The Positive Orthant Method

Jan de Leeuw

Version May 07, 2018

Abstract

The positive orthant method tries to find solutions to consistent systems of inequalities, and approximate solutions to inconsistent systems, by maximizing a fit measure based on the sign function and the absolute value function. We concentrate on systems of linear inequalities and develop a convergent majorization algorithm.

Contents

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

1 Introduction

In metric scaling methods we have data, which we take to be a vector $\zeta \in \mathbb{R}^n$, and we have a model, which we take to be a function $F : \mathbb{R}^p \Rightarrow \mathbb{R}^n$. We want to find $x \in \mathbb{R}^p$ such that $\zeta \approx F(x)$. In non-metric scaling methods this is more specifically formulated as $\mathbf{rank}(\zeta) \approx \mathbf{rank}(F(x))$. For our purposes it is convenient to define the rank $\mathbf{rank}(x)$ of a vector $x \in \mathbb{R}^n$ by

$$\mathbf{rank}_i(x) \triangleq \frac{1}{2}\sum_{j=1}^n \mathbf{sign}\ (x_i - x_j),$$

where

$$\mathbf{sign}\ (x) = \begin{cases} +1 & \text{if } x > 0, \\ -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Thus rank vectors are centered, and tied elements get an average rank number.

```
x<-c(1,2,3,4,4,5)
print (rowSums(sign(outer(x,x,"-")))/2)
```

```
## [1] -2.5 -1.5 -0.5  1.0  1.0  2.5
```

Many of the computational complications in non-metric scaling arise from the fact that $\mathbf{rank} : \mathbb{R}^n \Rightarrow \mathbb{R}^n$ is not smooth, in fact not even continuous.

There are other complications, however. Often the date do not consist of a single rank order, but of multiple rank orders. This is generally not difficult to incorporate into the various approaches that have been used so far. A more serious complication is that sometimes data are not given as rank orders, but as paired comparisons or triadic comparisons. Paired comparison data are not necessarily consistent with any given rank order, because they allow subjects to give asymmetric and intransitive judgments. In this paper we discuss non-metric scaling methods that deal with paired comparison data. Since rank orders, and more generally "rank k out of n" or "pick k out of n" data, can always be converted to paired comparisons, this generalizes treatment of rank ordered data. See, however, (**???**) for some thoughtful discussion how to handle triadic and similar data.

Suppose we have the following building blocks.

1. f_i with $i = 1, \cdots, n$ are real valued functions on $\mathbb{R}^p$.
2. $\Sigma = \{\sigma_{ij}\}$ is a square matrix of order n with elements $+1$, -1, or 0.

These components define a system of inequalities of the form

$$\sigma_{ij}(f_i(x) - f_j(x)) \geq 0. \tag{1}$$

We can assume, without loss of generality, that Σ is hollow (i.e. has zero diagonal elements). But we do *not* assume that Σ is anti-symmetric, i.e. that $\sigma_{ij} = -\sigma_{ji}$. Of course if $\sigma_{ij} = -\sigma_{ji}$ then $\sigma_{ij}(f_i(x) - f_j(x)) \geq 0$ and $\sigma_{ji}(f_j(x) - f_i(x)) \geq 0$ are the same inequality, and one of them is redundant. More generally, we allow for redundant inequalities in (1), and even for replications of the same inequality.

We also allow for the possibility that $\sigma_{ij} = \sigma_{ji} = 1$ or $\sigma_{ij} = \sigma_{ji} = -1$ for a pair (i, j), in which case (1) requires that $f_i(x) = f_j(x)$. In this way we can incorporate equality restrictions into our system of inequalities.

Of course if $\sigma_{ij} = 0$ the corresponding inequality is always trivially satisfied, whatever x is. By allowing for zero σ_{ij} we can proceed as if (1) is defined for all (i, j), even if we only have observed a subset of the possible binary comparisons in our data. Or if we decide to eliminate redundant inequalities. In that sense zero elements of Σ correspond with “missing data”.

Note that we only mention in passing the various coding decisions that have to be made. If $\mathbf{rank}(\zeta)$ is 1, 2, 3, 4, 4, 5, for example, we can decide to code all $\binom{6}{2}$ comparisons in the matrix Σ, using the sign function. Then Σ is

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    0   -1   -1   -1   -1   -1
## [2,]    1    0   -1   -1   -1   -1
## [3,]    1    1    0   -1   -1   -1
## [4,]    1    1    1    0    0   -1
## [5,]    1    1    1    0    0   -1
## [6,]    1    1    1    1    1    0
```

Note that this implies we use what Kruskal (1964) calls the “primary approach” to ties, where equality means we leave the order within tie-blocks unspecified. We can also use the “secondary approach”, where we require equality within tie-blocks. Then Σ becomes

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    0   -1   -1   -1   -1   -1
## [2,]    1    0   -1   -1   -1   -1
## [3,]    1    1    0   -1   -1   -1
## [4,]    1    1    1    0    1   -1
## [5,]    1    1    1    1    0   -1
## [6,]    1    1    1    1    1    0
```

We could also decide to remove redundant inequalities and wind up with

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    0   -1    0    0    0    0
## [2,]    0    0   -1    0    0    0
## [3,]    0    0    0   -1    0    0
## [4,]    0    0    0    0    1    0
## [5,]    0    0    0    1    0   -1
## [6,]    0    0    0    0    0    0
```

Using the sign function on pairs of model values has a long tradition in statistics and psychometrics. We review some of the history, with a slight emphasis on our own work, in a later section of this report.

2 Fit Function

We measure how well the inequalities (1) are satisfied by defining the fit function

$$\phi(x) \triangleq \frac{\alpha(x)}{\beta(x)}, \tag{2}$$

with

$$\alpha(x) \triangleq \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\sigma_{ij}(f_i(x) - f_j(x)), \tag{3}$$

$$\beta(x) \triangleq \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}|\sigma_{ij}(f_i(x) - f_j(x))|. \tag{4}$$

Because $w_{ij}\sigma_{ij} = w_{ij}$ the denominator β can also be written more simply as

$$\beta(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}|(f_i(x) - f_j(x))|.$$

Note that β is the weighted Gini Mean Difference of the $f_i(x)$ (Yitzhaki (2003), Yitzhaki and Schechtman (2013)).

Maximization of ϕ over x to find a solution of the system (1) if it is consistent, and an approximate solution if it is inconsistent, is the *Positive Orthant Method.* The name is chosen because we want the vector $\sigma_{ij}(f_i(x) - f_j(x))$ to be in the non-negative orthant.

Clearly

$$-|\sigma_{ij}(f_i(x) - f_j(x))| \leq \sigma_{ij}(f_i(x) - f_j(x)) \leq |\sigma_{ij}(f_i(x) - f_j(x))|,$$

and thus $-1 \leq \phi(x) \leq +1$ for all x for which the denominator is non-zero. Also $\phi(x) = 1$ if and only if $\sigma_{ij}(f_i(x) - f_j(x)) \geq 0$ for all (i, j) for which $\sigma_{ij} \neq 0$, again assuming the denominator is non-zero.

Again, the history of this loss function is reviewed towards the end of the report.

3 Analysis

3.1 General Case

If we define

$$\rho_i \triangleq \sum_{j=1}^{n}(w_{ij}\sigma_{ij} - w_{ji}\sigma_{ji}),$$

then

$$\alpha(x)=\sum_{i=1}^n\rho_if_i(x). \tag{5}$$

To verify, suppose $F(x)$ is

