Full-dimensional Scaling

Jan de Leeuw, Patrick Groenen, Patrick Mair

Version 005, January 25, 2016

Contents

Note: This is a working paper which will be expanded/updated frequently. The directory deleeuwpdx.net/pubfolders/full has a pdf copy of this article, the complete Rmd file with all code chunks, the bib file, the LaTeX style file, and the R source code.

1 Problem

We use the abbreviation *MDS* for the minimization of *Kruskal's stress*

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

over the $n \times p$ *configurations* X (Kruskal (1964a), Kruskal (1964b)). Symmetric matrix $W = \{w_{ij}\}$ has known non-negative *weights*, symmetric matrix $\Delta = \{\delta_{ij}\}$ has non-negative dissimilarities, and the symmetric matrix-valued function $D(X) = \{d_{ij}(X)\}$ has Euclidean distances between the *points*, i.e. the rows of X.

In the usual applications of MDS the user chooses *dimensionality p* to be much smaller than n, because the name of the game is *dimensionality reduction*. In this paper we give a detailed analysis of *full dimensional scaling* or *FDS*, which is MDS with $p = n$.

2   Notation

We introduce some notation which is useful for all Euclidean MDS problems, not just FDS problems. This notation has been more or less standard since De Leeuw (1977). Suppose the e_i are unit vectors of length n, i.e. e_i has all its elements equal to zero except element i, which is equal to one. Also define the matrices $A_{ij} := (e_i - e_j)(e_i - e_j)'$. Thus A_{ij} has four nonzero elements: elements (i, i) and (j, j) are $+1$, elements (i, j) and (j, i) are -1. With these definitions

$$d_{ij}^2(X) = (e_i - e_j)'XX'(e_i - e_j) = \mathbf{tr}\ X'A_{ij}X = \mathbf{tr}\ A_{ij}XX'. \tag{2}$$

Now expand stress. If we assume without loss of generality that

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

and we define

$$\rho(X) := \sum\sum_{1 \le i < j \le n} w_{ij}\delta_{ij}d_{ij}(X), \tag{3}$$

$$\eta(X) := \sum\sum_{1 \le i < j \le n} w_{ij}d_{ij}^2. \tag{4}$$

Now

$$\sigma(X) = 1 - \rho(X) + \frac{1}{2}\eta(X). \tag{5}$$

To develop some convenient matrix notation we define

$$V := \sum\sum_{1 \le i < j \le n} w_{ij}A_{ij}, \tag{6}$$

$$B(X) := \sum\sum_{1 \le i < j \le n} \begin{cases} w_{ij}\frac{\delta_{ij}}{d_{ij}(X)}A_{ij} & \text{if } d_{ij}(X) > 0, \\ 0 & \text{if } d_{ij}(X) = 0, \end{cases} \tag{7}$$

so that

$$\rho(X) = \mathbf{tr}\ X'B(X)X, \tag{8}$$

$$\eta(X) = \mathbf{tr}\ X'VX. \tag{9}$$

Note that V has off-diagonal elements equal to $-w_{ij}$ and $B(X)$ has off-diagonal elements

$$b_{ij}(X) = -w_{ij}\frac{\delta_{ij}}{d_{ij}(X)}$$

if $d_{ij}(X) > 0$ and zero otherwise. Diagonal elements are filled to make rows and columns sum to zero. Both V and $B(X)$ are doubly-centered and positive semi-definite. If we assume without loss of generality that the MDS problem is *irredicible* (De Leeuw (1977)) then $\mathbf{rank}(V) = n - 1$ and the only vectors in the null space are multiples of $e := (1, 1, \cdots, 1)$. If all weights are eual to one then $V = nI - ee'$.

3 Local minima of stress

De Leeuw (1984) proves a key necessary condition for a local minimum in MDS. Also see De Leeuw, Groenen, and Mair (2016) for a more recent proof. We also review some additional results on maxima and minima of stress.

Theorem 1: [De Leeuw 84, De Leeuw 93, De Leeuw and Groenen 97]

1. If σ has a local minimum at X then $d_{ij}(X) > 0$ for all i, j such that $w_{ij}\delta_{ij} > 0$.
2. If σ has a local minimum at X then σ is differentiable in a neighborhood of X. Thus $B(X)X = VX$.
3. If σ has a local minimum at a column-centered X and $\mathbf{rank}(X) = n-1$ then $\sigma(X) = 0$.
4. σ has a local maximum at X if and only if $X = 0$.

