Accelerated SMACOF Multidimensional Scaling

Jan de Leeuw - University of California Los Angeles

Started July 23 2024, Version of August 02, 2024

Abstract

TBD

Contents

Note: This is a working paper which will be expanded/updated frequently. All suggestions for improvement are welcome.

1 Introduction

In this paper we study minimization of the loss function

$$\sigma(X) := \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n w_{ij}(\delta_{ij}-d_{ij}(X))^2$$

over all $n\times p$ *configuration* matrices X. Here $W=\{w_{ij}\}$ and $\Delta=\{\delta_{ij}\}$ are known non-negative, symmetric, and hollow matrices of *weights* and *dissimilarities* and $D(X)=\{d_{ij}(X)\}$ is the matrix of *Euclidean distances* between the rows of X.

Define

$$\rho(X) := \sum_{i=1}^n\sum_{j=1}^n w_{ij}\delta_{ij}d_{ij}(X) = \text{tr } X'B(X)X,$$

where

$$B(X) := \sum_{i=1}^n\sum_{j=1}^n w_{ij}\frac{\delta_{ij}}{d_{ij}(X)}A_{ij}.$$

Aso define

$$\eta^2(X) := \sum_{i=1}^n\sum_{j=1}^n w_{ij}d_{ij}^2(X) = \text{tr } X'VX,$$

where

$$V := \sum_{i=1}^n\sum_{j=1}^n w_{ij}A_{ij}.$$

Thus

$$\sigma(X) = 1-\rho(X)+\frac{1}{2}\eta^2(X)$$

Both ρ and η are homogeneous convex functions.

$\rho(X)=0$ iff $d_{ij}(X)=0$ for all (i,j) for which $w_{ij}\delta_{ij}>0$.

2 One-point Methods

2.1 Basic Iteration

The *Guttman transform* of a configuration X, named after Guttman (1968), is defined as

$$\Phi(X) = V^{+}B(X)X,$$

with V^{+} the Moore-Penrose inverse of V. If $X = \Phi(X)$, i.e. if X is fixed point of Φ, then $VX - B(X)0$, which we can also write as $\mathcal{D}\sigma(X) = 0$. Thus X is a fixed point of Φ if and only if the gradient of σ vanishes at X, i.e. if and only if X is a stationary point if σ.

We have to be somewhat careful here. Subdifferential

$$\partial d_{ij}(X) = \frac{1}{d_{ij}(X)}A_{ij}X$$

$$\partial d_{ij}(X) = A_{ij}Z$$

Using the Guttman transform we can derive the basic smacof equality

$$\sigma(X) = 1 + \eta^2(X - \Phi(X)) - \eta^2(\Phi(X))$$

for all X and the inequality

$$\sigma(X) \leq 1 + \eta^2(X - \Phi(Y)) - \eta^2(\Phi(Y))$$

for all X and Y.

Taken together „ and „„ imply the sandwich inequality

$$\sigma(\Phi(Y)) \leq 1 - \eta^2(\Phi(Y)) \leq 1 + \eta^2(Y - \Phi(Y)) - \eta^2(\Phi(Y)) = \sigma(Y)$$

If Y is not a fixed point of Φ then the second inequality in the chain is strict and thus $\sigma(\Phi(Y)) < \sigma(Y)$. It also follows from … that $\eta^2(\Phi(Y)) \leq 1$.

Algorithm

$$\mathcal{D}\Phi_X(H) = V^{+}\sum w_{ij}\frac{\delta_{ij}}{d_{ij}(X)}\left\{A_{ij}H - \frac{\text{tr } X'A_{ij}H}{\text{tr } X'A_{ij}X}A_{ij}X\right\}$$

Thus $\mathcal{D}\Phi_X(X) = 0$ for all X. If X is a fixed point and S is anti-symmetric $\mathcal{D}\Phi_X(XS) = V^{+}B(X)XS = XS$, which means $\mathcal{D}\Phi_X$ has $\frac{1}{2}p(p-1)$ eigenvalues equal to one. At a fixed point all eigenvalues are between zero and one.