```
## [1] 3 4 2 5 1
```

Also choose W as

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    0    0    1    1    1
## [2,]    0    0    1    1    1
## [3,]    0    0    0    0    0
## [4,]    0    0    0    0    0
## [5,]    0    0    0    0    0
```

and Σ as

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    0    1    1    1    1
## [2,]   -1    0   -1   -1    1
## [3,]   -1    1    0    1   -1
## [4,]    1    1   -1    0    1
## [5,]    1    1    1   -1    0
```

Then ρ is

```
print (r <- rowSums (w * s - t(w * s)))
```

```
## [1]  3 -1  0  0 -2
```

From (3) we compute

```
sum (w * s * outer (f, f, "-"))
```

```
## [1] 3
```

while the right-hand side of $(5) is

```
sum (r * f)
```

```
## [1] 3
```

For the denominator of ϕ we find

$$\beta(x)=\sum_{i=1}^n\rho_i(x)f_i(x), \tag{6}$$

where

$$\rho_i(x) \overset{\Delta}{=} \sum_{j=1}^n (w_{ij} + w_{ji})\sigma_{ij}(x)$$

and $\sigma_{ij}(x) \overset{\Delta}{=} \textbf{sign}(f_i(x) - f_j(x))$.

As an example, we have for $\Sigma(x)$

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    0   -1    1   -1    1
## [2,]    1    0    1   -1    1
## [3,]   -1   -1    0   -1    1
## [4,]    1    1    1    0    1
## [5,]   -1   -1   -1   -1    0
```

From the definition in (4) we compute

```
sum (w * abs (outer (f, f, "-")))
```

```
## [1] 11
```

and the right-hand side of (6) is

```
sum (f * rowSums ((w + t(w)) * s))
```

```
## [1] 11
```

3.2 Rearrangement

The classical rearrangement theorem in Hardy, Littlewood, and Polya (1952), Section 10.2, Theorem 368 says that if we have two sorted vectors $x_1 \leq x_2 \cdots \leq x_n$ and $y_1 \leq y_2 \leq \cdots \leq y_n$ and if π is any permutation, i.e. any one-to-one mapping of $\mathcal{I}_n \overset{\Delta}{=} \{1, 2, \cdots, n\}$ into $\mathcal{I}_n$, then

$$\sum_{i=1}^n x_i y_i \geq \sum_{i=1}^n x_i y_{\pi(i)}.$$

The rearrangement theorem was used by De Leeuw (1973) to define a family of multidimensional scaling methods. There is a more recent discussion in De Leeuw (2017b). In this section we show that the positive orthant method can also be interpreted as a rearrangement method. Related material on alternative representations of the Gini Mean Difference are in Yitzhaki and Schechtman (2013), Chapter 2, which has the promising title "More Than a Dozen Alternative Ways of Spelling Gini".

For any sign matrix S

$$w_{ij}\ |f_i(x) - f_j(x)| \geq w_{ij}\ s_{ij}(f_i(x) - f_j(x))$$

for all pairs (i, j). Thus, adding up these inequalities,

$$\beta(x) \geq \sum_{i=1}^n f_i(x) \sum_{j=1}^n (w_{ij} + w_{ji}) s_{ij},$$

with equality if and only if $s_{ij} = \mathbf{sign}(f_i(x) - f_j(x))$ for all pairs for which $w_{ij}(f_i(x) - f_j(x)) \neq 0$. This implies

$$\beta(x) = \max_{S \in \mathcal{S}} \sum_{i=1}^n f_i(x) \sum_{j=1}^n (w_{ij} + w_{ji}) s_{ij}$$

This can also be written as

$$\beta(x) = \max_{r \in \mathcal{R}} \sum_{i=1}^n r_i f_i(x), \tag{7}$$

where $\mathcal{R}$ is the set of vectors of the form

$$r_i \triangleq \sum_{j=1}^n (w_{ij} + w_{ji}) s_{ij},$$

with $S = \{s_{ij}\}$ varying over the sign matrices (i.e. essentially over all permutations).

We check our result by using the `nextPermutation()` function from De Leeuw (2016), and apply it to our small example.

```
i <- 1:5
m <- -Inf
repeat {
  t <- sign(outer(i, i, "-"))
  m <- max (m, sum (f * rowSums ((w + t(w)) * t)))
  i <- nextPermutation (i)
  if (is.null (i)) break
}
print (m)
```

```
## [1] 11
```

which is the same as our previous result.

$$\phi(x) = \frac{\sum_{i=1}^n \rho_i f_i(x)}{\max_{r \in \mathcal{R}} \sum_{i=1}^n r_i f_i(x)} \tag{8}$$

The numerator of ϕ is a linear function of f. The denominator of ϕ is the maximum of a finite number of functions linear in f, which means it is convex in f. This does not mean, of course, that it is convex in x, except when the f_i itself are linear.

Also note that if the f_i are homogeneous, in the sense that $f(\lambda x) = \lambda^r f(x)$ for some $r > 0$ and for all x and $\lambda > 0$, then $\phi(\lambda x) = \phi(x)$ and maximizing ϕ can be done by maximizing α over all x for which $\beta \leq 1$. Note that $\beta(x) \leq 1$ if and only if $\sum_{i=1}^n f_i(x) \sum_{j=1}^n (w_{ij} + w_{ji}) s_{ij} \leq 1$ for all $S \in \mathcal{S}$.

If $w_{ij} = 1$ for all $i \neq j$ and S is a sign matrix with $s_{ij} = \mathbf{sign}(y_i - y_j)$, then

$$r_i = 2\sum_{j=1}^n s_{ij} = 4\ \mathbf{rank}_i(y)$$

In that case we have, from (7), by the rearrangement theorem

$$\beta(x) = 4\sum_{i=1}^n \mathbf{rank}_i(F(x))f_i(x) \tag{9}$$

Thus, if Σ is the sign matrix of the data ζ,

$$\phi(x) = \frac{\sum_{i=1}^n \mathbf{rank}_i(\zeta)f_i(x)}{\sum_{i=1}^n \mathbf{rank}_i(F(x))f_i(x)}, \tag{10}$$

where $F(x)$ is the vector with the $f_i(x)$.

