HOMOGENEITY ANALYSIS WITH RATIOS OF ALTERNATIVE NORMS

JAN DE LEEUW AND GEORGE MICHAILIDES

Abstract.

1. Homogeneity Analysis

1.1. **One-dimensional.** In (one-dimensional) *homogeneity analysis* (also known as *multiple correspondence analysis*, and under many other names) we minimize a function of the form

$$\sigma(x,y)\stackrel{\Delta}{=}\frac{1}{m}\sum_{j=1}^m(x-G_jy_j)'(x-G_jy_j),$$

where the G_j are $n\times k_j$ *indicator matrices* (or *dummies*) coding the n *observations* on *variable* j into k_j *categories*. In order to prevent trivial solutions, we require $x'x=1$. For an extensive discussion of the technique we refer to Gifi [1990] and Michailidis and Leeuw [1999], alhough the basics are already discussed in detail in Guttman [1941].

As a first step towards the solution, define, using the Moore-Penrose inverse,

$$P_j\stackrel{\Delta}{=}G_jG_j^-=G_j(G_j'G_j)^-G_j',$$

and

$$P\stackrel{\Delta}{=}\frac{1}{m}\sum_{j=1}^m P_j.$$

Then it is easy to see that minimizing $\sigma(x,y)$ over x and y can be done by maximizing the Rayleigh quotient

$$\lambda_P(x)\stackrel{\Delta}{=}\frac{x'Px}{x'x}$$

over x, and then setting $y_j=G_j^-x$.

Date: August 3, 2003.

2000 *Mathematics Subject Classification.* 62H25.

Key words and phrases. Multivariate Analysis, Correspondence Analysis, Principal Component Analysis, Binary Data.

It is trivial that we could also define $Q_j = I - P_j$ and Q as the average of the Q_j, and then minimize

$$\lambda_Q(x) \triangleq \frac{x'Qx}{x'x}.$$

Nevertheless it is important to state the equivalence of maximizing λ_P and minimizing λ_Q here, because we will introduce generalizations in which the two problems are different.

For interpretational purposes, we note that P_j is the orthogonal projector on the column space of G_j, and Q_j is the orthogonal projector on its null space. Thus $P_j x$ replaces elements of x by their category means on variable j and $Q_j x$ replaces x by deviations from those category means. The between-category variance is $x'Px$ and the within-category variance is $x'Qx$.

We should also note that

$$\lambda_P(x) = \frac{1}{m}\sum_{j=1}^m \frac{x'P_j x}{x'x},$$

$$\lambda_Q(x) = \frac{1}{m}\sum_{j=1}^m \frac{x'Q_j x}{x'x},$$

another trivial equivalence which will get lost in our later examples.

Although homogeneity analysis is a very useful technique, as many examples in many areas show, there are some disadvantages which can complicate the interpretations. In the one-dimensional context, the main one is that the technique is very sensitive to outliers.

1.2. **Multidimensional.** In the multidimensional case we look for more than one solution. There are three obvious ways to proceed.

- The first one is to observe that the stationary equations for homogeneity analysis $Px = \lambda x$ have more than one solution. Each eigenvalue of P defines a solution, and each solution can be used to define an additional dimension.
- We could also proceed by optimizing $\lambda_P(x)$ over all x, adding the requirement that x must be orthogonal to all the previous solutions. This is the *successive* strategy.
- Finally, there is the *simultaneous* strategy. We maximize $\mathbf{tr}\ X'PX$ over $X'X = I$, or equivalently we maximize $\mathbf{tr}\ (X'X)^{-1}X'PX$. Thus we compute a number of solutions at the same time, replacing the solution vector x by a solution matrix X.

The nice property of homogeneity analysis is that all three strategies give the same multidimensional solution, the eigenvectors of P corresponding with the largest eigenvalues. Again, this is a property which we will not be able to generalize to our later techniques.

