MINIMIZING RSTRESS USING NESTED MAJORIZATION

JAN DE LEEUW

ABSTRACT. We construct a majorization algorithm to minimize the sum of squares of discrepancies between dissimilarities and any positive power of Euclidean distances (including the logarithm). Iterations alternate Dinkelbach majorization with linear/quadratic majorization steps. Global convergence is guaranteed, although it typically is slow.

1. PROBLEM

Define the multidimensional scaling (MDS) loss function

$$\sigma_r(x) = \sum_{i=1}^{n} w_i(\delta_i - (x'A_ix)^r)^2$$

with $r > 0$ and the A_i positive semi-definite. We call this *rStress*. Special cases are *stress* [10, 11] for $r = \frac{1}{2}$, *sstress* [17] for $r = 1$, and the loss function used in MULTISCAL [13] for $r \rightarrow 0$.

In the usual MDS formulation uses Euclidean distances $d_{j\ell}(X)$ between points j and ℓ, which correspond with the rows of an $n \times p$ configuration matrix X. This fits into our formulation by setting $x = \mathbf{vec}(X)$ and by setting the A_i to matrices of the form $I_p \otimes E_{j\ell}$, where the matrix $E_{j\ell}$ has elements (j, j) and (ℓ, ℓ) equal to $+1$ and elements (j, ℓ) and (ℓ, j) equal to -1. Then $x'A_ix = d_{j\ell}^2(X)$.

Date: Saturday 28th June, 2014 — 14h 39min — Typeset in LUCIDA BRIGHT.

The problem we are trying to solve is to find an convergent iterative algorithm to minimize σ_r for all values of $r > 0$.

2. Use of Homogeneity

Following De Leeuw [1] we define

$$\rho_r(x) \triangleq \sum_{i=1}^{n} w_i \delta_i (x' A_i x)^r,$$

and

$$\eta_r^2(x) \triangleq \sum_{i=1}^{n} w_i (x' A_i x)^{2r}.$$

Without loss of generality we assume

$$\sum_{i=1}^{n} w_i \delta_i^2 = 1.$$

Thus

$$\sigma_r(x) = 1 - 2\rho_r(x) + \eta_r^2(x).$$

Now minimizing σ_r is equivalent to minimizing

$$\sigma_r(\alpha, x) = 1 - 2\alpha^r \rho_r(x) + \alpha^{2r} \eta_r^2(x).$$

over α and x with $x'x = 1$. The minimum over α for given x is

$$\min_{\alpha} \sigma_r(\alpha, x) = 1 - \frac{\rho_r^2(x)}{\eta_r^2(x)},$$

attained at

$$\hat{\alpha} = \sqrt[r]{\frac{\rho(x)}{\eta^2(x)}}.$$

If we define

$$\gamma_r(x) = \frac{\rho_r(x)}{\eta_r(x)},$$

then minimizing σ_r can be done by maximizing γ_r over x on the unit sphere and adjusting the scale afterwards. Of course γ_r is homogeneous of degree zero, which means the constraint $x'x = 1$ is used merely for identification and mathematical convenience.

3. Nested Majorization

We first use a famous result of Dinkelbach [4] to simplify the problem of maximizing the ratio γ_r.

Lemma 3.1 (Dinkelbach). *Suppose* $\gamma(x) = \frac{\rho(x)}{\eta(x)}$ *is* any *fractional function, with* $\eta(x) > 0$.

1. *If* $\rho(x) - \gamma(y)\eta(x) > \rho(y) - \gamma(y)\eta(y) = 0$ *then* $\gamma(x) > \gamma(y)$.
2. *If* $\mathcal{D}\rho(x) - \gamma(\hat{x})\mathcal{D}\eta(x) = 0$ *for* $x = \hat{x}$ *then* $\mathcal{D}\gamma(\hat{x}) = 0$.

Proof. The first part is obvious. The second part follows from the formula for differentiating the ratio of two functions. □

As a consequence of lemma 3.1 we can use an iterative algorithm that finds $x^{(k+1)}$ by maximizing, or at least increasing,

$$\tau_r(x, x^{(k)}) \stackrel{\Delta}{=} \rho_r(x) - \gamma_r(x^{(k)})\eta_r(x)$$

over x on the unit sphere. We call $\tau_r(x, y)$ the *Dinkelbach minorization* of γ_r at y.

In order to maximize, or increase, $\tau_r(x, x^{(k)})$ we again use minorization and majorization [2, 9, 12]. Suppose we have a function $\pi_r(x, y)$ such that $\rho_r(x) \geq \pi_r(x, y)$ for all x and y and $\rho_r(x) = \pi_r(x, x)$ for all x, as well as a function $\zeta_r(x, y)$ such that $\eta_r(x) \leq \zeta_r(x, y)$ for all x and y and $\eta_r(x) = \zeta_r(x, x)$ for all x. Then

$$\tau_r(x, x^{(k)}) \geq \pi_r(x, y) - \gamma_r(x^{(k)})\zeta_r(x, y).$$

Now use double superscripting for nested iterations. Set $x^{(k,0)} = x^{(k)}$ and find $x^{(k,\ell+1)}$ by maximizing

$$\pi_r(x, x^{(k,\ell)}) - \gamma_r(x^{(k)})\zeta_r(x, x^{(k,\ell)})$$

over x on the unit sphere. We perform one or more steps of this inner minorization algorithm (let's call them *M-steps*) before we compute a new Dinkelbach minorization (a *D-step*).

4. Powers of Quadratic Forms

We start with some lemmas we will need to construct the minorizations and majorizations.

Lemma 4.1. *$f_r(x) \overset{\Delta}{=} (x'Ax)^r$ is convex on $x'Ax > 0$ if and only if $r \geq \frac{1}{2}$.*

Proof. The first and second derivative are

$$\mathcal{D}f_r(x) = 2r(x'Ax)^{r-1}Ax,$$

and

$$\mathcal{D}^2 f_r(x) = 2r(x'Ax)^{r-1}\left(A + 2(r-1)\frac{Axx'A}{x'Ax}\right).$$

The matrix $H_r(x) \overset{\Delta}{=} A + 2(r-1)\frac{Axx'A}{x'Ax}$ is psd for $r = \frac{1}{2}$, and its eigenvalues increase with r. Thus it is psd for all $r \geq \frac{1}{2}$.

Also, if $0 < r < \frac{1}{2}$ then, by Sylvester's Law of Inertia, $\mathcal{D}^2 f_r(x)$ has precisely one negative eigenvalue, as well as $\mathbf{rank}(A) - 1$ positive eigenvalues, and $n - \mathbf{rank}(A)$ zero eigenvalues. Thus in this case f_r is not convex (and not concave either). □

Now write $\overline{\lambda}(X)$ or $\overline{\lambda}_X$ for the largest eigenvalue of a matrix X, and $\underline{\lambda}(X)$ or $\underline{\lambda}_X$ for the smallest eigenvalue. Note that if $A = I \otimes E_{j\ell}$ then $\overline{\lambda}_A = 2$ and $\underline{\lambda}(A) = 0$.

Lemma 4.2. *If $r \geq 1$ then*

$$\overline{\lambda}(\mathcal{D}^2 f_r(x)) \leq 2r(2r-1)\overline{\lambda}_A^r\,(x'x)^{r-1}.$$

If $x'x \leq 1$ then

$$\overline{\lambda}(\mathcal{D}^2 f_r(x)) \leq 2r(2r-1)\overline{\lambda}_A^r.$$

Proof. If $r \geq 1$, then

$$u'H_r(x)u = u'Au + 2(r-1)\frac{(u'Ax)^2}{x'Ax} \leq (2r-1)u'Au.$$

Thus

$$\overline{\lambda}(H_r(x)) \leq (2r-1)\overline{\lambda}_A,$$

and

$$\overline{\lambda}(\mathcal{D}^2 f_r(x)) \leq 2r(2r-1)\overline{\lambda}_A(x'Ax)^{r-1} \leq 2r(2r-1)\overline{\lambda}_A^r\ (x'x)^{r-1}.$$

□

Lemma 4.3. *If* $0 < r \leq 1$ *then*

$$f_r(x) \leq (1-r)f_r(y) + rf_{r-1}(y)x'Ax.$$

Proof. If $r \leq 1$ then $(x'Ax)^r$ is concave in $x'Ax$, although not in x. Thus

$$f_r(x) \leq f_r(y) + r(y'Ay)^{r-1}(x'Ax - y'Ay),$$

which simplifies to the required result. □

Lemma 4.4. *If* $0 < r < \frac{1}{2}$ *then*

$$\underline{\lambda}(\mathcal{D}^2 f_r(x)) \geq 2r(2r-1)\overline{\lambda}_A^r(x'x)^{r-1}.$$

If $x'x \leq 1$ *then*

$$\underline{\lambda}(\mathcal{D}^2 f_r(x)) \geq 2r(2r-1)\overline{\lambda}_A^r.$$

Proof. We have

$$\frac{(u'Ax)^2}{x'Ax} \leq u'Au$$

as before. Thus

$$u'H(x)u \geq (2r-1)u'Au \geq (2r-1)\overline{\lambda}_A u'u.$$

The result follows because in addition $x'Ax \leq \overline{\lambda}_A x'x$, and consequently $(x'Ax)^{r-1} \geq \overline{\lambda}_A^{r-1}(x'x)^{r-1}$. □

The following lemma, defining a type of uniform quadratic majorization [3], is an additional useful tool.

Lemma 4.5. *Suppose* ϕ *is homogeneous of degree* s, $x'x = y'y = 1$ *and* $\overline{\lambda}(\mathcal{D}^2\phi(z)) \leq \kappa$ *for all* $z \in [x, y]$. *Then*

$$\phi(x) \leq (1-s)\phi(y) + \kappa + x'(\mathcal{D}\phi(y) - \kappa y).$$

In the same way, if $\underline{\lambda}(\mathcal{D}^2\phi(z)) \geq \kappa$ *for all* $z \in [x, y]$ *we have*

$$\phi(x) \geq (1-s)\phi(y) + \kappa + x'(\mathcal{D}\phi(y) - \kappa y).$$

Proof. We only show the first part. The proof of the second part goes the same. By Taylor's theorem we have for all x and y

$$\phi(x) \leq \phi(y) + (x-y)'\mathcal{D}\phi(y) + \frac{1}{2}\kappa(x-y)'(x-y),$$

which simplifies to the stated result if $x'x = y'y = 1$ and ϕ is homogeneous of degree s. □

5. Majorizing/Minorizing ρ_r and η_r

We distinguish the two cases: case A, with $r \geq \frac{1}{2}$, and case B, with $0 < r \leq \frac{1}{2}$. Both cases use different lemmas and require different algorithms.

