Nonmetric Strain

Jan de Leeuw

February 20, 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/strain

1 Introduction

In Mutidimensional Scaling (MDS) the strain loss function is defined as

$$\sigma(X,\Delta) := \frac{1}{4}\text{tr}\ \{J(\Delta^2 - D^2(X))J\}^2 \tag{1}$$

Here J is the centering matrix of order n, i.e. $J := I - n^{-1}E$, with E a matrix with all elements equal to one. Matrix Δ^2 has squared dissimilarities and $D^2(X)$ contains the squared Euclidean distances between the rows of an $n \times p$ column-centered configuration matrix X.

We will also use another representation of strain in this paper. By applying classical embedding result of Schoenberg (1935) and Young and Householder (1938), in the version proposed by Torgerson (1958), we find

$$\sigma(X,\Delta) = \text{tr}\ (C - XX')^2. \tag{2}$$

with

$$C := -\frac{1}{2}J\Delta^2 J. \tag{3}$$

In *metric* MDS the matrix Δ is constant and given, and we minimize strain over X. From (2) the basic theorem of classical scaling (Torgerson (1958)), also known as principal coordinate analysis (Gower (1966)), tells us that the optimal X are the eigenvectors corresponding with the p largest eigenvalues of C_+, the positive definite part of C, with each eigenvector scaled to have norm equal to the square root of the corresponding eigenvalue. In the earlier literature there is some ambiguity on what to do if C has fewer than p positive eigenvalues. For rigorous proofs see Keller (1962) and De Leeuw (1974a). That classical scaling is equivalent to minimizing the loss function (1) was first pointed out by De Leeuw and Heiser (1982).

In non-metric MDS we minimize strain over both X and Δ, where Δ varies in some subset $\mathcal{D}$ of the space of non-negative, symmetric and hollow matrices of order n. Classical examples, treated in detail below in this paper, are the additive constant problem, the missing data problem, the use of bound constraints, and ordinal non-metric scaling.

The history of several tentative and/or incomplete attempts to construct a non-metric MDS technique based on strain is in De Leeuw and Heiser (1982) and Trosset (1998). The first systematic approach to the problem is De Leeuw (1974b), who discusses the additive constant, the missing data, and the ordinal forms of MDS. The paper has no software or examples, only equations. It uses what Trosset (1997) calls "variable reduction". The minimum of strain over X, which is the sum of squares of the discarded eigenvalues of C, is minimized over Δ.

Results from classical perturbation theory are used to give expressions for the first and second derivatives of the "reduced" loss function, with the implicit understanding that these derivatives can be used in a large variety of general-purpose optimization methods. An implementation of the same methodology for the additive constant problem is in Saito (1978). Since the "reduced" strain in this cases only depends on a single parameter (the additive constant) Saito uses a simple bisection-type search procedure on the positive real line. A more systematic and general approach was followed by Michael Trosset in a sequence of papers (Trosset (1997), Trosset (2000), Trosset (1998), Trosset (2002)). Trosset distinguishes "variable alternation" from "variable reduction" methods for the minimization of functions of two or more sets of variables. In minimizing a function such as $\sigma(X,\Delta)$ a variable alternation method uses iterations which consist of two substeps. The first substep minimizes over X while keeping Δ fixed at its current best value, and the second substep minimizes over Δ while keeping X fixed at its current best value. A variable reduction method defines a "reduced" function

$$\sigma_\star(\Delta):=\min_X\sigma(X,\Delta)=\sigma(X(\Delta),\Delta), \tag{4}$$

and then minimizes, using some iterative process, the function $\sigma_\star$. Here $X(\Delta)$ is the optimum X for given Δ, which, as we know, can be computed from the eigen-decomposition of C.

Trosset's papers use variable reduction, and the minimization of the reduced function uses general gradient/hessian based optimization methods. Software is discussed, but not explicitly provided, and examples are analyzed and reported.

This paper differs from Trosset's work because we use variable alternation instead of variable reduction. Software in R (R Core Team (2025)) is provided.

