



## Unidimensional Scaling

By Patrick Mair<sup>1</sup> and Jan De Leeuw<sup>2</sup>

**Keywords:** *fitting distances, multidimensional scaling, combinatorial optimization*

**Abstract:** We discuss the one-dimensional special case of multidimensional scaling, and the various algorithms that have been proposed to solve the corresponding computational problem. We concentrate on least squares unidimensional scaling and on the combinatorial nature of finding the best scaling.

Unidimensional scaling techniques are a popular tool in psychometrics within the context of nonparametric and parametric item response theory (see **Nonparametric Item Response Theory Models; Item Response Theory (IRT) Models for Dichotomous Data; Item Response Theory Models for Polytomous Response Data; Item Response Theory Models for Rating Scale Data**). In this article, we focus on unidimensional scaling as a special one-dimensional case of multidimensional scaling (MDS). It is often discussed separately because the unidimensional case is quite different from the general multidimensional case. It has been shown that the minimization of the stress target function with equal weights leads to a combinatorial problem when the number of dimensions of the target space is one<sup>[1]</sup>. Unidimensional scaling techniques are very different from multidimensional scaling techniques because they use very different algorithms to minimize their loss functions. If we perform a one-dimensional metric MDS with standard MDS algorithms, we have to be concerned about the fact that we end up in a local minimum after a few iterations. If we allow for transformations of the proximities, the local minimum problem may be less severe<sup>[2]</sup>.

Unidimensional scaling is applied in situations where we have a strong reason to believe that there is only one interesting underlying dimension, such as time, ability, or preference. We do not have to choose between different metrics, such as the Euclidean metric, the City Block metric, or the Dominance metric. The classical form of unidimensional scaling starts with a symmetric and nonnegative matrix  $\Delta = \{\delta_{ij}\}$  of dissimilarities and another symmetric and nonnegative matrix  $W = \{w_{ij}\}$  of weights. Both  $W$  and  $\Delta$  have a zero diagonal. Unidimensional scaling finds coordinates  $x_i$  for  $n$  points on the line such that the *stress*

$$\sigma(x) = \sum_{i < j} w_{ij} (\delta_{ij} - d_{ij}(x))^2$$

is minimized. The  $n$  coordinates in  $x$  define the scale we are looking for. Note that  $d_{ij}(x) = |x_i - x_j|$  and can be rewritten as  $d_{ij}(x) = (x_i - x_j)s_{ij}(x_i - x_j)$  with  $s_{ij}(x_i - x_j) = \text{sign}(x_i - x_j)$ . This term becomes 1 if  $x_i > x_j$ , 0 if  $x_i = x_j$ , and  $-1$  if  $x_i < x_j$ . Therefore, only the rank order of  $x$  determines  $s_{ij}(x_i - x_j)$ . The above-mentioned

<sup>1</sup>Harvard University, Cambridge, MA, USA

<sup>2</sup>University of California, Los Angeles, CA, USA

Update based on original article by Jan de Leeuw, Wiley StatsRef: Statistics Reference Online, © 2014, John Wiley & Sons, Ltd

Wiley StatsRef: Statistics Reference Online, © 2014–2015 John Wiley & Sons, Ltd.

This article is © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat06462.pub2





stress can be expressed as

$$\sigma(x) = \sum_{i < j} w_{ij} \delta_{ij}^2 + \sum_{i < j} w_{ij} (x_i - x_j)^2 - 2 \sum_{i < j} w_{ij} \delta_{ij} d_{ij}(x)$$

We see that the last term consists of one part that is linear in  $x$  and another part that depends on the rank order of the elements in  $x$ . Let us denote the rank order of  $x$  by  $\psi$  such that  $x_{\psi_1} \leq x_{\psi_2} \leq \dots \leq x_{\psi_n}$ .

Let  $V$  be the matrix with off-diagonal elements  $v_{ij} = -w_{ij}$  and diagonal elements  $v_{ii} = \sum_{j=1}^n w_{ij}$ . In addition, let  $R$  be the permutation matrix such that  $Rx$  represents the vector with the elements ordered monotonically. Furthermore, we define the vector  $l$  with  $l_i = \sum_{j < i} w_{\psi_i \psi_j} \delta_{\psi_i \psi_j}$  and the vector  $u$  with  $u_i = \sum_{j > i} w_{\psi_i \psi_j} \delta_{\psi_i \psi_j}$ . Using this notation, we can rewrite the stress as

$$\sigma(x) = \sum_{i < j} w_{ij} \delta_{ij}^2 + x' V x - 2x' R'(l - u)$$

For a given  $\psi$ , we see that the stress value has its minimum when  $x = V^+ R'(l - u)$ . We see that this Guttman transform uses the rank-order information of the previous permutation only and, consequently, the stress can be rewritten as (see Ref. 2 for details)

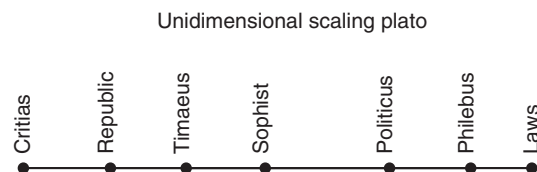
$$\sigma(x) = \sum_{i < j} w_{ij} \delta_{ij}^2 + \|x - V^+ R'(l - u)\|_V^2 - \|l - u\|_{R^+ V^+ R'}^2$$

The crucial term is the last one that we denote by  $f(\psi)$ . It is a function of the permutations only. Over the years, various combinatorial optimization strategies have been proposed to maximize  $f(\psi)$  over  $\psi$ . An overview is given in Ref. 3; more recent developments can be found in Refs 4 and 5.