5.1. **Case A.** If $r \geq \frac{1}{2}$ we use lemmas 4.1 and 4.2.

5.1.1. *Dealing with ρ_r.* Since

$$\rho_r(x) = \sum_{i=1}^{n} w_i \delta_i (x'A_i x)^r$$

we have

$$\rho_r(x) \geq (1-2r)\rho_r(y) + 2r x' B_r(y) y,$$

where

$$B_r(y) = \sum_{i=1}^{n} w_i \delta_i (y'A_i y)^{r-1} A_i.$$

5.1.2. *Dealing with η_r.* Now

$$\eta_r^2(x) = \sum_{i=1}^{n} w_i (x'A_i x)^{2r},$$

which is homogeneous of order $4r$. The upper bound on the eigenvalues of the second derivatives when $x'x = 1$ is, from lemma 4.2,

$$\kappa_r = 4r(4r-1)\sum_{i=1}^n w_i\overline{\lambda}^{2r}(A_i)$$

Thus, by lemma 4.5,

$$\eta_r^2(x) \leq (1-4r)\eta_r^2(y) + \kappa_r + x'(\mathcal{D}\eta_r^2(y) - \kappa_r y).$$

Now

$$\mathcal{D}\eta_r^2(y) = 4rC_r(y)y,$$

where

$$C_r(y) = \sum_{i=1}^n w_i(y'A_iy)^{2r-1}A_i,$$

and thus

$$\eta_r^2(x) \leq (1-4r)\eta_r^2(y) + \kappa_r + x'(4rC_r(y) - \kappa_r I)y.$$

Since, by the AM-GM inequality,

$$\eta_r(x) \leq \frac{1}{2\eta_r(y)}\left(\eta_r^2(x) + \eta_r^2(y)\right),$$

we see that

$$\eta_r(x) \leq \frac{1}{2\eta_r(y)}\left((2-4r)\eta_r^2(y) + \kappa_r + x'(4rC_r(y) - \kappa_r I)y\right).$$

5.1.3. *Putting it Together.* The last stage is collecting the various terms. We ignore terms that do not involve x. Define

$$\theta_r(y,z) \triangleq B_r(z) - \frac{\gamma_r(y)}{\eta_r(z)}\left[C_r(z) - (4r-1)\sum_{i=1}^n w_i\overline{\lambda}^{2r}(A_i)I\right]z$$

Our M-step is now to find $x^{(k,\ell+1)}$ by maximizing the linear function $x'\theta_r(x^{(k,0)}, x^{(k,\ell)})$ over $x'x = 1$, so

$$x^{(k,\ell+1)} = \frac{\theta_r(x^{(k,0)}, x^{(k,\ell)})}{\|\theta_r(x^{(k,0)}, x^{(k,\ell)}\|}.$$

After one or more M-steps we make another D-step.

5.2. **Case B.**

5.2.1. *Dealing with ρ_r.* Lemmas 4.4 and 4.5 are used to for a quadratic minorization of ρ_r, which is homogeneous of order $2r$. Thus

$$\rho_r(x) \geq (1-2r)\rho_r(y) + \kappa_r + x'(\mathcal{D}\rho_r(y) - \kappa_r y),$$

where

$$\kappa_r = 2r(2r-1)\sum_{i=1}^n w_i \overline{\lambda}^r(A_i).$$

Now

$$\mathcal{D}\rho_r(y) = 2rB_r(y)y,$$

with

$$D_r(y) \triangleq \sum_{i=1}^n w_i\delta_i(y'A_iy)^{r-1}A_i.$$

Thus

$$\rho_r(x) \geq (1-2r)\rho_r(y) + \kappa_r + x'(2rB_r(y) - \kappa_r I)y,$$

5.2.2. *Dealing with η_r.* From lemma 4.3

$$\eta_r^2(x) \leq (1-2r)\eta_r^2(y) + 2rx'C_r(y)x,$$

and thus

$$\eta_r(x) \leq \frac{1}{2\eta_r(y)}\left((1-2r)\eta_r^2(y) + 2rx'C_r(y)x + \eta_r^2(y)\right)$$

5.2.3. *Putting it Together.* Again we ignore terms that do not involve x. Define

$$g_r(y) \triangleq (B_r(y) - (2r-1)\sum_{i=1}^n w_i\overline{\lambda}^r(A_i)I)y,$$

and

$$E_r(y,z) \triangleq \frac{\gamma_r(z)}{\eta_r(y)}C_r(y).$$

We find $x^{(k,\ell+1)}$ by maximizing

$$x'g_r(x^{(k,\ell)}) - \frac{1}{2}x'E_r(x^{(k,0)},x^{(k,\ell)})x$$

over $x'x = 1$. This amounts to maximizing a concave quadratic form over the unit sphere, one of the classical secular equation problems [6, 16, 7, 8]. The theory is reviewed briefly in appendix A.

6. Example

We use the color data from Ekman [5], without weights w_i and with only a single M-step between D-steps. The code used and the tables and figures produced are in the appendices.

6.1. **Case A.** Results are computed in two dimensions, for $r = 0.5, 0.75, 1.00, 2.00, 3.00$. In all cases there is monotonic convergence, although for $r = 2$ and $r = 3$ the (strict) stop criterion is not reached after 100,000 iterations. Solutions for $r > 1$ are not really interesting for these data.

6.2. **Case B.** Results are computed for $r = 0.05, 0.10, 0.25$. Again we see monotone convergence, generally faster than for $r \geq 1$. As the Shepard plot and the fit statistics indicate, the results for $r = 0.25$ are most satisfactory. In that case we are fitting square roots of Euclidean distances to the dissimilarities.

7. Discussion

For increasing $r > 0.5$ the bound on the second derivatives used in majorizing $\eta_r(x)$ becomes less and less sharp, and as a consequence the convergence rate of the algorithm gets very close to sublinear. A more refined mathematical analysis may be needed to get a sharper bound, although such bounds are likely to increase the amount of work needed in each iteration. It may also be worthwhile to experiment with different numbers of M-steps between D-steps.

Values of $r < 0.25$ or $r > 1$ do not produce a good fit for these data. But comparing solutions for different values of r can be thought of as a parametric form of nonmetric scaling, where we allow for power function transformations. All Shepard plots for our example show strong monotonicity. And, contrary to typical multidimensional scaling methods, the different fits all use the numerical values of the actual dissimilarities to measure the loss, and consequently the loss function values are all on the same scale.

In our example we fitted powers of distances to the dissimilarities. We could also have fitted powered distances to powered dissimilarities, using the same power for both. This is similar to fitting distances to dissimilarities, using larger weights for larger distances. Statistically this type of weighting is somewhat counterintuitive [13, 14]. As a further generalization we could look at solutions which use different powers for the dissimilarities and distances, bringing us even closer to nonmetric scaling.

It would also, of course, be easy to combine fitting powers of distances with monotone transformations of the dissimilarities. This would result in a properly nonmetric version of rStress. It is unclear how monotone transformations of the distances, using power functions, and monotone transformations of the dissimilarities, using monotone regression, would interact.

Appendix A. Secular Equation

A.1. **Problem.** Consider the problem of minimizing the quadratic $g(x) \overset{\Delta}{=} \frac{1}{2}x'Ax - x'b$ over $x'x = 1$, where A is symmetric, but not necessarily positive semi-definite. Our treatment largely follows Hager [8].

A.2. **Necessary Conditions.** The stationary equations are, using a single Lagrange multiplier μ,

$$(A - \mu I)x = b, \tag{1a}$$

$$x'x = 1, \tag{1b}$$

Suppose $(\hat{x}, \hat{\mu})$ is a solution of (1). If $\hat{x}$ minimizes g over $x'x = 1$ then $\hat{x}'\hat{x} = 1$ and

$$\frac{1}{2}\hat{x}'A\hat{x} - b'\hat{x} \leq \frac{1}{2}x'Ax - b'x$$

for all $x'x = 1$. Now $(A - \hat{\mu}I)\hat{x} = b$. Thus

$$\frac{1}{2}\hat{x}'A\hat{x} - \hat{x}'(A - \hat{\mu}I)\hat{x} \leq \frac{1}{2}x'Ax - x'(A - \hat{\mu}I)\hat{x},$$

which can be written as

$$(x - \hat{x})'(A - \hat{\mu}I)(x - \hat{x}) \geq 0,$$

for all $x'x = 1$. Thus $A - \hat{\mu}I$ must be positive semi-definite, and $\hat{\mu} \leq \underline{\lambda}_A$. This argument is taken from Sorensen [15, lemma 2.4].

A.3. **Finding the root.** The eigen-decompoaition of A is $A = K\Lambda K'$, where the eigenvalues $\overline{\lambda}_A = \lambda_1 \geq \cdots \geq \lambda_n = \underline{\lambda}_A$ are not necessarily distinct. Define $y = K'x$ and $\beta = K'b$. Then (1) becomes

$$(\Lambda - \mu I)y = \beta, \tag{2a}$$

$$y'y = 1. \tag{2b}$$

We first look for solutions when β has no zero elements. This implies that μ cannot be equal to one of the λ_i. In this case we must have

$$y_i = \frac{1}{\lambda_i - \mu}\beta_i, \tag{3}$$

and

$$\sum_{i=1}^n y_i^2 = \sum_{i=1}^n \frac{\beta_i^2}{(\lambda_i - \mu)^2} = 1. \tag{4}$$

We can now solve (4) for μ, then use (3) to find y, and then use $x = Ky$ to compute the solution x.

