

# Convergence of SMACOF

Jan de Leeuw

First created on August 18, 2019. Last update on November 26, 2021

## Abstract

To study convergence of SMACOF we introduce a modification mSMACOF that rotates the configurations from each of the SMACOF iterations to principal components. This modification, called mSMACOF, has the same stress values as SMACOF in each iteration, but unlike SMACOF it produces a sequence of configurations that properly converges to a solution. We show that the modified algorithm can be implemented by iterating ordinary SMACOF to convergence, and then rotating the SMACOF solution to principal components. The speed of linear convergence of SMACOF and mSMACOF is the same, and is equal to the largest eigenvalue of the derivative of the Guttman transform, ignoring the trivial unit eigenvalues that result from rotational indeterminacy.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Notation</b>	<b>3</b>
<b>3</b>	<b>SMACOF</b>	<b>3</b>
<b>4</b>	<b>Global Convergence of SMACOF</b>	<b>4</b>
<b>5</b>	<b>Local Convergence of SMACOF</b>	<b>5</b>
<b>6</b>	<b>Modified SMACOF</b>	<b>7</b>
<b>7</b>	<b>Software</b>	<b>8</b>
<b>8</b>	<b>Examples</b>	<b>9</b>
8.1	Ekman . . . . .	9
8.2	De Gruijter . . . . .	11

<b>9 Conclusion</b>	<b>13</b>
<b>10 Appendix: Code</b>	<b>14</b>
10.1 smacof.R . . . . .	14
10.2 derivative.R . . . . .	15
<b>References</b>	<b>18</b>

```
## Warning in lemma_nums(name = "procrustus", display = FALSE): No caption
## supplied.

## Warning in lemma_nums(name = "solve_me", display = FALSE): No caption supplied.

## Warning in problem_nums(name = "procrustus", display = FALSE): No caption
## supplied.

## Warning in problem_nums(name = "ortho", display = FALSE): No caption supplied.

## Warning in problem_nums(name = "cmat", display = FALSE): No caption supplied.

## Warning in problem_nums(name = "solve", display = FALSE): No caption supplied.

## Warning in problem_nums(name = "rotate", display = FALSE): No caption supplied.
```

**Note:** This is a working paper which will be expanded/updated frequently. All suggestions for improvement are welcome. The directory [deleuwpdx.net/pubfolders/converge](http://deleuwpdx.net/pubfolders/converge) has a pdf version, the bib file, the complete Rmd file with the code chunks, and the R source code.

# 1 Introduction

In (Euclidean, metric, least squares) multidimensional scaling we minimize the real valued loss function

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

defined on  $\mathbb{R}^{n \times p}$ , the space of all  $n \times p$  matrices. We follow Kruskal (1964) and call  $\sigma(X)$  the *stress of configuration X*. Minimizing stress over  $p$ -dimensional configurations is the *pMDS problem*.

In (1) the matrices of *weights*  $W = \{w_{ij}\}$  and *dissimilarities*  $\Delta = \{\delta_{ij}\}$  are symmetric, non-negative, and hollow (zero diagonal). The matrix-valued function  $D(X) = \{d_{ij}(X)\}$  contains Euclidean distances between the rows of the *configuration X*, which are the coordinates of  $n$  *points* in  $\mathbb{R}^p$ . Thus  $D(X)$  is also symmetric, non-negative, and hollow.

## 2 Notation

First some convenient notation, first introduced in De Leeuw (1977). Vector  $e_i$  has  $n$  elements, with element  $i$  equal to  $+1$ , and all other elements zero.  $A_{ij}$  is the matrix  $(e_i - e_j)(e_i - e_j)'$ , which means elements  $(i, i)$  and  $(j, j)$  are equal to  $+1$ , while  $(i, j)$  and  $(j, i)$  are  $-1$ . Thus

$$d_{ij}^2(X) = (e_i - e_j)'XX'(e_i - e_j) = \text{tr } X'A_{ij}X = \text{tr } A_{ij}C, \quad (2)$$

with  $C = XX'$ .

We also define

$$V = \sum_{1 \leq i < j \leq n} w_{ij}A_{ij}, \quad (3)$$

and the matrix-valued function  $B$  with

$$B(X) = \sum_{d_{ij}(X) > 0} w_{ij} \frac{\delta_{ij}}{d_{ij}(X)} A_{ij}, \quad (4)$$

and  $B(X) = 0$  if  $X = 0$ . If we assume, without loss of generality, that

$$\frac{1}{2} \sum_{1 \leq i < j \leq n} w_{ij} \delta_{ij}^2 = 1,$$

then

$$\sigma(X) = 1 - \text{tr } X'B(X)X + \frac{1}{2} \text{tr } X'VX. \quad (5)$$

We also suppose, without loss of generality, that  $W$  is *irreducible*, so that the pMDS problem does not separate into a number of smaller pMDS problems. For symmetric matrices irreducibility means that we cannot find a permutation matrix  $\Pi$  such that  $\Pi'W\Pi$  is the direct sum of a number of smaller matrices.

