Differentiability of Functions of Distances

Jan de Leeuw

First created on March 31, 2019. Last update on November 26, 2021

Abstract

The stress loss function in metric multidimensional scaling is differentiable at local minima. In this note we generalize this result to more general functions of the distances.

Contents

Note: This is a working paper which will be expanded/updated frequently. All suggestions for improvement are welcome. The directory deleeuwpdx.net/pubfolders/distdiff has a pdf version, the bib file, and the complete Rmd file.

1 Introduction

In multidimensional scaling we minimize the stress measure proposed by Kruskal (1964a), Kruskal (1964b) and defined as

$$\sigma(x)=\sum_{k=1}^m w_k(\delta_k-\sqrt{x'A_kx})^2. \tag{1}$$

The w_k are known positive *weights*, the δ_k are known positive *dissimilarities*. The A_k are known positive semi-definite matrices. In multidimensional scaling the A_k are defined in such a way that the $x'A_kx$ are squared Euclidean distances between n points in $\mathbb{R}^p$, but this results in this note are valid for any positive semie-definite A_k.

Clearly stress is not differentiable at those x for which one or more of the $x'A_kx$ are zero. This can create problem for gradient-based algorithms, because the relevant derivatives may not exist and/or they may not vanish at stationary points.

It is shown, however, in De Leeuw (1984) that the directional derivative

$$\mathbf{D}\sigma(x,y)=\lim_{\epsilon\downarrow 0}\frac{\sigma(x+\epsilon y)-\sigma(x)}{\epsilon} \tag{2}$$

exists for all x and y. At a local minimum, i.e. an x for which $\mathbf{D}\sigma(x,y)\geq 0$ for each y, De Leeuw (1984), and more generally De Leeuw (2018), shows that $x'A_kx>0$, which implies that stress is differentiable at local minima.

2 Results

We now generalize the result from De Leeuw (1984) to a much more general class of functions of the $\sqrt{x'A_kx}$.

Theorem 1: [Differentiability] Suppose $\sigma(x)=F(\sqrt{x'A_1x},\cdots,\sqrt{x'A_mx})$ with F continuously differentiable and $\mathcal{D}_kF(z)<0$ if $z_k=0$. Then σ is differentiable at local minima, and at local minima $x'A_kx>0$.

Proof:

For the $\sqrt{x'A_kx}$ we have

$$\sqrt{(x+\epsilon y)'A_k(x+\epsilon y)}=\sqrt{x'A_kx}+\epsilon\begin{Bmatrix}\frac{1}{\sqrt{x'A_kx}}x'A_ky & \text{if } x'A_kx>0\\ \sqrt{y'A_ky} & \text{if } x'A_kx=0\end{Bmatrix}+o(\epsilon). \tag{3}$$

Thus we find for the directional derivative of σ

$$\begin{aligned}\mathbf{D}\sigma(x,y)=\sum_{x'A_kx>0}\mathcal{D}_kF(\sqrt{x'A_1x},\cdots,\sqrt{x'A_mx})\frac{1}{\sqrt{x'A_kx}}x'A_ky+\\ \sum_{x'A_kx=0}\mathcal{D}_kF(\sqrt{x'A_1x},\cdots,\sqrt{x'A_mx})\sqrt{y'A_ky}.\end{aligned} \tag{4}$$

If the assumption of the theorem is true there is an y such that the second term on the right in (4) is negative. If for that y the first term is positive we change y to $-y$. The first term becomes negative, the second term remains negative. Thus the derivative in the direction y is negative and x cannot be a local minimum. ■

Corollary 1: [Separable] Suppose $\sigma(x)=\sum_{k=1}^m f_k(\sqrt{x'A_kx})$ with the f_k continuously differentiable and $Df_k(0)<0$. Then σ is differentiable at local minima, and at local minima $x'A_kx>0$.

Proof: In this case

$$\sum_{x'A_kx=0}\mathcal{D}_kF(\sqrt{x'A_1x},\cdots,\sqrt{x'A_mx})\sqrt{y'A_ky}=\sum_{x'A_kx=0}\mathcal{D}f_k(0)\sqrt{y'A_ky},$$

which can be made negative for some y. ■

Corollary 2: **[Least Squares]** Suppose $\sigma(x) = \sum_{k=1}^m w_k(\delta_k - \phi(\sqrt{x'A_kx}))^2$ with ϕ continuously differentiable, $\phi(0) = 0$, and $\mathcal{D}\phi(0) > 0$. Then σ is differentiable at local minima, and at local minima $x'A_kx > 0$.

Proof: In this case

$$\sum_{x'A_kx=0} \mathcal{D}_kF(\sqrt{x'A_1x}, \cdots, \sqrt{x'A_mx})\sqrt{y'A_ky} = -\sum_{x'A_kx=0} w_k\delta_k\mathcal{D}\phi(0)\sqrt{y'A_ky},$$

which can be made negative for some y. ■

Remark 1: **[RStress]** The result of corollary 2 can be applied to rStress (De Leeuw, Groenen, and Mair (2016)), defined as

$$\sigma_r(x) = \sum_{k=1}^m w_k(\delta_k - (\sqrt{x'A_kx})^r)^2$$

with $r > 0$. Thus $\phi(z) = z^r$ and $\mathcal{D}\phi(z) = rz^{r-1}$. But if $r < 1$ then ϕ is not differentiable at zero and our results do not apply. If $r > 1$ then $\mathcal{D}\phi(0) = 0$, rStress is differentiable everywhere, and again our results do not apply (although obviously if $r > 1$ then rStress is differentiable at local minima, even if $x'A_kx = 0$). The only rStress covered by our result is Kruskal's stress, with $r = 1$.

References

De Leeuw, J. 1984. "Differentiability of Kruskal's Stress at a Local Minimum." *Psychometrika* 49: 111–13.

———. 2018. "Differentiability of Stress at Local Minima." 2018.

De Leeuw, J., P. Groenen, and P. Mair. 2016. "Differentiability of rStress at a Local Minimum." 2016.

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.