Let

$$f(\mu) \triangleq \sum_{i=1}^n \frac{\beta_i^2}{(\lambda_i - \mu)^2}.$$

Then $f(\mu) > 0$ and $\lim_{\mu\rightarrow\infty} f(\mu) = \lim_{\mu\rightarrow-\infty} f(\mu) = 0$.

Suppose $\overline{\lambda}_A = \lambda_1 < \cdots < \lambda_v = \underline{\lambda}_A$ are the v distinct eigenvalues of A. Then f is a rational analytic function on each of the $v + 1$ intervals bounded by these distinct eignevalues. We have $\lim_{\mu\rightarrow\lambda_i} f(\mu) = \infty$ for all i, which defines v vertical asymptotes. The horizontal axes is a horizontal asymptote.

Let's see how f behaves on the open intervals bounded by $\pm\infty$ and the smallest and largest eigenvalues.From

$$\mathcal{D}f(\mu) = 2\sum_{i=1}^n \frac{\beta_i^2}{(\lambda_i - \mu)^3}$$

we see that f increases from zero to ∞ as μ increases from $-\infty$ to $\underline{lambda}_A$. Thus $f(\mu) = 1$ has a unique root in that interval. In the same way f decreases from ∞ to zero as μ increases from $\overline{\lambda}_A$ to $+\infty$, which defines another unique root. But because $A - \mu I$ must be positive semi-definite at the minimum, we actually need the smallest root.

For the second derivatives we have

$$\mathcal{D}^2 f(\mu) = 6\sum_{i=1}^n \frac{\beta_i^2}{(\lambda_i - \mu)^4},$$

which means that in each of the $\nu + 1$ intervals between $\pm\infty$ and consecutive distinct eigenvalues f is convex, and consequently it a unique minimum in that interval.

For computational purposes it is useful to have bounds on the root better than $-\infty < \mu < \underline{\lambda}_A$. See Hager [8, p. 190].

A.4. **Degenerate case.** We still have to deal with the case in which some of the elements of β are zero.

A.5. **Matrix Version.**

Appendix B. Code

B.1. Main.

```
fStressMin <- function (delta, w = 1 - diag (nrow (delta)), p = 2, r = 0.5, eps = 1e-10, itmax = 100000,
  delta <- delta / enorm (delta, w)
  itel <- 1
        xold <- torgerson (delta, p = p)
        xold <- xold / enorm (xold)
        n <- nrow (xold)
        k <- sum (w) * ((4 * r) - 1) * (2 ^ (2 * r))
  l <- sum (w) * ((2 * r) - 1) * (2 ^ r)
        dold <- sqdist (xold)
        rold <- sum (w * delta * mkPower (dold, r))
        nold <- sqrt (sum (w * mkPower (dold, 2 * r)))
        lold <- rold / nold
        repeat {
                by <- mkBmat (w * delta * mkPower (dold, r - 1))
                cy <- mkBmat (w * mkPower (dold, (2 * r) - 1))
    if (r >= 0.5) {
                  my <- by - (lold / nold) * (cy - (k * diag(n)))
                  xnew <- my %*% xold
                  xnew <- xnew / enorm (xnew)
      }
    if (r < 0.5) {
      gy <- as.vector ((by - (l * diag (n))) %*% xold)
      ey <- kronecker (diag (p), (lold / nold) * cy)
      xnew <- matrix (secularEq (ey, gy), n, p)
      }
          dnew <- sqdist (xnew)
                rnew <- sum (w * delta * mkPower (dnew, r))
                nnew <- sqrt (sum (w * mkPower (dnew, 2 * r)))
                lnew <- rnew / nnew
                if (verbose) {
                cat (formatC (itel, width = 4, format = "d"),
                          formatC (lold, digits = 10, width = 13, format = "f"),
                          formatC (lnew, digits = 10, width = 13, format = "f"), "\n")
                }
                if ((itel == itmax) || ((lnew - lold) < eps)) break ()
                itel <- itel + 1
            xold <- xnew
            dold <- dnew
            lold <- lnew
        }
        return (list(x = xnew, gamma = c (lold, lnew), itel = itel))
}
```

B.2. Auxilaries.

```
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))
  v <- pmax(z$values,0)
  return(z$vectors[,1:p]%*%diag(sqrt(v[1:p])))
}

enorm <- function (x, w = 1) {
        return (sqrt (sum (w * (x ^ 2))))
}

sqdist <- function (x) {
        s <- tcrossprod (x)
        v <- diag (s)
        return (outer (v, v, "+") - 2 * s)
}

mkBmat <- function (x) {
        d <- rowSums (x)
        x <- -x
        diag (x) <- d
        return (x)
}

mkPower <- function (x, r) {
        n <- nrow (x)
        return (abs ((x + diag (n)) ^ r) - diag(n))
}
```

B.3. Seculat Equation.

```
secularEq <- function (a, b) {
    n <- dim(a)[1]
  eig <- eigen (a)
        eva <- eig $ values
        eve <- eig $ vectors
        beta <- drop (crossprod (eve, b))
        f <- function (mu) {
                return (sum ((beta / (eva + mu)) ^ 2) - 1)
        }
        lmn <- eva [n]
        uup <- sqrt (sum (b ^ 2)) - lmn
        ulw <- abs (beta [n]) - lmn
    rot <- uniroot (f, lower = ulw, upper = uup) $ root
    cve <- beta / (eva + rot)
    return (drop (eve %*% cve))
}
```

B.4. Run.

```
source ("fStress.R")

ekman <-
structure(c(0.86, 0.42, 0.42, 0.18, 0.06, 0.07, 0.04, 0.02, 0.07,
0.09, 0.12, 0.13, 0.16, 0.5, 0.44, 0.22, 0.09, 0.07, 0.07, 0.02,
0.04, 0.07, 0.11, 0.13, 0.14, 0.81, 0.47, 0.17, 0.1, 0.08, 0.02,
0.01, 0.02, 0.01, 0.05, 0.03, 0.54, 0.25, 0.1, 0.09, 0.02, 0.01,
0, 0.01, 0.02, 0.04, 0.61, 0.31, 0.26, 0.07, 0.02, 0.02, 0.01,
0.02, 0, 0.62, 0.45, 0.14, 0.08, 0.02, 0.02, 0.02, 0.01, 0.73,
0.22, 0.14, 0.05, 0.02, 0.02, 0, 0.33, 0.19, 0.04, 0.03, 0.02,
0.02, 0.58, 0.37, 0.27, 0.2, 0.23, 0.74, 0.5, 0.41, 0.28, 0.76,
0.62, 0.55, 0.85, 0.68, 0.76), Size = 14L, call = quote(as.dist.default(m = b)), clas
445, 465, 472, 490, 504, 537, 555, 584, 600, 610, 628, 651, 674
))

ekman <- as.matrix (1-ekman)
wave <- row.names (ekman)
e05 <- fStressMin (ekman, r = .10, verbose = FALSE)
e10 <- fStressMin (ekman, r = .10, verbose = FALSE)
e25 <- fStressMin (ekman, r = .25, verbose = FALSE)
ehalf <- fStressMin (ekman, r = .5, verbose = FALSE)
e34 <- fStressMin (ekman, r = .75, verbose = FALSE)
eone <- fStressMin (ekman, r = 1, verbose = FALSE)
etwo <- fStressMin (ekman, r = 2, verbose = FALSE)
ethree <- fStressMin (ekman, r = 3, verbose = FALSE)

save.image (file = "fstress.Rsave")
```