$V$  is symmetric with non-positive off-diagonal elements. It is doubly-centered (rows and columns add up to zero) and thus weakly diagonally dominant. It follows that it is positive semi-definite (see Varga (1962), section 1.5). Because of irreducibility it has rank  $n - 1$ , and the vectors in its null space are all proportional to  $e$ , the vector with all elements equal to  $+1$ . The matrix  $B(X)$  is also symmetric, positive semi-definite, and doubly-centered for each  $X$ . It may not be irreducible, because for example  $B(0) = 0$ .

## 3 SMACOF

Define the *Guttman Transform* of configuration  $X$  as

$$\Gamma(X) = V^+B(X)X \quad (6)$$

The *SMACOF algorithm* is

$$X^{(k+1)} = \Gamma(X^{(k)}) = \Gamma^k(X^0) \quad (7)$$

De Leeuw (1977) shows that the SMACOF iterations (7) tend (in a specific sense described later in this paper) to a fixed point of the Guttman transform, i.e. to an  $X$  with  $\Gamma(X) = X$ . If stress is differentiable at  $X$  then the fixed points of the Guttman transform are stationary points, where the derivative of stress vanishes. Also  $V^+B(X)X = X$  shows that the columns of  $X$  are eigenvectors of  $V^+B(X)$ , with eigenvalues equal to one.

The algorithm was proposed by Guttman (1968), by differentiating stress and setting the derivatives equal to zero. This ignores the problem that stress is not differentiable if one or more distances are zero, and it also does not say anything about convergence of the algorithm. In De Leeuw (1977) a new derivation of the algorithm was given, using a majorization argument based on the Cauchy-Schwarz inequality. This make it possible to avoid problems with differentiability, and it leads to a simple convergence proof.

## 4 Global Convergence of SMACOF

Following De Leeuw (1977) we also define

$$\rho(X) = \sum_{1 \leq i < j \leq n} w_{ij} \delta_{ij} d_{ij}(X) = \text{tr } X' B(X) X, \quad (8)$$

$$\eta^2(X) = \sum_{1 \leq i < j \leq n} w_{ij} d_{ij}^2(X) = \text{tr } X' V X, \quad (9)$$

and

$$\lambda(X) = \frac{\rho(X)}{\eta(X)}. \quad (10)$$

The main result in De Leeuw (1977) is:

1. The SMACOF iterates  $X^{(k)}$  are in the compact set  $\eta(X) \leq 1$ ,
2.  $\{\rho(X^k)\}$ ,  $\{\eta(X^k)\}$  and  $\{\lambda(X^k)\}$  are increasing sequences, with limits  $\rho_\infty, \eta_\infty$  and  $\lambda_\infty$ , where  $\eta_\infty = \lambda_\infty = \sqrt{\rho_\infty}$ .
3.  $\{\sigma(X^k)\}$  is a decreasing sequence converging to  $\sigma_\infty = 1 - \eta_\infty^2$ .
4. At any accumulation point  $X_\infty$  of  $\{\sigma(X^k)\}$  we have  $\sigma(X_\infty) = \sigma_\infty$  and  $X_\infty = \Gamma(X_\infty)$ .
5.  $\eta(X^{(k+1)} - X^{(k)}) \rightarrow 0$  and thus either  $\{\sigma(X^k)\}$  converges or the set of accumulation points of  $\{\sigma(X^k)\}$  is a continuum.

So, in words, the sequence of stress values converges monotonically. The sequence of configurations is asymptotically regular and has one or more accumulation points. Each accumulation point is a fixed point of the Guttman transform. All accumulation points have the same function value, which is also the limit of the stress sequence.

But it is important to emphasize that De Leeuw (1977) did not show that the configuration sequence  $\{X^{(k)}\}$  actually a Cauchy sequence, and converges to a configuration  $X_\infty$ . This is basically because of the rotational invariance of the pMDS problem. If  $K$  is a rotation matrix, i.e.  $K'K = KK' = I$ , then  $d_{ij}(XK) = d_{ij}(X)$  for all  $i$  and  $j$ , and thus  $\sigma(XK) = \sigma(X)$ .

If  $X$  is a fixed point of the Guttman transform, then so is  $XK$ . Consequently there are no isolated fixed points, each fixed point is part of a nonlinear continuum of fixed points.

It is also of interest that De Leeuw (1984) showed that if  $X$  is a local minimizer of stress then  $d_{ij}(X) > 0$  for all  $i$  and  $j$  such that  $w_{ij}\delta_{ij} > 0$ . Thus, if weights and dissimilarities are positive, stress is differentiable at a local minimum. De Leeuw (1993) shows that stress has only a single local maximum at  $X = 0$ , and consequently the only possible stationary points are local minima and saddle points. It is shown by De Leeuw (2019) that all fixed points of the Guttman transform which are not of full column rank are saddle points. Thus if  $X$  is a local minimizer of pMDS, then adding  $q$  zero columns to  $X$  produces a saddle point  $[X \mid 0]$  for  $(p+q)$ MDS. At saddle points there are always directions in which stress can be decreased, and consequently the SMACOF algorithm with enough iterations will eventually get close to a continuum of local minima.