2 Alternating Least Squares

Variable alternating in the context of minimizing least squaress loss functions is usually called Alternating Least Squares. We start with some initial estimate $\Delta^{(0)}$. Then, in each iteration, there are two subproblems. We first minimize strain over X with Δ fixed at its current value, and then minimize over Δ in $\mathcal{D}$ with X fixed at its current value, which is the optimal value from the first subproblem in the same iteration. Then we test for convergence, and if the loss function still decreases we go to the next iteration.

In iteration k this means

$$X^{(k)} = \underset{X}{\text{argmin}}\, \sigma(X, \Delta^{(k-1)}), \tag{5a}$$

$$\Delta^{(k)} = \underset{\Delta\in\mathcal{D}}{\text{argmin}}\, \sigma(X^{(k)}, \Delta) \tag{5b}$$

If a subproblem in an iteration is too complicated for some reason, then we merely make a move towards its minimum, making sure that

$$\sigma(X^{(k)}, \Delta^{(k)}) < \sigma(X^{(k)}, \Delta^{(k-1)}) < \sigma(X^{(k-1)}, \Delta^{(k-1)}). \tag{6}$$

In the first subproblem minimizing strain over X for fixed Δ is is a metric MDS problem (Torgerson (1958), Gower (1966), Mardia (1978)). We compute the diagonal matrix with p dominant eigenvalues Λ_p of C and the corresponding eigenvalues K_p. For this we use the R package RSpectra (Qiu and Mei (2024)). If $\Lambda_p \geq 0$ then the optimal X is $K_p\Lambda_p^{\frac{1}{2}}$. If some elements of Λ_p are negative, we first replace them by zeroes.

The second subproblem, minimizing strain of Δ for fixed X (and thus for fixed D), takes various forms depending on the constraints on Δ. We discuss some of the important special cases in the following sections, together with their implementation.

3 Special Cases

3.1 The Additive Constant

The simplest, and oldest, form of non-metric scaling estimates the additive constant. Thus $\Delta = \Delta_0 + \theta(E - I)$, where Δ_0 are the given dissimilarities and E is the matrix with all elements equal to one. Only one additional parameter needs to be estimated, with the side-condition that the adjusted δ_{ij} are non-negative.

Squaring the constraint gives

$$\Delta^2 = \Delta_0^2 + 2\theta\Delta_0 + \theta^2(E - I). \tag{7}$$

Because we want the Δ to be non-negative we require that $\theta \geq -\min \Delta_0$, where the minimum is taken over all off-diagonal elements of Δ.

From (7) we have, with

$$G := J\Delta_0 J, \tag{8a}$$

$$H := J(\Delta_0^2 - D^2), \tag{8b}$$

that strain is equal to

$$\sigma(\theta) := \frac{1}{4}\text{tr } \left\{J(\Delta_0^2 + 2\theta\Delta_0 + \theta^2(E - I) - D^2)J\right\}^2 = \frac{1}{4}\text{tr } \left\{H + 2\theta G - \theta^2 J\right\}^2, \tag{9}$$

which is a quartic in θ. Expanding and collecting terms gives

$$\sigma(\theta) = \frac{1}{4}\left\{\text{tr } H^2 + 4\theta \text{tr } HG + 2\theta^2(\text{tr } 2G^2 - \text{tr } H) - 4\theta^3 \text{tr } G + \theta^4(n-1)\right\}. \tag{10}$$

We minimize the quartic over the half-open interval $\theta \geq -\min \Delta$ using the formulas for the minimum and minimizers of a quartic in Jeffrey (1997). Note that in both substeps of the ALS iteration we compute the global minimum (except for precision considerations).

3.2 Imputing Missing Data

The next non-metric scaling problem has some of the δ_{ij} missing, either by accident or by design. Thus

$$\delta_{ij} = \begin{cases} \delta_{ij}^0 & \text{for non-missing,} \\ \theta_{ij} & \text{for missing.} \end{cases} \tag{11}$$

Thus the number of additional parameters is equal to the number of missing elements. We will require the θ_{ij} to be non-negative.