Let's look at a small example with four points, all dissimilarities equal, and all weights equal to one. There is a local minimum with four points in the corners of a square, with stress equal to 0.0285955. And there is another local minimum with three points forming an equilateral triangle, and the fourth point in the center. This has stress 0.0669873. We can compute stress for all points of the form $\alpha X + \beta Y$, where X and Y are the two local minima. Figure 1 has a contour plot of $\sigma(\alpha, \beta)$, showing the local minima at $(1, 0)$ and $(0, 1)$, and the local maximum at $(0, 0)$.
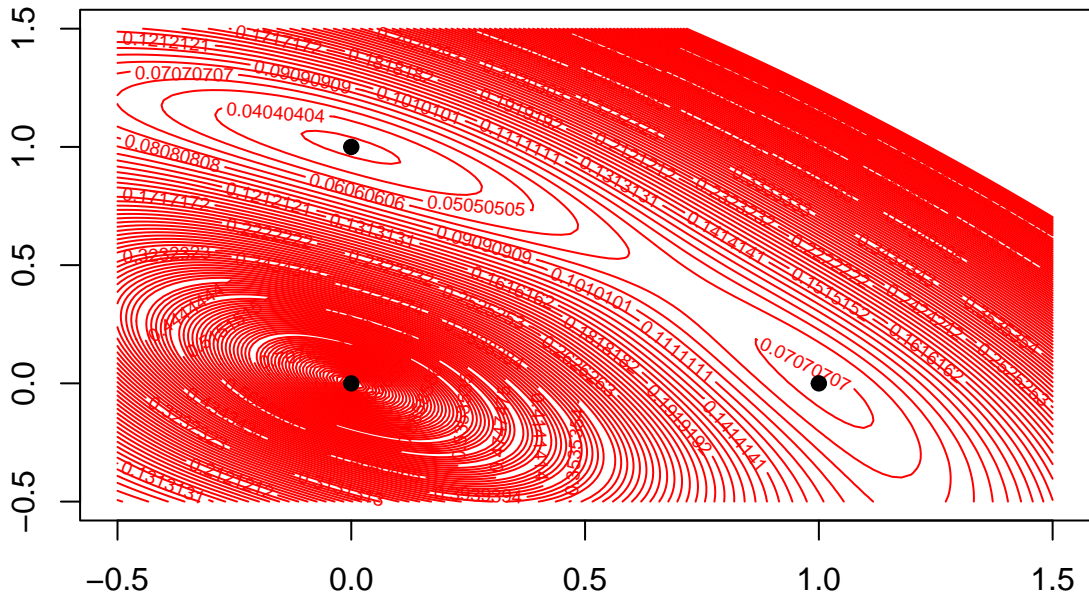
Figure 1: Plane spanned by two local minima, equal dissimilarities

Alternatively, we can plot stress on the line connecting X and Y. Note that although stress only has a local maximum at the origin in configuration space, it can very well have local maxima if restricted to lines.
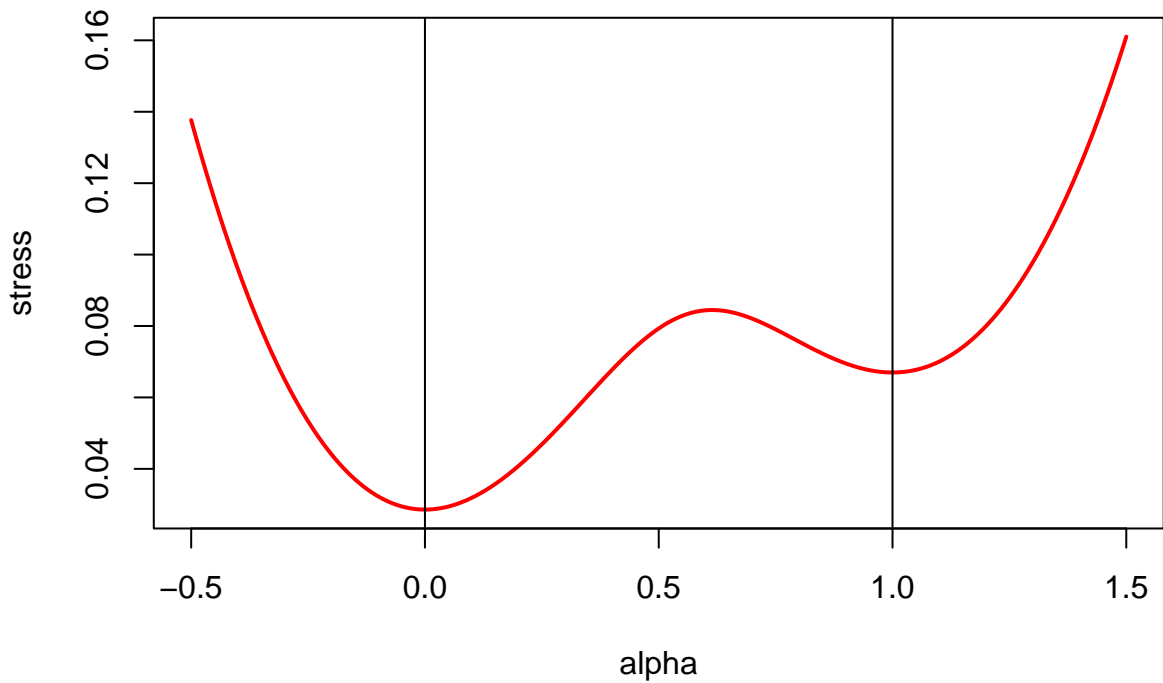

