# Metric/Nonmetric Elastic MDS

Jan de Leeuw

November 24, 2025

We present R and C implementations for metric (ratio) and non-metric (ordinal) versions of Elastic MDS, the multidimensional scaling technique proposed by McGee (1966). The R and C versions are compared for speed, with the C version anywhere from 15 to 100 times as fast as the R version.

## 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

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

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 $X$, $\delta_{ij}$ 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.

> 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 $\delta_{ij}$ 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 *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 $\sigma$ for any least squares MDS loss function.

The elastic MDS problem is minimization of stress over both $X$ and $\Delta$, where $\Delta$ 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 $\Delta$ is necessary. In the metric case minimization is over $X$ only, and the algorithm is much simpler.

## 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 $\delta_{ij}$ 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 $\delta_{ij}$ 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 and non-metric (ordinal) case. A majorization algorithm for the metric case is already available in the smacofx package (Rusch et al. (2025), Rusch et al. (In Press)). Our new non-metric algorithm is in the *alternating least squares* family, which means we alternate minimizing over $X$ for fixed $\Delta$ and over $\Delta$ for fixed $X$.

We will start our iterations with the classical metric solution (Torgerson (1958)) for $\Delta$. 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 $\lambda$. 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 $\Delta$ we rewrite loss as

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

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

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

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

which we must minimize over increasing $\gamma_{ij}$. 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 Software

The github repository contains R and C versions of the smacofSSElastic program. Both smacofSSElasticC() and smacofSSElasticR() have the same default values of the parameters and the two programs are structured the same way. We iterate a maximum of 1000 iterations until the change in stress from one iteration to the next is less than $10^{-6}$. A single majorization update is alternated with a single monotone regression.

smacofSSElasticC() has a driver in R that sets up the MDS data structure (De Leeuw (2025)). Both program use a number of R utilities for data manipulation and the initial configuration. One important difference is that the R version uses gpava() for monotone regression (De Leeuw, Hornik, and Mair (2009)), which is written in R, while the R version uses the C routine monotone() from Busing (2022).

# 5 Examples

We analyze two classical MDS examples: the color data of Ekman (1954) and the Morse code data of Rothkopf (1957). We are only interested in this paper in performance of the programs, not in the substantive interpretation of the results or in the comparison of elastic scaling and regular smacof.

Numerical elastic scaling takes 586 iterations to arrive at stress 2.3268637. For ordinal we need 437 iterations for stress 0.056998. Since the ordinal fit is very good the log-stress of (6) should be close to the elastic stress. It is 0.0581521.

The results of a comparison of the R and C implementations with microbenchmark (Mersmann (2024)) yields the times in microseconds in following table.

```
               min     lq   mean median     uq    max
R/Numerical  40.13  41.66  44.38  43.84  45.11 109.41
C/Numerical   1.44   1.53   1.60   1.60   1.64   1.81
R/Ordinal   163.52 168.14 172.10 169.50 172.28 242.27
C/Ordinal     4.06   4.39   4.57   4.46   4.71   7.37
```

Using median times we see that that C version is about 40 times as fast as the R version.

For the Morse code data numerical elastic scaling takes 212 iterations to arrive at stress 64.3828862, for ordinal we need 152 iterations for stress 29.7732277. Because the fit is much worse than in the Ekman case we do not expect log-stress to be close to elastic stress. For the ordinal case final log-stress is 42.8850382.

```
               min     lq   mean median     uq    max
R/Numerical  80.71  85.50  92.21  87.61  91.38 162.41
C/Numerical   6.38   6.62   6.77   6.71   6.80  10.96
R/Ordinal   711.81 744.80 767.76 764.32 784.45 862.41
C/Ordinal     7.25   7.60   7.82   7.74   7.83  11.35
```

In the numerical case the C version is about 15 times as fast, in the ordinal case about 100 times. The huge difference in the ordinal case is likely to be due in large part to the different monotone regression routiones used by the R and C programs.

Of course these microbenchmark results depend on the default parameters of the programs (same in R and C), on the speed of my computer (Mac Mini with Apple M4 Pro, 64 GB of Ram, Tahoe 26.2), and on my programming habits (same in R and C).

# References

Borg, Ingwer, and Patrick J. F. Groenen. 2005. *Modern Multidimensional Scaling*. Second Edition. Springer.

Busing, Frank M. T. A. 2022. "Monotone Regression: A Simple and Fast O(n) PAVA Implementation." *Journal of Statistical Software* 102 (Code Snippet 1). https://www.jstatsoft.org/index.php/jss/article/view/v102c01/4306.

De Leeuw, Jan. 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.

———. 2025. "Yet Another Smacof - Square Symmetric Case." 2025. https://jansweb.netlify.app/publication/deleeuw-e-25-b/.

De Leeuw, Jan, Kurt Hornik, and Patrick Mair. 2009. "Isotone Optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods." *Journal of Statistical Software* 32 (5): 1–24.

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

Ekman, Gosta. 1954. "Dimensions of Color Vision." *Journal of Psychology* 38: 467–74.

Mair, Patrick, Patrick J. F. Groenen, and Jan 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, Victor 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.

Mersmann, O. 2024. *microbenchmark: Accurate Timing Functions*. https://CRAN.R-project.org/package=microbenchmark.

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

Rothkopf, Ernst Z. 1957. "A Measure of Stimulus Similarity and Errors in some Paired-associate Learning." *Journal of Experimental Psychology* 53: 94–101.

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.

Rusch, Thomas, Jan De Leeuw, Patrick Mair, and Kurt Hornik. In Press. "Flexible Multidimensional Scaling with the r Package Smacofx." *Journal of Statistical Software*, In Press. https://jansweb.netlify.app/publication/rusch-deleeuw-mair-hornik-a-25/rusch-deleeuw-mair-hornik-a-25.pdf.

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