If there are missing data the classical scaling method to minimize strain must be adapted. In that case there is a set $\mathcal{M}$ of m index pairs that code for missing data. Matrix Δ_0^2 now has zeroes for missing data and squared dissimilarities for the non-missing ones. We have

$$\Delta^2 = \Delta_0^2 + \sum_{(i,j)\in\mathcal{M}} \theta_{(i,j)} E_{(i,j)}, \tag{12}$$

where $E_{(i,j)} := e_i e_j' + e_j e_i'$ and e_i has element i equal to one and all other elements zero. If D^2 are the current squared distances, then strain is the function of θ defined as

$$\sigma(\theta) := \text{tr}\,\{H - \sum_{(i,j)\in\mathcal{M}} \theta_{(i,j)\in\mathcal{M}} J E_{(i,j)\in\mathcal{M}} J)\}^2, \tag{13}$$

where H is as defined by in the previous section.

An important special case of the missing data problem is the unfolding problem, in which we have similarities between objects from two disjoint sets. The between-set distances are known, while both matrices of within-set distances are missing.

3.3 Bound Constraints

Bound constraints are of the form $\Delta_- \leq \Delta \leq \Delta_+$, where the inequality signs refer to elementwise comparisons. Thus each δ_{ij} is required to be in a given interval (which may be infinitely large). Updating Δ for given X is a quadratic programming problem. By choosing the interval to be the non-negative real axis this gives us another way, using majorization, to impute missing data. For non-missing data the interval has length zero and is equal to the dissimilarity δ_{ij}^0.

Strain for this problem By using the vec of our matrices Δ and D strain can also be written in the computationally more friendly form

$$\sigma(\delta) = \frac{1}{4}(\delta^2 - d^2)(J \otimes J)(\delta - d^2), \tag{14}$$

with $J \otimes J$ the Kronecker product.

Equation (21) shows that minimizing strain over δ is a weighted monotone regression problem. It is more complicated than standard monotone regression because of the weights $J \otimes J$, which are not diagonal. Thus strain as a function of δ, although convex, is not separable. In addition $J \otimes J$ is singular, having $(n-1)^2$ unit eigenvalues and $2n-1$ zero eigenvalues. Routines such as PAVA cannot be used, although active set methods are still a viable alternative (De Leeuw, Hornik, and Mair (2009)).

We will go an alternative route, using majorization to reduce the weighted least squares problem to a sequence of unweighted problems, for which we can use ordinary monotone regression. Related results on majorizing weighted by unweighted least squares are in Kiers (1997) and Groenen, Giaquinto, and Kiers (2003).

Suppose $\tilde{\Delta}$ is the current best value of the disparities. Expand strain around $\tilde{\Delta}$.

$$\sigma(\Delta) = \sigma(\tilde{\Delta}) + \frac{1}{2}\text{tr}\ J(\Delta - \tilde{\Delta})J(\tilde{\Delta} - D^2) + \frac{1}{4}\text{tr}\ J(\Delta - \tilde{\Delta})J(\Delta - \tilde{\Delta}), \qquad (15)$$

Because $J \otimes J \lesssim I$ we have $\sigma(\Delta) \leq \eta(\Delta, \tilde{\Delta})$, with the majorization function η defined as

$$\eta(\Delta, \tilde{\Delta}) := \sigma(\tilde{\Delta}) + \frac{1}{2}\text{tr}\ J(\Delta - \tilde{\Delta})J(\tilde{\Delta} - D^2) + \frac{1}{4}\text{tr}\ (\Delta - \tilde{\Delta})^2. \qquad (16)$$

The majorization, or MM, algorithm minimizes η over vectors Δ in the polyhedral convex cone of vectors withnincresing elements to find the update Δ^+.