3.3 Multiple Data Sources

If there are multiple data sources, for example multiple individuals in a paired comparisons design, then we simply combine the fit measures by adding them. This works well, because the denominators are all the same. We just add all numerators, and the result is still linear in f.

$$\phi(x) = \sum_{j=1}^m \left\{\frac{\sum_{i=1}^n \rho_{ij}f_i(x)}{\max_{r\in\mathcal{R}}\sum_{i=1}^n r_if_i(x)}\right\} \tag{11}$$

But now consider the method of triads, for example. We could take a single triad as a data source, with only three non-zero weights, and sum loss functions over triads. Since the weights for each triad are different, the denominators are different and we cannot just add numerators.

$$\phi(x) = \sum_{j=1}^m \left\{\frac{\sum_{i=1}^n \rho_{ij}f_i(x)}{\max_{r\in\mathcal{R}_j}\sum_{i=1}^n r_if_i(x)}\right\} \tag{12}$$

Individual difference models are an additional level of complication. The model F is different for each j, because there are parameters for individuals or replications, and the combined loss is consequently more involved.

$$\phi(x) = \sum_{j=1}^m \left\{\frac{\sum_{i=1}^n \rho_{ij}f_i(x,y_j)}{\max_{r\in\mathcal{R}_j}\sum_{i=1}^n r_if_i(x,y_j)}\right\} \tag{13}$$

4 Linear Nonmetric Problems

This section improves on the rather shaky early papers De Leeuw (1968b) and De Leeuw (1969). In the linear case we write $f_i(x) = f_i'x$. Note that this implies that $\sigma_{ij}(f_i'x - f_j'x) \geq 0$ is automatically satisfied if $f_i = f_j$. Also note that because we take differences we cannot estimate an intercept. It also means that $\alpha(x) = x'F'\rho$. We can require without loss of generality that $\alpha(x) \geq 0$.

4.1 Linear Programming

The positive orthant method is equivalent to the constrained optimization problem

$$\max \{\sum_{i=1}^n \rho_i f_i(x) \mid \sum_{i=1}^n r_i f_i(x) \leq 1, \forall r \in \mathcal{R}\}. \tag{14}$$

If the f_i are linear in x, as they are in non-metric regression and discriminant analysis, then (14) is a linear programming problem, with a large, but finite, number of linear constraints. There are $n!$ sign matrices, and thus equally many linear constraints. It should be possible to cleverly add constraints one at a time, and update the solution. This gives an iterative procedure which converges to the global maximum of ϕ after a finite number of steps. Developing this algorithm will have to wait for another publication.

4.2 Majorization

Alternatively, we can apply majorization starting from an initial estimate, and improving the fit iteratively. Optimizing functions that depend on the absolute value function can have run into difficulties, because of lack of differentiability at the orign. It makes sense to replace the absolute values by an approximation which can be made arbitrary close, and which remains smooth. The most common choice is $|x| \approx \sqrt{x^2 + \epsilon}$, which has been shown to have some optimal properties (Ramirez et al. (2014)). More precise smooth approximations have been studied (Voronin, Ozkaya, and Yoshida (2015), Bagul (2017)) but they are computationally more complicated to deal with.

Thus, instead of maximizing ϕ, we maximize ϕ_ϵ, which is

$$\phi_\epsilon(x) \triangleq \frac{\alpha(x)}{\beta_\epsilon(x)},$$

where

$$\beta_\epsilon(x) \triangleq \sum_{i=1}^n \sum_{j=1}^n w_{ij} \sqrt{((f_i - f_j)'x)^2 + \epsilon}$$

Note that $\beta_\epsilon(x) > \beta(x)$, and thus $\phi(x) < \phi_\epsilon(x)$. Remember we always choose x such that $\alpha(x) \geq 0$.

By the AM-GM inequality we have

$$\sqrt{((f_i - f_j)'x)^2 + \epsilon} \leq \frac{1}{2} \frac{1}{\sqrt{((f_i - f_j)'y)^2 + \epsilon}} \left\{((f_i - f_j)'x)^2 + ((f_i - f_j)'y)^2 + 2\epsilon\right\}.$$

for all x and y. Consequently

$$\beta_\epsilon(x) \leq \gamma_\epsilon(x,y) \stackrel{\Delta}{=} \frac{1}{2}x'B_\epsilon(y)x + \frac{1}{2}y'B_\epsilon(y)y + \epsilon\sum\sum w_{ij}, \tag{15}$$

where

$$B_\epsilon(y) \stackrel{\Delta}{=} \sum_{i=1}^n\sum_{j=1}^n \frac{w_{ij}}{\sqrt{((f_i - f_j)'y)^2 + \epsilon}}(f_i - f_j)(f_i - f_j)' \tag{16}$$

Note that $B_\epsilon(y)$ is positive semi-definite for all y. Define $x^{(k+1)}$ by maximizing $\delta_\epsilon(x, x^{(k)})$, where

$$\delta_\epsilon(x,y) \stackrel{\Delta}{=} \frac{\alpha(x)}{\gamma_\epsilon(x,y)}.$$

Now

$$\phi_\epsilon(x^{(k+1)}) \geq \delta_\epsilon(x^{(k+1)}, x^{(k)}) \geq \delta_\epsilon(x^{(k)}, x^{(k)}) = \phi_\epsilon(x^{(k)}),$$

which means we have a proper majorization algorithm. We maximize $\delta_\epsilon(x, x^{(k)})$ by setting the derivative equal to zero. This gives

$$x^{(k+1)} = \lambda B_\epsilon(x^{(k)})^{-1}F'r$$

for some appropriate λ. Note that $\delta_\epsilon(x, x^{(k)})$ is not invariant under multiplication of x by a constant, unlike $\phi(x)$, so we have to find the optimal λ. We must maximize

$$\delta_\epsilon(x^{(k+1)}, x^{(k)}) = \frac{\lambda r'FB_\epsilon(x^{(k)})^{-1}F'r}{\frac{1}{2}\lambda^2 r'FB_\epsilon(x^{(k)})^{-1}F'r + \frac{1}{2}x^{(k)'}B_\epsilon(x^{(k)})x^{(k)} + \epsilon\sum\sum w_{ij}}$$

from which

$$\lambda = \sqrt{\frac{x^{(k)'}B_\epsilon(x^{(k)})x^{(k)} + 2\epsilon\sum\sum w_{ij}}{r'FB_\epsilon(x^{(k)})^{-1}F'r}}$$

4.3 Initial Estimate

We can find an approximate solution, or an initial point for an iterative maximization process, by maximizing

$$\psi(x) = \frac{\sum_{i=1}^n\sum_{j=1}^n w_{ij}\sigma_{ij}(f_i - f_j)'x}{\sqrt{\sum_{i=1}^n\sum_{j=1}^n w_{ij}((f_i - f_j)'x)^2}}$$

Define

$$V \stackrel{\Delta}{=} \sum_{i=1}^n\sum_{j=1}^n w_{ij}(f_i - f_j)(f_i - f_j)'.$$

Then the optimal x is just $x = V^{-1}F'\rho$.