## 5 Local Convergence of SMACOF

De Leeuw (1988) studies the local convergence of SMACOF, i.e. the *rate of convergence* of the SMACOF iterations. There is, however, a problem not adequately addressed in that article, which has quite a bit of hand-waiving, and some incorrect statements. If there is no convergence to a single configuration then the usual rate of convergence is not defined either.

We know the sequence  $\{\eta(X^{(k+1)} - X^{(k)})\}$  tends to zero. We can measure its rate of convergence by the *ratio factor*

$$q_k = \frac{\eta(X^{(k+1)} - X^{(k)})}{\eta(X^{(k)} - X^{(k-1)})}, \quad (11)$$

and the *root factor*

$$r_k = \eta(X^{(k+1)} - X^{(k)})^{1/k}. \quad (12)$$

There is no guarantee that either  $\{q_k\}$  or  $\{r_k\}$  converges, but

$$\liminf_{k \rightarrow \infty} q_k \leq \liminf_{k \rightarrow \infty} r_k \leq \limsup_{k \rightarrow \infty} r_k \leq \limsup_{k \rightarrow \infty} q_k \quad (13)$$

Thus if  $\lim_{k \rightarrow \infty} q_k$  exists then  $\lim_{k \rightarrow \infty} r_k$  exist and the two limit values are equal. See Rudin (1976), p. 68, theorem 3.37.

The rate of convergence of the SMACOF iterations at a fixed point  $X$  is computed in De Leeuw (1988) as the spectral norm  $\kappa(X) = \|\mathcal{D}\Gamma(X)\|_\infty$ , i.e. as the modulus of the largest eigenvalue of the derivative of the Guttman transform. There are good reasons to compute this quantity. From Ostrowski's Theorem (Ostrowski (1973), theorem 22.1, p. 151, Ortega and Rheinboldt (1970), theorem 10.1.3, p. 300) we know that if  $\kappa(X) < 1$ , then  $X$  is a *point of attraction* of the SMACOF iteration. If we start close enough then the iterations will converge to  $X$ . We also know (Ostrowski (1973), theorem 22.2, p. 152) that if  $\kappa(X) > 1$  then  $X$  is a *point of repulsion*, which means the iterations will diverge when started in a certain solid angle with  $X$  at its apex.

If we define, with Ortega and Rheinboldt (1970), chapter 9, for a sequence  $\{X^{(k)}\}$  converging to  $X$ , the ratio and root convergence factors

$$Q_1(\{X^{(k)}\}) = \limsup_{k \rightarrow \infty} \frac{\eta(X^{(k+1)} - X)}{\eta(X^{(k)} - X)}, \quad (14)$$

and

$$R_1(\{X^{(k)}\}) = \limsup_{k \rightarrow \infty} \eta(X^{(k)} - X)^{1/k}, \quad (15)$$

then  $\kappa(X) < 1$  implies  $R_1(X^{(k)}) = \kappa(X)$  (Ortega and Rheinboldt (1970), theorem 10.1.4, p. 301). Also  $R_1(X^{(k)})$  is independent of the norm we have chosen, which is  $\eta$  in the SMACOF case (Ortega and Rheinboldt (1970), theorem 9.9.2, p. 288). Because  $Q_1(X^{(k)})$  depends on the norm we can only assert that  $Q_1(X^{(k)}) \geq R_1(X^{(k)}) = \kappa(X)$ , although it is guaranteed in the SMACOF case that some norm exists for which there is equality (Ortega and Rheinboldt (1970), Notes and Remarks NR 10.1-5, p. 306).

There is a problem, however, with applying the Ostrowski and Ortega-Rheinboldt results in our case. Because of the invariance of distances under rotation we have  $\kappa(X) = 1$ , no matter what  $X$  is. And, basically for the same reason, we have not shown that the SMACOF iterations converge to a fixed point at all. De Leeuw (1988) on page 171 acknowledges the problem. In fact, he proposes a way around the problem in the proof of theorem 3 on page 175, but the proof is not very convincing, although convincing enough to get past the reviewers. I will try to be more specific and precise.

Observe that

$$\mathcal{D}\Gamma(X) = V^+ \mathcal{D}^2 \rho(X), \quad (16)$$

and

$$\mathcal{D}^2 \sigma(X) = V - \mathcal{D}^2 \rho(X) = V(I - \mathcal{D}\Gamma(X)). \quad (17)$$

Because  $\rho$  is convex  $\mathcal{D}^2 \rho(X)$  is symmetric and positive semi-definite. The eigenvalues of  $\mathcal{D}\Gamma(X)$  are the generalized eigenvalues of the positive semi-definite matrix pair  $(\mathcal{D}^2 \rho(X), V)$ .