To simplify the expression for η we "complete the square". Relying heavily on the idempotency of J, and thus of $J \otimes J$, we find

$$\eta(\Delta, \tilde{\Delta}) := \frac{1}{4}\text{tr}\ (\Delta - \hat{\Delta})^2, \qquad (17)$$

with

$$\hat{\Delta} := \tilde{\Delta} - J(\tilde{\Delta} - D^2)J. \qquad (18)$$

Thus $\hat{\Delta}$ is a matrix-weighted average of the "conservative" part $\tilde{\Delta}$ and the "progressive" part D^2 (Chamberlain and Leamer (1976)).

Finally the update Δ^+ is the least squares projection of $\hat{\Delta}$ on the cone, i.e. the unweighted monotone regression on $\hat{\Delta}$. We now have the sandwich inequality

$$\sigma(\Delta^+) \leq \eta(\Delta^+, \tilde{\Delta}) \leq \eta(\tilde{\Delta}, \tilde{\Delta}) = \sigma(\tilde{\Delta}). \qquad (19)$$

In the substep of ALS in which we improve Δ we make a number of "inner" majorization iterations. In the later "outer" ALS iterations we will already have a good initial estimate to start these "inner" iterations.

3.4 Ordinal MDS

A third important form of non-metric MDS is ordinal MDS. We require the δ_{ij} to non-negative and monotone with the δ_{ij}^0. More precisely if $\delta_{ij}^0 \geq \delta_{kl}^0$ then we must have $\delta_{ij} \geq \delta_{kl} \geq 0$.

For the additive constant problem and the missing data problem the constraint sets are convex polyhedrons that do not contain the origin, but for ordinal MDS $\Delta = 0$ satisfies the ordinal constraints. Choosing $X = 0$ and $\Delta = 0$ produces $\sigma = 0$ which is clearly the global minimum. To prevent this trivial solution from happening we use a normalization condition on either X or Δ thate keeps both of them away from the origin.

For fixed D strain is a function of the disparities Δ, which are required to be monotone with Δ_0. Thus

$$\sigma(\Delta) = \frac{1}{4}\text{tr } J(\Delta^2 - D^2)J(\Delta - D^2)J \tag{20}$$

Note that we fit disparities Δ to squared distances. By using the vec of our matrices Δ and D strain can also be written in the computationally more friendly form

$$\sigma(\delta) = \frac{1}{4}(\delta - d^2)(J \otimes J)(\delta - d^2), \tag{21}$$

with $J \otimes J$ the Kronecker product.

Equation (21) shows that minimizing strain over δ is a weighted monotone regression problem. It is more complicated than standard monotone regression because of the weights $J \otimes J$, which are not diagonal. Thus strain as a function of δ, although convex, is not separable. In addition $J \otimes J$ is singular, having $(n-1)^2$ unit eigenvalues and $2n - 1$ zero eigenvalues. Routines such as PAVA cannot be used, although active set methods are still a viable alternative (De Leeuw, Hornik, and Mair (2009)).

We will go an alternative route, using majorization to reduce the weighted least squares problem to a sequence of unweighted problems, for which we can use ordinary monotone regression. Related results on majorizing weighted by unweighted least squares are in Kiers (1997) and Groenen, Giaquinto, and Kiers (2003).

Suppose $\tilde{\Delta}$ is the current best value of the disparities. Expand strain around $\tilde{\Delta}$.

$$\sigma(\Delta) = \sigma(\tilde{\Delta}) + \frac{1}{2}\text{tr } J(\Delta - \tilde{\Delta})J(\tilde{\Delta} - D^2) + \frac{1}{4}\text{tr } J(\Delta - \tilde{\Delta})J(\Delta - \tilde{\Delta}), \tag{22}$$

Because $J \otimes J \lesssim I$ we have $\sigma(\Delta) \leq \eta(\Delta, \tilde{\Delta})$, with the majorization function η defined as

$$\eta(\Delta, \tilde{\Delta}) := \sigma(\tilde{\Delta}) + \frac{1}{2}\text{tr } J(\Delta - \tilde{\Delta})J(\tilde{\Delta} - D^2) + \frac{1}{4}\text{tr } (\Delta - \tilde{\Delta})^2. \tag{23}$$

The majorization, or MM, algorithm minimizes η over vectors Δ in the polyhedral convex cone of vectors withnincresing elements to find the update Δ^+.