4.4 Example: Neumann

Our first example are the data from Neumann, used previously to illustrate nonmetric linear regression by Gifi (1990), De Leeuw and Mair (2009), and De Leeuw (2017a).

```
data(neumann, package="Gifi")
```

Our first analysis uses the complete sign matrix, with primary approach to ties.

```
g <- neumann$density
f <- cbind(neumann$temperature,neumann$pressure)
ss <- sign(outer(g, g, "-"))
```

The results of the `linpom()` function, for different values of ϵ, are given below. In each case we iterate until the increase of the fit function (phi_e in the table below) is less than 10^{-10}.

```
## eps: 1.0e-01 itel: 14 phi_e: 0.881518 phi: 0.992111 x: -0.023688 0.002955

## eps: 1.0e-02 itel: 21 phi_e: 0.969609 phi: 0.992127 x: -0.020392 0.002536

## eps: 1.0e-03 itel: 29 phi_e: 0.989094 phi: 0.992156 x: -0.020120 0.002487

## eps: 1.0e-04 itel: 25 phi_e: 0.991780 phi: 0.992168 x: -0.020103 0.002473

## eps: 1.0e-05 itel: 20 phi_e: 0.992119 phi: 0.992168 x: -0.020104 0.002471

## eps: 1.0e-06 itel: 17 phi_e: 0.992162 phi: 0.992169 x: -0.020108 0.002472
```

For this example, in which there are only a few ties, we do not expect the secondary approach to be that different from the primary one.

```
ss[ss==0] <- 1
diag(ss) <- 0
```

And indeed, the results are virtually the same, although the fit function is slightly different.

```
## eps: 1.0e-06 itel: 17 phi_e: 0.990859 phi: 0.990866 x: -0.020101 0.002472
```

We can also eliminate redundant inequalities and only have $\sigma_{ij} = +1$ when ζ_j is the largest of all data elements strictly less than ζ_i, i.e. the next element in the rank order. If there are ties we have $\sigma_{ij} = +1$ for more than one element, and withn tie blocks we have $\sigma_{ij} = 0$.

```
n <- length (g)
st<-matrix(0, n, n)
for (i in 1:n) {
  ind <- which (g < g[i])
  if (length(ind) == 0) next
  mnd <- max(g[ind])
  jnd <- which(g[ind] == mnd)
  st[i, ind[jnd]] <- 1
}
```

The `linpom()` analysis gives results on a quite different scale.

```
dim(st)
```

```
## [1] 65 65
```

