GENERALIZATIONS OF OSTROWSKI'S THEOREM

JAN DE LEEUW

1. Introduction

Ostrowski [1966], see also Ortega and Rheinboldt [1970, Chapter 10], studied stationary one-point iterative processes of the form

(1) $$x^{(k+1)} = G(x^{(k)}).$$

Here $G : D \subset \mathbb{R}^n \Rightarrow \mathbb{R}^n$. In particular, Ostrowski's theory deals with *points of attraction* and *points of repulsion* of such iterations, and with their speed of convergence.

Definition 1.1. A point $x^\star$ is a poa if there is a neighborhood of $x^\star$ such that the iteration G converges to $x^\star$ when started at any point x_0 in the neighborhood.

2. Nonmetric Discriminant Analysis

2.1. Loss Function.

In nonmetric discriminant analysis [De Leeuw, 1968] the problem is to minimize a loss function of the form

(2) $$\sigma(\beta, y) = \|X\beta - y\|^2$$

over all β such that $\|X\beta\|^2 = 1$ and over all $y \geq 0$. Here X is a given $n \times p$ matrix, which we suppose (without loss of generality) to satisfy $X'X = I$. This is just one method to find an approximate solution to the (possibly inconsistent) system of linear inequalities $X\beta \geq 0$.

Date: December 2, 2004.

2000 *Mathematics Subject Classification.* 62H25.

Key words and phrases. Multivariate Analysis, Correspondence Analysis.

2.2. **Algorithm.** De Leeuw [1968] proposes to minimize the loss function by block relatation [De Leeuw, 1994]. We start with some initial β, then find the optimal $y \geq 0$ for the given β, then update β by computing the optimal β for the current y, and so on.

Finding $y(\beta)$, the optimal y for given β, is very simple, because it merely involves taking the non-negative part of $X\beta$. Thus

$$y_i(\beta) = \begin{cases} x_i'\beta & \text{if } x_i'\beta \geq 0, \\ 0 & \text{if } x_i'\beta < 0 \end{cases}$$

It is convenient, for our purposes, to define the binary diagonal matrix $M(\beta)$ by

$$m_{ii}(\beta) = \begin{cases} 1 & \text{if } x_i'\beta \geq 0, \\ 0 & \text{if } x_i'\beta < 0 \end{cases}$$

Clearly $y(\beta) = M(\beta)X\beta$. It should be emphasized that if $X\beta$ has zero elements, then there are other binary diagonal matrices M such that $y(\beta) = MX\beta$

In this notation our algorithm alternates the steps

$$y^{(k+1)} = M(\beta^{(k)})X\beta^{(k)}, \tag{3a}$$

$$\lambda^{(k+1)}\beta^{(k+1)} = X'y^{(k+1)}. \tag{3b}$$

with $\lambda^{(k+1)}$ chosen such that $\|\beta^{(k+1)}\| = 1$.

2.3. **Condensed Form.** If we combine the two steps of Algorithm 7 we find

$$\lambda^{(k+1)}\beta^{(k+1)} = X'M(\beta^{(k)})X\beta^{(k)} \tag{4}$$

and this is the iteration in $\mathbb{R}^p$ that we study in this example.

The map $\beta^{(k)} \Rightarrow \beta^{(k+1)}$ is continuous, but in general not differentiable. Thus Ostrowski's theorem does not apply directly, and a more detailed analysis is needed.

2.4. **Fixed Points.** What are the fixed points of this map ?

Let us partition the set of all possible values of β by using the set $\mathbb{M}$ of binary diagonal matrices M. For each of the 2^n matrices $M \in \mathbb{M}$ define the cones

$$K_M = \{\beta \mid MX\beta \geq 0\}.$$

The cones K_M consist of all β that produce the same sign pattern in $X\beta$. Of course some of the cones may be degenerate cone $K_M = \{0\}$. In particular

Now find the fixed points in each of the K_M, skipping the ones for which $K_M = \{0\}$. This means we have to solve

$$X'MX\beta = \lambda X'X\beta, \tag{5a}$$

$$\beta \in K_M. \tag{5b}$$

Equation (5a) says β is an eigenvector of the generalized eigenvalue problem defined by M, so we merely have to check if there is indeed an eigenvector in the cone K_M. If there is such an eigenvector $\hat{\beta}$, normalize it by $\hat{\beta}'X'X\hat{\beta} = 1$, and set $\hat{y} = MX\hat{\beta}$. For this solution $\hat{\lambda} = \hat{\beta}'X'MX\hat{\beta} = 1 - \sigma(\hat{\beta}, \hat{y})$.