It seems perhaps more straightforward to define

$$\lambda_P(X) = \frac{\mathbf{tr}\ X'PX}{\mathbf{tr}\ X'X},$$

and then maximize this over X. Unfortunately this does not really give a multidimensional solution, because it simply sets each dimension equal to the eigenvector corresponding with the dominant eigenvalue. This does not happen for

$$\lambda_P(X) = \frac{\mathbf{det}(X'PX)}{\mathbf{det}(X'X)},$$

so it becomes interesting to look for other generalizations of these ratio measures of homogeneity that do not collapse the points into one-dimensional subspaces.

Multidimensional generalizations have another problematic aspect. In many cases it produces what is commonly known as "horseshoes", in which subsequent solutions are quadratic and cubic functions of the first solution, and consequently do not contribute independent information. See Schriever [1985] and Bekker and Leeuw [1988] for extensive discussions of horseshoes.

2. Reformulation

2.1. **One-dimensional.** For positive semi-definite A, define the semi-norm

$$\|x\|_A \triangleq \sqrt{x'Ax}.$$

Clearly

$$\lambda_P(x) = \frac{\|x\|_P^2}{\|x\|_I^2},$$

and we see that the one-dimensional homogeneity analysis problem can be defined as maximizing the ratio of two norms (more precisely, of a semi-norm and a norm). Of course

$$|||x|||_P = |||P^{1/2}x|||_I = |||Px|||_{P^-},$$

using the Moore-Penrose inverse P^-.

2.2. **Multidimensional.** It is more complicated to give a satisfactory reformulation of the multidimensional problem in these terms. We can, of course, define a family of matrix norms $\|X\|_A \triangleq \sqrt{\mathbf{tr}\ X'AX}$, and set

$$\lambda(X) \triangleq \frac{\|X\|_P}{\|X\|_I},$$

but unfortunately this does not lead to a useful technique. The stationary equations are $PX = \lambda X$, which shows that the columns of X will all be equal to the dominant eigenvector of P. This is why we had to require $X'X = I$ in the multidimensional version of homogeneity analysis, which means that we really deal with the multivariate Rayleigh quotient

$$\lambda(X) = \mathbf{tr}\ (X'X)^- X'PX.$$

This does not readily translate into maximizing a ratio of norms. Nevertheless we think the norm formulation is still interesting, because it can be readily generalized and may lead to approaches for dealing with the outlier and horseshoe problems.

3. Generalizations

3.1. **One-dimensional.** In this paper we will deal with the following generalization of one-dimensional homogeneity analysis. We want to maximize

$$\lambda_{P,\psi}(x) = \frac{\|x\|_P}{\psi(x)},$$

where ψ is any vector norm. In fact, we will also study the even more general problem in which we maximize

$$\lambda_{\phi,\psi}(x) = \frac{\phi(Px)}{\psi(x)},$$

or minimize

$$\lambda_{\phi,\psi}(x) = \frac{\phi(Qx)}{\psi(x)},$$

where ϕ and ψ are both arbitrary vector norms.

There is a further generalization possible: in HOMALS we really minimize

$$\frac{1}{m}\sum_{j=1}^m \min_{y_j} \|x - G_j y_j\|_2^2 = \frac{1}{m}\sum_{j=1}^m \|x - P_j x_j\|_2^2 = \|x\|_2^2 - \|x\|_P^2.$$

This suggests

$$\lambda_{\phi,\psi}(x) = \max_x \frac{\sum_{j=1}^m \phi(P_j x)}{m\psi(x)},$$

or we minimize the same criterion with Q_j instead of P_j. Observe that HOMALS is not even a special case of this generalization, because in HOMALS we add squared norms in the numerator.

Indeed, we can go even further than this. The operators P_j and Q_j are projections in the least squares norm. If we use general norms it is more consistent to define

$$\lambda_{\phi,\psi}(x) = \min_x \frac{\sum_{j=1}^m \min_{y_j} \phi(x - G_j y_j)}{m\psi(x)},$$

3.2. **Multidimensional.** There are two obvious ways to generalize these approaches to more than one dimension. The first one is a *successive* approach: after we have computed a solution we compute the next solution with the added requirement that it is orthogonal to the previous one. And then the third one with the requirements that it is orthogonal to the previous two, and so on. The second is a *simultaneous* approach: we reformulate our criteria using matrix norms, and optimize over all dimensions at the same time.