Appendix C. Figures

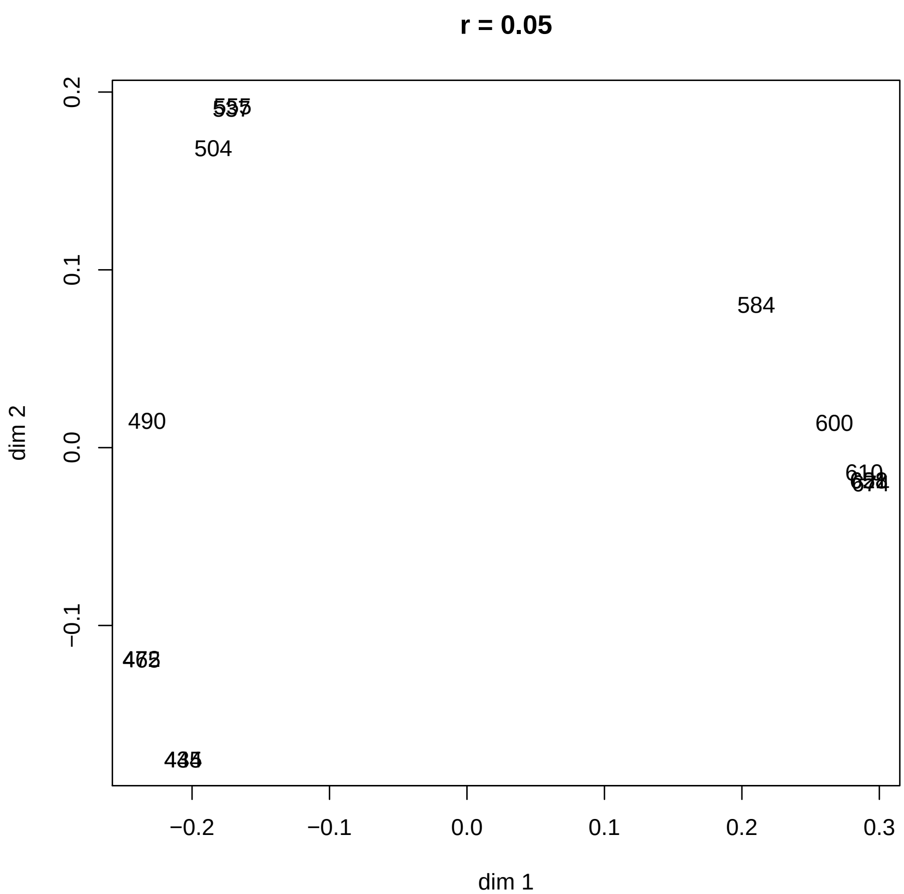

Figure 1. Configuration plot for r=0.05

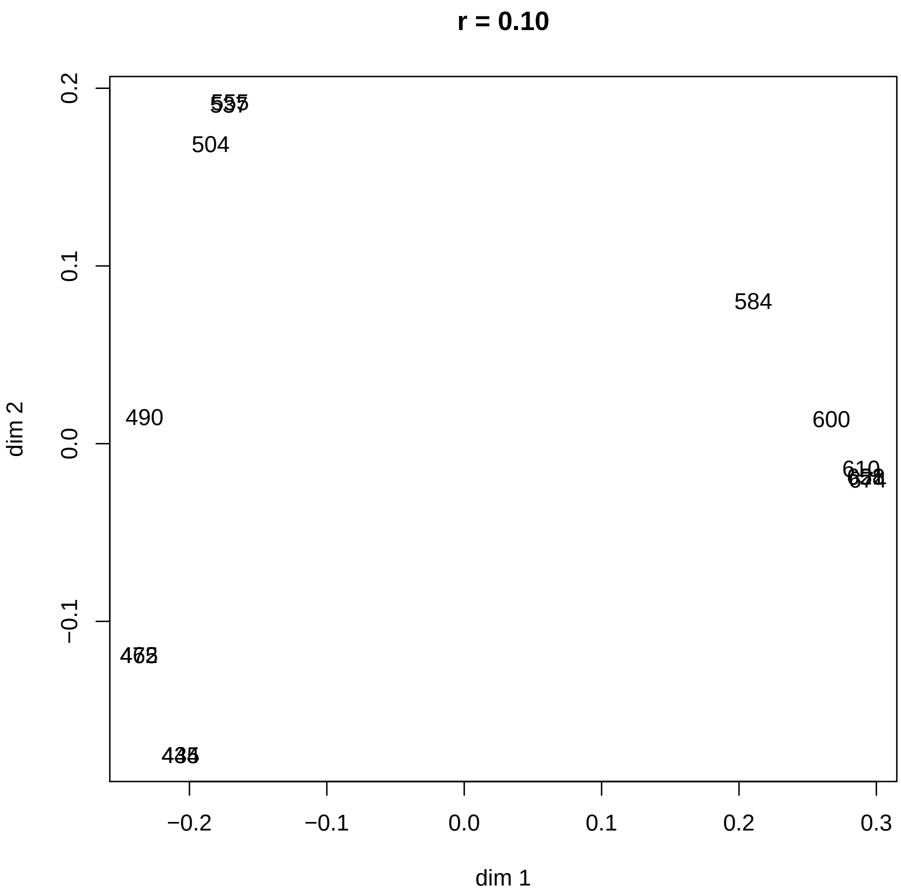

FIGURE 2. Configuration plot for r=0.10

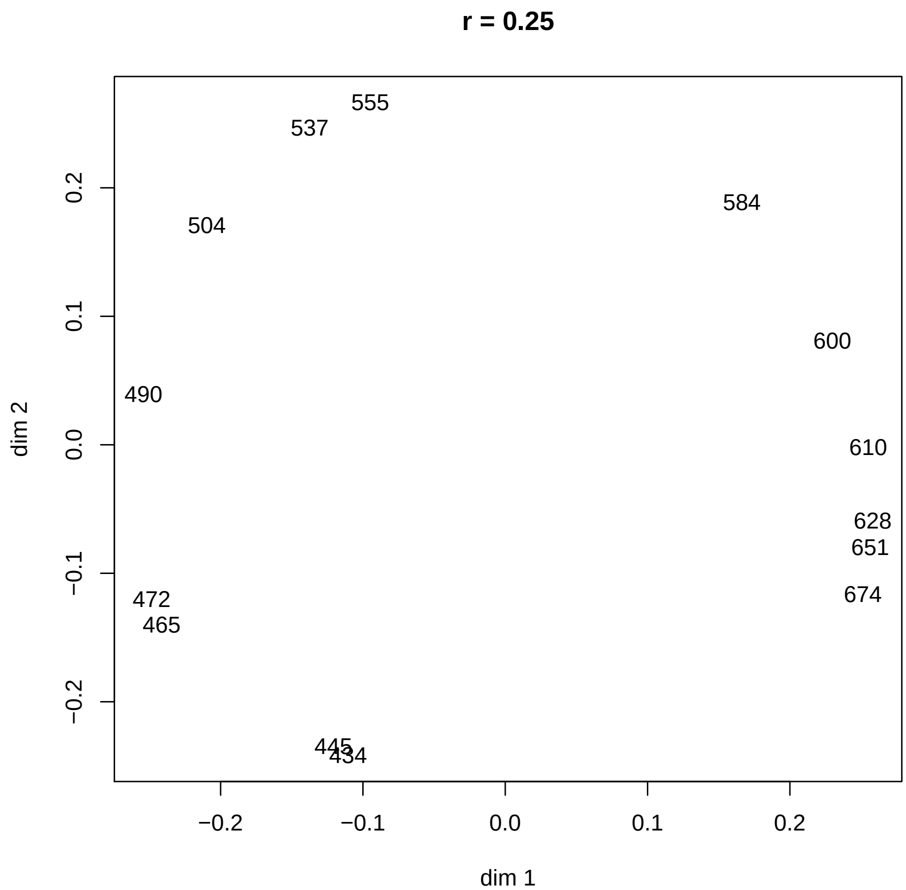

FIGURE 3. Configuration plot for r=0.25

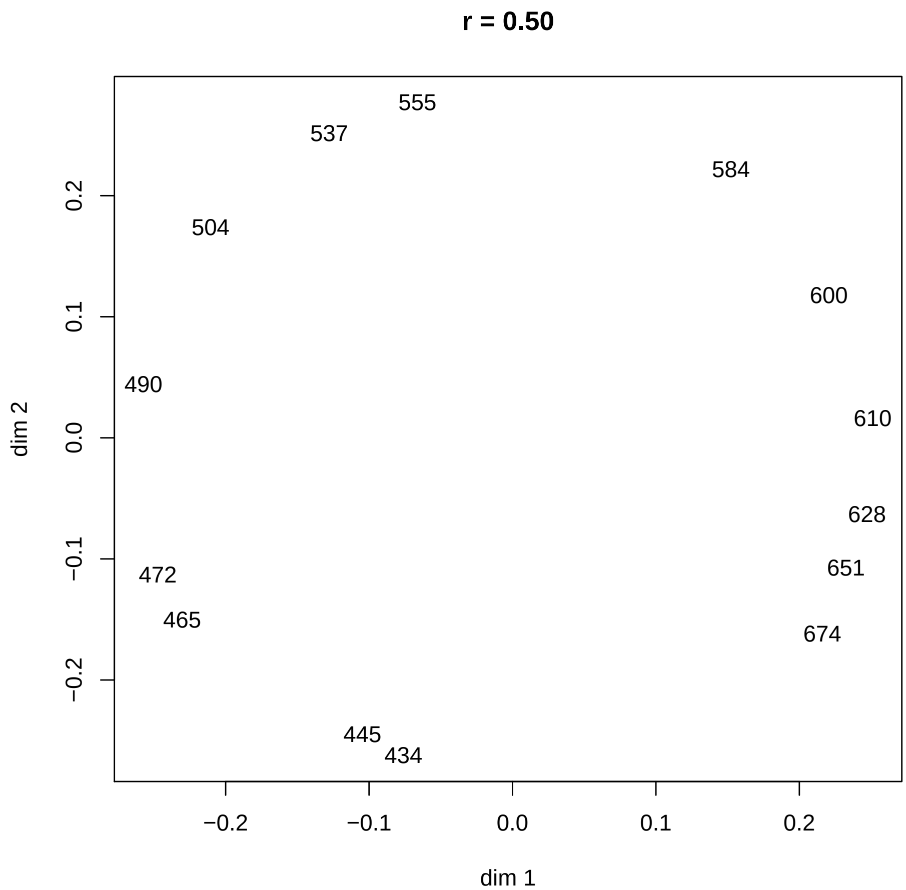

FIGURE 4. Configuration plot for r=0.50

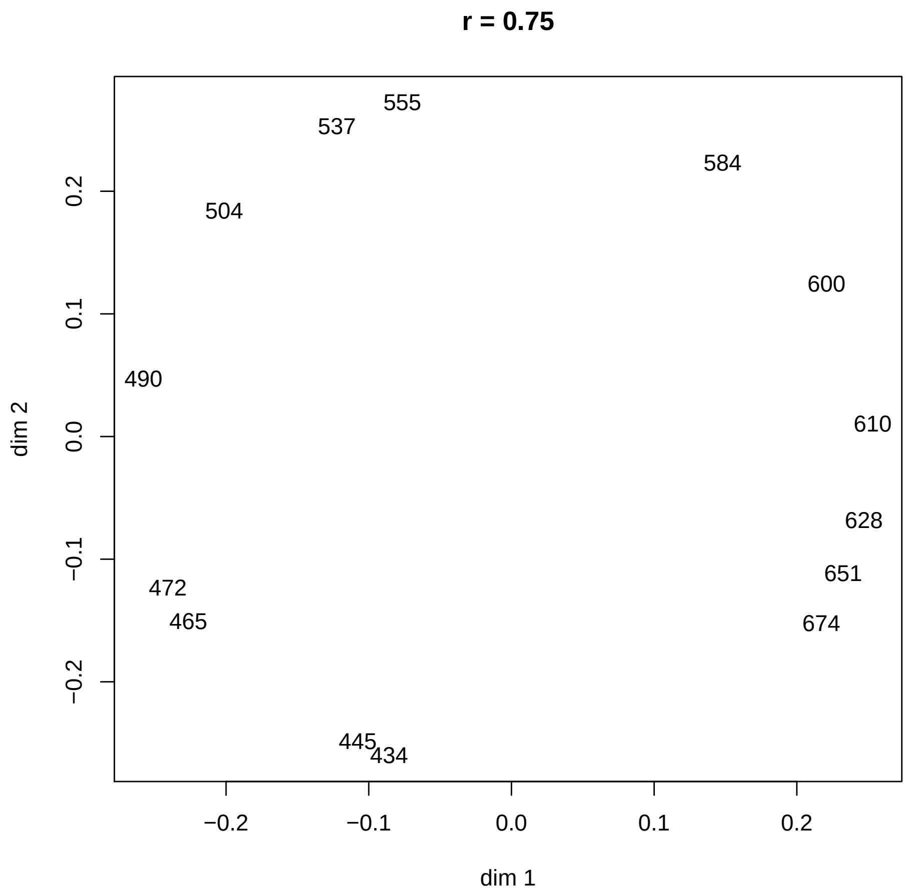

FIGURE 5. Configuration plot for r=0.75

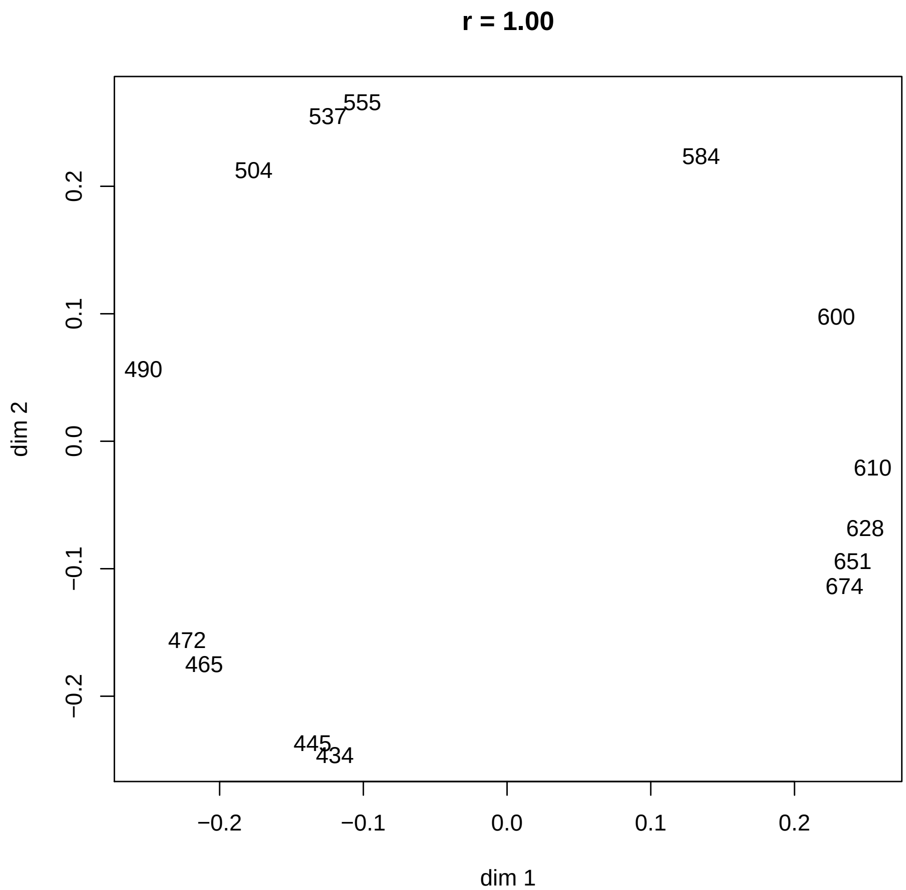