To simplify the expression for η we "complete the square". Relying heavily on the idempotency of J, and thus of $J \otimes J$, we find

$$\eta(\Delta, \tilde{\Delta}) := \frac{1}{4}\text{tr }(\Delta - \hat{\Delta})^2, \tag{24}$$

with

$$\hat{\Delta} := \tilde{\Delta} - J(\tilde{\Delta} - D^2)J. \tag{25}$$

Thus $\hat{\Delta}$ is a matrix-weighted average of the "conservative" part $\tilde{\Delta}$ and the "progressive" part D^2 (Chamberlain and Leamer (1976)).

Finally the update Δ^+ is the least squares projection of $\hat{\Delta}$ on the cone, i.e. the unweighted monotone regression on $\hat{\Delta}$. We now have the sandwich inequality

$$\sigma(\Delta^+) \leq \eta(\Delta^+, \tilde{\Delta}) \leq \eta(\tilde{\Delta}, \tilde{\Delta}) = \sigma(\tilde{\Delta}). \tag{26}$$

In the substep of ALS in which we improve Δ we make a number of "inner" majorization iterations. In the later "outer" ALS iterations we will already have a good initial estimate to start these "inner" iterations.

Normalization: $G = J\Delta^2 J \,\text{tr}\, G^2 = 1$.

4 Software

For the four special cases we have discussed there are four R programs (R Core Team (2025)) in the github repository.

- strainSSAddOne()
- strainSSMissing()
- strainSSBound()
- strainSSOrdinal()

We implemented Jeffrey's solution in the C function jeffrey(), which is in the shared library jeffrey.so. The R program strainSSAddOne() loads the shared library and iteratively executes the two ALS sub problems.

Minimizing σ from (13) over the m element vector $\theta \geq 0$ is a non-negative linear least squares problem that we solve by using the R function nnls() from the package with the same name (Mullen and van Stokkum (2024)). The R function strainSSMissing() calls nnls(), as well as eigs_sym from RSpectra, to solve the two ALS subproblems.