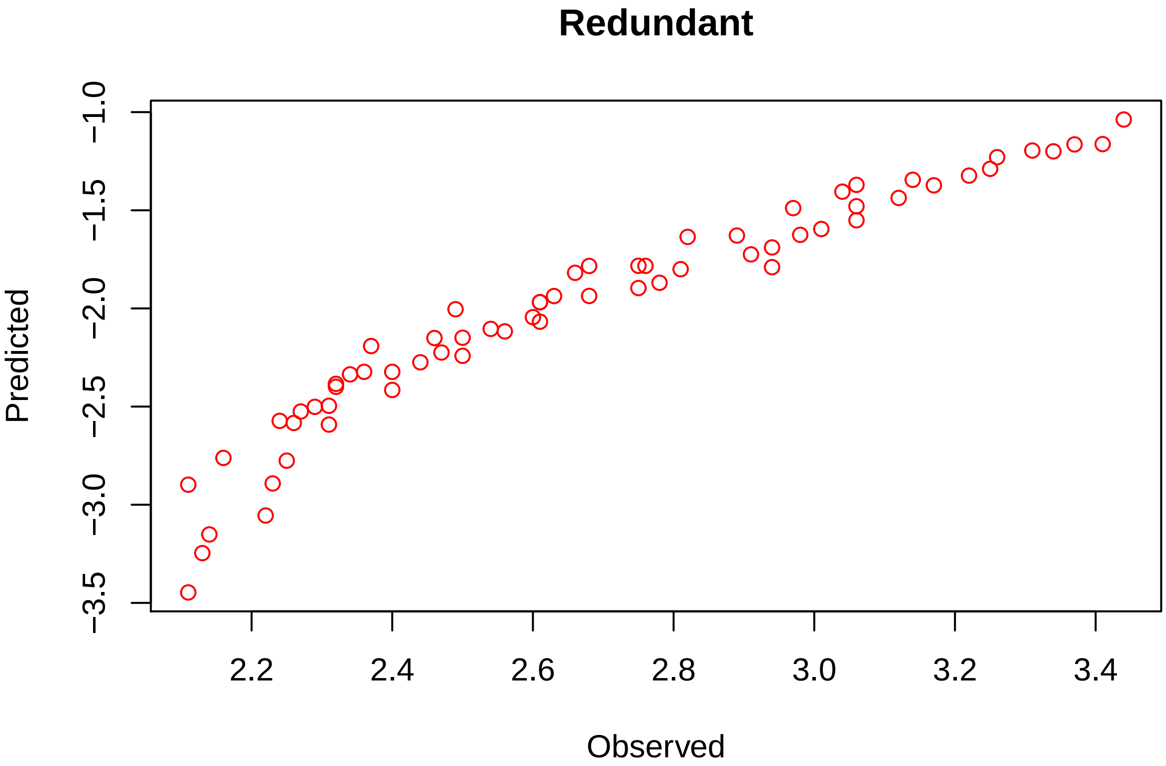
```
dim(f)
```

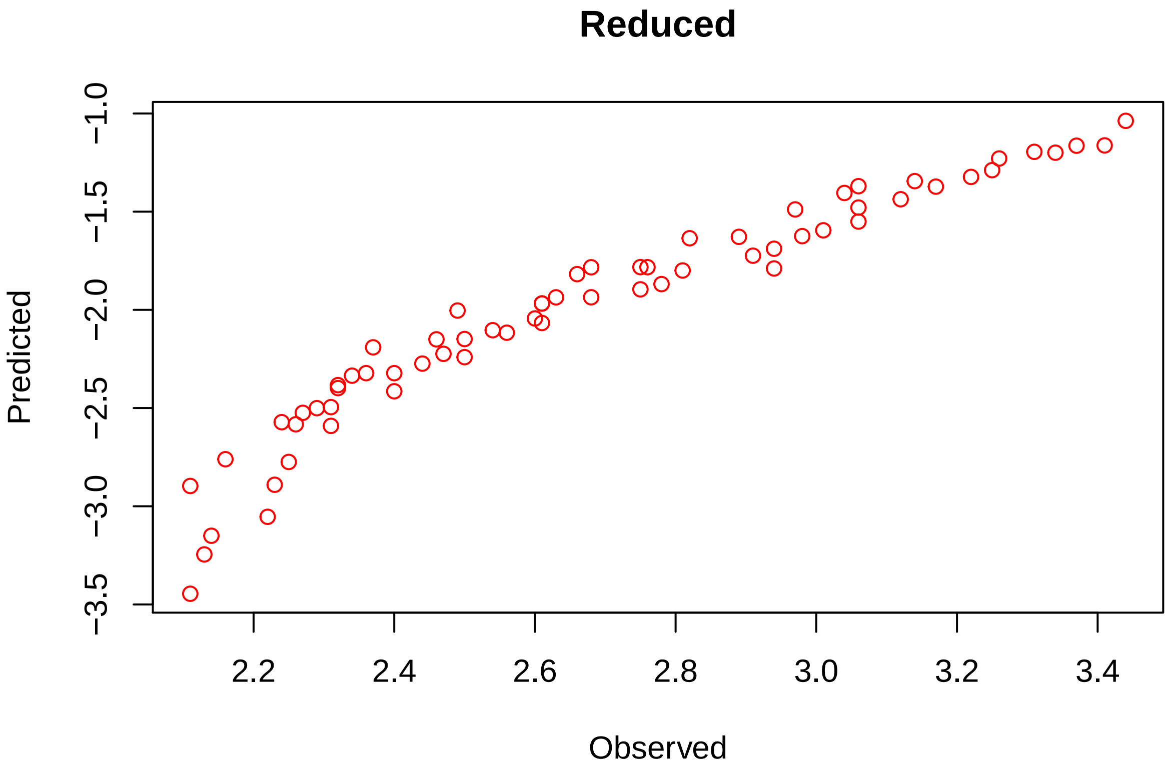
```
## [1] 65  2
```

```
hnr <- hns
#hnr <- linpom (f, st, aeps = 1e-6, xold = NULL, eps = 1e-10, itmax = 100, verbose = F
```

```
## eps: 1.0e-06 itel: 17 phi_e: 0.990859 phi: 0.990866 x: -0.020101 0.002472
```

Although the coefficients and fit function values are very different from the ones we had before, the plots of "observed" vs "predicted" are very similar.





For the "redundant" soution the correlation between observed and predicted is 0.9573195841, and Kendall's tau is 0.957319564. For the "reduced" solution this is 0.9348822832 and 0.9348822832.

5 Binary Regression

If the data are binary (success/failure, yes/no, alive/death, trump/clinton) then obviously treatment of ties becomes rather essential. It seems the primary approach is more natural. It tries to find a separating hyperplane, and thus it is similar to logistic regression, support vector machines, and discriminant analysis.

As an example we use the breast cancer data from Newman et al. (1998), included in the R package mlbench (Leisch and Dimitriadou (2010))

```
data(BreastCancer)
bccom <- BreastCancer[complete.cases(BreastCancer), ]
g<-as.numeric(bccom$Class)
g <- 2 * g - 3
h<-as.matrix(bccom[,c(-1,-11)])
f <- array(0, dim(h))
for (j in 1:9) f[,j] <- as.numeric(h[,j])
s <- sign(outer(g, g, "-"))
```

```
hbin <- linpom (f, s, aeps = 1e-6, eps = 1e-10, itmax = 1000, verbose = FALSE)
```

```
## eps: 1.0e-06 itel: 418 phi_e: 0.998821 phi: 0.998821 x: 0.041302 -0.002514 0.041981 0
```

The secondary approach is conceptually not very appealing. We require the predicted values for all negative instances to be equal, and similar for the predicted values of the positive instances.

```
s[s==0]<-1
diag(s)<-0
hus <- linpom (f, s, aeps = 1e-6, eps = 1e-10, itmax = 1000, verbose = FALSE)
```

```
## eps: 1.0e-06 itel: 82 phi_e: 0.839712 phi: 0.839754 x: 0.004107 0.044646 0.014937 0.0
```

Both the primary and secondary approach have the disadvantage that the matrix $\Sigma = \{\sigma_{ij}\}$ is quite large. To make computations more efficient, and more like the usual computations in logistic regression and the like, we define our positive orthant inequalities as

$$\sigma_i f_i'x \geq 0,$$

where $\sigma_i = +1$ if i is is positive response and $\sigma_i = -1$ if the instance is negative. We also add a column with ones to F to estimate a cutoff value. The fit function becomes

$$\phi(x) = \frac{\sum_{i=1}^n w_i \sigma_i f_i'x}{\sum_{i=1}^n w_i |f_i'x|},$$

where we assume without any real loss of generality that both all w_i are positive and all σ_i are non-zero. We can now apply exactly the same results as before to find a majorization algorithm, which we have implemented in the R function `binpom()`.

The inequality (15) is still valid, but instead of (16) we now have

$$B_\epsilon(y) \triangleq \sum_{i=1}^n \frac{w_i}{\sqrt{(f_i'y)^2 + \epsilon}} f_i\, f_i'. \tag{17}$$

```
hbin <- binpom (f, g, aeps = 1e-6, eps = 1e-10, itmax = 500, verbose = FALSE)
```

```
## eps: 1.0e-06 itel: 111 phi_e: 0.984996 phi: 0.984999 x: -4.960047 0.244466 -0.077994
```

6 Paired Comparisons

In a typical paired comparison experiment each of m subjects generates a matrix Σ_k with elements -1 and $+1$. If we have units weights, the vector ρ_k has elements

$$\rho_{ik} = \sum_{j=1}^n (\sigma_{ijk} - \sigma_{jik}).$$

The results of paired comparison experiments are often presented in aggregated form, summed over subjects or occasions. An an example, consider the vegetable data of Guilford, from the psych package (Revelle (2018)).

```r
library(psych)
data(vegetables)
veg
```

```
##          Turn   Cab  Beet   Asp   Car  Spin S.Beans  Peas  Corn
## Turn    0.500 0.818 0.770 0.811 0.878 0.892   0.899 0.892 0.926
## Cab     0.182 0.500 0.601 0.723 0.743 0.736   0.811 0.845 0.858
## Beet    0.230 0.399 0.500 0.561 0.736 0.676   0.845 0.797 0.818
## Asp     0.189 0.277 0.439 0.500 0.561 0.588   0.676 0.601 0.730
## Car     0.122 0.257 0.264 0.439 0.500 0.493   0.574 0.709 0.764
## Spin    0.108 0.264 0.324 0.412 0.507 0.500   0.628 0.682 0.628
## S.Beans 0.101 0.189 0.155 0.324 0.426 0.372   0.500 0.527 0.642
## Peas    0.108 0.155 0.203 0.399 0.291 0.318   0.473 0.500 0.628
## Corn    0.074 0.142 0.182 0.270 0.236 0.372   0.358 0.372 0.500
```

In the aggregated table N cell n_{ij} indicates the number of individuals who prefer i to j, i.e. the number or proportion of individuals for which $\sigma_{ijk} = +1$. Thus

$$\sum_{k=1}^{m} \sigma_{ijk} = n_{ij} - n_{ji},$$

and

$$\sum_{k=1}^{m} \rho_{ik} = 2\sum_{j=1}^{n}(n_{ij} - n_{ji}).$$

```r
s <- as.matrix(veg)
s <- s - t(s)
hpc<-pcpom(s, aeps = 1e-6, eps = 1e-10, itmax = 1000, verbose = FALSE)
```

```
## eps: 1.0e-06 itel: 78 phi_e: 0.721500 phi: 0.721500 x: 1.000000 -0.125000 -0.125000 -
```

In this case we find a scale with a single positive element, corresponding to the largest element of ρ, while the other $n-1$ elements are all negative and equal. For such solutions the fit is $\max(\rho)/(n-1)$, or in our case 0.7215. Clearly such a solution is disappointing, because we do not need elaborate majorization iterations to invariably come up with the largest element in a vector.