FIGURE 6. Configuration plot for r=1.00

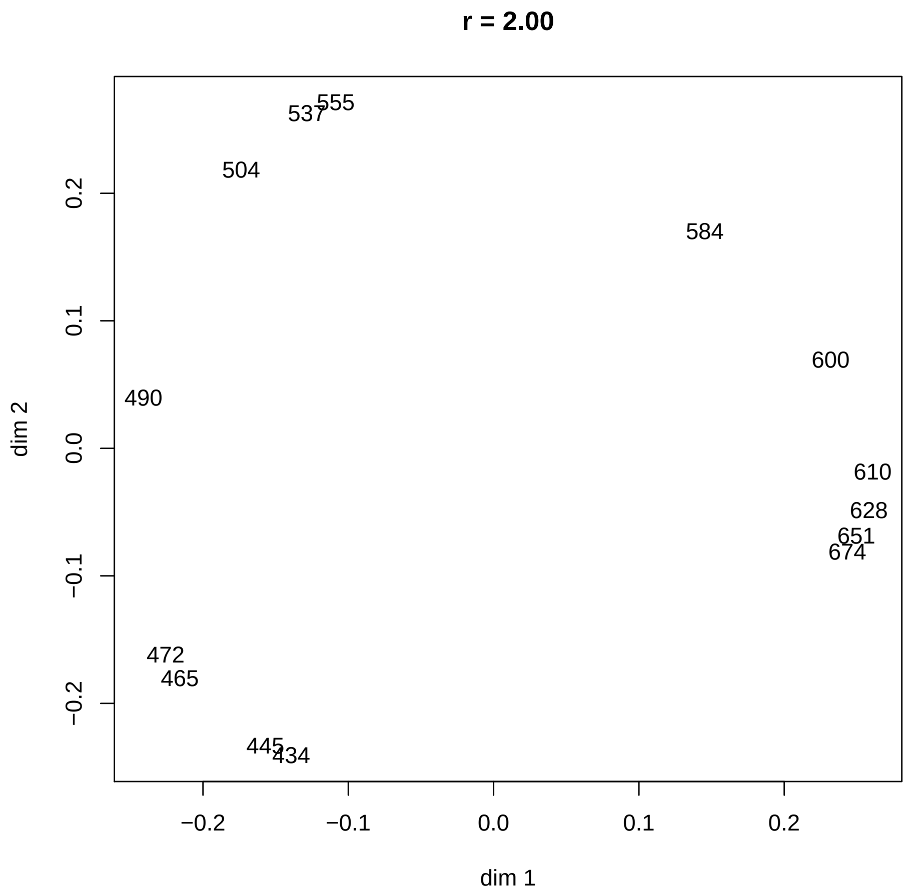

FIGURE 7. Configuration plot for r=2.00

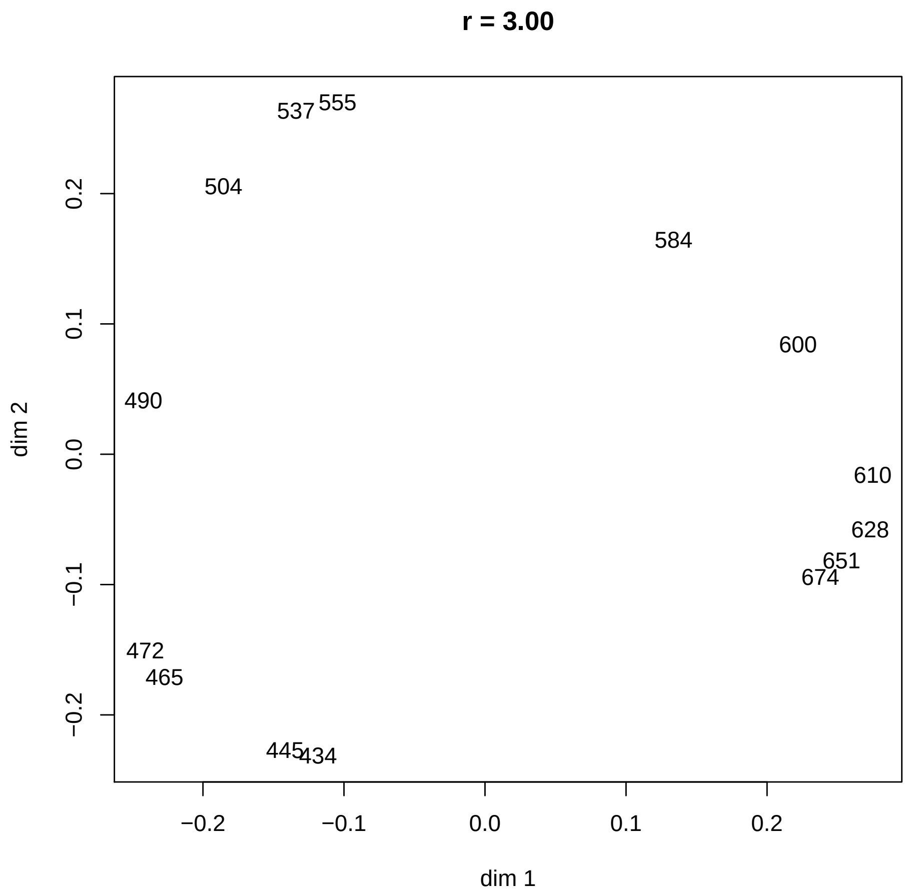

FIGURE 8. Configuration plot for r=3.00

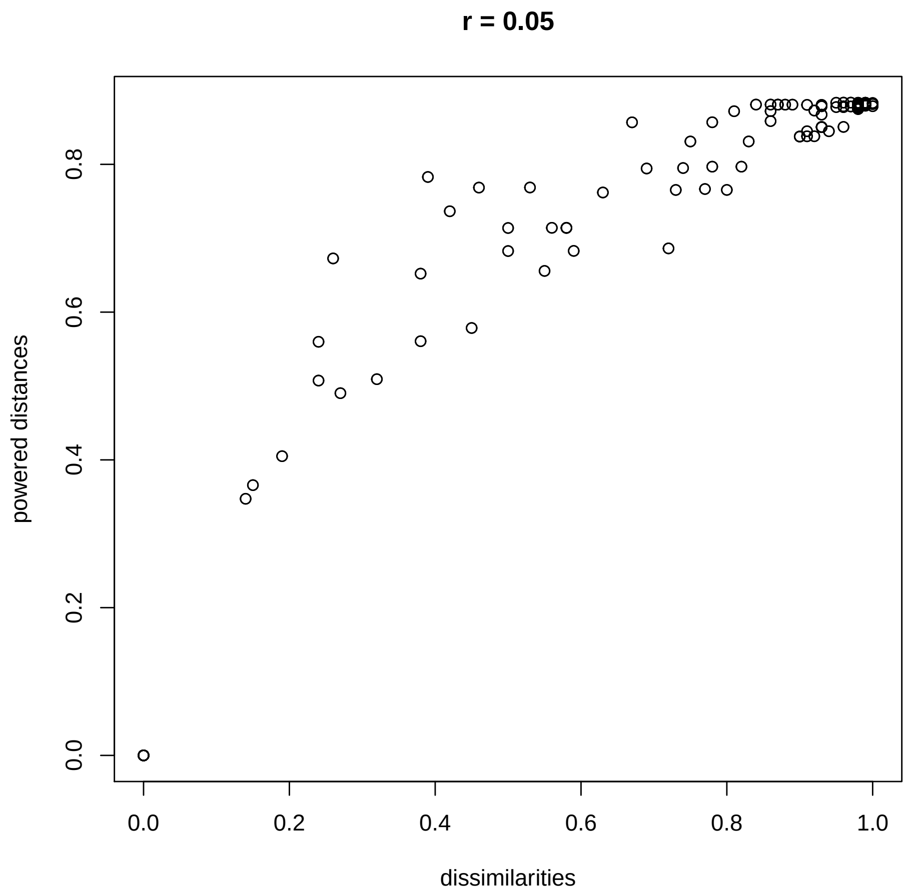

FIGURE 9. Shepard plot for r=0.05

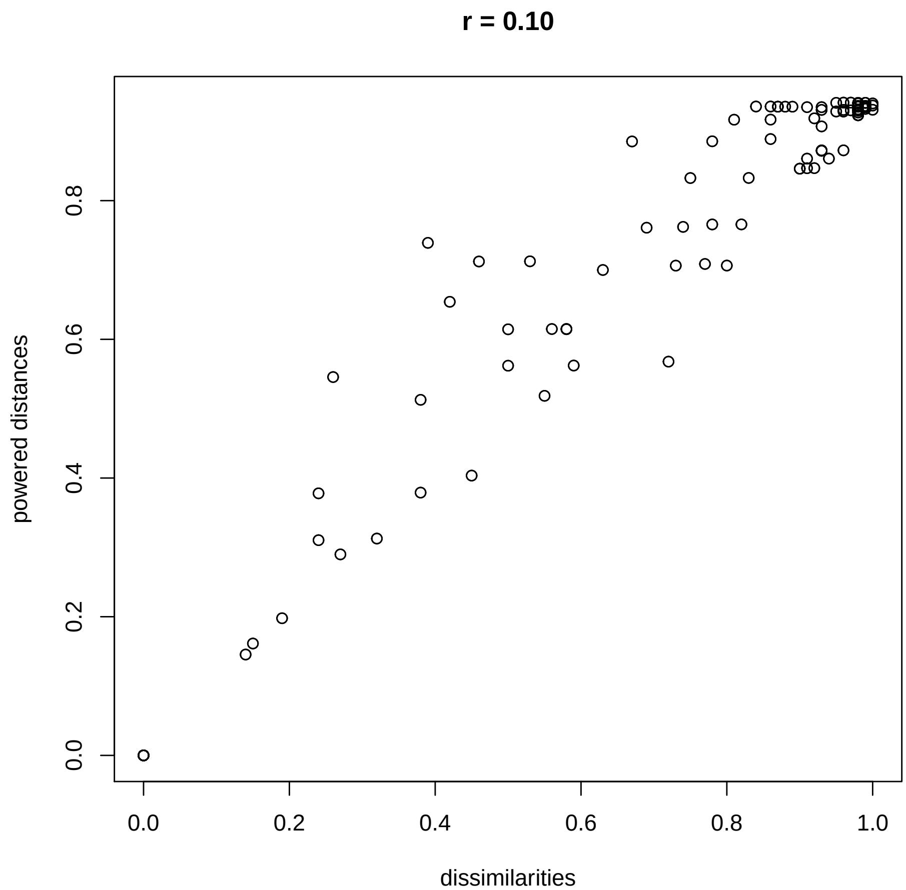

FIGURE 10. Shepard plot for r=0.10

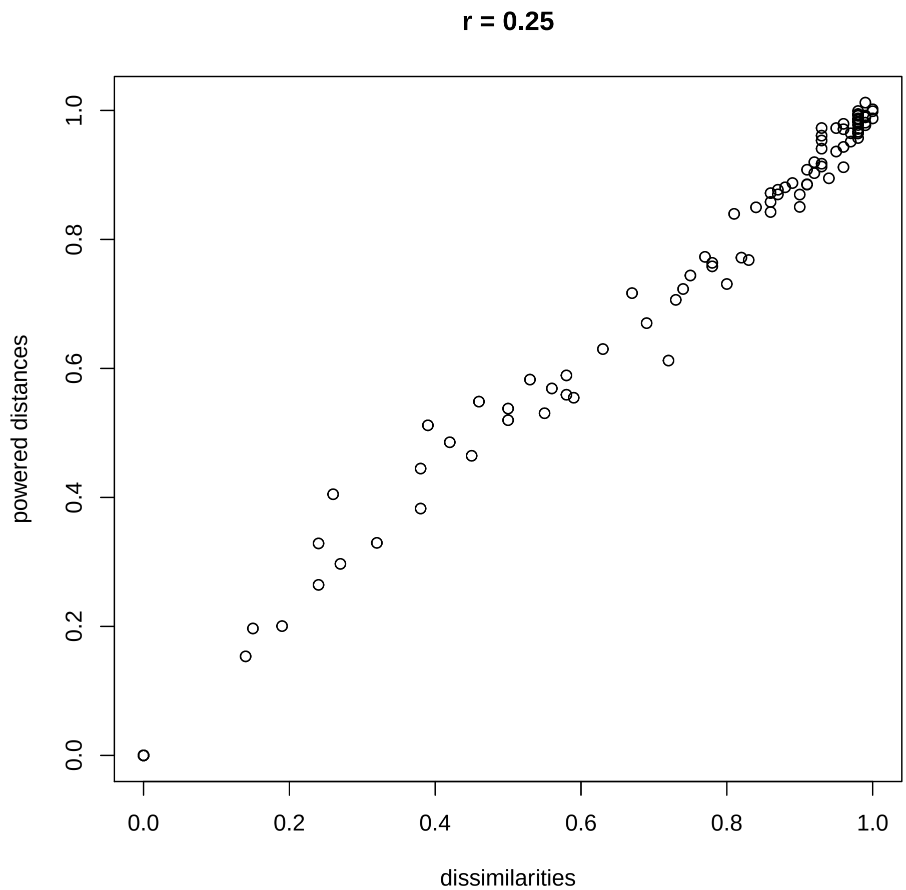

FIGURE 11. Shepard plot for r=0.25

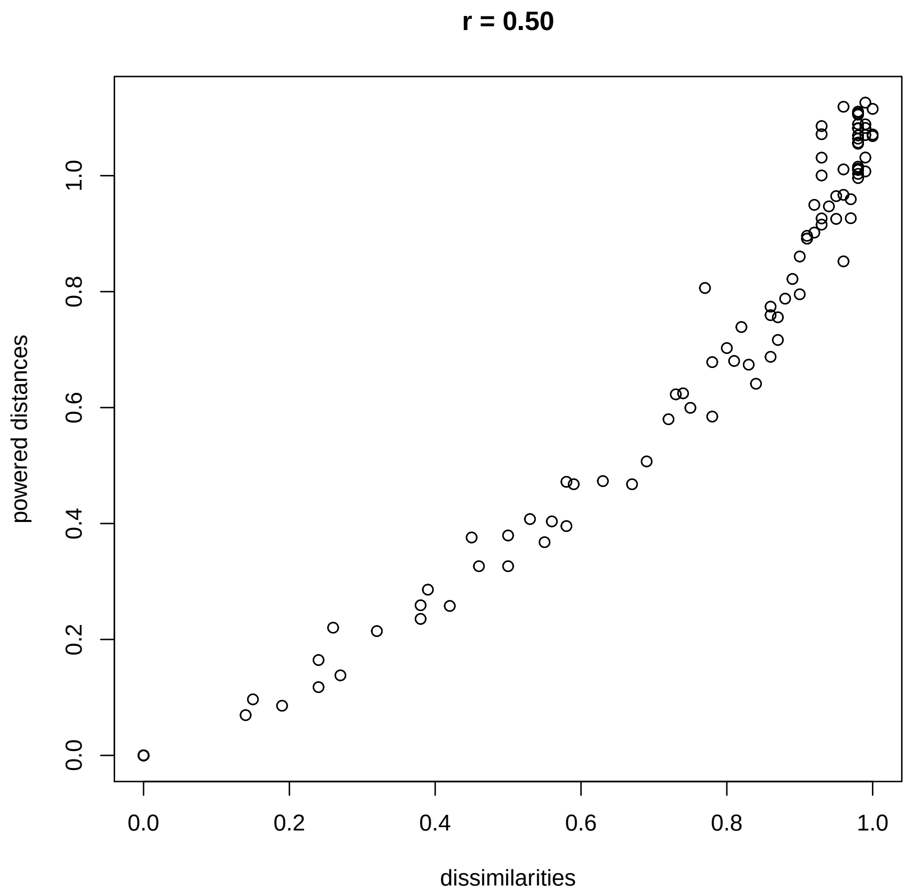

FIGURE 12. Shepard plot for r=0.50