An explicit formula for the derivatives of the Guttman transform, or equivalently for the second derivatives of stress, was first given by De Leeuw (1988). We write  $\mathcal{D}\Gamma(X)[Y]$  for the derivative evaluated at  $Y$ . Partial derivatives can be obtained by choosing  $Y = e_i e'_s$ . Of course we assume here that  $d_{ij}(X) > 0$  for all  $i < j$  with  $w_{ij} \delta_{ij} > 0$ .

$$\mathcal{D}\Gamma(X)[Y] = V^+ \{B(X)Y - H(X, Y)X\}, \quad (18)$$

with

$$H(X, Y) = \sum_{1 \leq i < j \leq n} w_{ij} \frac{\delta_{ij}}{d_{ij}(X)} \frac{\text{tr } X' A_{ij} Y}{d_{ij}^2(X)} A_{ij}. \quad (19)$$

These formulas, together with equations (16) and (17), prove the main results in De Leeuw (1988).

1. All eigenvalues of  $\mathcal{D}\Gamma(X)$  are non-negative.
2. (**Homogeneity**)  $X$  is an eigenvector of  $\mathcal{D}\Gamma(X)$  with eigenvalue zero.

3. **(Translation)**  $\mathcal{D}\Gamma(X)$  has at least  $p$  additional eigenvalues equal to zero, corresponding with eigenvectors of the form  $ee'_s$ , where  $e$  has all elements equal to one and  $e_s$  has one of its elements equal to one and the others equal to zero.
4. **(Rotation)** If  $X$  is a fixed point of the Guttman transform then  $\mathcal{D}\Gamma(X)$  has at least  $\frac{1}{2}p(p-1)$  eigenvalues equal to one, corresponding with eigenvectors of the form  $XA$ , with  $A$  anti-symmetric.
5. At a local minimum point  $X$  of stress all eigenvalues of  $\mathcal{D}\Gamma(X)$  are less than or equal to one. At a strict local minimum they are strictly less than one.
6. At a saddle point  $X$  of stress the largest eigenvalue of  $\mathcal{D}\Gamma(X)$  is larger than one.

## 6 Modified SMACOF

To avoid the problems caused by a continuum of fixed points De Leeuw (1988) proposes to rotate each SMACOF iterate to a unique position, for example to principal components. For any  $X$  with singular value decomposition  $X = K\Lambda L'$  we define  $\Pi(X) = K\Lambda = XL$ . Note that if one or more singular values are equal then  $\Pi$  is not unique and becomes a point-to-set map. We can rotate within each of the spaces corresponding to multiple singular values, and thus we have not completely eliminated indeterminacy due to rotation. In the sequel we simply assume that the  $p$  singular values of  $X$  are different.

Now consider the iterations

$$\tilde{X}^{(k+1)} = \Pi(\Gamma(\tilde{X}^{(k)})) = (\Pi\Gamma)^k(\tilde{X}^0). \quad (20)$$

Let's call this modified SMACOF, or mSMACOF for short.

It looks initially as if mSMACOF require a great deal of extra computation, but this is actually not the case. If  $M$  is a rotation matrix we have  $\Pi(XM) = \Pi(X)$  and  $\Gamma(XM) = \Gamma(X)M$ . This implies that  $\Pi\Gamma\Pi(X) = \Pi\Gamma(X)$  for all  $X$ . As a result the sequences  $\{X^{(k)}\}$  and  $\{\tilde{X}^{(k)}\}$  have a simple relationship. If we start the  $\{\tilde{X}^{(k)}\}$  sequence with  $X^0$  then for each  $k$  we have  $\sigma(\tilde{X}^{(k)}) = \sigma(X^{(k)})$  and  $\tilde{X}^{(k)} = \Pi(X^{(k)})$ . Thus the two sequences of stress values are exactly the same for SMACOF and mSMACOF and the two configurations, and consequently also the accumulation points of the two sequences of configurations, just differ by a rotation. To get  $\tilde{X}^{(k)}$  we can just compute the SMACOF iterate  $X^{(k)}$  and rotate it to principal components.

The derivative of  $\Pi\Gamma$  is, by the chain rule,

$$\mathcal{D}\Pi\Gamma(X) = \mathcal{D}\Pi(\Gamma(X))\mathcal{D}\Gamma(X). \quad (21)$$

After some computation we find

$$\mathcal{D}\Pi(X)[Y] = YL + K\Lambda M, \quad (22)$$

where  $X = K\Lambda L'$  and  $M$  is the anti-symmetric matrix with off-diagonal elements

$$m_{ij} = -\frac{\lambda_i u_{ij} + \lambda_j u_{ji}}{\lambda_i^2 - \lambda_j^2} \quad (23)$$

with  $U = K'YL$ .

If  $X$  satisfies  $\Pi(X) = X$  then  $L = I$  and thus  $\mathcal{D}\Pi(X)[Y] = Y + XM$  and  $U = K'Y$ . We proceed to compute the eigenvectors and eigenvalues of the Jacobian at an  $X$  with  $\Pi(X) = X$ .