Figure 2: Line connecting two local minima, equal dissimilarities

4 Cross Product Space

So far we have formulated the MDS problem in *configuration space*. Stress is a function of X, the $n \times p$ configuration matrix. We now consider an alternative formulation, where stress is a function of a positive semi-definite C or order n. The relevant definitions are

$$\sigma(C) := 1 - 2\rho(C) + \eta(C), \tag{10}$$

where

$$\rho(C) := \mathbf{tr}\ B(C)C,$$
$$\eta(C) := \mathbf{tr}\ VC,$$

with

$$B(C) := \sum\sum_{1 \leq i < j \leq n} \begin{cases} w_{ij} \frac{\delta_{ij}}{d_{ij}(C)} A_{ij} & \text{if } d_{ij}(C) > 0, \\ 0 & \text{if } d_{ij}(C) = 0. \end{cases}$$

and $d_{ij}^2(C) := \mathbf{tr}\ A_{ij}C$.

We call the space of all positive semi-definite $n \times n$ matrices *cross product space*. The problem of minimizing σ over $n \times p$-dimensional configuration space is equivalent to the problem of minimizing σ over the set of matrices C in $n \times n$-dimensional cross product space that have rank less than or equal to p. The corresponding solutions are related by the simple relationship $C = XX'$.

Theorem 2: [De Leeuw 93] Stress is convex on cross product space.

Proof: First, η is linear in C. Second,

$$\rho(C) = \sum\sum_{1 \leq i < j \leq n} w_{ij}\delta_{ij}\sqrt{\mathbf{tr}\ A_{ij}C}.$$

This is the weighted sum of square roots of non-negative functions that are linear in C, and it is consequently concave. Thus σ is convex. **QED**

Unfortunately the subset of cross product space of all matrices with rank less than or equal to p is far from simple (see Datorro (2015)), so computational approaches to MDS prefer to work in configuration space.

5 Full-dimensional Scaling

Cross product space, the set of all positive semi-definite matrices, is a closed convex cone $\mathcal{K}$ in the linear space of all $n \times n$ symmetric matrices. This has an interesting consequence.

Theorem 3: [De Leeuw 93] Full-dimensional scaling, i.e. minimizing σ over $\mathcal{K}$, is a convex programming problem. Thus in FDS all local minima are global. If $w_{ij}\delta_{ij} > 0$ for all i, j then the minimum is unique.

This result has been around since about 1985. De Leeuw (1993) gives a proof, but the report it appeared in remained unpublished. A published proof is in "Inverse Multidimensional Scaling" (2007). Another treatment of FDS, with a somewhat different emphasis, is in De Leeuw (2014).

Now, by a familiar theorem (Theorem 31.4 in Rockafellar (1970)), a matrix C minimizes σ over $\mathcal{K}$ if and only if

$$C \in \mathcal{K}, \tag{11}$$
$$V - B(C) \in \mathcal{K}, \tag{12}$$
$$\mathbf{tr}\ C(V - B(C)) = 0. \tag{13}$$

We give a computational proof of this result for FDS that actually yields a bit more.

Theorem 4: [Expansion] For $\Delta \in \mathcal{K}$ we have

$$\sigma(C + \epsilon\Delta) = \sigma(C) - 2\epsilon^{\frac{1}{2}} \sum_{\mathbf{tr}\ A_i C = 0} w_i \delta_i \sqrt{\mathbf{tr}\ A_i \Delta} + \epsilon\ \mathbf{tr}\ (V - B(C))\Delta + o(\epsilon). \tag{14}$$

Proof: Simple expansion. **QED**

Theorem 5: [Convex]: Suppose C is a solution to the problem of minimizing σ over $\mathcal{K}$. Then

1. $\mathbf{tr}\ A_{ij} C > 0$ for all i, j for which $w_{ij}\delta_{ij} > 0$.
2. $V - B(C)$ is positive semi-definite.
3. $\mathbf{tr}\ C(V - B(C)) = 0$.
4. If C is positive definite then $V = B(C)$ and $\sigma(C) = 0$.