$$\mathcal{D}^2\rho_X(G,H) = \sum w_{ij}\frac{\delta_{ij}}{d_{ij}(X)}\left\{\text{tr } G'A_{ij}H - \frac{\text{tr } H'A_{ij}X \text{ tr } G'A_{ij}X}{d_{ij}^2(X)}\right\}$$

```
## itel  57 sold 2.1114112739 snew 2.1114112739 chng 0.0000000000 labd 0.7669812392
```

stress is 2.1114112739076

```
##            [,1]
##  [1,]  +1.0000000000
##  [2,]  +0.7669965027
##  [3,]  +0.7480939418
##  [4,]  +0.7185926293
##  [5,]  +0.7007452300
##  [6,]  +0.6920114811
##  [7,]  +0.6859492532
##  [8,]  +0.6593334523
##  [9,]  +0.6541779410
## [10,]  +0.6477573342
## [11,]  +0.6237683212
## [12,]  +0.6178713315
## [13,]  +0.5735285948
## [14,]  +0.5483330654
## [15,]  +0.5260355535
## [16,]  +0.5112510731
## [17,]  +0.5064703617
## [18,]  +0.5059294793
## [19,]  +0.4919752629
## [20,]  +0.4827646549
## [21,]  +0.4782034983
## [22,]  +0.4757907684
## [23,]  +0.4682965897
## [24,]  +0.4619226490
## [25,]  +0.4559704883
## [26,]  +0.0000000000
## [27,]  +0.0000000000
## [28,]  -0.0000000000
```

$$\mathcal{D}\Phi_X(H)=V^+\sum w_{ij}\frac{\delta_{ij}}{d_{ij}(X)}\left\{A_{ij}H-\frac{\text{tr }X'A_{ij}H}{\text{tr }X'A_{ij}X}A_{ij}X\right\}$$

Thus $\mathcal{D}\Phi_X(X)=0$ for all X. If X is a fixed point and S is anti-symmetric $\mathcal{D}\Phi_X(XS)=V^+B(X)XS=XS$, which means $\mathcal{D}\Phi_X$ has $\frac{1}{2}p(p-1)$ eigenvalues equal to one. At a fixed point all eigenvalues are between zero and one.

$$\mathcal{D}^2\rho_X(G,H)=\sum w_{ij}\frac{\delta_{ij}}{d_{ij}(X)}\left\{\text{tr }G'A_{ij}H-\frac{\text{tr }H'A_{ij}X\text{ tr }G'A_{ij}X}{d^2_{ij}(X)}\right\}$$

2.2 Rotated Basic Iteration

```
## itel  54 sold 2.1114112739 snew 2.1114112739 chng 0.0000000000 labd 0.7669843896
```

stress is 2.1114112739076

```
##        [,1]
##  [1,]   +0.7669964894
##  [2,]   +0.7480939420
##  [3,]   +0.7185926297
##  [4,]   +0.7007452335
##  [5,]   +0.6920114817
##  [6,]   +0.6859492534
##  [7,]   +0.6593334543
##  [8,]   +0.6541779412
##  [9,]   +0.6477573345
## [10,]   +0.6237683217
## [11,]   +0.6178713316
## [12,]   +0.5735285959
## [13,]   +0.5483330651
## [14,]   +0.5260355534
## [15,]   +0.5112510730
## [16,]   +0.5064703618
## [17,]   +0.5059294793
## [18,]   +0.4919752632
## [19,]   +0.4827646574
## [20,]   +0.4782035029
## [21,]   +0.4757907649
## [22,]   +0.4682965885
## [23,]   +0.4619226493
## [24,]   +0.4559704884
## [25,]   -0.0000000000
## [26,]   -0.0000000000
## [27,]   +0.0000000000
## [28,]   +0.0000000000
```

```
## itel  56 sold 2.1114112739 snew 2.1114112739 chng 0.0000000000 labd 0.7669940008
```

stress is 2.1114112739076

```
##        [,1]
##  [1,]   +0.7669964993
##  [2,]   +0.7480939419
##  [3,]   +0.7185926294
##  [4,]   +0.7007452309
##  [5,]   +0.6920114813
##  [6,]   +0.6859492533
##  [7,]   +0.6593334529
##  [8,]   +0.6541779410
##  [9,]   +0.6477573343
## [10,]   +0.6237683213
```

```
## [11,]   +0.6178713317
## [12,]   +0.5735285948
## [13,]   +0.5483330653
## [14,]   +0.5260355535
## [15,]   +0.5112510731
## [16,]   +0.5064703617
## [17,]   +0.5059294792
## [18,]   +0.4919752629
## [19,]   +0.4827646550
## [20,]   +0.4782034999
## [21,]   +0.4757907672
## [22,]   +0.4682965894
## [23,]   +0.4619226495
## [24,]   +0.4559704888
## [25,]   -0.0000000006
## [26,]   -0.0000000000
## [27,]   -0.0000000000
## [28,]   +0.0000000000
```