Thus in the simultaneous case we are interested in maximizing

$$\lambda_{P,\psi}(X) = \frac{\|X\|_P}{\psi(X)},$$

where ψ is any matrix norm, or in maximizing

$$\lambda_{\phi,\psi}(X) = \frac{\phi(PX)}{\psi(X)}.$$

Observe that we have already established that maximizing $\lambda_{P,\psi}$ with $\psi(X) = \|X\|_I$ does not work. Hopefully, we can find other matrix norms that are better behaved.

4. Robert's Algorithm

Robert's algorithm [Robert, 1967], also studied in more specific cases by Boyd [1974] and Tao [1976, 1975, 1984, 1985], maximizes the ratio of two norms

$$\lambda_{\phi,\psi}(x) = \frac{\phi(x)}{\psi(x)}.$$

It was first applied in data analysis by De Leeuw [1977], in the original derivation of the majorization method for multidimensional scaling.

The algorithm starts with an arbitrary x_1 and then generates two sequences x_k and y_k according to the rules

$$\begin{aligned} y_k &\in \partial\phi(x_k),\\ x_{k+1} &\in \partial\psi^\circ(y_k). \end{aligned}$$

Theorem 4.1.

Proof. Clearly $\phi^\circ(y_k) = \psi(x_k) = 1$ for all k. Also $\phi(x_k) = \langle x_k, y_k\rangle$ and $\psi^\circ(y_k) = \langle x_{k+1}, y_k\rangle$. Therefor

$$\begin{aligned} \phi(x_k) = \langle x_k, y_k\rangle &\leq \psi^\circ(y_k)\psi(x_k) = \psi^\circ(y_k),\\ \psi^\circ(y_k) = \langle x_{k+1}, y_k\rangle &\leq \phi(x_{k+1})\phi^\circ(y_k) = \phi(x_{k+1}), \end{aligned}$$

and

$$\phi(x_k) \leq \psi^\circ(y_k) \leq \phi(x_{k+1}).$$

If we define

$$\lambda^\circ_{\phi,\psi}(x) = \frac{\psi^\circ(x)}{\phi^\circ(x)}$$

then the chain above can also be written as

$$\lambda_{\phi,\psi}(x_k) \leq \lambda^\circ_{\phi,\psi}(y_k) \leq \lambda_{\phi,\psi}(x_{k+1})$$

which shows that $\lambda_{\phi,\psi}(x_k)$ and $\lambda^\circ_{\phi,\psi}(y_k)$ are bounded increasing sequences, converging to the same value. Since subdifferentials are closed maps and all iterates are in a compact set, we actually see that any accumulation point of the algorithm is a stationary point [Zangwill, 1969]. □

Robert's algorithm can be tied in with general block relaxation theory [De Leeuw, 1994] by observing that

$$\lambda(x) = \frac{\phi(x)}{\psi(x)} = \max_y \frac{\langle x, y\rangle}{\psi(x)\phi^\circ(y)}.$$

Thus, letting

$$\xi(x, y) \triangleq \frac{\langle x, y\rangle}{\psi(x)\phi^\circ(y)},$$

we see that

$$\max_x \lambda(x) = \max_{x,y} \xi(x, y).$$

Thus ξ is an *augmentation* of λ, and alternating maximization ξ over x for fixed y and over y for fixed x is exactly Robert's algorithm.

Now minimizing $\lambda(x)$ is clearly equivalent to maximizing $1/\lambda(x)$, and thus for minimization we can use the algorithm

$$y_k \in \partial\psi(x_k),$$
$$x_{k+1} \in \partial\phi^\circ(y_k).$$

5. Results for specific norms

5.1. **Schatten p norms.** In this section we look at maximizing

$$\lambda_{P,\psi} = \frac{\sqrt{\mathbf{tr}\ X'PX}}{\psi_p(X)},$$

over all $X \in \mathbb{R}^{n\times r}$, where ψ_p is the Schatten p-norm, i.e. the ℓ_p norm of the singular values.