7 Multidimensional Scaling

Multidimensional scaling of squared distances means $f_i(x) = x'A_ix$. Thus globally maximizing ϕ means maximizing a non-convex quadratic form over a region defined by $n!$ non-convex quadratic inequality constraints. This seems a pretty hopeless task. A mental picture of the problem also suggests how serious the local maximum problem is bound to be.

Scaling of distances has $f_i(x) = \sqrt{x'A_ix}$, which is even more complicated than the squared distance case. So the best we can do is start an infinite iterative procedure with an initial solution that is as close as possible, using some form of classical scaling, and then improve the solution with gradient or majorization steps until we find a local maximum. Again, this is a topic for further research.

8 History

The history of positive orthant type loss functions is actually more complicated than our previous discussion suggests. I review that history here, which will give me the opportunity to reconstruct my thoughts on the matter of fifty years ago

The idea of measuring fit by looking at differences between model values was inspired by early work of Kendall (1938) and Daniels (1944). Just to refresh your memory: suppose x and y are two vectors of length n and we want to measure some form of correlation between the two. Here is a nice classical quotation.

"Let us assign to each pair (x_i, x_j) what for convenience will be called a score a_{ij} and to each $(y_i, y_j$ a score b_{ij}, where

$$a_{ij} = -a_{ji}, \quad b_{ij} = -b_{ji}.$$

Denote by Γ the number

$$\Gamma = \frac{\sum a_{ij}b_{ij}}{\sqrt{\sum a_{ij}^2 \sum b_{ij}^2}},$$

the summation extending over all i and j from 1 to n." (Daniels, 1944, p. 129).

Daniels goes on to point out that $a_{ij} = x_i - x_j$ gives the Pearson correlation, $a_{ij} = \mathbf{sign}(x_i - x_j)$ gives Kendall's tau, and $a_{ij} = \mathbf{rank}(x)_i - (\mathbf{rank}(x)_j$ gives Spearman's rho.

I am quite sure that this idea inspired my early work. In De Leeuw (1968a) fit measures for a model $F = \{f_i\}$\$ are introduced in the form

$$\sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}(f_i(x) - f_j(x))$$

where the anti-symmetric weights σ_{ij} are derived from the data ζ and are either Kendall, Spearman, or Pearson weights.

There was another very important influence, which we will discuss next. Guttman's work on scale construction, while he was working for the army during WW II, still looms over much of psychometrics. It was elaborated on and tinkered with, but it was never improved, because it was perfect. For our purposes the key publication is Guttman (1946), which curiously enough appeared the Annals of Mathematical Statistics.

"The new criterion adopted is that the numerical values be determined so as best to distinguish between those things judged higher and thoise judged lower *for each individual*." (Guttman (1946), p 144)

Guttman's notation for paired comparisons is somewhat different from ours. He codes paired comparisons using zero and one, instead of plus and minus one. Thus for subject or replication

k define

$$\epsilon_{ijk} \overset{\Delta}{=} \frac{1}{2}(\sigma_{ijk}+1).$$

Now suppose x is the scale we are looking for, in deviations of the mean. The average value of objects that subject k judges higher or larger or better than something else is

$$t_k = \frac{\sum_{j=1}^n \sum_{i=1}^n \epsilon_{ijk}x_j}{\sum_{j=1}^n \sum_{i=1}^n \epsilon_{ijk}},$$

and the average of objects judged smaller is

$$u_k = \frac{\sum_{j=1}^n \sum_{i=1}^n \epsilon_{jik}x_j}{\sum_{j=1}^n \sum_{i=1}^n \epsilon_{jik}}.$$

We now want to maximize

$$\sum_{k=1}^m (t_k-u_k)^2$$

over x normalized by $x'x=1$. In a complete paired comparison experiment in which $\epsilon_{ijk}+\epsilon_{jik}=1$ for all $i \neq j$ we have $t_k=-u_k$. We also have

$$t_k = \frac{\frac{1}{2}\sum_{j=1}^n x_j \sum_{i=1}^n \sigma_{ijk}}{\frac{1}{2}n(n-1)} = \frac{2\ x'\rho_k}{n(n-1)},$$

where the ρ_k are the ranks corresponding to the σ_{ijk}. Thus we can maximize

$$\sum_{k=1}^m t_k^2 = x'\left\{\sum_{k=1}^n \rho_k\rho_k'\right\}x$$

over $x'x=1$, which mean we find the largest eigenvalue and corresponding eigenvector of the cross-product matrix of the centered rank numbers. This means that Guttman's signature based method is equivalent to the method of Slater (1960) or Carroll and Chang (1964), which would later become to be identified with the MDPREF program (Carroll and Chang (1968)).

We start with De Leeuw (1968b) on nonmetric discriminant analysis. This proposes the loss function

$$\tau(x) \overset{\Delta}{=} \frac{\sum_{i=1}^n (|t_i|-t_i)^2}{4\sum_{i=1}^n t_i^2},$$

where $t_i=\sigma_i f_i'x$. The partial derivatives are given, and a gradient method is suggested.

The actual algorithm in the report, implemented in PL/I for the IBM 360/50, minimizes

$$\kappa(x,y) \overset{\Delta}{=} (Fx-y)'(Fx-y)$$

over x and $y \geq 0$ by alternating least squares (in fact, this is possibly the first use of that term). Of course

$$\min_{y\geq 0} \kappa(x,y) = \frac{1}{4}\sum_{i=1}^n (|t_i|-t_i)^2,$$

which is the numerator of τ. Although this is not mentioned in the report, this result also shows that τ is a differentiable function of f, despite the absolute value signs. After each update of x it is normalized by $x'F'Fx = 1$. The ALS process converges to a local minimum of both κ and τ.

A note in the back of the report says: "Although there is no program for NDA in the G-L-series as yet, professor Louis Guttman informs me that he would favor the absolute value principle. From the (up to now) rather sketchy accounts of this principle I gather it is quite similar to my positive orthant method."

