Squared Distance Scaling

Jan de Leeuw

June 9, 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/elegant

1 Introduction

In *squared distance multidimensional scaling* we minimize the least squares loss function[1]

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

over $n \times p$ configurations X. Here the δ_k are known non-negative *pseudo-distances*, the w_k are known positive *weights*, and the $d_k(X)$ are Euclidean distances. Each index k in (1) corresponds with a pair of indices (i, j), with both $1 \leq i \leq n$ and $1 \leq j \leq n$. Thus we try to find a configuration of n points on $\mathbb{R}^p$ such that the distances between the points approximate the corresponding pseudo-distances in the data.

Loss function (1) is traditionally known as *sstress*. In the *ALSCAL* program for squared distance scaling (Takane, Young, and De Leeuw (1977)) a coordinate descent algorithm, in which each iteration cycle consists of minimizing np univariate quartics, is used to minimize loss. There have been quite a few alternative algorithms proposed, both in multidimensional scaling (De Leeuw (1975), Browne (1987), Kearsley, Tapia, and Trosset (1994)) and in low-rank distance matrix completion (Mishra, Meyer, and Sepulchre (2011)).

The reference section of the present paper does not have publication information on De Leeuw (1975), in fact not even a URL, because that paper somehow got lost in the folds of time (Takane (2016)). Proof of its existence are references to it in Takane (1977) and Browne (1987). At the time it was concluded that the algorithm proposed in De Leeuw (1975), which was proudly baptized *ELEGANT*, converged too slowly to be practical. Recent attempts to revive and improve it are De Leeuw, Groenen, and Pietersz (2016) and De Leeuw (2016). This paper is another such attempt.

[1] The symbol := is used for defintions.

2 Majorization

The original derivation of the algorithm in De Leeuw (1975) was based on *augmentation*. The derivation is reviewed in De Leeuw, Groenen, and Pietersz (2016). For a general discussion of augmentation, see De Leeuw (1994). Improvements of *ELEGANT* are possible if we discuss it in the general framework of majorization , currently more widely known as MM (De Leeuw (1994), Heiser (1995), Lange (2016)).

We start by changing variables from X to $C = XX'$. Thus

$$\sigma(C) := \sum_{k=1}^m w_k(\delta_k^2 - \text{tr } A_k C)^2. \tag{2}$$

If k indexes pair (i, j) then $A_k := (e_i - e_j)(e_i - e_j)'$, with e_i and e_j unit vectors[2]. Thus squared distances can be expressed as $\|x_i - x_j\|^2 = \text{tr } X'A_{ij}X = \text{tr } A_{ij}C$. In these new variables the MDS problem is now to minimize (2) over all C with $\text{rank}(C) \leq p$.

It is convenient to define

$$B := \sum_{k=1}^m w_k \delta_k^2 A_k, \tag{3a}$$

and

$$V := \sum_{k=1}^m w_k a_k a_k' \tag{3b}$$

with $a_k := \text{vec}(A_k)$. Then $\sigma(c) := K - 2b'c + c'Vc$, with $c := \text{vec}(C)$ and $b := \text{vec}(B)$.

To start the quadratic majorization, use $c = \tilde{c} + (c - \tilde{c})$. Then

$$\sigma(c) = \sigma(\tilde{c}) - 2(c - \tilde{c})'(b - V\tilde{c}) + (c - \tilde{c})'V(c - \tilde{c}), \tag{4a}$$

and thus

$$\sigma(c) \leq \sigma(\tilde{c}) + \lambda((c - \tilde{c}) - g)'((c - \tilde{c}) - g) - \lambda g'g, \tag{4b}$$

with λ the largest eigenvalue of $V$and $g := \lambda^{-1}(b - V\tilde{c})$. In a majorization step we minimize $(c - \overline{c})'(c - \overline{c})$, where $\overline{c} := \tilde{c} + \lambda^{-1}(b - V\tilde{c})$.]

Now

$$V\tilde{c} = \sum_{k=1}^m w_k a_k a_k' \tilde{c} = \sum_{k=1}^m w_k a_k \text{tr } A_k \tilde{C} = \sum_{k=1}^m w_k a_k d_k^2(\tilde{C}), \tag{5}$$

and thus

$$\text{vec}^{-1}(B - V\tilde{c}) = \sum_{k=1}^m w_k(\delta_k^2 - d_k^2(\tilde{C})A_k. \tag{6}$$

Equation (6) shows that in the majorization stepm we need to minimize tr $(\overline{C} - XX')^2$ with $\overline{C} := \text{vec}^{-1}(\overline{c})$ over X.