Now

$$\psi_p(X) = [\mathbf{tr}(X'X)^{p/2}]^{1/p}$$

It is clear that the problem of maximizing $\mathbf{tr}\ X'PX$ over all X such that $\mathbf{tr}\ (X'X)^{p/2} = 1$ is equivalent to the Schatten-p homogeneity analysis problem we formulated above. It turns out that this problem can be solved by conventional eigenvalue-eigenvector methods.

The stationary equations are $PX = \mu X(X'X)^{\frac{p-2}{2}}$ and $\mathbf{tr}\ (X'X)^{p/2} = 1$. It follows that at a solution $\mu = \mathbf{tr}\ X'PX$. Suppose $X = K\Lambda L'$ is the singular value decomposition of X. Then the stationary equations become $PK\Lambda = \mu K\Lambda^{p-1}$ and $\mathbf{tr}\ \Lambda^p = 1$. We must solve this for $K \in \mathbb{R}^{n\times r}$ with $K'K = I$ and for diagonal $\Lambda \geq 0$. The diagonal elements of Λ are also non-increasing along the diagonal. Observe that the right singular vectors L are irrelevant for the Schatten problem, because both norms are invariant under right multiplication of X by a orthonormal matrix. In fact we can require without loss of generality that the columns of X are orthogonal.

We shall actually solve r sets of stationary equations, each set defined by the additional condition that the first s elements of Λ are positive and the last $r - s$ elements are equal to zero, with $s = 1, \cdots, r$. Of course we only solve for the first s columns of K and for the remaining s positive diagonal elements of Λ. To indicate this, we write the parts we are solving for as K_s and Λ_s. Thus, for problem s, we solve $PK_s = \mu K_s\Lambda_s^{p-2}$ and $\mathbf{tr}\ \Lambda_s^p = 1$.

Let us first deal with the situation $p = 2$, in which the Schatten norm is the Frobenius norm. Then we must have $PK_s = \mu K_s$, which is only possible if P has an eigenvalue of multiplicity at least equal to s. Since μ is equal to that eigenvalue, and $\mathbf{tr}\ X'PX = \mu$, we see that the maximum

is attained by choosing the largest eigenvalue of P. K_s is a corresponding set of eigenvectors. Generally, the multiplicity of the largest eigenvalue will be one, and thus X will be one-dimensional. In any case for $p = 2$ the maximum of $\lambda_{P,\psi}$ is the largest eigenvalue of P. If we collect all eigenvalues in the diagonal matrix in a vector Ω, then the optimum is the spectral norm, or the Schatten ∞-norm, of Ω.

If $p \neq 2$ it follows that K_s consists of $s \leq r$ orthogonal eigenvectors of P, and $\mu\Lambda_s^{p-2}$ is equal to the corresponding nonzero eigenvalues, say collected in a diagonal matrix Ω_s . Thus

$$\mathbf{tr}\ \Lambda_s^p = \mathbf{tr}\ \left[\frac{1}{\mu}\Omega_s\right]^{\frac{p}{p-2}} = 1,$$

and thus $\mu = [\mathbf{tr}\ \Omega_s^q]^{1/q}$, where $q = p/(p-2)$. Thus the optimum is the Schatten q-norm of Ω_s. Since $\mathbf{tr}\ X'PX = \mu$, it follows we must choose the eigenvalues such that this norm is maximized.

Again, there are two cases to consider. If $p > 2$ then q is positive. The optimum is found by choosing the s largest eigenvalues, i.e. the optimum is the Ky Fan (s, q) norm of P. The very best solution is the Ky Fan (r, q) norm. And, moreover, the solution for X is identical to the usual homogeneity analysis solution based on alternating least squares. If $p < 2$ then q is negative. We want to choose s eigenvalues such that $\mathbf{tr}\ \Omega_s^q$ is a small as possible. This means that we again choose the s largest eigenvalues, but if we vary s we see that the best solution is obtained for $s = 1$, in which case we have the largest eigenvalue only, and μ is again equal to the spectral norm of P.

