COMMUN. STATIST.-THEOR. METH., A7(13), 1233-1241 (1978)

SOME PROPERTIES OF CLASSICAL MULTI-DIMENSIONAL SCALING

K.V. Mardia

University of Leeds,
Leeds, U.K.

Key Words & Phrases: *classical solution, principal coordinates, Euclidean representation, linear model, missing values.*

ABSTRACT

The paper gives a new optimal property of the classical method of multi-dimensional scaling when the distance matrix is non-Euclidean. We also examine robustness of the method under a linear model. A technique to estimate missing values is also given.

1. INTRODUCTION

Let $\underset{\sim}{D} = (d_{ij})$ be an $n \times n$ distance matrix (i.e. $d_{ii} = 0$, $d_{ij} = d_{ji} \geq 0$). The multi-dimensional scaling (MDS) solution consists of obtaining a configuration of n points in p dimensional Euclidean space. The classical method of Torgerson (1952,1958) can be constructed from the following theorem.

Theorem 1.1. Let

$$\underset{\sim}{B} = \underset{\sim\sim\sim}{HAH}, \tag{1.1}$$

where

$$\underset{\sim}{H} = \underset{\sim n}{I} - n^{-1}\underset{\sim\sim}{11'}, \quad \underset{\sim}{A} = (a_{ij}), \quad a_{ij} = -\frac{1}{2}d_{ij}^2 . \tag{1.2}$$

Suppose that $\underset{\sim}{B}$ is positive semi-definite (p.s.d.) with rank

p ($p \leq n-1$). Let $\lambda_1, \ldots, \lambda_p$ be the p non-zero eigenvalues of $\underset{\sim}{B}$ with corresponding eigenvectors

$$\underset{\sim}{x}'_{(k)} = (x_{1k}, x_{2k}, \ldots, x_{nk}), \quad k=1,\ldots,p , \tag{1.3}$$

normalized by

$$\underset{\sim}{x}'_{(k)} \underset{\sim}{x}_{(k)} = \lambda_k, \quad k=1,\ldots,p . \tag{1.4}$$

Then the points P_i with co-ordinates $\underset{\sim}{x}'_i = (x_{i1}, \ldots, x_{ip})$, $i=1,\ldots,n$ constitute an MDS solution in p-dimensional space. Further, the centre of gravity of the points P_i is at the origin. Conversely, if $\underset{\sim}{D}$ is Euclidean (i.e. $d_{ij}^2 = \sum_{k=1}^{p} (x_{ik}-x_{jk})^2$ then $\underset{\sim}{B}$ is positive semi-definite (p.s.d.)

For a proof see Gower (1966). Also see Schoenberg (1935) and Young and Householder (1938). These co-ordinates $\underset{\sim}{x}'_i$, are called <u>principal co-ordinates</u> in p-dimensions.

2. A NEW OPTIMAL PROPERTY

Let $\underset{\sim}{B}$ be p.s.d. with rank $\underset{\sim}{B} = p$ so that we can obtain principal co-ordinates $\underset{\sim}{x}_i$, $i=1,\ldots,n$ from Theorem 1.1. We have the following optimal property when $\underset{\sim}{D}$ is Euclidean (Gower, 1966). Let $\underset{\sim}{x}^*_i$, $i=1,\ldots,n$ be a projection of $\underset{\sim}{x}_i$ onto an r-dimensional subspace, $r \leq p$, and let d^*_{ij} be the Euclidean distance between the i^{th} and j^{th} projected points. Then

$$\Sigma\Sigma(d_{ij}^2 - d_{ij}^{*2})$$

is a minimum when the co-ordinates are the rows of $(\underset{\sim}{x}_{(1)}, \ldots, \underset{\sim}{x}_{(r)})$, i.e. the co-ordinates are principal co-ordinates in r-dimensions. However, if $\underset{\sim}{D}$ is <u>not</u> Euclidean, it is still possible to find an optimality property in terms of $\underset{\sim}{B}$.

Consider $\underset{\sim}{B} = \underset{\sim\sim\sim}{HAH}$ to be a symmetric matrix (not necessarily p.s.d.) with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$. Fix r, $1 \leq r < n$. We propose to find a p.s.d. matrix $\underset{\sim}{B}^* = (b^*_{ij})$ of rank at most r such that

$$\Sigma\Sigma(b_{ij}-b^*_{ij})^2 = \mathrm{tr}\,(\underset{\sim}{B}-\underset{\sim}{B}^*)^2 \qquad (2.1)$$

is a minimum. Note that $d^{*2}_{ij} = b^*_{ii}+b^*_{jj}-2b^*_{ij}$ for the resulting configuration.

