Nonmetric Elastic MDS

Jan de Leeuw

November 16, 2025

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/elastic

1 Introduction

Early in the history of computerized Multidimensional Scaling (MDS), between the seminal contributions of Shepard/Kruskal and Guttman/Lingoes/Roskam, there was the "elastic method" proposed by Victor E. McGee in a series of papers (McGee (1965), McGee (1966), McGee (1967), McGee (1968)). The method has been largely forgotten, but it is worth remembering, because it is different in some important aspects from the more well-known methods.

The least squares loss function in most MDS methods (Borg and Groenen (2005)) can be written as

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

where X is the *configuration* of n *points* in p *dimensions*, $d_{ij}(X)$ is the *Euclidean distance* between points i and j in configuration X, δ_{ij} is the *dissimilarity* between points i and j, and w_{ij} is a *weight*. Weights are non-negative.

In the *metric* version of the MDS method the dissimilarities are observed and fixed, and minimization is over configurations only. The denominator in (1) is irrelevant for the minimization problem, but it normalizes the problem in the sense that the minimum of stress is between zero and one. In the *nonmetric* version minimization is over both configurations and dissimilarities, with the constraint that the dissimilarities are monotonic with an observed set of dissimilarities.

It is true that in the original Kruskal formulation the denominator is the sum of the squared distances instead. But De Leeuw (1975) shows that normalizing by using either the sum of squared distances or the sum of squared dissimilarities leads to the same solution (up to a scalar proportionality factor). The smacof program (De Leeuw and Mair (2009), Mair, Groenen, and De Leeuw (2022)) minimizes the numerator of (1) over configurations and dissimilarities, with the additional constraint that the sum of squares of the dissimilarities is equal to a constant. Again, this *explicit normalization* gives the same solution, up to proportionality, as the original Kruskal formulation that uses *implicit normalization* by dividing by the sum of squared distances.

In McGee's elastic method the loss is constructed to satisfy two requirements. The first one is

> Psychological judgments which indicated relatively great separation of stimuli should be allowed greater error than judgments indicating close proximity (l.c. p. 182)

And the second requirement is the basic MDS requirement that dimensionality of X must as low as possible, while still providing a good fit.

2

A criterion which suggested itself in response to the first requirement was one based on the physical work done on an elastic spring to stretch or compress it from an initial length δ_{ij} to a final length d_{ij}. (l.c. p. 183)

This leads to the loss function

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

which McGee calls the loss in (2) *work*. We shall just call it *stress* (and *elastic stress* or *McGee stress*), using the more familiar MDS name for loss. We will also continue to use the symbol σ for any least squares MDS loss function.

The elastic MDS problem is minimization of stress over both X and Δ, where Δ must be monotone with the given dissimilarities. In (2) the weight w_{ij} is interpreted as the modulus of elasticity of the spring (i, j).

In McGee (1967) the alternative loss function

$$\sigma(X, \Delta) := \sum_{1 \leq i < j \leq n} \sum w_{ij} \frac{(\delta_{ij} - d_{ij}(X))^2}{d_{ij}^2(X)}, \tag{3}$$

is proposed. Minimizing (3) seems more complicated, and we will postpone studying algorithms to minimize it. Thus we will work with (2) in this paper.

In McGee's papers the actual algorithm and its implementation are not described in sufficient detail. Part of the problem is that he is dealing exclusively with the nonmetric case, in which minimization over both X and Δ is necessary. In the metric case minimization is over X only, and the algorithm is much simpler.

3

2 Properties

If we compare (1) and (2) we see one important difference. The normalization in (1) is by the sum of squared dissimilarities, while the normalization in (2) is term-by-term, by the squared dissimilarities themselves. An alternative way of writing the elastic loss function makes this more clear.

$$\sigma(X, \Delta) := \sum_{1 \leq i < j \leq n} \sum w_{ij} \left(1 - \frac{d_{ij}(X)}{\delta_{ij}}\right)^2 \tag{4}$$

Thus we see that instead of minimizing the difference of dissimilarities and distances from zero the elastic method minimizes the deviations of their ratios from one.

If δ_{ij} is multiplied by a positive constant, then so is the optimum configuration X and the $D(X)$. As a consequence, the minimum of elastic stress does not change, i.e. is homogeneous of degree zero in the dissimilarities.

It is clear from (4) that the loss function is undefined if one or more of the dissimilarities are zero. If one of the distances is zero, then the corresponding term in the loss function is equal to the corresponding weight, irrespective of what the corresponding dissimilarity is. Thus the minimum of elastic stress is always less than or equal to the sum of the weights, its value at $X = 0$.