5 Examples

5.1 The Additive Constant

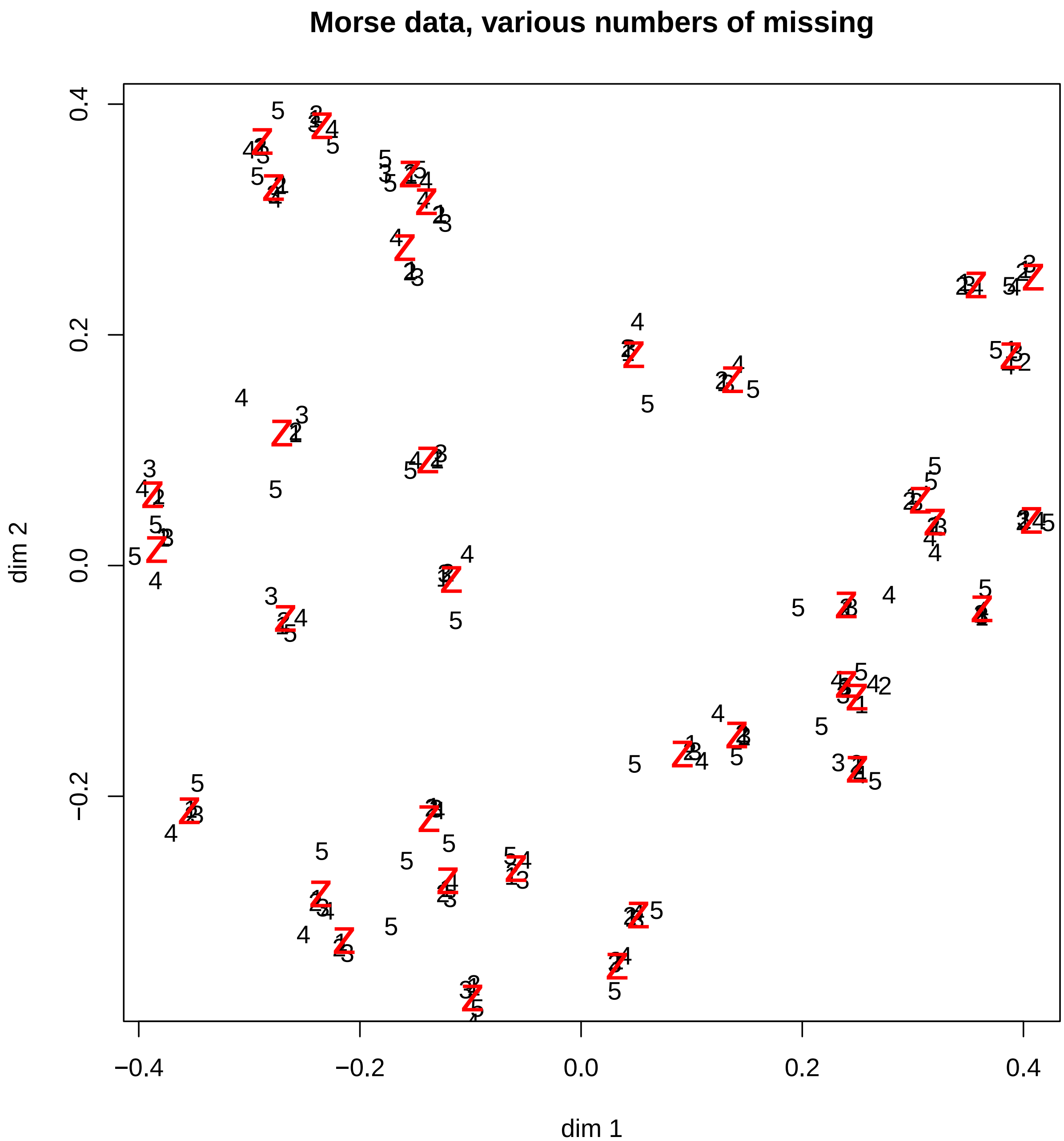
Our example is taken from Torgerson (1958), pages 280-290. Dissimilarty judgments of nine Munsell colors of the same red hue, but differing in brightness and saturation, where collected from 38 subjects using the method of triads. The matrix Δ_0, which Torgerson calls the "comparative distances" is taken from his Table 5 on p 286.

```
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9]
 [1,]  0.00 -2.37 -0.12 -0.62  0.23  1.56  1.09  2.02  2.23
 [2,] -2.37  0.00 -1.01 -1.93 -0.90  0.80 -0.47  1.05  0.78
 [3,] -0.12 -1.01  0.00  0.70 -1.32 -0.67  1.07  0.70  2.62
 [4,] -0.62 -1.93  0.70  0.00 -0.78  1.25 -1.75  0.28 -0.72
 [5,]  0.23 -0.90 -1.32 -0.78  0.00 -1.02 -1.23 -1.65  0.49
 [6,]  1.56  0.80 -0.67  1.25 -1.02  0.00  0.57 -0.67  1.88
 [7,]  1.09 -0.47  1.07 -1.75 -1.23  0.57  0.00 -1.18 -1.30
 [8,]  2.02  1.05  0.70  0.28 -1.65 -0.67 -1.18  0.00  0.42
 [9,]  2.23  0.78  2.62 -0.72  0.49  1.88 -1.30  0.42  0.00
```

Torgerson (p, 286-287) uses an elegant, but elaborate, method to find an additive constant value of 3.60. His technique is not based on minimization of an explicit loss function. Iterating strainSSAddOne() until strain changes less than 10^{-10} from one iteration to the next, starting with $\theta = 0$ uses 196 iterations and find $\theta = 2.85$. Starting with Torgerson's 3.6 value for the additive constant gives the same θ in 184 iterations. Of course the 10^{-10} stopping criterion is much too strict for most applications, although a run with even higher precision shows the value for the optimal additive constant only has three stable decimals. Using a more realistic 0.001 requires 105 iterations and gives two stable decimals. In other words, convergence is slow.