3 Two Point Iteration

De Leeuw and Heiser (1980) suggested the update

$$\Psi(X) = 2\Phi(X) - X$$

The reasoning here is two-fold. The smacof inequality says

$$\sigma(X) \leq 1 + \eta^2(X - \Phi(Y)) - \eta^2(\Phi(Y))$$

If $X = \alpha\Phi(Y) + (1-\alpha)Y$ then this becomes

$$\sigma(\alpha\Phi(Y) + (1-\alpha)Y) \leq 1 + (1-\alpha)^2\eta^2(Y - \Phi(Y)) - \eta^2(\Phi(Y))$$

If $(1-\alpha)^2 \leq 1$ then

$$1 + (1-\alpha)^2\eta^2(Y - \Phi(Y)) - \eta^2(\Phi(Y)) \leq 1 + \eta^2(Y - \Phi(Y)) - \eta^2(\Phi(Y)) = \sigma(Y)$$

Thus updating with $X^{(k+1)} = \alpha\Phi(X^{(k)}) + (1-\alpha)X^{(k)}$ is a stable algorithm as long as $0 \leq \alpha \leq 2$.

It turns out that applying the relaxed update

$$X^{(k+1)} = 2V^+B(X^{(k)})X^{(k)} - X^{(k)}$$

has some unintended consequences.

```
## itel  23 sold 3.9946270666 snew 3.9946270666 chng 3.7664315853 labd 1.0000000000
```

stress is 2.1114112739076

We see that $\eta^2(X^{(k+1)} - X^{(k)})$ does not converge to zero, and that σ_k converges to a value which does not even correspond to a local minimum of σ.

```
##        [,1]
##  [1,]   -1.0000000000
##  [2,]   +1.0000000000
##  [3,]   -1.0000000000
##  [4,]   -1.0000000000
##  [5,]   +0.5339929781
##  [6,]   +0.4961878839
##  [7,]   +0.4371852597
##  [8,]   +0.4014904677
##  [9,]   +0.3840229636
## [10,]   +0.3718985068
## [11,]   +0.3186669088
## [12,]   +0.3083558824
## [13,]   +0.2955146690
## [14,]   +0.2475366434
## [15,]   +0.2357426633
```

```
## [16,]   +0.1470571918
## [17,]   +0.0966661301
## [18,]   -0.0880590224
## [19,]   -0.0761547015
## [20,]   -0.0634068231
## [21,]   +0.0520711068
## [22,]   -0.0484184707
## [23,]   -0.0435929937
## [24,]   -0.0344706856
## [25,]   +0.0225021460
## [26,]   -0.0160494738
## [27,]   +0.0129407235
## [28,]   +0.0118589584
```

A more thorough analysis of the results show that the method produces a sequence $X^{(k)}$ with two subseqences. If $\overline{X}$ is a fixed point of Φ then there is a $\tau > 0$ such that the subsequence with k even converges to $\tau\overline{X}$ while the subsequence with k odd converges to $(2-\tau)\overline{X}$.

```
## itel  18 sold 3.9946270666 snew 3.9946270666 chng 0.0000000000 labd 0.2737973829
```