If δ_{ij} and $d_{ij}(X)$ are non-zero and close then a first order Taylor series expansion gives

$$\log d_{ij}(X) - \log \delta_{ij} \approx \frac{1}{\delta_{ij}}(d_{ij}(X) - \delta_{ij}), \tag{5}$$

from which it follows that

$$\sigma(X, \Delta) \approx \sum_{1 \leq i < j \leq n} \sum w_{ij}(\log \delta_{ij} - \log d_{ij}(X))^2. \tag{6}$$

If the fit is good the elastic stress will be approximately equal to the logarithmic stress from Ramsay (1977). Or, to put it differently, minimizing elastic stress can serve as an approximation to minimizing logarithmic stress.

3 Algorithm

In this paper we will give an iterative algorithm for minimizing loss from (2) in the metric an non-metric (ordinal) case. A majorization algorithm for the metric case is already available in the smacofx package (Rusch et al. (2025)). Our new non-metric algorithm is in the *alternating least squares* family, which means we alternate minimizing over X for fixed Δ and over Δ for fixed X.

We will start our iterations with the classical metric solution (Torgerson (1958)) for Δ. We actually scale that solution by minimizing

$$\sigma(\lambda) := \sum_{k=1}^{m} \frac{w_k}{\delta_k^2} (\delta_k - \lambda d_k(X))^2 \tag{7}$$

over λ. The minimum is attained at

$$\hat{\lambda} := \frac{\sum_{k=1}^{m} \frac{w_k}{\delta_k} d_k(X)}{\sum_{k=1}^{m} \frac{w_k}{\delta_k^2} d_k^2(X)}. \tag{8}$$

In iteration k we perform one majorization step to replace $X^{(k)}$ by $X^{(k+1)}$ to improve loss for fixed $\Delta^{(k)}$ and one monotone regression step to replace $\Delta^{(k)}$ by $\Delta^{(k+1)}$ for fixed $X^{(k+1)}$.

To minimize (2) over X for fixed Δ we rewrite loss as

$$\sigma(X, \Delta) = \sum\sum_{1 \leq i < j \leq n} \frac{w_{ij}}{\delta_{ij}^2} (\delta_{ij} - d_{ij}(X))^2. \tag{9}$$

This can be minimized (or decreased) by using the standard smacof majorization step with the weights w_{ij}/δ_{ij}^2.

To minimize over Δ for given X we define $\gamma_{ij} := -\delta_{ij}^{-1}$ abd $c_{ij}(X) := -d_{ij}^{-1}(X)$. Rewrite loss as

$$\sigma(X, \Gamma) = \sum\sum_{1 \leq i < j \leq n} w_{ij} d_{ij}^2(X) (\gamma_{ij} - c_{ij}(X))^2, \tag{10}$$

which we must minimize over increasing γ_{ij}. This is just monotone regression with the $c_{ij}(X)$ as the targets and with weights $w_{ij} d_{ij}^2(X)$. After we have found the optimal $\hat{\gamma}_{ij}$ we transform back to $\hat{\delta}_{ij} = -\hat{\gamma}_{ij}^{-1}$.

4 Examples

We start with a simple artificial example to illustrate our iterations. Suppose

```
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    0    1    2    3    4    5
[2,]    1    0    6    7    8    9
[3,]    2    6    0   10   11   12
[4,]    3    7   10    0   13   14
[5,]    4    8   11   13    0   15
[6,]    5    9   12   14   15    0
```

Moreover all weights are equal to one. The Torgerson solution in two dimensions is

```
            [,1]         [,2]
[1,]   0.2189030  -0.2977892
[2,]  -0.2069156   0.2069172
[3,]  -0.7627856   1.0397322
[4,]  -2.0514988   6.7680097
[5,]  -5.8874415  -5.6144769
[6,]   8.6897386  -2.1023931
```

and the distances $d(X)$ for this solution are

```
           1          2          3          4          5
2  0.6603408
3  1.6591190  1.0012853
4  7.4216061  6.8154546  5.8714517
5  8.0965802  8.1336956  8.3988450 12.9630410
6  8.6609267  9.1914836  9.9610825 13.9304783 14.9942960
```

and stress is 26.6531626.

If we apply the scaling from … and … we find $\hat{\lambda}$ equal to 0.3889323 and a stress value of 5.8259173.

We now make a majorization step. The configuration becomes

6

```
            [,1]           [,2]
[1,]   0.1503740 -0.2690647
[2,] -0.3926772  0.3529975
[3,] -1.1967821  1.5863548
[4,] -1.2320447  4.1405188
[5,] -4.1596387 -4.0936940
[6,]  6.8307688 -1.7171126
```