5.2 Imputing Missing Data

We give two examples. The first uses the classical Morse code data from Rothkopf (1957). The data are confusion probabilities between 36 auditory Morse code signals, collected from 598 airmen. For details of the design we refer to Rothkopf. The data are available, for example, in the smacof package (De Leeuw and Mair (2009)). With 36 objects there are 630 dissimilarities. We ran the strainSSMissing() function with 0, 5, 10, 50, and 100 random dissimilarities missing. The five configurations are matched with procrustus() from procrustus.R, which minimizes $\sum_k \|Z - X_k M_k\|^2$ over Z and the square orthonormal M_k.



The plot shows a stable solution. Points are labeled “0” for no missing, “1” for 5 missing, and so on. The red points are for the matched Z matrix. The number of iterations for 0, 5, 10, 50, 100 missing data is 0, 10, 9, 15, 23 and strain is 6.7478705, 6.7240824, 6.6684388, 6.2187352, 5.8854484.

roskam

6 Discussion

Most MDS loss functions are of the form

$$\sigma(X) := \sum_k w_k(f(\delta_k) - f(d_k(X)))^2 \tag{27}$$

for some given function f. Groenen, De Leeuw, and Mathar (1995) call loss function (27) fStress. For f equal to the identity we have stress (Kruskal (1964)), for f the square we have sstress (Takane, Young, and De Leeuw (1977)), and for f the logarithm we have log-stress (Ramsay (1977)). If we compare (27) and strain we see that strain is more complicated because of the non-diagonal Kronecker weights. This makes the usual optimal transformations in MDS, such as monotone regression in the ordinal case, more complicated and costly. As a consequence non-metric MDS techniques minimizing strain have not been implemented and used to the same degree as techniques based on variations of fStress.

On the other hand in fStress the minimization over X for given Δ is more costly than it is for strain, in which we simply have to find the p largest eigenvalues and corresponding eigenvectors of a matrix. Thus, the ALS framework, for fStress finding optimal Δ for fixed x is cheap and finding optimal X for fixed Δ is expensive. For strain it is exactly the other way around. To put it even more sharply, in fStress finding the optimal X is impossible, because it requires an inifinite iterative process which can converge to a local minimum. For strain, although solving the eigen-problem is also an infinite iterative process, it quickly gets us as close as we want to the global minimum (for given $Delta$, that is).

Trosset (2002)

Trosset (1998) discusses an implementation or ordinal MDS uing strain. His aproach is quite different from ours, because it does not use ALS but gradient projection. More importantly, Trosset projects out the optimization over X. He calls this "variable reduction", to contrast it with "variable alternation". Thus his strain is a function of Δ only, defined as

$$\sigma(\Delta) = \min_X \text{tr}\,(C(\Delta) - XX')^2 = \sum_{r=p+1}^{n} \lambda_r^2(C_+(\Delta)), \tag{28}$$

where the λ_r are the $n-p$ smallest eigenvalues of its argument, and with C_+ the positive definite part of C. This loss function is then minimized over Δ by gradient projection, constrained by the linear inequalities that impose monotonicity. It is unclear how this compares to ALS, although we can say that in general projecting out a set of unknowns tends to accelerate convergence, at the cost of more complicated subproblems in each iteration. Also, unlike ALS, gradient projection implies choosing a step-size procedure to guarantee convergence. Projection seems especially suitable for the additive constant problem in which (28) is a function of a single parameter, which can be minimized in a number of ways.

weighted strain

References

De Leeuw, Jan. 1974a. "Approximation of a Real Symmetric Matrix by a Positive Semidefinite Matrix of Rank r." Technical Memorandum TM-74-1229-10. Murray Hill, N.J.: Bell Telephone Laboratories. https://jansweb.netlify.app/publication/deleeuw-r-74-c/deleeuw-r-74-c.pdf.

———. 1974b. "Finding a Positive Semidefinite Matrix of Prescribed Rank r in a Nonlinear Differentiable Manifold." Technical Memorandum. Murray Hill, N.J.: Bell Telephone Laboratories. https://jansweb.netlify.app/publication/deleeuw-r-74-b/deleeuw-r-74-b.pdf.