Proof: The $\epsilon^{\frac{1}{2}}$ term in (14) needs to vanish at a local minimum. This proves the first part. It follows that at a local minimum

$$\sigma(C + \epsilon\Delta) = \sigma(C) + \epsilon\ \mathbf{tr}\ (V - B(C))\Delta + o(\epsilon).$$

If $V - B(C)$ is not positive semi-definite, then there is a $\Delta \in \mathcal{K}$ such that $\mathbf{tr}\ (V - B(C))\Delta < 0$. Thus C cannot be the minimum, which proves the second part. If we choose $\Delta = C$ we find

$$\sigma((1 + \epsilon)C) = \sigma(C) + \epsilon\ \mathbf{tr}\ (V - B(C))C + o(\epsilon).$$

and choosing ϵ small and negative shows we must have $\mathbf{tr}\ (V - B(C))C = 0$ for C to be a minimum. This proves the third part. Finally, if σ has a minimum at C, and C is positive definite, then from parts 2 and 3 we have $V = B(C)$. Comparing off-diagonal elements shows $\Delta = D(C)$, and thus $\sigma(C) = 0$. **QED**

If C is the solution of the FDS problem, then $\mathbf{rank}(C)$ defines the *Gower rank* of the dissimilarities. The number of positive eigenvalues of the negative of the doubly-centered

matrix of squared dissimilarities, the matrix factored in classical MDS, defines the *Torgerson rank* of the dissimilarities. The *Gower conjecture* is that the Gower rank is less than or equal to the Torgerson rank. No proof and no counter examples have been found.

We compute the FDS solution using the smacof algorithm

$$X^{(k+1)} = V^+ B(X^{(k)}) \tag{15}$$

in the space of all $n \times n$ configurations, using the identity matrix as a default starting point. Since we work in configuration space, not in crossproduct space, this does not guarantee convergence to the unique FDS solution, but after convergence we can easily check the necessary and sufficient conditions of theorem 5.

As a small example, consider four points with all dissimilarities equal to one, except δ_{14} which is equal to three. Clearly the triangle inequality is violated, and thus there certainly is no perfect fit mapping into Euclidean space.

The FDS solution turns out to have rank two, thus the Gower rank is two. The singular values of the FDS solution are

```
## [1]  0.4508464709 0.2125310645 0.0000001303
```

Gower rank two also follows from the eigenvalues of the matrix $B(C)$, which are

```
## [1]  1.0000000000 1.0000000000 0.9205543464
```

6 Another planar example

Suppose X and Y are two linearly independent configurations and consider all configurations $Z = \alpha X + \beta Y$ in the plane spanned by X and Y. Define

$$B(\gamma) := \sum\sum_{1 \leq i < j \leq n} \begin{cases} w_{ij} \frac{\delta_{ij}}{d_{ij}(\alpha X + \beta Y)} \begin{bmatrix} \mathbf{tr}\ X'A_{ij}X & \mathbf{tr}\ X'A_{ij}Y \\ \mathbf{tr}\ Y'A_{ij}X & \mathbf{tr}\ Y'A_{ij}Y \end{bmatrix} & \text{if } d_{ij}(\alpha X + \beta Y) > 0, \\ 0 & \text{if } d_{ij}(\alpha X + \beta Y) = 0, \end{cases}$$

where $\gamma := \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$. Also

$$U := \begin{bmatrix} \mathbf{tr}\ X'VX & \mathbf{tr}\ X'VY \\ \mathbf{tr}\ Y'VX & \mathbf{tr}\ Y'VY \end{bmatrix}.$$

Then

$$\sigma(\gamma) := 1 - 2\gamma'B(\gamma)\gamma + \gamma'U\gamma, \tag{16}$$

and the stationary equations are

$$B(\gamma)\gamma = U\gamma.$$

We have already seen an example of σ on what can be called *combination space* in figure 1, where X and Y are two local minima. Another interesting example uses a configuration X and its gradient $Y = \mathcal{D}\sigma(X)$. In that case finding the minimum of σ over combination space means finding the optimal step size in the steepest descent gradient method.