1. If  $Y = XA$  with  $A$  antisymmetric then  $M = -L'AL$  and  $\mathcal{D}\Pi(X)[Y] = 0$ . This corresponds with  $\frac{1}{2}p(p-1)$  zero eigenvalues
2. If  $Y = K\Lambda^{-1}A$  with  $A$  antisymmetric then  $M = 0$ . This gives  $\frac{1}{2}p(p-1)$  eigenvectors with eigenvalue one.
3. If  $Y = K\Lambda^{-1}D$  with  $D$  diagonal then  $M = 0$ . This gives another  $p$  eigenvectors with eigenvalue one.
4. If  $Y = K_{\perp}S$  with  $K_{\perp}$  a basis for the orthogonal complement of  $X$  then  $M = 0$  and we have another  $p(n-p)$  eigenvalues equal to one.

This is a complete set of eigenvectors. It turns out all eigenvalues are equal to one, except for  $\frac{1}{2}p(p-1)$  zero eigenvalues. It follows that fixed points of  $\Pi\Gamma$  are of the form  $\tilde{X} = XL$ , with  $X$  a fixed point of  $\Gamma$  and  $L$  the rotation of  $X$  to principal components). Thus  $\Gamma(\tilde{X}) = \Gamma(X)L = XL = \tilde{X}$ , and  $\tilde{X}$  is a fixed point of both  $\Gamma$  and  $\Pi\Gamma$ . The eigenvalues of  $\mathcal{D}\Pi\Gamma(\tilde{X})$  are the same as the eigenvalues of  $\mathcal{D}\Gamma(X)$ , except for the  $\frac{1}{2}p(p-1)$  eigenvalues equal to one, corresponding with  $XA$  for antisymmetric  $A$ , which are replaced by zero eigenvalues. The Ostrowski and Ortega-Rheinboldt results now apply directly to  $\Pi\Gamma$ . If there are no zero distances, if the singular values of  $X$  are different, and if the largest eigenvalue of  $\mathcal{D}\Pi\Gamma(X)$  is less than one, then  $X$  is a point of attraction, the SMACOF iterations converge to  $X$  if started sufficiently close to it (in the *attraction ball* of  $X$ ), and  $R_1(X)$  is equal to the largest eigenvalue.

## 7 Software

The appendix has an ad-hoc version of the `smacof` function. Its arguments are

```
args(smacof)
```

```
## function (delta, w, p = 2, xold = torgerson(delta, p), pca = FALSE,
##     verbose = FALSE, eps = 1e-15, itmax = 10000)
## NULL
```

The function expects both dissimilarities and weights to be of class `dist`. The default initial configuration uses classical MDS (Torgerson (1958)). We iterate until either the distance  $\eta(X^{(k)} - X^{(k-1)})$  between successive configurations is less than `eps`, which has the ridiculously small default value of  $10^{-15}$ , or until the maximum number of iterations `itmax` is reached, which defaults to the ridiculously large default value of 10000. If `pca` is `TRUE` all iterates



are rotated to principal components. If `verbose` is TRUE we print, for each iteration, the iteration number  $k$ , the stress  $\sigma(X^{(k)})$ , the change  $\eta(X^{(k)} - X^{(k-1)})$ , and the root and ratio convergence factors.

The second file `derivative.R` computes the Jacobian of the iteration functions at the limit (i.e. at the point where SMACOF stops). There are actually six functions in the file. The three functions `dGammaA`, `dPiA` and `dPiGammaA` use the analytical expressions we have derived earlier for  $\mathcal{D}\Gamma(X)$ ,  $\mathcal{D}\Pi(X)$  and  $\mathcal{D}\Pi\Gamma(X)$ . The corresponding functions `dGammaN`, `dPiN` and `dPiGammaN` numerically compute the Jacobian using the `jacobian` function from the `numDeriv` package (Gilbert and Varadhan (2019)). They are basically used to check the formulas.

## 8 Examples

### 8.1 Ekman

Time for a numerical example. We will use the color similarity data of Ekman (1954), taken from the SMACOF package (De Leeuw and Mair (2009)). We transform Ekman's similarities  $s_{ij}$  to dissimilarities  $\delta_{ij}$  using  $\delta_{ij} = (1 - s_{ij})^3$ , and we use unit weights in  $W$ .

SMACOF requires 52 iterations for convergence. The minimum of stress is 0.0110248119, the root factor is equal to 0.5099244097 and the ratio factor is 0.6093903884. The eigenvalues of  $\mathcal{D}\Gamma(X)$  are

```
## 0.9999999999999999
## 0.538510668196408
## 0.532498554224166
## 0.529669334191043
## 0.525541097754551
## 0.519602718218806
## 0.516533687841055
## 0.513727908691609
## 0.510295310409166
## 0.506357695934493
## 0.495649713446001
## 0.481518780073304
## 0.469548625536452
## 0.445038589020916
## 0.410479252654483
## 0.394284315478173
## 0.357670750114585
## 0.348395413045201
## 0.330341477528788
## 0.313520384427864
```

```
## 0.287598015841956
## 0.280399104121623
## 0.267278474596759
## 0.247631495462991
## 0.216576009083046
## 0.000000000000000
## 0.000000000000000
## 0.000000000000000
```