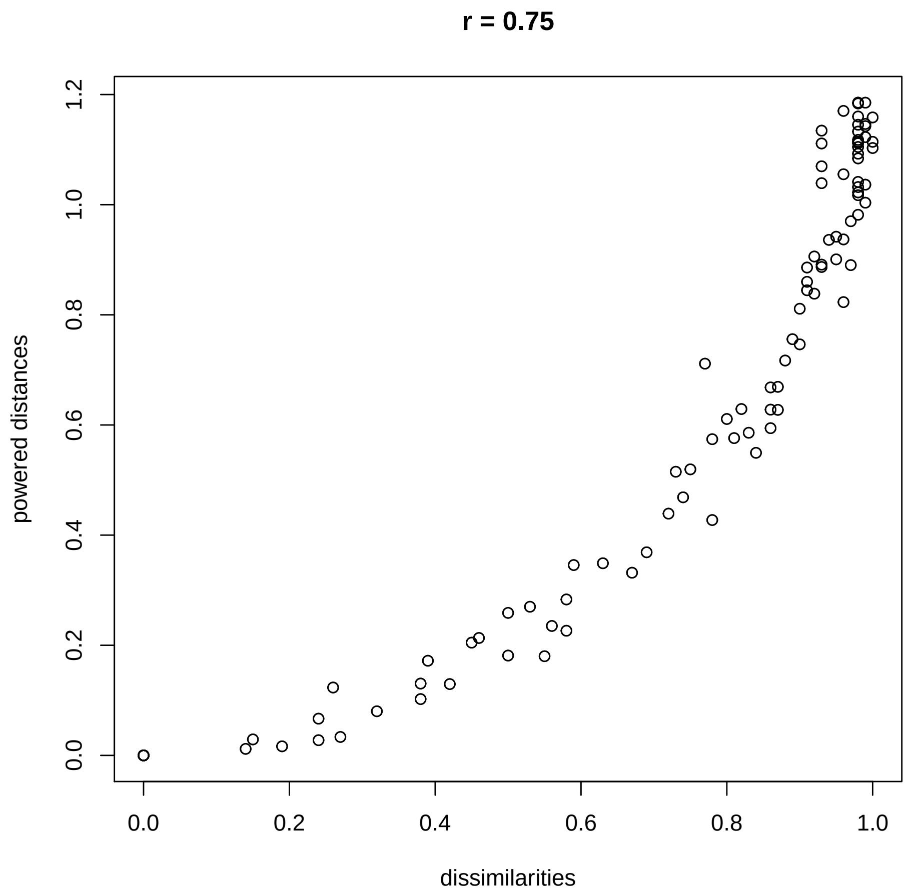

FIGURE 13. Shepard plot for r=0.75

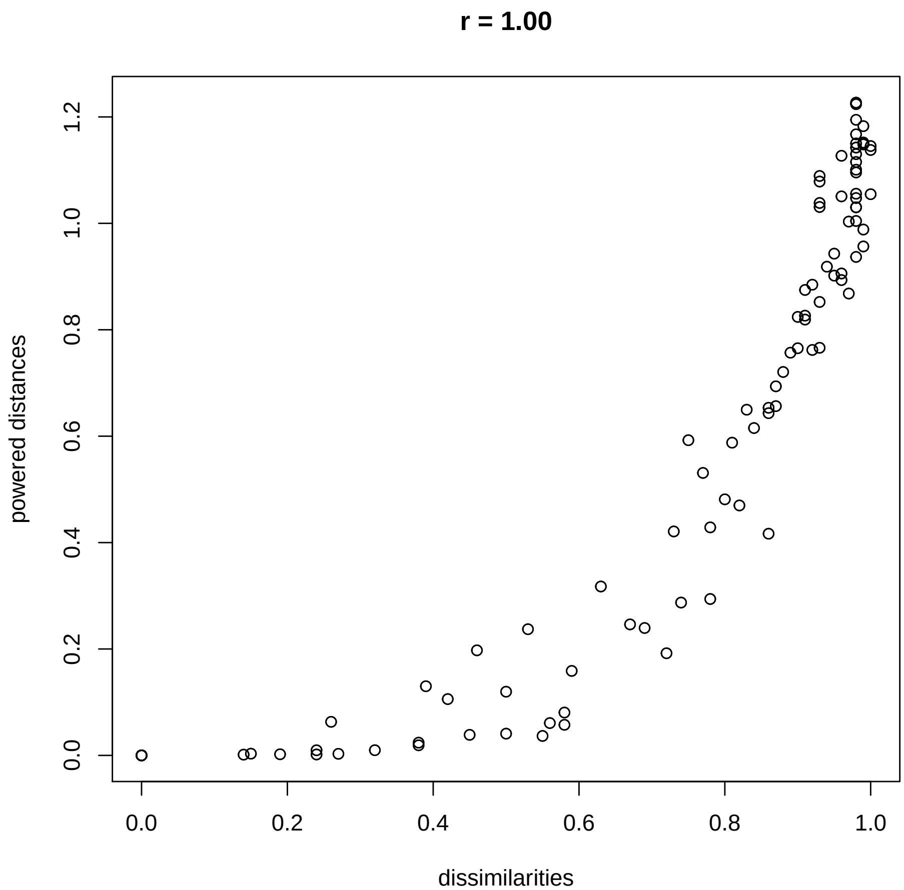

FIGURE 14. Shepard plot for r=1.00

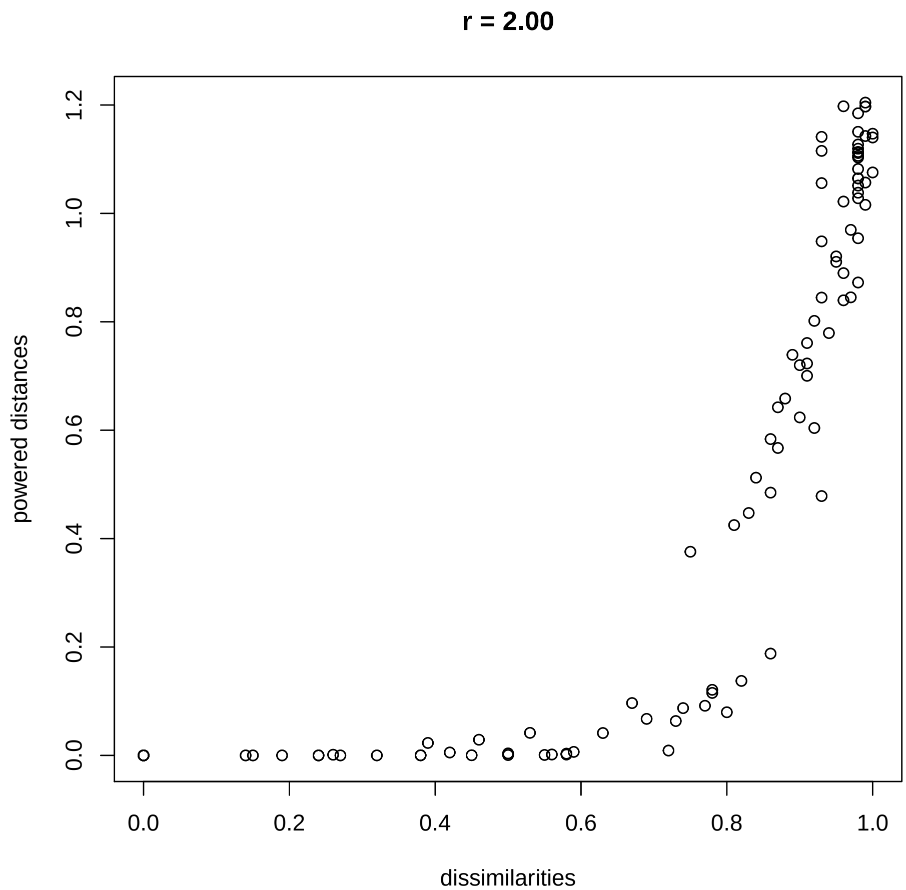

FIGURE 15. Shepard plot for r=2.00

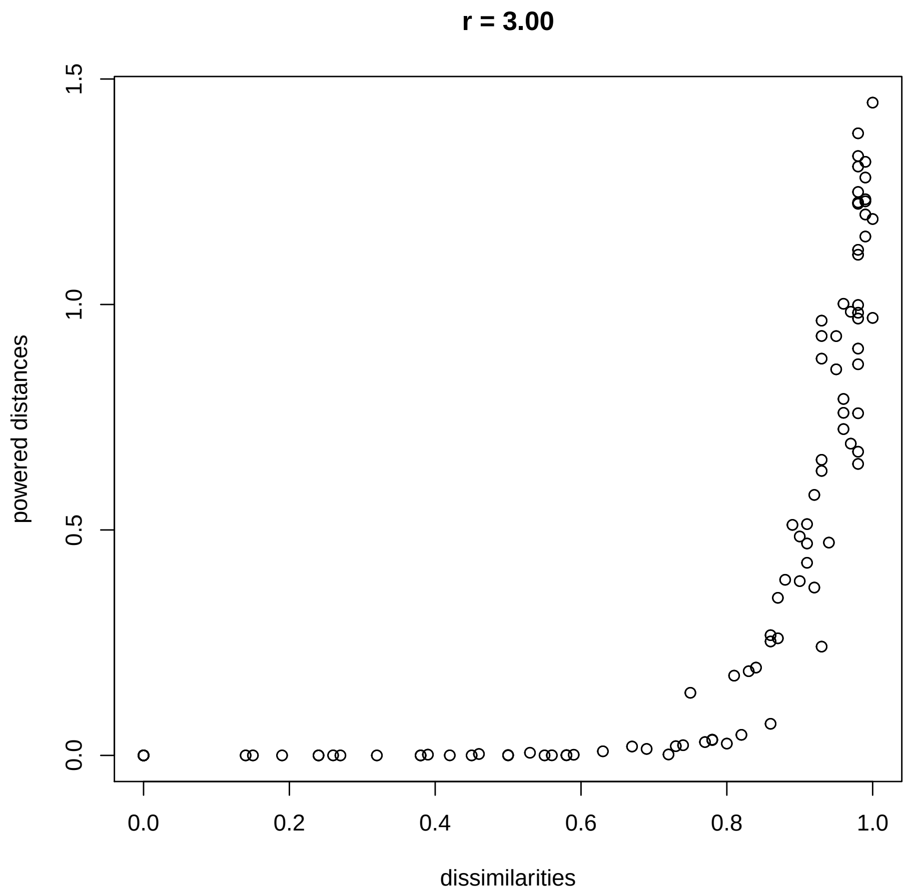

FIGURE 16. Shepard plot for r=3.00

Appendix D. Tables

```
## $x
##                [,1]          [,2]
##  [1,] -0.2064186420 -0.17542647884
##  [2,] -0.2064434825 -0.17538592228
##  [3,] -0.2366875770 -0.11906619847
##  [4,] -0.2367421135 -0.11885197985
##  [5,] -0.2327086797  0.01487970711
##  [6,] -0.1845354404  0.16846404013
##  [7,] -0.1713144661  0.19069202358
##  [8,] -0.1704538460  0.19191264992
##  [9,]  0.2105530417  0.08029678880
## [10,]  0.2672383678  0.01377692126
## [11,]  0.2890410530 -0.01396958093
## [12,]  0.2923788898 -0.01849852392
## [13,]  0.2924259065 -0.01856306452
## [14,]  0.2936669888 -0.02026038175
##
## $gamma
## [1] 0.9942858536 0.9942858537
##
## $itel
## [1] 2078
```