For a configuration with three points equally spaced on a horizontal line and the fourth point above the middle of the three

```
##                [,1]        [,2]
## [1,] -0.1006209 -0.1006209
## [2,]  0.0000000 -0.1006209
## [3,]  0.1006209 -0.1006209
## [4,]  0.0000000  0.3018627
```

we have gradient

```
##            [,1]          [,2]
## 1   0.3250910 -0.08772788
## 2   0.3638044 -0.07804082
## 3  -0.6888953 -0.08772788
## 4   0.0000000  0.25349657
```

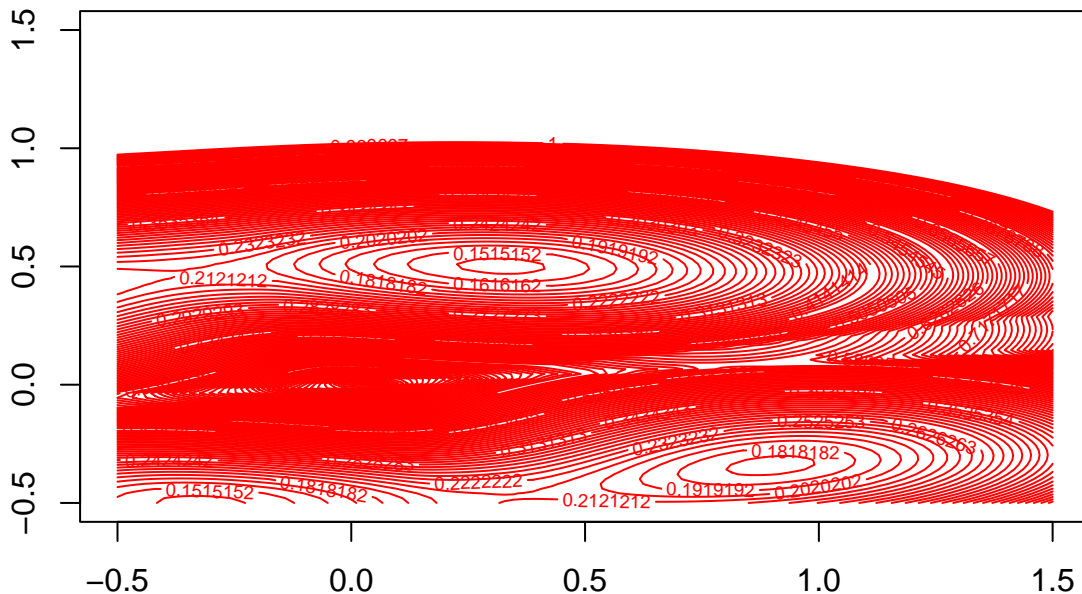The contour plot for the linear combinations is

Figure 3: Plane spanned by configuration and gradient, equal dissimilarities

Note that the plot shows various ridges, where stress is not differentiable because some of the distances in the linear combination are zero.

in this case, where we are interested in step size, we should look at linear combinations with $\alpha = 1$ and $\beta < 0$, i.e. of the form $X - |\beta|\mathcal{D}\sigma(X)$. Figure Figure 4 shows the initial descent and an optimal step size around .3.
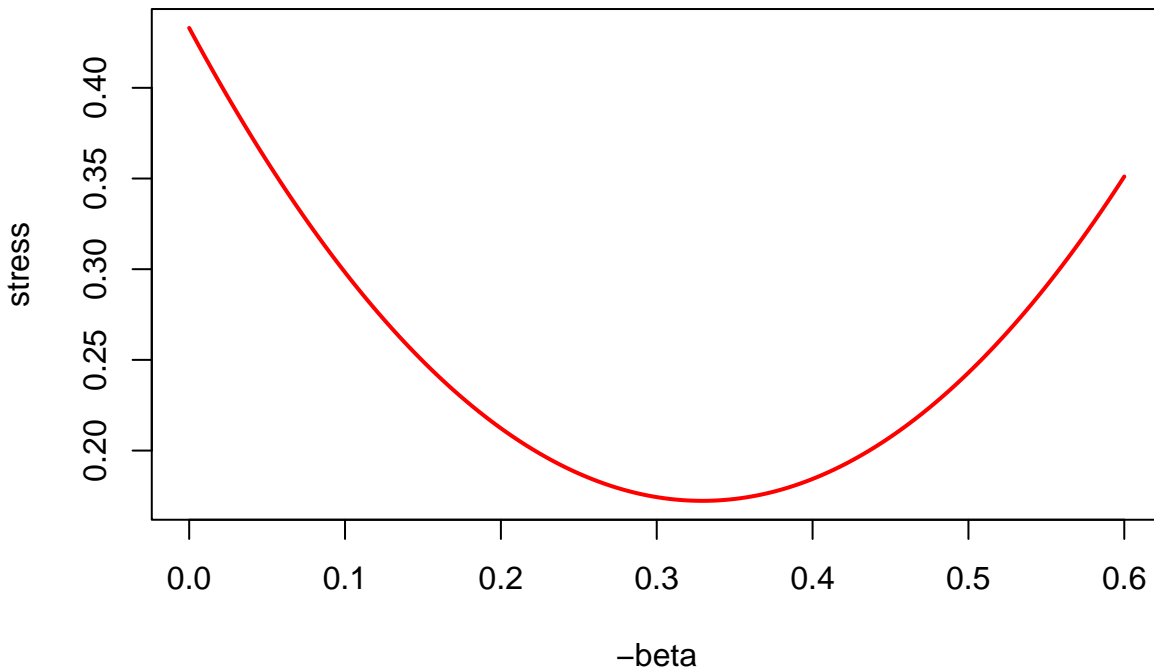

