 1: $D\sigma(X, Y) = \mathbf{tr} Y'(V - B(X))X - \theta(X, Y)$

Proof: We have

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

The rest is simple computation. **QED**

Theorem 2: If σ has a local minimum at X then $B(X)X = VX$ and $\theta(X, Y) = 0$ for all Y .

Proof: Suppose $(V - B(X))X \neq 0$. Then we can find Y such that $\mathbf{tr} Y'(V - B(X))X < 0$, and because $\theta(X, Y) \geq 0$ we have $D\sigma(X, Y) < 0$. Suppose $\theta(X, Y) > 0$ for some Y . If $\mathbf{tr} Y'(V - B(X))X \leq 0$ we have $D\sigma(X, Y) < 0$, and if $\mathbf{tr} Y'(V - B(X))X > 0$ we replace Y by $-Y$, and again $D\sigma(X, Y) < 0$. **QED**

Corollary 1: If σ has a local minimum at X then $d_{ij}(X) > 0$ for all (i, j) with $w_{ij} \delta_{ij} > 0$.

Proof: If there is an (i, j) such that $w_{ij}\delta_{ij}\alpha_{ij}(X) > 0$ then there is a Y such that $\theta(X, Y) > 0$. **QED**

Corollary 2: If σ has a local minimum at X then $w_{ij}\delta_{ij} = 0$ for all (i, j) with $d_{ij}(X) = 0$.

Proof: This is just another way of saying that $\theta(X, Y) = 0$ for all Y . **QED**

Corollary 3: If $w_{ij}\delta_{ij} > 0$ for all $i \neq j$ then σ is differentiable at a local minimum.

3 Final Result

Corollary 3 in the previous section allows for the possibility that σ is not differentiable at local minima if $w_{ij}\delta_{ij} = 0$ for some index pairs (i, j) . This is important, for example, in unfolding where indices $1, 2, \dots, n$ are partitioned into two disjoint subsets, and all within-subset weights are zero. It turns out that a fairly trivial manipulation allows us to find the appropriate generalization of our result.

Theorem 2: σ is differentiable at local minima.

Proof: An essentially equivalent way to formulate the MDS problem is to minimize

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

where $d_k(X) = \sqrt{\mathbf{tr} X' A_{ij} X}$ for some pair $1 \leq i < j \leq n$. Thus we fit distances to some, but not necessarily all, of the dissimilarities. In this formulation we can clearly assume without loss of generality that $w_k > 0$ for all $1 \leq k \leq K$. Suppose \mathcal{K}_0 and \mathcal{K}_1 are the subsets of indices for which, respectively, $\delta_k = 0$ and $\delta_k > 0$. Then

$$\sigma(X) = \sum_{k \in \mathcal{K}_1} w_k(\delta_k - d_k(X))^2 + \sum_{k \in \mathcal{K}_0} w_k d_k^2(X).$$

By the same reasoning as before at a local minimum X we have $d_k(X) > 0$ for all $k \in \mathcal{K}_1$, which implies that σ is differentiable at that local minimum. **QED**

Note that $d_k(X) > 0$ for all $k \in \mathcal{K}_1$ actually implies more: stress is infinitely many times differentiable in an open neighborhood of each local minimum.

References

- De Leeuw, J. 1984. "Differentiability of Kruskal's Stress at a Local Minimum." *Psychometrika* 49: 111–13. http://www.stat.ucla.edu/~deleeuw/janspubs/1984/articles/deleeuw_A_84f.pdf.
- De Leeuw, J., P. Groenen, and P. Mair. 2016. "Singularities and Zero Distances in Multidimensional Scaling." 2016. <https://doi.org/10.13140/RG.2.1.3339.2403>.