As an aside, the Ekman example (with this particular transformation of the similarities) is special, because the two-dimensional solution is actually the global minimum over all configurations (i.e. the global minimum of pMDS for all  $1 \leq p \leq n$ ). As shown in De Leeuw (2014) this follows from the fact that the two unit eigenvalues of  $V^+B(X)$  are actually its two largest eigenvalues, and consequently  $X$  is the solution of the convex full-dimensional nMDS relaxation of the pMDS problem (De Leeuw, Groenen, and Mair (2016)). The eigenvalues of  $V^+B(X)$  are

```
## 1.000000000000000
## 0.999999999999999
## 0.923497086367287
## 0.907901212970889
## 0.862936584878895
## 0.852692003103344
## 0.829803620857356
## 0.814556167662381
## 0.793238576338501
## 0.791651722456648
## 0.786442678118770
## 0.747679475705025
## 0.728268247434340
## 0.000000000000001
```

We now set `pca=TRUE` and run `mSMACOF`. It requires 52 iterations for convergence. The minimum of stress is 0.0110248119, and the root factor is equal to 0.5099406626 and the ratio factor is 0.6146787643. The eigenvalues of  $\mathcal{D}\Pi\Gamma(X)$  at the fixed point  $X$  are

```
## 0.538510668196406
## 0.532498554224166
## 0.529669334191042
## 0.525541097754553
## 0.519602718218806
## 0.516533687841055
## 0.513727908691609
## 0.510295310409167
```

```

## 0.506357695934493
## 0.495649713446000
## 0.481518780073304
## 0.469548625536451
## 0.445038589020917
## 0.410479252654483
## 0.394284315478172
## 0.357670750114584
## 0.348395413045201
## 0.330341477528788
## 0.313520384427864
## 0.287598015841956
## 0.280399104121622
## 0.267278474596759
## 0.247631495462990
## 0.216576009083047
## 0.000000000000001
## 0.000000000000001
## 0.000000000000000
## 0.000000000000000

```

They are equal to the eigenvalues of  $\mathcal{D}\Gamma(X)$  and the root convergence factor  $R_1(\{X^{(k)}\})$  is 0.5385106682.

## 8.2 De Gruijter

The second example are dissimilarities between nine Dutch political parties, collected a long time ago by De Gruijter (1967).

```

##      KVP PvdA  VVD  ARP  CHU  CPN  PSP   BP
## PvdA 2.63
## VVD  2.27 3.72
## ARP  1.60 2.64 2.46
## CHU  1.80 3.22 1.97 0.20
## CPN  4.54 2.12 5.13 4.84 4.80
## PSP  3.73 1.59 4.55 3.73 4.08 1.08
## BP   4.18 4.22 3.90 4.28 3.96 3.34 3.88
## D66  3.17 2.47 1.67 3.13 3.04 4.42 3.36 4.36

```

We analyze the data with SMACOF using three dimensions.

SMACOF requires 779 iterations for convergence. The minimum of stress is 0.003442194, the root factor is equal to 0.9565805061, and the ratio factor is 0.9580406219. The eigenvalues of  $\mathcal{D}\Gamma(X)$  are

```

## 1.0000000000000000
## 1.0000000000000000
## 1.0000000000000000
## 0.965505429805660
## 0.940592046981168
## 0.919047686446446
## 0.863993920924432
## 0.822263696226350
## 0.810020971414307
## 0.771947328045072
## 0.728539213663602
## 0.712621684571534
## 0.659318624263221
## 0.638485365688917
## 0.624552464148481
## 0.616085432872672
## 0.557480497708779
## 0.501954186928525
## 0.458630667850016
## 0.450598367846592
## 0.355243400669833
## 0.309186479174060
## 0.247708397109092
## 0.0000000000000002
## 0.0000000000000001
## 0.0000000000000000
## 0.0000000000000000

```

In this case there is no guarantee that we have the three-dimensional global minimum. The unit eigenvalues of the matrix  $V^+B(X)$  are not the three largest ones.

```

## 1.079524009371954
## 1.032606649163671
## 1.0000000000000000
## 1.0000000000000000
## 1.0000000000000000
## 0.986706272372898
## 0.971839080877692
## 0.906211919383163
## 0.0000000000000002

```

mSMACOF in three dimensions requires 783 iterations for convergence. The minimum of stress is 0.003442194, the root factor is equal to 0.9568196354, and the ratio factor is 0.8713099253. The eigenvalues of  $\mathcal{D}\Pi\Gamma(X)$  are

```
## 0.965505429805661
## 0.940592046981166
## 0.919047686446444
## 0.863993920924432
## 0.822263696226350
## 0.810020971414307
## 0.771947328045072
## 0.728539213663603
## 0.712621684571535
## 0.659318624263221
## 0.638485365688918
## 0.624552464148482
## 0.616085432872673
## 0.557480497708779
## 0.501954186928524
## 0.458630667850017
## 0.450598367846592
## 0.355243400669832
## 0.309186479174062
## 0.247708397109091
## 0.0000000000000001
## 0.0000000000000001
## 0.0000000000000001
## 0.0000000000000000
## 0.0000000000000000
## 0.0000000000000000
## 0.0000000000000000
## 0.0000000000000000
```

Again, as expected, the eigenvalues for SMACOF and mSMACOF are the same and the root convergence factor is the spectral norm of the mSMACOF derivative, which is 0.9655054298.