Figure 4: Along the path of steepest descent, equal dissimilarities

7   Ekman example

The Ekman (1954) color data give similarities between 14 colors.

```
##       434   445   465   472   490   504   537   555   584   600   610   628   651
## 445 0.86
## 465 0.42 0.50
## 472 0.42 0.44 0.81
## 490 0.18 0.22 0.47 0.54
## 504 0.06 0.09 0.17 0.25 0.61
## 537 0.07 0.07 0.10 0.10 0.31 0.62
## 555 0.04 0.07 0.08 0.09 0.26 0.45 0.73
## 584 0.02 0.02 0.02 0.02 0.07 0.14 0.22 0.33
## 600 0.07 0.04 0.01 0.01 0.02 0.08 0.14 0.19 0.58
## 610 0.09 0.07 0.02 0.00 0.02 0.02 0.05 0.04 0.37 0.74
## 628 0.12 0.11 0.01 0.01 0.01 0.02 0.02 0.03 0.27 0.50 0.76
## 651 0.13 0.13 0.05 0.02 0.02 0.02 0.02 0.02 0.20 0.41 0.62 0.85
## 674 0.16 0.14 0.03 0.04 0.00 0.01 0.00 0.02 0.23 0.28 0.55 0.68 0.76
```

We use three different transformations of the similarities to dissimilarities. The first is $1-x$, the second $(1-x)^3$ and the third $\sqrt[3]{1-x}$. We need the following iterations to find the FDS solution (up to a change in loss of 1e-15).

```
## power =  1.00  itel =   6936  stress =  0.0000875293
## power =  3.00  itel =    171  stress =  0.0110248119
## power =  0.33  itel =    423  stress =  0.0000000000
```

For the same three solutions we compute singular values of the thirteen-dimensional FDS solution.

```
##  [1] 0.1797609824 0.1454675297 0.0843865491 0.0777136109 0.0486123551
##  [6] 0.0393576522 0.0236290817 0.0162344515 0.0072756171 0.0000031164
## [11] 0.0000000009 0.0000000000 0.0000000000
##
##  [1] 0.2159661347 0.1549184093 0.0000000727 0.0000000041 0.0000000000
##  [6] 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
## [11] 0.0000000000 0.0000000000 0.0000000000
##
##  [1] 0.1336126813 0.1139019875 0.0880453752 0.0851609618 0.0710424935
##  [6] 0.0664988952 0.0561005006 0.0535112029 0.0492295395 0.0479964575
## [11] 0.0468628701 0.0410193579 0.0388896490
```

Thus the Gower ranks of the transformed dissimilarities are, repectively, nine (or ten), two, and thirteen. Note that for the second set of dissimilarities, with Gower rank two, the first two principal components of the thirteen-dimensional solution are the global minimizer in two dimensions. To illustrate the Gower rank in yet another way we give the thirteen non-zero eigenvalues of $V^+B(X)$, so that the Gower rank is the number of eigenvalues equal to one. All three solutions satisfy the necessary and sufficient conditions for a global FDS solution.

```
##  [1] 1.0000000432 1.0000000222 1.0000000012 1.0000000005 1.0000000002
##  [6] 1.0000000001 1.0000000000 1.0000000000 0.9999993553 0.9989115116
## [11] 0.9976821885 0.9942484083 0.9825147154
##
##  [1] 1.0000000000 1.0000000000 0.9234970864 0.9079012130 0.8629365849
##  [6] 0.8526920031 0.8298036209 0.8145561677 0.7932385763 0.7916517225
## [11] 0.7864426781 0.7476794757 0.7282682474
##
##  [1] 1.0000000820 1.0000000241 1.0000000047 1.0000000009 1.0000000004
##  [6] 1.0000000003 1.0000000001 1.0000000001 1.0000000001 1.0000000000
## [11] 0.9999999999 0.9999999689 0.9999999005
```