Table 1. Results for $r = 0.05$

```
## $x
##                 [,1]           [,2]
##  [1,] -0.2064186420 -0.17542647884
##  [2,] -0.2064434825 -0.17538592228
##  [3,] -0.2366875770 -0.11906619847
##  [4,] -0.2367421135 -0.11885197985
##  [5,] -0.2327086797  0.01487970711
##  [6,] -0.1845354404  0.16846404013
##  [7,] -0.1713144661  0.19069202358
##  [8,] -0.1704538460  0.19191264992
##  [9,]  0.2105530417  0.08029678880
## [10,]  0.2672383678  0.01377692126
## [11,]  0.2890410530 -0.01396958093
## [12,]  0.2923788898 -0.01849852392
## [13,]  0.2924259065 -0.01856306452
## [14,]  0.2936669888 -0.02026038175
##
## $gamma
## [1] 0.9942858536 0.9942858537
##
## $itel
## [1] 2078
```

TABLE 2. Results for $r = 0.10$

```
## $x
##                [,1]           [,2]
##  [1,] -0.1104446970 -0.241750497123
##  [2,] -0.1207101869 -0.234617544923
##  [3,] -0.2414170911 -0.140125766952
##  [4,] -0.2484165949 -0.119993374094
##  [5,] -0.2541291789  0.039422637057
##  [6,] -0.2096043961  0.170948937659
##  [7,] -0.1373829811  0.246972385797
##  [8,] -0.0948325444  0.266312107988
##  [9,]  0.1662726884  0.188792906607
## [10,]  0.2298247442  0.081190784243
## [11,]  0.2550316661 -0.001997583851
## [12,]  0.2581397137 -0.059181256689
## [13,]  0.2564283819 -0.079642866965
## [14,]  0.2512404761 -0.116330868755
##
## $gamma
## [1] 0.9990442973 0.9990442974
##
## $itel
## [1] 577
```

TABLE 3. Results for $r = 0.25$

```
## $x
##                [,1]           [,2]
## 434 -0.07599404565 -0.26228730067
## 445 -0.10461346632 -0.24489607207
## 465 -0.23032705865 -0.15036175338
## 472 -0.24732138111 -0.11271759212
## 490 -0.25723816154  0.04424561217
## 504 -0.21048008181  0.17386276495
## 537 -0.12773410097  0.25149442994
## 555 -0.06622083933  0.27685477868
## 584  0.15236478966  0.22171049993
## 600  0.22053773701  0.11783931629
## 610  0.25111363573  0.01615762520
## 628  0.24721580375 -0.06303987768
## 651  0.23260695929 -0.10727855931
## 674  0.21609020994 -0.16158387193
##
## $gamma
## [1] 0.9913560127 0.9913560127
##
## $itel
## [1] 503
```

TABLE 4. Results for $r = 0.5$

```
## $x
##                [,1]           [,2]
## 434 -0.08578315819 -0.25999589889
## 445 -0.10773645423 -0.24822586714
## 465 -0.22639619641 -0.15075537651
## 472 -0.24083259085 -0.12326097617
## 490 -0.25782081656  0.04697178768
## 504 -0.20126009229  0.18393622462
## 537 -0.12235044897  0.25294637546
## 555 -0.07643950216  0.27224848978
## 584  0.14800040288  0.22332009667
## 600  0.22078367719  0.12452330665
## 610  0.25312900153  0.01026943981
## 628  0.24699448669 -0.06835371402
## 651  0.23248946162 -0.11128201112
## 674  0.21722222976 -0.15234187683
##
## $gamma
## [1] 0.9722297233 0.9722297234
##
## $itel
## [1] 3276
```

TABLE 5. Results for $r = 0.75$

```
## $x
##               [,1]          [,2]
## 434 -0.1197262846 -0.24645034261
## 445 -0.1353370668 -0.23690853438
## 465 -0.2106976500 -0.17486700375
## 472 -0.2225972691 -0.15594542031
## 490 -0.2529193680  0.05641222164
## 504 -0.1763176332  0.21269742705
## 537 -0.1247634391  0.25483197151
## 555 -0.1007389132  0.26563050370
## 584  0.1350794584  0.22359888630
## 600  0.2290799435  0.09777498637
## 610  0.2543807774 -0.02081959527
## 628  0.2491688991 -0.06804934811
## 651  0.2405326083 -0.09423784053
## 674  0.2348559375 -0.11366791161
##
## $gamma
## [1] 0.9523319539 0.9523319540
##
## $itel
## [1] 13660
```

TABLE 6. Results for $r = 1.00$

```
## $x
##               [,1]           [,2]
## 434 -0.1392215974 -0.24079496978
## 445 -0.1570059462 -0.23324201342
## 465 -0.2159364583 -0.18016104236
## 472 -0.2257076333 -0.16177674564
## 490 -0.2408905693  0.03980796025
## 504 -0.1736970345  0.21847458932
## 537 -0.1284836416  0.26284662477
## 555 -0.1086113557  0.27109615903
## 584  0.1453867185  0.17022892494
## 600  0.2319463331  0.06944576179
## 610  0.2608441504 -0.01819262207
## 628  0.2582071325 -0.04837043239
## 651  0.2496864774 -0.06838773666
## 674  0.2434834244 -0.08097445778
##
## $gamma
## [1] 0.9045693554 0.9045694314
##
## $itel
## [1] 1e+05
```

TABLE 7. Results for $r = 2.00$

```
## $x
##               [,1]          [,2]
## 434 -0.1183498633 -0.23140513679
## 445 -0.1418032977 -0.22703246154
## 465 -0.2272949437 -0.17088079681
## 472 -0.2408899045 -0.15064064627
## 490 -0.2421309541  0.04144020149
## 504 -0.1854257853  0.20574731496
## 537 -0.1339492433  0.26355907350
## 555 -0.1045582842  0.26976362404
## 584  0.1337444304  0.16439788379
## 600  0.2219694242  0.08426823708
## 610  0.2748923517 -0.01583997876
## 628  0.2731521800 -0.05765519522
## 651  0.2528793559 -0.08147936021
## 674  0.2377645338 -0.09424275927
##
## $gamma
## [1] 0.8451640000 0.8451640093
##
## $itel
## [1] 1e+05
```

TABLE 8. Results for $r = 3.00$

REFERENCES

[1] J. De Leeuw. "Applications of Convex Analysis to Multidimensional Scaling". In: *Recent developments in statistics*. Ed. by J.R. Barra et al. Amsterdam, The Netherlands: North Holland Publishing Company, 1977, pp. 133–145. URL: http://www.stat.ucla.edu/~deleeuw/janspubs/1977/chapters/deleeuw_C_77.pdf.

[2] J. De Leeuw. "Block Relaxation Algorithms in Statistics". In: *Information Systems and Data Analysis*. Ed. by H.H. Bock, W. Lenski, and M.M. Richter. Berlin: Springer Verlag, 1994, pp. 308–324. URL: http://www.stat.ucla.edu/~deleeuw/janspubs/1994/chapters/deleeuw_C_94c.pdf.

[3] J. De Leeuw and K. Lange. "Sharp Quadratic Majorization in One Dimension". In: *Computational Statistics and Data Analysis* 53 (2009), pp. 2471–2484. URL: http://www.stat.ucla.edu/~deleeuw/janspubs/2009/articles/deleeuw_lange_A_09.pdf.

[4] W. Dinkelbach. "On Nonlinear Fractional Programming". In: *Management Science* 13 (1967), pp. 492–498.

[5] G. Ekman. "Dimensions of Color Vision". In: *Journal of Psychology* 38 (1954), pp. 467–474.

[6] G.E. Forsythe and G.H. Golub. "On the Stationary Values of a Second Degree Polynomial on the Unit Sphere". In: *Journal of the Society for Industrial and Applied Mathematics* 13 (1965), pp. 1050–1068.

[7] G.H. Golub. "Some Modified Matrix Eigenvalue Problems". In: *SIAM Review* 15 (1973), pp. 318–334.

[8] William W. Hager. "Minimizing a Quadratic over a Sphere". In: *SIAM Journal of Optimization* 12 (2001), pp. 188–208.

[9] W.J. Heiser. "Convergent Computing by Iterative Majorization: Theory and Applications in Multidimensional Data Analysis". In: *Recent Advantages in Descriptive Multivariate Analysis*. Ed. by W.J. Krzanowski. Oxford: Clarendon Press, 1995, pp. 157–189.

[10] J. B. Kruskal. "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis". In: *Psychometrika* 29 (1964), pp. 1–27.

[11] J.B. Kruskal. "Nonmetric Multidimensional Scaling: a Numerical Method". In: *Psychometrika* 29 (1964), pp. 115–129.

[12] K. Lange, D.R. Hunter, and I. Yang. "Optimization Transfer Using Surrogate Objective Functions". In: *Journal of Computational and Graphical Statistics* 9 (2000), pp. 1–20.

[13] J. O. Ramsay. "Maximum Likelihood Estimation in MDS". In: *Psychometrika* 42 (1977), pp. 241–266.

[14] J. O. Ramsay. "Some statistical approaches to multidimensional scaling data". In: *J. Roy. Statist. Soc. Ser. A* 145.3 (1982). With discussion and a reply by the author, pp. 285–312. ISSN: 0035-9238.

[15] D.C. Sorensen. "Newton's Method with a Model Trust Region Mdification". In: *SIAM Journal of Numerical Analysis* (1982), pp. 409–426.

[16] E. Spjøtvoll. "A Note on a Theorem of Forsythe and Golub". In: *SIAM Journal on Applied Mathematics* 23 (1972), pp. 307–311.

[17] Y. Takane, F.W. Young, and J. De Leeuw. "Nonmetric Individual Differences in Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features". In: *Psychometrika* 42 (1977), pp. 7–67. URL: `http://www.stat.ucla.edu/~deleeuw/janspubs/1977/articles/takane_young_deleeuw_A_77.pdf`.

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095-1554

E-mail address, Jan de Leeuw: `deleeuw@stat.ucla.edu`

URL, Jan de Leeuw: `http://gifi.stat.ucla.edu`