## 9 Conclusion

From a practical point of view there is no need to ever use modified SMACOF. We only use the Guttman transform, and compute the largest eigenvalue of its derivative at the solution  $X$ , ignoring the  $\frac{1}{2}p(p-1)$  trivial unit eigenvalues. That largest eigenvalue, if it is strictly smaller than one, is the root convergence factor. In that case SMACOF can be said to converge to  $\Pi(X)$ , which is the SMACOF solution rotated to principal components.

## 10 Appendix: Code

### 10.1 smacof.R

```
torgerson <- function (delta, p = 2) {
  h <- as.matrix(delta ^ 2)
  n <- nrow (h)
  j <- diag (n) - (1 / n)
  h <- -(j %*% h %*% j) / 2
  e <- eigen (h)
  return (e$vector[, 1:p] %*% diag(sqrt (e$values[1:p])))
}

smacof <-
function (delta,
        w,
        p = 2,
        xold = torgerson (delta, p),
        pca = FALSE,
        verbose = FALSE,
        eps = 1e-15,
        itmax = 10000) {
  n <- round ((1 + sqrt (1 + 8 * length (delta))) / 2)
  v <- -as.matrix(w)
  diag(v) <- -rowSums(v)
  vinv <- solve(v + (1 / n)) - (1 / n)
  delta <- delta / sqrt (sum (w * delta ^ 2) / 2)
  dold <- dist (xold)
  sold <- sum (w * (delta - dold) ^ 2) / 2
  eold <- Inf
  itel <- 1
  repeat {
    b <- as.matrix (-w * delta / dold)
    diag (b) <- -rowSums(b)
    xnew <- vinv %*% b %*% xold
    if (pca) {
      xsvd <- svd (xnew)
      xnew <- xnew %*% xsvd$v
    }
    dnew <- dist (xnew)
    snew <- sum (w * (delta - dnew) ^ 2) / 2
    enew <- sqrt (sum (v * tcrossprod(xold - xnew)))
    rnew <- enew ^ (1 / itel)
  }
}
```

```

qnew <- enew / eold
if (verbose) {
  cat(
    "itel ",
    formatC(itel, digits = 4, format = "d"),
    "loss ",
    formatC(snew, digits = 15, format = "f"),
    "chan ",
    formatC(enew, digits = 15, format = "f"),
    "rcnf ",
    formatC(rnew, digits = 15, format = "f"),
    "qcnf ",
    formatC(qnew, digits = 15, format = "f"),
    "\n"
  )
}
if ((enew < eps) || (itel == itmax))
  break
xold <- xnew
dold <- dnew
sold <- snew
eold <- enew
itel <- itel + 1
}
out <-
  list (
    itel = itel,
    x = xnew,
    s = snew,
    q = qnew,
    r = rnew,
    b = vinv %*% b
  )
return (out)
}

```

## 10.2 derivative.R

```

dGammaA <- function (x, delta, w) {
  n <- nrow (x)
  p <- ncol (x)
  h <- matrix (0, n * p, n * p)

```

```

d <- dist (x)
delta <- delta / sqrt (sum (w * delta ^ 2) / 2)
dmat <- as.matrix (d)
wmat <- as.matrix (w)
emat <- as.matrix (delta)
v <- -wmat
diag(v) <- -rowSums (v)
b <- as.matrix(-w * delta / d)
diag(b) <- -rowSums (b)
vinv <- solve(v + (1 / n)) - (1 / n)
for (s in 1:p) {
  for (t in 1:p) {
    gst <- matrix (0, n, n)
    for (i in 1:n) {
      for (j in 1:n) {
        if (i == j)
          next
        gs <- x[i, s] - x[j, s]
        gt <- x[i, t] - x[j, t]
        gst[i, j] <-
          -wmat[i, j] * emat[i, j] * gs * gt / (dmat[i, j] ^ 3)
      }
    }
    diag(gst) <- -rowSums (gst)
    h[(s - 1) * n + 1:n, (t - 1) * n + 1:n] <- -vinv %*% gst
  }
}
for (s in 1:p) {
  kn <- (s - 1) * n + 1:n
  h[kn, kn] <- h[kn, kn] + vinv %*% b
}
return (h)
}

dPiA <- function (x) {
  n <- nrow (x)
  p <- ncol (x)
  sx <- svd (x)
  xu <- sx$u
  xv <- sx$v
  xd <- sx$d
  h <- matrix (0, n * p, n * p)
  for (s in 1:p) {
    for (i in 1:n) {