We also plot the first two principal components of the thirteen-dimensional FDS solution. Not surprisingly, they look most circular and regular for the solution with power three, because this actually is the global minimum over two-dimensional solutions. The other configurations still have quite a lot of variation in the remaining dimensions.
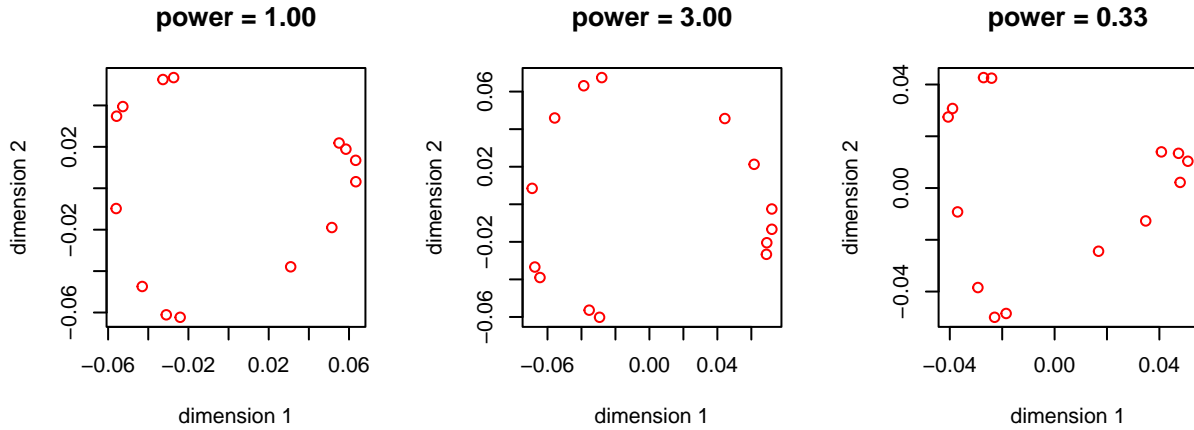

Figure 5: Ekman data, configurations for three powers

Figure 6 illustrates that the FDS solution with power 3 is quite different from power 1 and power one 1/3 Basically the transformations with lower powers result in dissimilarity measures that are very similar to Euclidean distances in a high-dimensional configuration, while power equal to 3 makes the dissimilarties less Euclidean. This follows from metric transform theory, where concave increasing transforms of finite metric spaces tend to be Euclidean. In particular the square root transformation of a finite metric space has the Euclidean four-point property, and there is a $c > 0$ such that the metric transform $f(t) = ct/(1 + ct)$ makes a finite metric space Euclidean (Maehara (1986)).
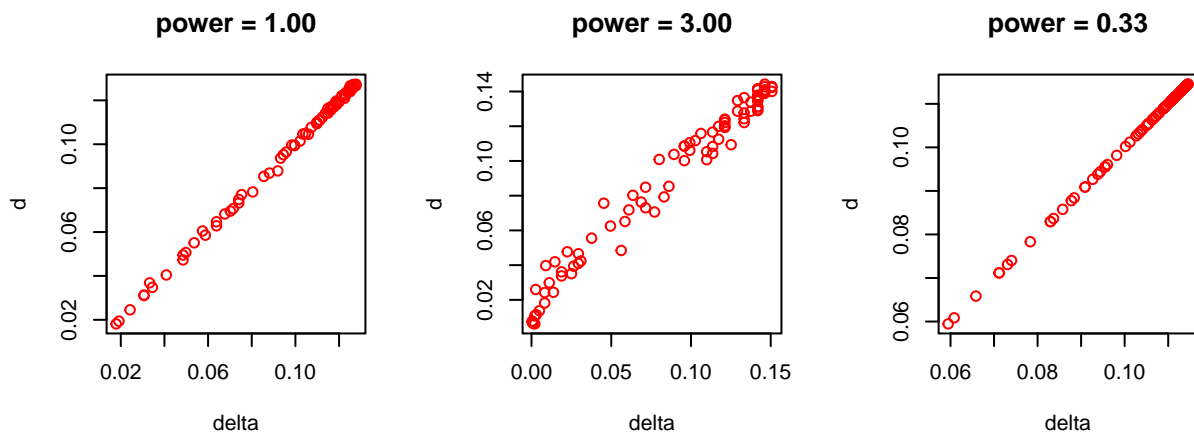