De Leeuw, Jan, and Willem J. Heiser. 1982. "Theory of Multidimensional Scaling." In *Handbook of Statistics, Volume II*, edited by P. R. Krishnaiah and L. Kanal. Amsterdam, The Netherlands: North Holland Publishing Company.

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.

Gower, John C. 1966. "Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis." *Biometrika* 53: 325–38.

Groenen, Patrick J. F., Jan De Leeuw, and Rudolf Mathar. 1995. "Least Squares Multidimensional Scaling with Transformed Distances." In *From Data to Knowledge: Theoretical and Practical Aspects of Classification, Data Analysis and Knowledge Organization*, edited by W. Gaul and D. Pfeifer. Berlin, Germany: Springer Verlag. https://jansweb.netlify.app/publication/groenen-deleeuw-mathar-c-95/groenen-deleeuw-mathar-c-95.pdf.

Groenen, Patrick J. F., Patrizia Giaquinto, and Henk A. L Kiers. 2003. "Weighted Majorization Algorithms for Weighted Least Squares Decomposition Models." Econometric Institute Report EI 2003-09. Econometric Institute, Erasmus University Rotterdam. https://repub.eur.nl/pub/1700.

Jeffrey, David J. 1997. "Formulae, Algorithms, and Quartic Extrema." *Mathematics Magazine* 70 (5): 341–48.

Keller, Joseph B. 1962. "Factorization of Matrices by Least Squares." *Biometrika* 49: 239–42.

Kiers, Henk A. L. 1997. "Weighted Least Squares Fitting Using Iterative Ordinary Least Squares Algorithms." *Psychometrika* 62: 251–66.

Kruskal, Joseph B. 1964. "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis." *Psychometrika* 29: 1–27.

Mardia, Kanti V. 1978. "Some Properties of Classical Multidimensional Scaling." *Communications in Statistics - Theory and Methods* 7 (13): 1233–41.

Mullen, Katherine M., and Ivo H. M. van Stokkum. 2024. *nnls: The Lawson-Hanson algorithm for non-negative least squares (NNLS)*. %7Bhttps://CRAN.R-project.org/package=nnls%

7D.
Qiu, Yixuan, and Jiali Mei. 2024. *RSpectra: Solvers for Large-Scale Eigenvalue and SVD Problems*. https://CRAN.R-project.org/package=RSpectra.
R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
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.
Saito, Takayuki. 1978. “The Problem of the Additive Constant and Eigenvalues in Metric Multidimensional Scaling.” *Psychometrika* 43 (2): 193–201.
Schoenberg, Isaac J. 1935. “Remarks to Maurice Frechet’s article: Sur la Definition Axiomatique d’une Classe d’Espaces Vectoriels Distancies Applicables Vectoriellement sur l’Espace de Hllbert.” *Annals of Mathematics* 36: 724–32.
Takane, Yoshio, Forrest W. Young, and Jan De Leeuw. 1977. “Nonmetric Individual Differences in Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features.” *Psychometrika* 42: 7–67.
Torgerson, Warren S. 1958. *Theory and Methods of Scaling*. New York: Wiley.
Trosset, Michael W. 1997. “Numerical Algorithms for Multidimensional Scaling.” In *Classification and Knowledge Organization*, edited by Rüdiger Klar and Otto Opitz, 80–92. Springer.
———. 1998. “A New Formulation of the Nonmetric Strain Problem in Multidimensional Scaling.” *Journal of Classification* 15: 15–35.
———. 2000. “Distance Matrix Completion by Numerical Optimization.” *Computational Optimization and Applications* 17: 11–22.
———. 2002. “Extensions of Classical Multidimensional Scaling via Variable Reduction.” *Computational Statistics* 17: 147–63.
Young, Gale, and Alston S. Householder. 1938. “Discussion of a Set of Points in Terms of Their Mutual Distances.” *Psychometrika* 3 (19-22).