Thus, in summary, maximizing the ratio with the usual norm $\sqrt{\mathbf{tr}\ X'PX}$ in the numerator and the Schatten p norm in the denominator gives the same solution as classical least squares homogeneity analysis if $p > 2$ and gives the eigenvector corresponding to the largest eigenvalue of P as a solution if $p \leq 2$. For $p > 2$ the optimal value is the Ky Fan (r, q) norm of P, where $q = p/(p-2)$. For $p \leq 2$ it is the spectral norm.

From the data analysis point of view this means that using Schatten p-norms to normalize does not give us anything new. It is interesting, however, that classical homogeneity analysis solutions can be recovered from maximization of a ratio of two matrix norms, without imposing orthonormality restrictions.

5.2. **Orthogonally invariant norms.**

6. Algorithm

We have versions of the basic algorithm in both R and Matlab. The R version is build on top of the existing homogeneity analysis implementation [de Leeuw and Ouwehand, 2003], and it uses some subroutines from that implementation. R code is in the appendix. Since the algorithm is so simple, the implementation is straightforward. The two norms are passed as parameters, so that users can add more norms to the repertoire of the algorithm.. The matrix P, which can be both large and dense, is never actually computed and stored.

7. Examples

Appendix A. Norms

In this Appendix we collect, for easy reference, some material about vector and matrix norms. Most of it is classical, but some other results, directly related to the discussion in our paper, are relatively recent.

A.1. **Generalities.** First some terminology [Rockafellar, 1970, Section 15]. A *gauge* is a non-negative positively homogeneous convex function. Thus η is a gauge if

- $\eta(x) \geq 0, \forall x,$
- $\eta(\lambda x) = \lambda\eta(x), \forall x, \forall \lambda > 0,$
- $\eta(x + y) \leq \eta(x) + \eta(y), \forall x, \forall y$

A gauge is a *semi-norm* if it is finite everywhere and

- $\eta(x) \geq 0, \forall x,$
- $\eta(\lambda x) = |\lambda|\eta(x), \forall x, \forall \lambda,$
- $\eta(x + y) \leq \eta(x) + \eta(y), \forall x, \forall y,$

and it is a *norm* if it is finite everywhere and

- $\eta(x) > 0, \forall x \neq 0,$
- $\eta(\lambda x) = |\lambda|\eta(x), \forall x, \forall \lambda,$
- $\eta(x + y) \leq \eta(x) + \eta(y), \forall x, \forall y,$

A gauge, or norm, or semi-norm is *symmetric* if, in addition,

- $\eta(x) = \eta(\Pi x), \forall x, \forall \Pi,$ with Π a permutation matrix,
- $\eta(x) = \eta(\Sigma x), \forall x, \forall \Sigma,$ with Σ a sign matrix, i.e. a diagonal matrix with elements ± 1.

We write η° for the *dual* (or *polar*) gauge, i.e.

$$\eta^\circ(y) = \inf \{\lambda \geq 0 \mid \langle x, y\rangle \leq \lambda\eta(x), \forall x\},$$

which can be written for a norm as

$$\eta^\circ(y) = \sup_{x \neq 0} \frac{\langle x, y\rangle}{\eta(x)}.$$

This definition implies

$$\langle x, y\rangle \leq \eta(x)\eta^\circ(y),$$

which is the generalization of Hölder's inequality we use to prove convergence of our algorithm.

For a norm η (actually, for any convex function) a vector y is a *subgradient* at x if, for all z,

$$\eta(z) \geq \eta(x) + \langle y, z - x \rangle.$$

See Rockafellar [1970, Section 23]. The set of all subgradients of η at x is the *subdifferential* at x, written as $\partial\eta$(x). Because norms are finite, their subdifferential is a nonempty compact convex set. If the norm η is differentiable at x, then its only subgradient is the gradient at x.

By looking at the ray $z = \theta x$, with $\theta > 0$, and by using the homogeneity of the norm, we see that if $y \in \partial\eta(x)$ then $\eta(x) = \langle y, x \rangle$. Thus subgradients y of η at x are also defined by the condition that $\eta(z) \geq \langle y, z \rangle$ for all z and $\eta(x) = \langle y, z \rangle$. In other words

$$\eta^\circ(y) = \max_{z \neq 0} \frac{\langle y, z \rangle}{\eta(z)} = 1,$$

and the maximum is attained at $z = x$.

