A PLANE IN MDS CONFIGURATION SPACE

JAN DE LEEUW

1. Problem

The *Multidimensional Scaling* or *MDS* problem (of locating n point in p dimensions) is to minimize the function

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

over all $n \times p$ matrices X. Here w_{ij} are known *weights* and δ_{ij} are known *dissimilarities*. ToAlso

$$d_{ij}(X) = \|x_i - x_j\| = \sqrt{(x_i - x_j)'(x_i - x_j)}$$

is the Euclidean distance between rows i and j of X.

2. Transformation

By choosing a basis $Y_1, \cdots, Y_m$ for the space of all $n \times p$ matrices $\mathbb{R}^{n\times p}$, we can write X in the form

$$X = \sum_{\nu=1}^{m} \gamma_\nu Y_\nu. \tag{2}$$

This also makes it possible to consider the more general problem in which X varies in an m-dimensional subspace of $\mathbb{R}^{n\times p}$. The MDS problem now becomes minimization of the loss function $\sigma(\gamma)$ over the coefficients in (2).

Date: November 29, 2008.

2000 *Mathematics Subject Classification.* 62H25.

Key words and phrases. Multivariate Analysis, Correspondence Analysis.

Let us introduce some notation to simplify this problem [De Leeuw, 1993]. We can write

$$d_{ij}^2(y) = \mathbf{tr}(\sum_{\nu=1}^{m} y_\nu Y_\nu)' A_{ij} (\sum_{\nu=1}^{m} y_\nu Y_\nu),$$

where

$$A_{ij} = (e_i - e_j)(e_i - e_j)'$$

and e_i and e_j are unit vectors (columns of the identity matrix). Define the $m \times m$ matrices C_{ij}, with elements

$$(C_{ij})_{\nu\mu} = \mathbf{tr}\; Y_\nu' A_{ij} Y_\mu.$$

Observe the C_{ij} are positive semi-definite. Now

$$d_{ij}^2(y) = y' C_{ij} y$$

Suppose, without loss of generality, that the basis Y_ν is chosen in such a way that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} C_{ij} = I.$$

Then

$$\sigma(y) = 1 + \frac{1}{2} y' y - \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \delta_{ij} \sqrt{y' C_{ij} y}, \tag{3}$$

where we have also assumed, without loss of generality, that

$$\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \delta_{ij}^2 = 1.$$

The MDS problem, which we call problem $\mathcal{P}$ from now on, is clearly equivalent to minimizing (3) over y.

3. Necessary Conditions for a Local Minimum

In a neighborhood of each local minimum we have $d_{ij}(y) > 0$ for all $i \neq j$ [De Leeuw, 1984]. Thus the loss function is differentiable at local minima and

$$\frac{\partial \sigma}{\partial y} = (I - B(y))y,$$

with

$$B(\gamma)=\sum_{i=1}^n\sum_{j=1}^n w_{ij}\frac{\delta_{ij}}{d_{ij}(\gamma)}C_{ij}.$$

Also

$$\frac{\partial^2\sigma}{\partial\gamma\partial\gamma}=I-H(\gamma),$$

where

$$H(\gamma)=\sum_{i=1}^n\sum_{j=1}^n w_{ij}\frac{\delta_{ij}}{d_{ij}(\gamma)}\left[C_{ij}-\frac{C_{ij}\gamma\gamma'C_{ij}}{\gamma'C_{ij}\gamma}\right].$$

Thus at local minimum we must have

$$B(\gamma)\gamma=\gamma, \tag{4a}$$

and in addition

$$H(\gamma)\lesssim I. \tag{4b}$$

Or, in words, γ must be an eigenvector of $B(\gamma)$ corresponding with an eigenvalue equal to one, and all eigenvalues of $H(\gamma)$ must be less than or equal to one. Observe that both $B(\gamma)$ and $H(\gamma)$ are positive semi-definite. Because $H(\gamma)\gamma=0$ the matrix $H(\gamma)$ is always singular. Also note that if $H(\gamma)<I$ then γ is a local minimum.

The majorization algorithm [De Leeuw, 1977] for MDS takes the simple form

$$\gamma^{(k+1)}=B(\gamma^{(k)})\gamma^{(k)},$$

while the Newton-Raphson method can be written in the form

$$\gamma^{(k+1)}=(I-H(\gamma^{(k)}))^{-1}B(\gamma^{(k)})\gamma^{(k)}.$$

4. Relaxation

There is yet another way of rewriting the problem. Define

$$\sigma(\Gamma)=1+\frac{1}{2}\mathbf{tr}\,\Gamma-\sum_{i=1}^n\sum_{j=1}^n w_{ij}\delta_{ij}\sqrt{\mathbf{tr}\,C_{ij}\Gamma}, \tag{5}$$

Problem $\mathcal{P}$ is now minimizing (5) over all positive semi-definite Γ of rank one, i.e. all Γ of the form $\Gamma=\gamma\gamma'$.