De Leeuw (1968c) defines the positive orthant method, in the context of nonmetric multidimensional scaling, as the minimization of one of a family of loss functions . As in (1) we define

$$t_{ij}(x) \overset{\Delta}{=} \sigma_{ij}(f_i(x) - f_j(x)),$$

and then, for some $q > 0$,

$$\tau_q(x) \overset{\Delta}{=} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(|t_{ij}(x)| - t_{ij}(x))^q}{2^q \sum_{i=1}^n \sum_{j=1}^n w_{ij}|t_{ij}(x)|^q}. \tag{18}$$

The matrix $W = \{w_{ij}\}$ has non-negative weights. We assume, without loss of generality, that $w_{ij}|\sigma_{ij}| = w_{ij}$, i.e. if $\sigma_{ij} = 0$ then also $w_{ij} = 0$. Clearly $0 \leq \tau_q(x) \leq 1$ with $\tau_q(x) = 0$ if and only if $t_{ij}(x) \geq 0$ for all i and j.

Using $w_{ij}|\sigma_{ij}| = w_{ij}$ and $\mathbf{sign}(\sigma_{ij}) = \sigma_{ij}$ we can show that $\tau_q(x) = \frac{1}{2}(1 - \phi_q(x))$, where

$$\phi_q(x) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}\sigma_{ij}(f_i(x) - f_j(x))|f_i(x) - f_j(x)|^{q-1}}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}|f_i(x) - f_j(x)|^q}.$$

Thus minimizing τ_q is equivalent to maximizing ϕ_q. See also Young (1975) or Hartmann (1979) for additional accounts of this method, and some historical context.

De Leeuw (1968c) and De Leeuw (1970) discussed the general case $q > 0$ and the limiting cases with $q \rightarrow 0$ and $q \rightarrow +\infty$, but the report is computationally rather naieve, because it just suggest to apply the gradient method and it ignores the fact that the absolute value function is not differentiable. The report discusses a

From a computational point of view the most interesting members of the family (18) are τ_1 and τ_2. The case $q = 2$ has been discussed in Guttman (1969), as the "absolute value principle", and in Johnson (1973) as the "pairwise method", again just suggesting to use gradient methods.

$$\phi_2(x) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}\sigma_{ij}\mathbf{sign}(f_i(x) - f_j(x))(f_i(x) - f_j(x))^2}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(f_i(x) - f_j(x))^2}.$$

$$\phi_1(x) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}\sigma_{ij}(f_i(x) - f_j(x))}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}|f_i(x) - f_j(x)|}.$$

The positive orthant method is introduced in De Leeuw (1968c). Chapter 3 in that report introduces the loss function

$$\tau_q(x)\triangleq\frac{\sum_{i=1}^n w_i(|t_i|-t_i)^q}{2^q\sum_{i=1}^n w_i|t_i|^q}$$

The PL/I program NMSPOM for minimizing τ_q in the case of multidimensional scaling, using various gradient methods, is decribed, and some examples are given.

The next report in the sequence is De Leeuw (1969). In our notation it suggests maximizing

$$\psi(x)=\frac{\sum_{i=1}^n\sum_{j=1}^n\sigma_{ij}(f_i'x-f_j'x)}{\sqrt{\sum_{i=1}^n(f_i'x)^2}}$$

9 Code

9.1 linpom.R

```
linpom <- function(f,
                   s,
                   w = array (1, dim(s)),
                   xold = NULL,
                   aeps = 1e-06,
                   itmax = 100,
                   eps = 1e-06,
                   verbose = TRUE)
{
  w[s == 0] <- 0
  r <- rowSums(w * s - t(w * s))
  wsum <- sum(w)
  v <- -(w + t(w))
  diag(v) <- -rowSums(v)
  g <- crossprod(f, v %*% f)
  u <- colSums(r * f)
  if (is.null(xold)) {
    xold <- solve(g, u)
  }
  yold <- drop(f %*% xold)
  nold <- sqrt((outer(yold, yold, "-") ^ 2) + aeps)
  pold <- sum(xold * u) / sum(w * nold)
  itel <- 1
  repeat {
    v <- w / nold
    v <- -(v + t(v))
    diag(v) <- -rowSums(v)
    h <- crossprod(f, v %*% f)
    m <- sum (xold * (h %*% xold))
    xnew <- solve (h, u)
    lbd <- sqrt ((m + (2 * aeps * wsum))/ sum (u * xnew))
    xnew <- lbd * xnew
    ynew <- drop(f %*% xnew)
    nnew <- sqrt((outer(ynew, ynew, "-") ^ 2) + aeps)
    pnew <- sum(xnew * u) / sum(w * nnew)
    if (verbose)
      cat(
        "Iteration: ",
        formatC(itel, width = 3, format = "d"),
        "pold: ",
        formatC(
          pold,
          digits = 8,
          width = 12,
          format = "f"
        ),
        "pnew: ",
        formatC(
          pnew,
          digits = 8,
          width = 12,
          format = "f"
        ),
        "\n"
      )
    if (((pnew - pold) < eps) || (itel == itmax))
      break
    nold <- nnew
    pold <- pnew
    itel <- itel + 1
  }
  fend <- sum (u * xnew) / sum (w * abs (outer (ynew, ynew, "-")))
  return(list(x = xnew, p = pnew, f = fend, itel = itel))
}
```

9.2 binpom.R

```
binpom <- function(f,
                   s,
                   w = rep (1, length(s)),
                   xold = NULL,
                   aeps = 1e-06,
                   itmax = 100,
                   eps = 1e-06,
                   verbose = TRUE)
{
  f <- cbind(1, f)
  r <- crossprod (f, w * s)
  wsum <- sum(w)
  g <- crossprod(f, w * f)
  if (is.null(xold)) {
    xold <- solve (g, r)
  }
  yold <- drop(f %*% xold)
  nold <- sqrt(yold ^ 2 + aeps)
  pold <- sum(xold * r) / sum(w * nold)
  itel <- 1
  repeat {
    v <- w / nold
    h <- crossprod(f, v * f)
    m <- sum (xold * (h %*% xold))
    xnew <- solve (h, r)
    lbd <- sqrt ((m + (2 * aeps * wsum)) / sum (r * xnew))
    xnew <- lbd * xnew
    ynew <- drop(f %*% xnew)
    nnew <- sqrt(ynew ^ 2 + aeps)
    pnew <- sum(xnew * r) / sum(w * nnew)
    if (verbose)
      cat(
        "Iteration: ",
        formatC(itel, width = 3, format = "d"),
        "pold: ",
        formatC(
          pold,
          digits = 8,
          width = 12,
          format = "f"
        ),
        "pnew: ",
        formatC(
          pnew,
          digits = 8,
          width = 12,
          format = "f"
        ),
        "\n"
      )
    if (((pnew - pold) < eps) || (itel == itmax))
      break
    nold <- nnew
    pold <- pnew
    itel <- itel + 1
  }
  fend <- sum (r * xnew) / sum (w * abs (ynew))
  return(list(
    x = xnew,
    p = pnew,
    f = fend,
    itel = itel
  ))
}
```