Appendix B. Matrix Norms

There are excellent reviews of the theory and application of matrix norms in Householder [1964, Chapter 2], Horn and Johnson [1985, Chapter 5], and Stewart [2001, Chapter 1, Section 2] . There is some disagreement over terminology, however.

Both Householder and Horn and Johnson emphasize square matrices, and call a norm η a *matrix norm* if it satisfies, in addition to the usual conditions,

- $\eta(AB) \leq \eta(A)\eta(B), \forall A, B$

Norms on the space of square matrices which do not satisfy this condition are called *generalized matrix norms*.

We prefer Stewart's terminology. A matrix norm is simply a norm defined on $\mathbb{R}^{n \times m}$. And three matrix norms η_1, η_2, and η_3 are *consistent* if

- $\eta_1(AB) \leq \eta_2(A)\eta_3(B), \forall A, B$

B.1. **Vector norms.**

B.2. **Orthogonally invariant norms.**

This is an interesting class of norms, studied first by Neumann [1937]. They have the defining characteristic that $\eta(UXV) = \eta(X)$ for all square orthonormals U and V. Some obvious examples are the *Frobenius norm*, which we write as $|||X|||_I$, the *trace norm*, which is the sum of the singular values, and the *spectral norm*, which is the

largest singular value. Von Neumann also proved the basic theorem that any orthogonally invariant norm is a symmetric gauge function of the singular values (also see Mirsky [1960]). We use the notation $\psi(X) = \phi(\sigma(X))$, with ϕ a symmetric gauge, and with $\sigma(X)$ the vector of singular values of X. A good overview of the properties of orthogonally invariant norms is in Horn and Johnson [1991, Section 3.5].

The Frobenius norm, the trace norm, and the spectral norm are all examples of the *Schatten p-norms*, which are simply the ℓ_p norms of the vectors of singular values. For Frobenius we have $p = 2$, for trace $p = 1$, and for the spectral norm $p = +\infty$.

Another class of orthogonally invariant norms are the *Ky Fan k-norms*, which are the sums of the k largest singular values. They play a central role in the theory of orthogonally invariant norms. Clearly both the spectral and the trace norm are in this class.

We can combine the Schatten and Ky Fan classes to find the class of *Ky Fan* (k, p)*-norms*. These are simply the ℓ_p norms of the k largest singular values.

There are several general results available for subgradients of orthogonally invariant norms. [Watson, 1992; Zietak, 1988, 1993; Sá, 1994; Lewis, 1995]. Watson [1992, Theorem 2] shows that for any orthogonally invariant norm ψ corresponding to a symmetric gauge ϕ we have

$$\partial\psi(X) = \mathbf{conv}\{K\,\mathbf{diag}(d)L' \mid X = K\,\mathbf{diag}(\lambda)L', d \in \partial\phi(\lambda)\}.$$

It is shown by Zietak [1993] and Sá [1994] that it is not necessary to actually take the convex hull, because the set inside the brackets is already convex. Lewis [1995, Corollary 2.5] proves the result for general orthogonally invariant functions, which are not necessarily norms or gauges. It is also shown by Lewis [1995, Theorem 3.1] that ψ is differentiable at A if and only if ϕ is differentiable at $\sigma(A)$ and that