Now define problem $\overline{\mathcal{P}}$, which is the problem of minimizing (5) over all positive semi-definite matrices Γ, without imposing the rank one constraint. Problem $\overline{\mathcal{P}}$ is a *convex relaxation* of problem $\mathcal{P}$, because the loss function (5) is convex in Γ and in $\overline{\mathcal{P}}$ we require Γ to vary in a convex set. The necessary and sufficient conditions for a local (and thus global) minimum in $\overline{\mathcal{P}}$ are

$$I - B(\Gamma) \gtrsim 0, \tag{6a}$$

$$\Gamma \gtrsim 0, \tag{6b}$$

$$\mathbf{tr}\ \Gamma(I - B(\Gamma)) = 0, \tag{6c}$$

where

$$B(\Gamma) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\frac{\delta_{ij}}{d_{ij}(\Gamma)}C_{ij},$$

and $d_{ij}(\Gamma) = \sqrt{\mathbf{tr}\ C_{ij}\Gamma}$.

If the solution $\hat{\Gamma}$ of $\overline{\mathcal{P}}$ is of rank one, then $\hat{\Gamma} = \hat{\gamma}\hat{\gamma}'$, and obviously $\hat{\gamma}$ solves $\mathcal{P}$. This means that not only (4a) and (4b) are satisfied, but actually $\hat{\gamma}$ gives the global minimum of the loss function (3).

If $\hat{\gamma}$ is a solution of (4a), then $\hat{\Gamma} = \hat{\gamma}\hat{\gamma}'$ satisfies both (6b) and (6c). If $B(\hat{\gamma}) \lesssim I$, i.e. if $\hat{\gamma}$ is the eigenvector of $B(\hat{\gamma})$ corresponding to the largest eigenvalue, then (6a) is satisfied as well. Thus $\hat{\Gamma}$ solves $\overline{\mathcal{P}}$, and $\hat{\gamma}$ gives the global minimum of (3). Observe that $H(\gamma) \lesssim B(\gamma)$ for all γ, which means that for loss function (3) the sufficient condition for a local minimum $H(\gamma) < I$ is weaker than the sufficient condition for global minimum $B(\gamma) \lesssim I$ (as it should be, of course).

5. A Plane in Configuration Space

We can obtain some additional results if $m = 2$, i.e. if X is in a plane in $\mathbb{R}^{n\times p}$ spanned by just two matrices Y_1 and Y_2. In that case the solution $\hat{\Gamma}$ of $\overline{\mathcal{P}}$ is a matrix of order two. The rank of $\hat{\Gamma}$ can be either zero, one, or two. We can safely exclude $\Gamma = 0$, because we know that at a local minimum all $d_{ij}(\Gamma)$ must be positive. If the rank of $\hat{\Gamma}$ is two, then $B(\hat{\Gamma}) = I$. Because the C_{ij} are linearly independent this implies that $d_{ij}(\hat{\Gamma}) = \delta_{ij}$, i.e. $\sigma(\hat{\Gamma}) = 0$. It follows that if $\sigma(\hat{\Gamma}) > 0$ then $\hat{\Gamma}$ is of rank one and gives the global minimum $\hat{\gamma}$ of $\mathcal{P}$.

$B(\gamma)\gamma = \gamma$ means that γ is an eigenvector of both $B(\gamma)$ and $H(\gamma)$, with eigenvalues one and zero, repectively. If $m = 2$ then $\overline{\gamma}$, which is orthogonal to γ, is also an eigenvector of both $B(\gamma)$ and $H(\gamma)$, with eigenvalues, say, μ and ξ. We know that $0 \leq \xi \leq \mu$. If $0 \leq \xi \leq 1 \leq \mu$ we have a local minimum, if $0 \leq \xi \leq \mu \leq 1$ we have a global minimum.

References

J. De Leeuw. Differentiability of Kruskal's Stress at a Local Minimum. *Psychometrika*, 49:111–113, 1984.

J. De Leeuw. Fitting Distances by Least Squares. Technical Report UCLA Statistics Series 130, Interdivisonal Program in Statistics, UCLA, Los Angeles, California, 1993.

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.

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