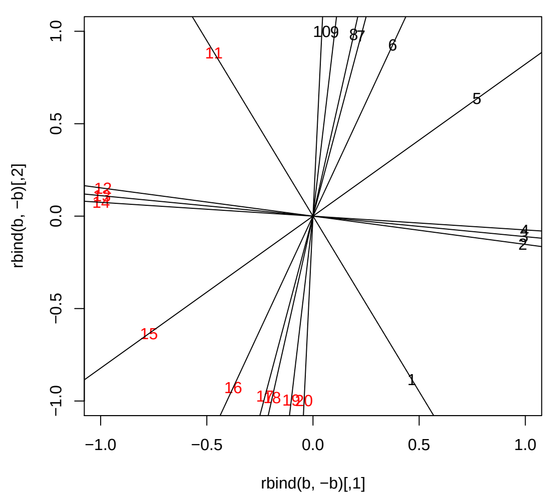
It is also easy to see that if $(\hat{\beta}, \hat{\lambda})$ is a solution of the stationary equations for cone K_M, then so is $(-\hat{\beta}, 1 - \hat{\lambda})$ for cone K_{I-M}. Stationary values occur in pairs.

2.5. **Numerical Example.** Consider the following small example with $n = 10$.

There are twenty non-trivial cones K_M in this case, because each of the ten lines $\ell_i = \{\beta \mid x_i'\beta = 0\}$ divides the plane into two half-spaces. Ordering the lines clockwise defines the cones as the pieces between adjacent half-lines. The twenty cones occur in pairs, with one member of the pair the negative of the other.

By exhaustively searching the ten cones we find the following pairs of stationary values.

-1.35197864003772	0.697985577308151
-0.0874873643060007	-0.94704322032545
0.0536054871679144	0.863455120675496
-0.36186071455119	-2.83999386122057
1.05205110884098	0.638117926006177
-1.96427832254002	0.298018854643880
0.500615236829422	-0.153907184745402
-0.522611258664262	0.780199767226874
0.559141386252437	-0.0439217343211844
0.181360391793109	-0.0477995853847141



loss	β_1	β_2
0.8154202	0.9628697	-0.2699664
0.1845798	-0.9628697	0.2699664
0.7478425	0.9009728	0.4338756
0.2521575	-0.9009728	-0.4338756
0.8174072	0.2157659	0.9764451
0.1825928	-0.2157659	-0.9764451

3. Nonhomogeneous Inequalities

There is an interesting variation

$$\sigma(\beta, y) = \|X\beta - z - y\|^2 \tag{6}$$

with z a known non-zero vector. This must be minimized over β, which we do not have to normalize any more, and over $y \geq 0$. We can think of this as a method to approximately solve the system $X\beta \geq z$. Alternatively, we can think of this as a homogeneous problem in which we do not normalize by $\|\beta\| = 1$ but by fixing one of the elements of β to be -1.

Define $M(\beta)$ as any binary diagonal matrix such that $M(\beta)(X\beta - z) \geq 0$. The block iterations are

$$y^{(k+1)} = M(\beta^{(k)})(X\beta^{(k)} - z), \tag{7a}$$

$$\beta^{(k+1)} = X'(z + y^{(k+1)}), \tag{7b}$$

which is in the condensed form

$$\beta^{(k+1)} = X'[z + M(\beta^{(k)})(X\beta^{(k)} - z)]. \tag{8}$$

At a stationary point we have the normal equations

$$X'(I - M(\beta))X\beta = X'(I - M(\beta))z. \tag{9}$$

Instead of solving an eigenvalue problem for each cone, we now have to solve these normal equations for each closed convex polyhedron C_M.

Appendix A. Code

```
require ("MASS")

lineq<-function(x,beta=rep(1,dim(x)[2])) {
xinv<-ginv(x)
y<-x%*%beta
s<-sqrt(sum(y^2))
beta<-beta/s; y<-y/s; itel<-1
repeat {
        yhat<-y
        yhat[which(y<0)]<-0
        print(c(0,itel,sum((y-yhat)^2)))
        beta<-xinv%*%yhat
        y<-x%*%beta
        s<-sqrt(sum(y^2))
        beta<-beta/s; y<-y/s
        print(c(1,itel,sum((y-yhat)^2)))
        if (itel==10) break()
                else itel<-itel+1
        }
yhat
}
```

References

J. De Leeuw. Nonmetric Discriminant Analysis. Technical Report RN006-1968, Department of Data Theory, University of Leiden, 1968.

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. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, N.Y., 1970.

A. M. Ostrowski. *Solution of Equations and Systems of Equations*. Academic Press, New York, N.Y., 1966.

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