$$\nabla\psi(X) = K\,\mathbf{diag}(\nabla\phi(\sigma(X))L'$$

where $X = K\,\mathbf{diag}(\sigma(X))L'$. Moreover

$$\sigma(\nabla\psi(X)) = \nabla\phi(\sigma(X).$$

B.3. **Operator norms.**

Appendix C. Code

```
require(homals)

normHom<-function(mydat,datanorm,dualnorm,ndim=2,eps=1e-6,homit=10) {
x<-startX(mydat,ndim,homit)
while(1) {
        y<-datanorm(x,mydat)
        z<-dualnorm(y)
        print(c(sum(z*datanorm(z,mydat)),sum(y*dualnorm(y))))
        if (max(abs(x-z))<eps) break
        x<-z
        }
normHomPlot(x,mydat,deparse(substitute(mydat)),
        deparse(substitute(datanorm)),deparse(substitute(dualnorm)))
}

normHomPlot<-function(x,mydat,name,datanorm,dualnorm) {
xlim <- c(min(x[,1]),max(x[,1]))
ylim <- c(min(x[,2]),max(x[,2]))
name<-paste(name,datanorm,dualnorm)
pdf(file=paste(name,"pdf",sep="."))
graphplot(name,mydat,x,1,2)
for (i in 1:dim(mydat)[2]) {
        catplot(name,labels(mydat)[[2]][i],mydat[,i],computeY(mydat[,i
            ],x),xlim,ylim,1,2)
        starplot(name,labels(mydat)[[2]][i],mydat[,i],FALSE,x,1,2)
        }
dev.off()
}

startX<-function(mydat,ndim,homit) {
n<-dim(mydat)[1]; x<-matrix(rnorm(n*ndim),n,ndim)
for (i in 1: homit) x<-newX(mydat,x)
x<-apply(x, 2, function(z) z - mean(z))
svd(x)$u
}

newX<-function(mydat,x){
m<-dim(mydat)[2]; n<-dim(mydat)[1]; z<-matrix(0,n,dim(x)[2])
for (j in 1:m){
                g<-mydat[,j]; y<-apply(x, 2, function(z) tapply(z, g,
                    mean));
                z<-z+y[g,]
```

```
        }
    z<-as.matrix(apply(z, 2, function(z) z - mean(z)))/m
    }

sumS<-function(x) {s<-svd(x); s$u%*%t(s$v)}
maxS<-function(x) {s<-svd(x); outer(s$u[,1],s$v[,1])}
sumL<-function(x) sign(x)
maxL<-function(x) ifelse(abs(x)==max(abs(x)),sign(x),0)
quad<-function(x) x/sqrt(sum(x^2))
powr<-function(x) sign(x)*(abs(x)^(pp-1))*(sum(abs(x)^pp)^((1/pp)-1))
pows<-function(x) {s<-svd(x); d<-s$d[1:kk];
        (sum(d^pp)^((1/pp)-1))*(s$u[,1:kk]%*%diag(d^(pp-1))%*%t(s$v
            [,1:kk]))                    }

maxLP<-function(x,mydat) {z<-newX(mydat,x); sumL(z)}
sumLP<-function(x,mydat) {z<-newX(mydat,x); maxL(z)}
sumSP<-function(x,mydat) {z<-newX(mydat,x); sumS(z)}
maxSP<-function(x,mydat) {z<-newX(mydat,x); maxS(z)}
quadP<-function(x,mydat) {z<-newX(mydat,x); a<-sqrt(sum(x*z)); z/a}
powrP<-function(x,mydat) {z<-newX(mydat,x); powr(z)}
quitP<-function(x,mydat) {
m<-dim(mydat)[2]; n<-dim(mydat)[1]; z<-matrix(0,n,dim(x)[2])
for (j in 1:m){
                g<-mydat[,j]; y<-apply(x, 2, function(z) tapply(z, g,
                    mean));
                h<-y[g,]; a<-sqrt(sum(x*h)); z<-z+h/a
        }
    z<-as.matrix(apply(z, 2, function(z) z - mean(z)))/m
    }
```

References

P. Bekker and J. De Leeuw. Relation between variants of nonlinear principal component analysis. In J.L.A. Van Rijckevorsel and J. De Leeuw, editors, *Component and Correspondence Analysis*. Wiley, Chichester, England, 1988.

D.W. Boyd. The Power Method for ℓ^p Norms. *Linear Algebra and Its Applications*, 9:95–101, 1974.

J. De Leeuw. Applications of convex analysis to multidimensional scaling. In J.R. Barra, F. Brodeau, G. Romier, and B. Van Cutsem, editors, *Recent developments in statistics*, pages 133–145, Amsterdam, The Netherlands, 1977. North Holland Publishing Company.