Figure 6: Ekman data, fit plots for three powers

11

8   Code

```
library(MASS)

torgerson <- function(delta, p = 2) {
  doubleCenter <- function(x) {
    n <- dim(x)[1]
    m <- dim(x)[2]
    s <- sum(x) / (n * m)
    xr <- rowSums(x) / m
    xc <- colSums(x) / n
    return((x - outer(xr, xc, "+")) + s)
  }
  z <-
    eigen(-doubleCenter((as.matrix (delta) ^ 2) / 2), symmetric = TRUE)
  v <- pmax(z$values, 0)
  return(z$vectors[, 1:p] %*% diag(sqrt(v[1:p])))
}

u <- function (i, n) {
  return (ifelse (i == 1:n, 1, 0))
}

e <- function (i, j, n) {
  d <- u (i, n) - u (j, n)
  return (outer (d, d))
}

directSum <- function (x) {
  m <- length (x)
  nr <- sum (sapply (x, nrow))
  nc <- sum (sapply (x, ncol))
  z <- matrix (0, nr, nc)
  kr <- 0
  kc <- 0
  for (i in 1:m) {
    ir <- nrow (x[[i]])
    ic <- ncol (x[[i]])
    z[kr + (1:ir), kc + (1:ic)] <- x[[i]]
    kr <- kr + ir
    kc <- kc + ic
  }
  return (z)
```

```r
}

repList <- function(x, n) {
  z <- list()
  for (i in 1:n)
    z <- c(z, list(x))
  return(z)
}

makeA <- function (n) {
  m <- n * (n - 1) / 2
  a <- list()
  for (j in 1:(n - 1))
    for (i in (j + 1):n) {
      d <- u (i, n) - u (j, n)
      e <- outer (d, d)
      a <- c(a, list (e))
    }
  return (a)
}

makeD <- function (a, x) {
  return (sapply (a, function (z)
    sqrt (sum (x * (z %*% x)))))
}

makeB <- function (w, delta, d, a) {
  n <- length (a)
  m <- nrow (a[[1]])
  b <- matrix (0, m , m)
  for (i in 1:n)
    b <- b + w[i] * (delta[i] / d[i]) * a[[i]]
  return (b)
}

makeV <- function (w, a) {
  n <- length (a)
  m <- nrow (a[[1]])
  v <- matrix (0, m, m)
  for (i in 1:n)
    v <- v + w[i] * a[[i]]
  return (v)
}
```

```r
inBetween <- function (alpha, beta, x, y, w, delta, a) {
  z <- alpha * x + beta * y
  d <- makeD (a, z)
  return (sum (w * (delta - d) ^ 2))
}

biBase <- function (x, y, a) {
  biBi <- function (x, y, v) {
    a11 <- sum (x * (v %*% x))
    a12 <- sum (x * (v %*% y))
    a22 <- sum (y * (v %*% y))
    return (matrix (c(a11, a12, a12, a22), 2, 2))
  }
  return (lapply (a, function (u) biBi (x, y, u)))
}

fullMDS <-
  function (delta,
            w = rep (1, length (delta)),
            xini,
            a,
            itmax = 100,
            eps = 1e-6,
            verbose = TRUE) {
    m <- length (a)
    v <- makeV (w, a)
    vv <- ginv (v)
    xold <- xini
    dold <- makeD (a, xini)
    sold <- sum ((delta - dold) ^ 2)
    bold <- makeB (w, delta, dold, a)
    itel <- 1
    repeat {
      xnew <- vv %*% bold %*% xold
      dnew <- makeD (a, xnew)
      bnew <- makeB (w, delta, dnew, a)
      snew <- sum ((delta - dnew) ^ 2)
      if (verbose) {
        cat (
          formatC (itel, width = 4, format = "d"),
          formatC (
            sold,
            digits = 10,
            width = 13,
```

```
        format = "f"
      ),
      formatC (
        snew,
        digits = 10,
        width = 13,
        format = "f"
      ),
      "\n"
    )
  }
  if ((itel == itmax) || (abs(sold - snew) < eps))
    break
  itel <- itel + 1
  xold <- xnew
  dold <- dnew
  sold <- snew
  bold <- bnew
}
return (list (x = xnew, d = dnew, delta = delta, s = snew, b = bnew, v = v, itel = i
}
```

9   NEWS

001 01/21/16 Made a start

002 01/23/16 Reorganization and further material

003 01/24/16 First upload. Figures, tables, examples.

004 01/25/16 Corrected small errors

005 01/25/16 Reference to metric transform theory

References

Datorro, J. 2015. *Convex Optimization and Distance Geometry*. Second Edition. Palo Alto, CA: Meebo Publishing.

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.

———. 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.

De Leeuw, J., P. Groenen, and P. Mair. 2016. "Differentiability of rStress at a Local Minimum." 2016.

Ekman, G. 1954. "Dimensions of Color Vision." *Journal of Psychology* 38: 467–74.

"Inverse Multidimensional Scaling." 2007. *Journal of Classification* 14: 3–21.

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

———. 1964b. "Nonmetric Multidimensional Scaling: a Numerical Method." *Psychometrika* 29: 115–29.

Maehara, H. 1986. "Metric Transforms of Finite Spaces and Connected Graphs." *Discrete Mathematics* 61: 235–46.

Rockafellar, R. T. 1970. *Convex Analysis.* Princeton University Press.