9.3 pcpom.R

```
pcpom <- function(s,
                  w = array (1, dim(s)),
                  xold = NULL,
                  aeps = 1e-06,
                  itmax = 100,
                  eps = 1e-06,
                  verbose = TRUE)
{
  w[s == 0] <- 0
  r <- rowSums(w * s - t(w * s))
  wsum <- sum(w)
  r <- r - mean (r)
  n <- length(r)
  wsum <- sum(w)
  if (is.null(xold)) {
    xold <- r
  }
  nold <- sqrt((outer(xold, xold, "-") ^ 2) + aeps)
  pold <- sum(xold * r) / sum(w * nold)
  itel <- 1
  repeat {
    v <- w / nold
    v <- -(v + t(v))
    diag(v) <- -rowSums(v)
    m <- sum (xold * (v %*% xold))
    xnew <- solve(v + 1 / n, r)
    lbd <- sqrt ((m + (2 * aeps * wsum)) / sum (r * xnew))
    xnew <- lbd * xnew
    nnew <- sqrt((outer(xnew, xnew, "-") ^ 2) + aeps)
    pnew <- sum(xnew * r) / sum(w * nnew)
    if (verbose)
      cat(
        "Iteration: ",
        formatC(itel, width = 3, format = "d"),
        "pold: ",
        formatC(
          pold,
          digits = 8,
          width = 12,
          format = "f"
        ),
        "pnew: ",
        formatC(
          pnew,
          digits = 8,
          width = 12,
          format = "f"
        ),
        "\n"
      )
    if (((pnew - pold) < eps) || (itel == itmax))
      break
    nold <- nnew
    pold <- pnew
    itel <- itel + 1
  }
  fend <- sum (r * xnew) / sum (w * abs (outer (xnew, xnew, "-")))
  return(list(
    x = xnew,
    p = pnew,
    f = fend,
    itel = itel
  ))
}
```

References

Bagul, Y.J. 2017. "A Smooth Transcendental Approximation to |x|." *International Journal of Mathematical Sciences and Engineering Applications* 11: 213–17.

Carroll, J. D., and J.J. Chang. 1964. "Non-parametric Multidimensional Analysis of Paired Comparison Data."

———. 1968. "How to Use MDPREF, a Computer Program for the Multidimensional Analysis of Preference Data." Murray Hill, N.J.: Bell Telephone Laboratories.

Daniels, H.E. 1944. "The Relation Between Measures of Correlation in the Universe of Sample Permutations." *Biometrika* 33 (2): 129–35.

De Leeuw, J. 1968a. "Canonical Discriminant Analysis of Relational Data." Research Note 007-68. Department of Data Theory FSW/RUL. http://deleeuwpdx.net/janspubs/1968/reports/deleeuw_R_68e.pdf.

———. 1968b. "Nonmetric Discriminant Analysis." Research Note 06-68. Department of Data Theory, University of Leiden. http://deleeuwpdx.net/janspubs/1968/reports/deleeuw_R_68d.pdf.

———. 1968c. "Nonmetric Multidimensional Scaling." Research Note 010-68. Department of Data Theory FSW/RUL. http://deleeuwpdx.net/janspubs/1968/reports/deleeuw_R_68g.pdf.

———. 1969. "The Linear Nonmetric Model." Research Note 003-69. Department of Data Theory FSW/RUL. http://deleeuwpdx.net/janspubs/1969/reports/deleeuw_R_69c.pdf.

———. 1970. "The Positive Orthant Method for Nonmetric Multidimensional Scaling." Research Report 001-70. Leiden, The Netherlands: Department of Data Theory FSW/RUL. http://deleeuwpdx.net/janspubs/1970/reports/deleeuw_R_70a.pdf.

———. 1973. "Nonmetric Scaling with Rearrangement Methods." Technical Memorandum. Murray Hill, N.J.: Bell Telephone Laboratories.

———. 2016. "Exceedingly Simple Permutations and Combinations." http://deleeuwpdx.net/pubfolders/nextPC/nextPC.pdf.

———. 2017a. "Computing and Fitting Monotone Splines." http://deleeuwpdx.net/pubfolders/splines/splines.pdf.

———. 2017b. "Shepard Non-metric Multidimensional Scaling." http://deleeuwpdx.net/pubfolders/shepard/shepard.pdf.

De Leeuw, J., and P. Mair. 2009. "Homogeneity Analysis in R: The Package Homals." *Journal of Statistical Software* 31 (4): 1–21. http://deleeuwpdx.net/janspubs/2009/articles/deleeuw_mair_A_09a.pdf.

Gifi, A. 1990. *Nonlinear Multivariate Analysis*. New York, N.Y.: Wiley.

Guttman, L. 1946. "An Approach for Quantifying Paired Comparisons and Rank Order."

Annals of Mathematical Statistics 17: 144–63.

———. 1969. “Smallest Space Analysis by the Absolute Value Principle.” In *Proceedings of the XIX International Congress of Psychology, London.*

Hardy, G.H., J.E. Littlewood, and G. Polya. 1952. *Inequalities.* Second Edition. Cambridge University Press.

Hartmann, W. 1979. *Geometrische Modelle zur Analyse empirischer Data.* Akademie Verlag.

Johnson, R.M. 1973. “Pairwise Nonmetrc Multidimensional Scaling.” *Psychometrika* 38 (12–18).

Kendall, M.G. 1938. “A New Measure of Rank Correlation.” *Biometrika* 30 (1-2): 81–93.

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

Leisch, F., and E. Dimitriadou. 2010. *Mlbench: Machine Learning Benchmark Problems.*

Newman, D.J., S. Hettich, C.L. Blake, and C.J. Merz. 1998. “UCI Repository of Machine Learning Databases.” University of California, Irvine, Dept. of Information; Computer Sciences. http://www.ics.uci.edu/~mlearn/MLRepository.html.

Ramirez, C., R. Sanchez, V. Kreinovich, and M. Argaez. 2014. “$\sqrt{x^2+\mu}$ is the Most Computationally Efficient Smooth Approximation to |x|.” *Journal of Uncertain Systems* 8: 205–10.

Revelle, W. 2018. *psych: Procedures for Psychological, Psychometric, and Personality Research.* Evanston, Illinois: Northwestern University. {https://CRAN.R-project.org/package=psych}.

Slater, P. 1960. “The Analysis of Personal Preferences.” *British Journal of Mathematical and Statistical Psychology* 13 (2): 119–35.

Voronin, S., G. Ozkaya, and Y. Yoshida. 2015. “Convolution Based Smooth Approximations to the Absolute Value Function with Application to Non-smooth Regularization.” https://arxiv.org/abs/1408.6795.

Yitzhaki, S. 2003. “‘Gini’s Mean Difference: A Superior Measure of Variability for Non-Normal Distributions’.” *Metron* 61 (2): 285–316.

Yitzhaki, S, and E. Schechtman. 2013. *The Gini Methodology: A Primer on Statistical Methodology.* Springer Series in Statistics. Springer Verlag.

Young, F.W. 1975. “Methods for Describing Ordinal Data with Cardinal Models.” *Journal of Mathematical Psychology* 12: 416–36.