Derivation of Kruskal's Negative Gradient Algorithm for Nonmetric and Metric unweighted MDS

1. Define the fit measure to be minimized (Stress, which we denote S):

$$S = \sqrt{\frac{\sum_i \sum_j (d_{ij} - o_{ij})^2}{\sum_i \sum_j o_{ij}^2}} \quad .$$

2. Kruskal solved for partial derivatives of S^2 relative to coordinate x_{he}.

$$\partial S^2_{x_{he}} = \partial \left(\frac{\sum_i \sum_j (d_{ij} - o_{ij})^2}{\sum_i \sum_j o_{ij}^2} \right) \quad .$$

Since $\sum_i \sum_j o_{ij}^2$ is constant we define $B = \sum_i \sum_j o_{ij}^2$ and note that

$$\partial S^2_{x_{he}} = \frac{1}{B} \partial \left[\sum_i \sum_j (d_{ij} - o_{ij})^2 \right] \quad ,$$

$$\partial S^2_{x_{he}} = \frac{1}{B} \left[\sum_i \sum_j \partial (d_{ij} - o_{ij})^2 \right] \quad ,$$

$$\partial S^2_{x_{he}} = \frac{2}{B} \left[\sum_i \sum_j (d_{ij} - o_{ij}) \partial (d_{ij} - o_{ij}) \right] \ ,$$

$$\partial S^2_{x_{he}} = \frac{2}{B} \left[\sum_i (d_{ih} - o_{ih}) \partial d_{ih} \right] \quad .$$

The last simplification is due to the fact that the derivatives of d_{ij} are zero except when $j = h$, thus the summation over j can be dropped and the subscript j changed to h.

3. We note that Kruskal defined distance using the Minkowski formula :

$$d_{ih} = \left[\sum_{a=1}^{r} \left|x_{ia} - x_{ha}\right|^{p}\right]^{\frac{1}{p}}.$$

So Kruskal solved for the derivatives of the Minkowski distances d_{ih} relative to a coordinate x_{he}. They are:

$$\partial d_{ih_{x_{he}}} = \partial\left[\sum_{a=1}^{r} \left|x_{ia} - x_{ha}\right|^{p}\right]^{\frac{1}{p}}$$

$$\partial d_{ih_{x_{he}}} = \frac{1}{p}\left[\sum_{a=1}^{r} \left|x_{ia} - x_{ha}\right|^{p}\right]^{\frac{1}{p}-1} \partial\left[\sum_{a=1}^{r} \left|x_{ia} - x_{ha}\right|^{p}\right]$$

$$\partial d_{ih_{x_{he}}} = \frac{1}{p} d_{ih}^{(1-p)}\left[\sum_{a=1}^{r} \partial\left|x_{ia} - x_{ha}\right|^{p}\right]$$

$$\partial d_{ih_{x_{he}}} = \frac{p}{p} d_{ih}^{(1-p)} \sum_{a=1}^{r} \left|x_{ia} - x_{ha}\right|^{p-1} \partial\left|x_{ia} - x_{ha}\right|$$

$$\partial d_{ih_{x_{he}}} = d_{ih}^{(1-p)} \left|x_{ie} - x_{he}\right|^{p-1} \mathrm{Sign}[x_{ie} - x_{he}] \partial[x_{ie} - x_{he}]$$

$$\partial d_{ih_{x_{he}}} = -\frac{1}{d_{ih}^{p-1}} \left|x_{ie} - x_{he}\right|^{p-1} \mathrm{Sign}[x_{ie} - x_{he}]$$

The second to last simplification is due to the fact that the derivatives of x_{ha} are zero except when $a = e$, thus the summation over a can be dropped and the subscript a changed to e.

4. We can now substitute the derivatives of the Minkowski distances (step 3) into the derivatives of Stress (step 2) to obtain the final formula for the derivatives of Stress as a function of the coordinates:

$$\partial S^2_{x_{he}} = -\frac{2}{B}\left[\sum_i \frac{(d_{ih} - o_{ih})}{d_{ih}^{p-1}} |x_{ie} - x_{he}|^{p-1} \mathrm{Sign}[x_{ie} - x_{he}]\right] .$$

5. When the distances are Euclidean, the formula simplifies to:

$$\partial S^2_{x_{he}} = -\frac{2}{B}\left[\sum_i \left[\frac{(d_{ih} - o_{ih})}{d_{ih}}\right][x_{ie} - x_{he}]\right]$$

6. The iterative algorithm to obtain a better fit uses the iteration equation:

$$x^+_{he} = x_{he} - \frac{2\alpha}{B}\sum_{i=1}^{n}\left[\frac{(d_{ih} - o_{ih})}{d_{ih}^{p-1}} |x_{ie} - x_{he}|^{p-1} \mathrm{Sign}[x_{ie} - x_{he}]\right]$$

   Here x^+_{he} denotes the value of the coordinate on the next iteration, and α is the unknown "step-size" that must be "guestimated" on each iteration. The quantity inside the summation is called the "correction value". We can denote it c_{ihe} and write the iteration equation as:

$$x^+_{he} = x_{he} - \frac{2\alpha}{B}\sum_{i=1}^{n} c_{ihe} .$$

   The c_{ihe} are coordinates of a "correction vector" which tells the direction and relative distance that x_{he} must be moved to improve fit.

7. Note that x^+_{he} *is not* the exact location that optimizes fit, since we do not know the step size. It is only a guess as to a better location.