stress is 2.1114112739076

```
##         [,1]
##   [1,]   +1.0000000000
##   [2,]   -1.0000000000
##   [3,]   -1.0000000000
##   [4,]   -0.9999999999
##   [5,]   +0.5339930271
##   [6,]   +0.4961878833
##   [7,]   +0.4371852579
##   [8,]   +0.4014904541
##   [9,]   +0.3840229612
##  [10,]   +0.3718985063
##  [11,]   +0.3186669016
##  [12,]   +0.3083558816
##  [13,]   +0.2955146680
##  [14,]   +0.2475366415
##  [15,]   +0.2357426629
##  [16,]   +0.1470571877
##  [17,]   +0.0966661315
##  [18,]   -0.0880590242
##  [19,]   -0.0761547024
##  [20,]   -0.0634068192
##  [21,]   +0.0520711070
##  [22,]   -0.0484184575
##  [23,]   -0.0435930109
##  [24,]   -0.0344706938
```

```
## [25,]   +0.0225021463
## [26,]   -0.0160494743
## [27,]   +0.0129407234
## [28,]   +0.0118589585
```

```
## Warning in microbenchmark(smacofAccelerate(delta, ndim = 2, opt = 1, halt = 2,
## : less accurate nanosecond times to avoid potential integer overflows
```

```
## Unit: milliseconds
##                                                                     expr      min
##  smacofAccelerate(delta, ndim = 2, opt = 1, halt = 2, verbose = FALSE) 3.188775
##  smacofAccelerate(delta, ndim = 2, opt = 2, halt = 2, verbose = FALSE) 3.151875
##  smacofAccelerate(delta, ndim = 2, opt = 3, halt = 2, verbose = FALSE) 3.847276
##  smacofAccelerate(delta, ndim = 2, opt = 4, halt = 2, verbose = FALSE) 1.421019
##  smacofAccelerate(delta, ndim = 2, opt = 5, halt = 2, verbose = FALSE) 1.564027
##  smacofAccelerate(delta, ndim = 2, opt = 6, halt = 2, verbose = FALSE) 1.639467
##  smacofAccelerate(delta, ndim = 2, opt = 7, halt = 2, verbose = FALSE) 1.526307
##        lq     mean   median       uq      max neval
##  3.304825 3.602914 3.358987 3.476656 6.070009   100
##  3.241993 3.521317 3.294924 3.374894 5.645618   100
##  3.939567 4.335232 4.029644 4.215682 6.634046   100
##  1.466795 1.601369 1.498222 1.541600 3.919846   100
##  1.618536 1.762409 1.650004 1.712775 3.995286   100
##  1.692829 1.827641 1.720585 1.775444 6.503174   100
##  1.578336 1.718087 1.618106 1.683132 6.439214   100
```

```
## Unit: milliseconds
##                                                                     expr      min
##  smacofAccelerate(delta, ndim = 2, opt = 1, halt = 2, verbose = FALSE) 43.06710
##  smacofAccelerate(delta, ndim = 2, opt = 2, halt = 2, verbose = FALSE) 44.09382
##  smacofAccelerate(delta, ndim = 2, opt = 3, halt = 2, verbose = FALSE) 54.69146
##  smacofAccelerate(delta, ndim = 2, opt = 4, halt = 2, verbose = FALSE) 20.16495
##  smacofAccelerate(delta, ndim = 2, opt = 5, halt = 2, verbose = FALSE) 14.30547
##  smacofAccelerate(delta, ndim = 2, opt = 6, halt = 2, verbose = FALSE) 22.89670
##  smacofAccelerate(delta, ndim = 2, opt = 7, halt = 2, verbose = FALSE) 20.67150
##        lq     mean   median       uq      max neval
##  44.02676 45.24634 44.87503 45.65168 66.89129   100
##  45.18950 46.15677 46.23545 46.68055 50.64677   100
##  56.24936 56.93894 56.64076 57.10158 74.41721   100
##  20.85358 22.01609 22.03133 22.24851 40.76454   100
##  14.63911 15.66496 14.95206 16.28350 34.73516   100
##  24.80289 25.22137 25.04815 25.37377 44.09050   100
##  22.44541 22.52770 22.71785 22.93573 24.76105   100
```

References

De Leeuw, J., and W. J. Heiser. 1980. "Multidimensional Scaling with Restrictions on the Configuration." In *Multivariate Analysis, Volume V*, edited by P. R. Krishnaiah, 501–22. Amsterdam, The Netherlands: North Holland Publishing Company.

Guttman, L. 1968. "A General Nonmetric Technique for Fitting the Smallest Coordinate Space for a Configuration of Points." *Psychometrika* 33: 469–506.