MDS by Majorizing Gauss-Newton

Jan de Leeuw

January 24, 2026

TBD

Table of 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

1 Loss

rStress with $r > 0$ is defined as

$$\sigma_r(x) := \sum_k w_k(\delta_k^r - d_k^r(x))^2$$

2 First-order Taylor approximation to powered distances

$$d_k^r(x) \approx d_k^r(y) + rd_k^{r-1}(y)(d_k(x) - d_k(y))$$

Note that if $r = 1$ both sides are equal to $d_k(x)$. If $r = 2$

$$d_k^2(x) \approx d_k^2(y) + 2d_k(y)(d_k(x) - d_k(y)) = 2d_k(x)d_k(y) - d_k^2(y).$$

3 Gauss-Newton approximation to rStress

$$\sigma_r(x) \approx \sum_k w_k(\delta_k^r - d_k^r(y) - rd_k^{r-1}(y)(d_k(x) - d_k(y)))^2 = \tag{1}$$

$$\sum_k w_k\{rd_k^{r-1}(y)\}^2 (\frac{\delta_k^r - d_k^r(y)}{rd_k^{r-1}(y)} - (d_k(x) - d_k(y)))^2 \tag{2}$$

Let

$$\tilde{w}_k(y) := r^2 w_k d_k^{2(r-1)}(y)$$

and

$$\tilde{\delta}_k(y) := \frac{\delta_k^r - d_k^r(y)}{rd_k^{r-1}(y)} + d_k(y) = \frac{\delta_k^r + (r-1)d_k^r(y)}{rd_k^{r-1}(y)}$$

then

$$\sigma_r(x; y) \approx \sum_k \tilde{w}_k(y)(\tilde{\delta}_k(y) - d_k(x))^2$$

which can be minimized by majorization. Also if $x = y$ then $\sigma_r(x; x) = \sigma_r(x)$.

For $r \geq 1$ we have $\delta_k(y) \geq 0$.

For $r = 1$ we have $\tilde{w}_k(y) = w_k$ and $\tilde{\delta}_k(y) = \delta_k$. This is regular smacof.

For $r = 2$ we have $\tilde{w}_k(y) = 4w_k d_k^2(y)$ and $\tilde{\delta}_k(y) = \frac{\delta_k^2 + d_k^2(y)}{2d_k(y)}$.

4 Algorithm

$$\sigma_r(x^+) \approx \sigma_r(x^+; x) \leq \sigma_r(x; x) = \sigma_r(x)$$

5 fStress

$$\sigma_f(x) = \sum_k w_k(f(\delta_k) - f(d_k(x))^2$$

$$f(d_k(x)) \approx f(d_k(y)) + \mathcal{D}f(d_k(y))(d_k(x) - d_k(y))$$

$$\tilde{w}_k(y) = w_k\{\mathcal{D}f(d_k(y))\}^2$$

$$\tilde{\delta}_k(y) = \frac{f(\delta_k) - f(d_k(y))}{\mathcal{D}f(d_k(y))} + d_k(y)$$

The Gauss-Newton approximation will be good if $f(\delta_k) \approx f(d_k(y))$, which will tend to be true for smooth monotone f (or for Lipschitz f) if $\delta_k \approx d_k(y)$. This suggests starting with some regular smacof iterations (although Torgerson may be good enough).

6 Negative $\tilde{\delta}_k$

If some of the $\tilde{\delta}_k$ are negative we may usae the AM/GM inequality for majorization as in Heiser (1991).

References

Heiser, W. J. 1991. "A Generalized Majorization Method for Least Squares Multidimensional Scaling of Pseudodistances that May Be Negative." *Psychometrika* 56 (1): 7–27.