J. De Leeuw. Block Relaxation Methods in Statistics. In H.H. Bock, W. Lenski, and M.M. Richter, editors, *Information Systems and Data Analysis*, Berlin, 1994. Springer Verlag.

J. de Leeuw and A. Ouwehand. Homogeneity Analysis in R. Technical report, UCLA Statistics, 2003.

A. Gifi. *Nonlinear multivariate analysis*. Wiley, Chichester, England, 1990.

L. Guttman. The quantification of a class of attributes: A theory and method of scale construction. In P. Horst, editor, *The prediction of personal adjustment*. Social Science Research Council, New York, New York, 1941.

R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

A.S. Householder. *The Theory of Matruces in Numerical Analysis*. Blaisdell Publising Company, New York - Toronto - London, 1964.

A.S. Lewis. The Convex Analysis of Unitarily Invariant Matrix Functions. *Journal of Convex Analysis*, 2:173–183, 1995.

G. Michailidis and J. De Leeuw. The Gifi system for descriptive multivariate analysis. *Statistical Science*, 13:307–336, 1999.

L. Mirsky. Symmetric gauge Functions and Unitarily Invariant Norms. *Quaterly Journal of Mathematics (Oxford)*, 11:50–59, 1960.

J. Von Neumann. Some Matrix Inequalities and Matrix Norms. *Tomsk University Review*, 1:286–300, 1937.

F. Robert. Calcul du Rapport Maximal de Deux Normes sur $\mathbb{R}^n$. *Revue Francaise d'Automatique, d'Informatique Et De Recherche Operationelle*, 1: 97–118, 1967.

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

E. Marques De Sá. Exposed Faces and Duality for Symmatrix and Unitarily Invariant Norms. *Linear Algebra and Its Applications*, 197:429–450, 1994.

B.F. Schriever. *Order Dependence*. PhD thesis, University of Amsterdam, The Netherlands, 1985. Also published in 1985 by CWI, Amsterdam, The Netherlands.

G.W. Stewart. *Matrix Algorithms. Volume II: Eigenssystems*. SIAM, Philadelphia, 2001.

Pham Dinh Tao. Elements Homoduaux d'une Matrice A Relatifs a un Couple de Normes (ϕ, ψ). Applications au Calcul de $S_{\phi\psi}(A)$. Seminaire d'Analyse Numérique 236, Université Scientifique et Médicale de Grenoble, 1975.

Pham Dinh Tao. Calcul du Maximum d'une Forme Quadratique Définie Positive sur la Boule Unité de ψ_∞. Seminaire d'Analyse Numérique 247,

Université Scientifique et Médicale de Grenoble, 1976.

Pham Dinh Tao. Convergence of a Subgradient Method for Computing the Bound Norm of Matrices. *Linear Algebra and Its Applications*, 62:163–182, 1984.

Pham Dinh Tao. Algorithmes de Calculdu Maximum des Formes Quadratiques sur la Boule Unité de la Norme Maximum. *Numerische Mathematik*, 45:377–440, 1985.

G.A. Watson. Characterization of the Subdifferentials of Some Matrix Norms. *Linear Algebra and Its Applications*, 170:33–45, 1992.

W. I. Zangwill. *Nonlinear Programming: a Unified Approach*. Prentice-Hall, Englewood-Cliffs, N.J., 1969.

K. Zietak. On the Characterization of the Extreme Points of the Unit Sheree of Matrices. *Linear Algebra and Its Applications*, 106:57–75, 1988.

K. Zietak. Subdifferentials, Faces, and Dual Matrices. *Linear Algebra and Its Applications*, 185:125–141, 1993.

Jan de Leeuw, Department of Statistics , University of California, Los Angeles

E-mail address: deleeuw@stat.ucla.edu

URL: http://gifi.stat.ucla.edu

George Michailides, Department of Statistics, University of Michigan, Ann Arbor

E-mail address: gmichail@umich.edu

URL: http://www.stat.lsa.umich.edu/ gmichail/