Let $\lambda^*_1 \geq \ldots \geq \lambda^*_n$ be the eigenvalues of $\underset{\sim}{B}^*$. Following Eckart and Young (1936), it can be shown that

$$\min \mathrm{tr}\,(\underset{\sim}{B}-\underset{\sim}{B}^*)^2 = \min \sum_{k=1}^{n} (\lambda_k-\lambda^*_k)^2 , \qquad (2.2)$$

where the minimum on the RHS is taken over non-negative λ^*_i's for which at least $n-r$ of them are zero. In general this minimum is given by

$$\lambda^*_k = \max(\lambda_k, 0), \quad k=1,\ldots,r,$$
$$= 0, \quad k=r+1,\ldots,n.$$

Thus if there are r positive eigenvalues in $\underset{\sim}{B}$, we have $\lambda^*_k = \lambda_k$ for $k=1,\ldots,r$. Further, if $\underset{\sim}{B}$ is p.s.d. of rank $p \geq r$, then under this optimisation, only the first r principal co-ordinates are to be used.

The minimum value of (2.1) defines a measure of distortion of $\underset{\sim}{B}$ from r-dimensional Euclidean space and in this case is given by

$$\sum_{k=r+1}^{p} \lambda^2_k . \qquad (2.3)$$

Linear Model. Lingoes (1971) has shown that given any real symmetric $\underset{\sim}{C}$, there exists a Euclidean representation in $r \leq n-2$ dimensions with interpoint distances d^*_{ij}, such that

$$d^{*2}_{ij} = 2(c_{ij}-a), \quad i \neq j,$$
$$= 0, \quad i=j, \qquad (2.4)$$

for some real number a. Thus $c_{ij} < c_{kl}$ implies $d^{*2}_{ij} < d^{*2}_{kl}$, and d^{*2}_{ij} is a linear function of c_{ij}. Hence the method preserves the rank order.

To see how Lingoes' method works on an empirical distance matrix $\underset{\sim}{D} = (d_{ij})$ with $\underset{\sim}{A} = (-\frac{1}{2}d_{ij}^2)$, let a be a real number and define

$$\underset{\sim}{A}^* = (-\tfrac{1}{2}d_{ij}^{*2}) = \underset{\sim}{A} - a(\underset{\sim}{I}-\underset{\sim}{J}) ,$$

which implies

$$d_{ij}^{*2} = d_{ij}^2 - 2a, \quad i \neq j,$$

$$= 0, \qquad i=j. \tag{2.5}$$

Define also $\underset{\sim}{B}^* = \underset{\sim}{H}\underset{\sim}{A}^*\underset{\sim}{H}$ so that

$$\underset{\sim}{B}^* = \underset{\sim}{H}\underset{\sim}{A}^*\underset{\sim}{H} = \underset{\sim}{H}\underset{\sim}{A}\underset{\sim}{H} - a\underset{\sim}{H} = \underset{\sim}{B} - a\underset{\sim}{H} . \tag{2.6}$$

If $\lambda_1 \geq \ldots \geq \lambda_n$ are the eigenvalues of $\underset{\sim}{B}$, they can be rewritten as $\lambda_1 \geq \ldots \geq \lambda_r > 0 \geq \lambda'_1 \geq \ldots \geq \lambda'_s$ with corresponding eigenvectors $\underset{\sim}{u}_1, \ldots, \underset{\sim}{u}_r, \underset{\sim}{1}, \underset{\sim}{v}_1, \ldots, \underset{\sim}{v}_s$, say.

Lingoes' choice for a is $a = \lambda'_s$. The eigenvectors of $\underset{\sim}{B}$ and $\underset{\sim}{B}^*$ remain the same, but with a change of order. Note that $\lambda_i = 0$ maps into $\lambda_i^* = 0$ with eigenvector $\underset{\sim}{1}$, whilst the remaining eigenvalues of $\underset{\sim}{B}^*$ become $\lambda_1 - \lambda'_s \geq \ldots \geq \lambda_r - \lambda'_s \geq \lambda'_1 - \lambda'_s \geq \ldots \geq \lambda'_{s-1} - \lambda'_s \geq 0$ and consequently all the eigenvalues of $\underset{\sim}{B}^*$ are non-negative. Hence we have the eigenvectors $\underset{\sim}{u}_1, \ldots, \underset{\sim}{u}_r, \underset{\sim}{v}_1, \ldots, \underset{\sim}{v}_s, \underset{\sim}{1}$, corresponding to the eigenvalues $\lambda_1 - \lambda'_s, \ldots, \lambda_r - \lambda'_s, \lambda'_1 - \lambda'_s, \ldots, \lambda'_{s-1} - \lambda'_s, 0, 0$.

