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

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

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 is proposed to minimize loss, in which each iteration cycle consists of minimizing np univariate quartics. There have been quite a few alternative algorithms proposed, both in the area of multidimensional scaling (De Leeuw (1975), Browne (1987), Kearsley, Tapia, and Trosset (1994)) and in that of low-rank distance matrix completion (Mishra, Meyer, and Sepulchre (2011)).

The reference section of this paper does not have publication information on De Leeuw (1975), in fact not even a URL, because the 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 as *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.

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). But improvements are possible if we discuss it in the general framework of majorization (De Leeuw (1996)), currently more familiarly known as MM ((**?**)).

We change variables from X to $C = XX'$. Thus

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

Define

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

and

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

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

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

and thus

$$\sigma(c) \leq \sigma(\tilde{c}) - 2(c - \tilde{c})'(b - V\tilde{c}) + \lambda(c - \tilde{c})'(c - \tilde{c}) = \sigma(\tilde{c}) + \lambda((c - \tilde{c}) - g)'((c - \tilde{c}) - g) - \lambda g'g,$$

with λ the largest eigenvalue of V and $g := \lambda^{-1}(b - V\tilde{c})$.

In a majozation step we minimize $(c - \overline{c})'(c - \overline{c})$, where

$$\overline{c} := \tilde{c} + \lambda^{-1}(b - V\tilde{c})$$

.

Taking the inverse of vec show that we must minimize

$$\text{tr } (\overline{C} - XX')^2$$

over X.

3 Details

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 non-negative. Also h_{kl} is equal to 4 if $k = l$ and equal to 1 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, $\lambda \leq 2n$.

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

and thus

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

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

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

This is a major improvement ofthe bound that is us3ed, either explicitly or implicitly, in (**?**) and (**?**), which is

$$\lambda \leq \text{tr } W^{\frac{1}{2}}HW^{\frac{1}{2}} = 4\sum_{k-1}^m w_k.$$

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.

———. 1996. "Series Editors Introduction." In *T. Heinen: Latent Class and Discrete Latent Trait Models*. Advanced Quantitative Techniques in the Social Sciences 6. Thousand Oaks, CA: Sage Publications.

———. 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.

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.

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.