[2]Unit vector e_i has element i equal to one and all other elements zero.

3 Bound

Computing λ is simplified by noting that the largest eigenvalue of V is equal to the largest eigenvalue of $W^{\frac{1}{2}}HW^{\frac{1}{2}}$, where H has elements $h_{kl} = a_k'a_l$.

The elements of H are all non-negative. Also h_{kl} is equal to four if $k = l$ and equal to one if A_k and A_l have one index in common, otherwise it is zero. It follows that in the complete case, with $m = n(n-1)/2$, and in addition if there are unit weights, $\lambda = 2n$. In the incomplete case, still with unit weights, $\lambda \leq 2n$.

If it is too expensive to calculate the largest eigenvalue, we can use the bound

$$\lambda = \max_x \frac{x'W^{\frac{1}{2}}HW^{\frac{1}{2}}x}{x'x} = \max_x \frac{x'W^{\frac{1}{2}}HW^{\frac{1}{2}}x}{x'Wx}\frac{x'Wx}{x'x} \leq 2n \max_k w_k. \tag{7}$$

This is a major improvement of the bound that is used, either explicitly or implicitly, in De Leeuw (1975) and De Leeuw, Groenen, and Pietersz (2016), which is

$$\lambda \leq \text{tr}\, WH = 4\sum_{k=1}^m w_k. \tag{8}$$

4 Code

References

Browne, M. W. 1987. “The Young-Householder Algorithm and the Least Squares Multidimensional Scaling of Squared Distances.” *Journal of Classification* 4: 175–90.

De Leeuw, J. 1975. “An Alternating Least Squares Approach to Squared Distance Scaling.” Department of Data Theory FSW/RUL.

———. 1994. “Block Relaxation Algorithms in Statistics.” In *Information Systems and Data Analysis*, edited by H. H. Bock, W. Lenski, and M. M. Richter, 308–24. Berlin: Springer Verlag. https://jansweb.netlify.app/publication/deleeuw-c-94-c/deleeuw-c-94-c.pdf.

———. 2016. “Convergence Rate of ELEGANT Algorithms.” 2016. https://jansweb.netlify.app/publication/deleeuw-e-16-o/deleeuw-e-16-o.pdf.

De Leeuw, J., P. Groenen, and R. Pietersz. 2016. “An Alternating Least Squares Approach to Squared Distance Scaling.” https://jansweb.netlify.app/publication/deleeuw-groenen-pietersz-e-16-m/deleeuw-groenen-pietersz-e-16-m.pdf.

Heiser, W. J. 1995. “Convergent Computing by Iterative Majorization: Theory and Applications in Multidimensional Data Analysis.” In *Recent Advantages in Descriptive Multivariate Analysis*, edited by W. J. Krzanowski, 157–89. Oxford: Clarendon Press.

Kearsley, A. J., R. A. Tapia, and M. W. Trosset. 1994. “The Solution of the Metric STRESS and SSTRESS Problems in Multidimensional Scaling Using Newton’s Method.” TR94-44. Houston, TX: Department of Computational and Applied Mathematics, Rice University. https://apps.dtic.mil/sti/pdfs/ADA445621.pdf.

Lange, K. 2016. *MM Optimization Algorithms*. SIAM.

Mishra, B., G. Meyer, and R. Sepulchre. 2011. “Low-Rank Optimization for Distance Matrix Completion.” In *50th IEEE Conference on Decision and Control and European Control Conference, Orlando, FL, USA, 2011*, 4455–60. https://arxiv.org/pdf/1304.6663.

Takane, Y. 1977. “On the Relations among Four Methods of Multidimensional Scaling.” *Behaviormetrika* 4: 29–42.

———. 2016. “My Early Interactions with Jan and Some of His Lost Papers.” *Journal of Statistical Software* 73 (7): 1–14. https://www.jstatsoft.org/article/view/v073i07.

Takane, Y., F. W. Young, and J. De Leeuw. 1977. “Nonmetric Individual Differences in Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features.” *Psychometrika* 42: 7–67.