with elastic stress 2.6002809.

```
o <- order(delta)
print(o)
```

```
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
```

```
print(d)
```

```
          1           2           3           4           5
2  0.8257517
3  2.2929045  1.4723298
4  4.6212020  3.8794144  2.5544074
5  5.7622911  5.8277839  6.4063620  8.7391686
6  6.8355334  7.5142216  8.6806953  9.9659825 11.2444296
```

```
y <- 1 / d[o]
omat <- cbind(2:15,1:14)
print(omat)
```

```
      [,1] [,2]
 [1,]    2    1
 [2,]    3    2
 [3,]    4    3
 [4,]    5    4
 [5,]    6    5
 [6,]    7    6
 [7,]    8    7
 [8,]    9    8
 [9,]   10    9
```
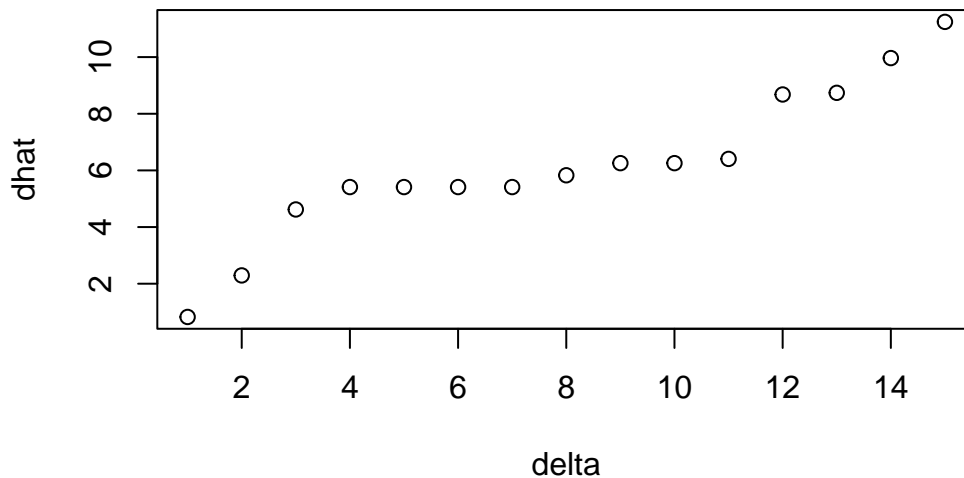
```
[10,]    11    10
[11,]    12    11
[12,]    13    12
[13,]    14    13
[14,]    15    14
```

```r
w <- d[o]^2
idel <- activeSet(omat, lsSolver, weights = w, y = y)
dhat <- 1 / idel$x
plot(delta, dhat)
```



```r
r <- d / dhat
s <- sum((1 - r)^2)
print(s)
```

```
[1] 1.074024
```

Unit: milliseconds expr min lq mean median uq smacofSSElasticR(morseData, ordinal = FALSE, verbose = FALSE) 106.137110 108.788006 114.907009 109.952508 111.618687 smacofSSElasticC(morseData, ordinal = FALSE, verbose = FALSE) 6.183292 6.481526 6.539732 6.542575 6.614325 smacofSSElasticR(morseData, ordinal = TRUE, verbose = FALSE) 709.953376 720.502860 738.153851 728.890579 751.493530 smacofSSElasticC(morseData, ordinal = TRUE, verbose = FALSE) 6.684804 6.952637 7.041597 7.018318 7.134574 max neval 189.327053 100 6.714078 100 883.220729 100 9.207616 100

References

Borg, I., and P. J. F. Groenen. 2005. *Modern Multidimensional Scaling*. Second Edition. Springer.

De Leeuw, J. 1975. "A Normalized Cone Regression Approach to Alternating Least Squares Algorithms." Department of Data Theory FSW/RUL. https://jansweb.netlify.app/publication/deleeuw-u-75-a/deleeuw-u-75-a.pdf.

De Leeuw, J., and P. Mair. 2009. "Multidimensional Scaling Using Majorization: SMACOF in R." *Journal of Statistical Software* 31 (3): 1–30. https://www.jstatsoft.org/article/view/v031i03.

Mair, P., P. J. F. Groenen, and J. De Leeuw. 2022. "More on Multidimensional Scaling in R: smacof Version 2." *Journal of Statistical Software* 102 (10): 1–47. https://www.jstatsoft.org/article/view/v102i10.

McGee, V. E. 1965. "More on an "Elastic" Multidimensional Scaling Procedure." *Perceptual and Motor Skills* 21: 81–82.

———. 1966. "The Multidimensional Analysis of 'Elastic' Distances." *British Journal of Mathematical and Statistical Psychology* 19 (2): 181–96.

———. 1967. "A Reply to Some Criticisms of Elastic Multidimensional Scaling." *British Journal of Mathematical and Statistical Psychology* 20 (2): 243–47.

———. 1968. "Multidimensional Scaling of N Sets of Similarity Measures: A Nonmetric Individual Differences Approach." *Multivariate Behavioral Research* 3 (2): 233–48.

Ramsay, J. O. 1977. "Maximum Likelihood Estimation in Multidimensional Scaling." *Psychometrika* 42: 241–66.

Rusch, Thomas, Jan de Leeuw, Lisha Chen, and Patrick Mair. 2025. *Smacofx: Flexible Multidimensional Scaling and 'Smacof' Extensions*. https://doi.org/10.32614/CRAN.package.smacofx.

Torgerson, W. S. 1958. *Theory and Methods of Scaling*. New York: Wiley.