```



```

    ir <- (s - 1) * n + i
    e <- matrix (0, n, p)
    m <- matrix (0, p, p)
    e[i, s] <- 1
    u <- crossprod (xu, e %*% xv)
    for (k in 1:p) {
      for (l in 1:p) {
        if (k == l)
          next
        m[k, l] <-
          (xd[k] * u[k, l] + xd[l] * u[l, k]) / (xd[k] ^ 2 - xd[l] ^ 2)
      }
    }
    h[, ir] <- as.vector (e %*% xv - xu %*% diag (xd) %*% m)
  }
}
return (h)
}

```

```

dGammaN <- function (x, delta, w) {
  n <- nrow (x)
  p <- ncol (x)
  delta <- delta / sqrt (sum (w * delta ^ 2) / 2)
  v <- - as.matrix (w)
  diag (v) <- - rowSums(v)
  vinv <- solve (v + (1 / n)) - (1 / n)
  guttman <- function (x) {
    d <- dist (matrix (x, n, p))
    b <- - as.matrix (w * delta / d)
    diag (b) <- - rowSums (b)
    z <- vinv %*% b %*% matrix (x, n, p)
    return (as.vector (z))
  }
  return (jacobian (guttman, as.vector(x)))
}

```

```

dPiGammaN <- function (x, delta, w) {
  n <- nrow (x)
  p <- ncol (x)
  delta <- delta / sqrt (sum (w * delta ^ 2) / 2)
  v <- - as.matrix (w)
  diag (v) <- - rowSums(v)
  vinv <- solve (v + (1 / n)) - (1 / n)
  guttman <- function (x) {

```

```

    d <- dist (matrix (x, n, p))
    b <- - as.matrix (w * delta / d)
    diag (b) <- - rowSums (b)
    z <- vinv %% b %% matrix (x, n, p)
    z <- z %% svd(z)$v
    return (as.vector (z))
  }
  return (jacobian (guttman, as.vector(x)))
}

dPiGammaA <- function (x, delta, w) {
  n <- nrow (x)
  guttman <- function (x, delta, w) {
    d <- dist (x)
    b <- as.matrix (- w * delta / d)
    diag (b) <- - rowSums (b)
    v <- - as.matrix (w)
    diag (v) <- - rowSums (v)
    vinv <- solve(v + (1 / n)) - (1 / n)
    return (vinv %% b %% x)
  }
  h <- dGammaA (x, delta, w)
  g <- dPiA (guttman (x, delta, w))
  return (g %% h)
}

```

## References

- De Gruijter, D. N. M. 1967. “The Cognitive Structure of Dutch Political Parties in 1966.” Report E019-67. Psychological Institute, University of Leiden.
- De Leeuw, J. 1977. “Applications of Convex Analysis to Multidimensional Scaling.” In *Recent Developments in Statistics*, edited by J. R. Barra, F. Brodeau, G. Romier, and B. Van Cutsem, 133–45. Amsterdam, The Netherlands: North Holland Publishing Company.
- . 1984. “Differentiability of Kruskal’s Stress at a Local Minimum.” *Psychometrika* 49: 111–13.
- . 1988. “Convergence of the Majorization Method for Multidimensional Scaling.” *Journal of Classification* 5: 163–80.
- . 1993. “Fitting Distances by Least Squares.” Preprint Series 130. Los Angeles, CA: UCLA Department of Statistics.
- . 2014. “Bounding, and Sometimes Finding, the Global Minimum in Multidimensional Scaling.” UCLA Department of Statistics.
- . 2019. “Fitting Distances by Least Squares.” 2019.

- De Leeuw, J., P. Groenen, and P. Mair. 2016. “Full-Dimensional Scaling.” 2016.
- De Leeuw, J., and P. Mair. 2009. “Multidimensional Scaling Using Majorization: SMACOF in R.” *Journal of Statistical Software* 31 (3): 1–30.
- Ekman, G. 1954. “Dimensions of Color Vision.” *Journal of Psychology* 38: 467–74.
- Gilbert, P., and R. Varadhan. 2019. *numDeriv: Accurate Numerical Derivatives*. <https://CRAN.R-project.org/package=numDeriv>.
- Guttman, L. 1968. “A General Nonmetric Technique for Fitting the Smallest Coordinate Space for a Configuration of Points.” *Psychometrika* 33: 469–506.
- Kruskal, J. B. 1964. “Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis.” *Psychometrika* 29: 1–27.
- Ortega, J. M., and W. C. Rheinboldt. 1970. *Iterative Solution of Nonlinear Equations in Several Variables*. New York, N.Y.: Academic Press.
- Ostrowski, A. M. 1973. *Solution of Equations in Euclidean and Banach Spaces*. Third Edition of Solution of Equations and Systems of Equations. Academic Press.
- Rudin, W. 1976. *Principles of Mathematical Analysis*. Third Edition. McGraw-Hill.
- Torgerson, W. S. 1958. *Theory and Methods of Scaling*. New York: Wiley.
- Varga, R. S. 1962. *Matrix Iterative Analysis*. Englewood Cliffs: Prentice Hall.