Now, we present two examples. The first example is quite simple. The data set we use is taken from Ref. 6 and contains statistical information about Plato's seven works. Within a unidimensional scaling context, it has been analyzed in Ref. 7. The underlying problem to this data set is the fact that the chronological order of Plato's works is unknown. Scholars only know that *Republic* was his first work and *Laws* his last work. For each work, Ref. 6 extracted the last five syllables of each sentence. Each syllable is classified as long or short, which gives  $2^5 = 32$  types. Consequently, we obtain a percentage distribution across the 32 scenarios for each of the seven works. We compute a  $7 \times 7$  dissimilarity matrix that gives the Euclidean distances between each pair of works based on the percentage vector. This matrix acts as input matrix for unidimensional scaling with the underlying dimension being time. We investigate all  $7! = 5040$  permutations and use the one with the lowest stress value. The result is shown in Figure 1.

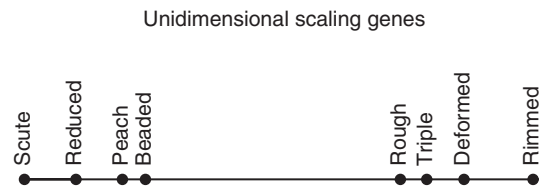
The axis represents the timeline and the works are scaled accordingly. We obtain the chronological order *Critias* < *Republic* < *Timaeus* < *Sophist* < *Politicus* < *Philebus* < *Laws*. We see that *Republic* is not scaled as Plato's first work. *Laws*, however, is scaled where it should be: as his last work.

The second example is quite different. It has weights and incomplete information. We take it from an early paper by Fisher<sup>[8]</sup> in which he studies crossover percentages of eight genes on the sex chromosome of *Drosophila willistoni*. He takes the crossover percentage as a measure of distance, and supposes that the number  $n_{ij}$  of crossovers in  $N_{ij}$  observations is binomial. Although there are 8 genes, and thus 28 possible dissimilarities, there are only 15 pairs that are actually observed. Thus, 13 of the off-diagonal weights are zero, and the other weights are set to the inverses of the standard errors of the proportions.



**Figure 1.** Unidimensional scaling solution of Plato's works.





**Figure 2.** Unidimensional scaling solution of gene dataset.

We investigate all  $8! = 40320$  permutations, and we find 78 local minima. The solution given by Fisher, computed by solving linearized likelihood equations, has *Reduced* < *Scute* < *Peach* < *Beaded* < *Rough* < *Triple* < *Deformed* < *Rimmed*. This order corresponds with a local minimum of  $\sigma(x)$  equal to 40.16. The global minimum is obtained for the permutation that interchanges *Reduced* and *Scute*, with value 35.88. In Figure 2, we see the result of our unidimensional scaling solution.

In this article, we have discussed least squares metric unidimensional scaling. The first obvious generalization is to replace the least squares loss function, for example, by a  $L_1$  loss function as given in Ref. 4. The second generalization is to look at nonmetric unidimensional scaling (see Ref. 3). The combinatorial nature of the problem remains intact.

## References

- [1] De Leeuw, J. and Heiser, W.J. (1977) Convergence of correction matrix algorithms for multidimensional scaling, in *Geometric Representations of Relational Data* (ed. J.C. Lingoes), Mathesis Press, Ann Arbor, MI, pp. 735–752.
- [2] Borg, I. and Groenen, P.J.F. (2005) *Modern Multidimensional Scaling: Theory and Applications*, 2nd edn, Springer-Verlag.
- [3] Hubert, L.J., Arabie, P., and Meulman, J.J. (2002) Linear unidimensional scaling in the  $L_2$ -norm: basic optimization methods using MATLAB. *J. Classif.*, **19**, 303–328.
- [4] Murillo, A., Vera, J.F., and Heiser, W.J. (2005) A permutation-translation simulated annealing algorithm for  $L_1$  and  $L_2$  unidimensional scaling. *J. Classif.*, **22**, 119–138.
- [5] Brusco, M.J. and Stahl, S. (2005) Optimal least-squares unidimensional scaling: improved branch-and-bound procedures and comparison to dynamic programming. *Psychometrika*, **70**, 253–270.
- [6] Cox, D.R. and Brandwood, L. (1959) On a discriminatory problem connected with the work of Plato. *J. R. Stat. Soc., Ser. B*, **21**, 195–200.
- [7] Cox, T.F. and Cox, M.A.A. (2001) *Multidimensional Scaling*, 2nd edn, Chapman and Hall/CRC.
- [8] Fisher, R.A. (1922) The systematic location of genes by means of crossover observations. *Am. Nat.*, **56**, 406–411.