Let $\underset{\sim}{D}$ denote a general distance matrix. For $a < \lambda_n$, $\underset{\sim}{D}^*$ will be a Euclidean distance matrix and from (2.2) the distortion is

$$\operatorname{tr}(\underset{\sim}{B}-\underset{\sim}{B}^*)^2 = \operatorname{tr}(a\underset{\sim}{H})^2 = (n-1)a^2 .$$

If $\lambda_n < 0$ then the distortion from the Lingoes class of models is minimal for $a = \lambda_n$ and equals $(n-1)\lambda_n^2$. Similarly, if all the eigenvalues are non-negative then no distortion results by taking $a = 0$ and using the full Torgerson solution.

Note that if $\lambda_1 \geq \ldots \geq \lambda_r \geq 0 \geq \lambda'_1 \geq \ldots \geq \lambda'_s$, the distortion resulting from the r-dimensional Torgerson solution is $\sum_{k=1}^{s} \lambda'^2_k$,

which is in general considerably less than the Lingoes distortion, $(n-1)\lambda_n^2$. Thus, we conclude that the Lingoes transformation (if used as a method) gives an exact Euclidean fit to the data at the cost of increased distortion and increased dimension. The Torgerson solution based on the positive eigenvalues of B is to be preferred.

Measures of Agreement. Again suppose D to be a general distance matrix and let r be the number of positive eigenvalues of B. For $k \leq r$ two possible agreement measures for the 'proportion of D explained' by the k-dimensional Torgerson solution are

$$\alpha_{1,k} = \Big(\sum_{\ell=1}^{k} |\lambda_\ell| \Big/ \sum_{\ell=1}^{n} |\lambda_\ell|\Big)\times 100\%, \quad \alpha_{2,k} = \Big(\sum_{\ell=1}^{k} \lambda_\ell^2 \Big/ \sum_{\ell=1}^{n} \lambda_\ell^2\Big)^{\frac{1}{2}}\times 100\%.$$

We write $\alpha_{1,k}^*$ and $\alpha_{2,k}^*$ when only positive eigenvalues are used in the denominators of $\alpha_{1,k}$ and $\alpha_{2,k}$, respectively. A justification for $\alpha_{2,k}$ as a measure has already been described. Possible measures of the 'Euclideanness' of D are obtained by taking $k=r$, $\beta_1 = \alpha_{1,r}$, $\beta_2 = \alpha_{2,r}$. Note that $\beta_1 = \beta_2 = 100\%$ when there is no distortion.

3. LINEAR MODEL

The Torgerson method is expected to be sensitive to extreme distortions, but the question arises whether the resulting configurations from this method remain robust under monotone transformations of the distances. We now study the effect on the configuration under the model

$$d_{ij}^{*2} = d_{ij}^2 - 2a, \qquad i \neq j , \tag{3.1}$$

where a is small so that $d_{ij}^2 - 2a \geq 0$ for all $i \neq j$. Let $X(n\times p)$ and $Y(n\times p)$ be two matrices with centre of gravity zero, whose rows are the estimated coordinates of an unknown configuration. A measure of the differences in the configurations (Schönemann and Carroll, 1970; Gower, 1971) is given by

$$m^2 = \operatorname{tr} XX' + \operatorname{tr} YY' - 2 \operatorname{tr} (X'YY'X)^{\frac{1}{2}} .$$

We will consider the case when the configuration of interest is 2-dimensional but the method can easily be extended. Let $\underset{\sim}{X} = (\underset{\sim}{x}_{(1)}, \underset{\sim}{x}_{(2)})$ and $\underset{\sim}{Y} = (\underset{\sim}{y}_{(1)}, \underset{\sim}{y}_{(2)})$ be the principal coordinates under $a=0$ and (3.1) respectively in two dimensions, and let (λ_1, λ_2) and (μ_1, μ_2) be the corresponding eigenvalues of $\underset{\sim}{B}$ for the two cases. Then we have

(i) $\mu_i = \lambda_i - a, \quad i=1,2;$

(ii) $\underset{\sim}{y}_{(i)} = \left(\frac{\lambda_i - a}{\lambda_i}\right)^{\frac{1}{2}} \underset{\sim}{x}_{(i)}, \; \underset{\sim}{x}'_{(i)}\underset{\sim}{x}_{(i)} = \lambda_i, \; \underset{\sim}{y}'_i\underset{\sim}{y}_i = \lambda_i - a .$

Hence, it can be shown that

$$m^2 = 2(\lambda_1+\lambda_2-a) - 2[\{\lambda_1(\lambda_1-a)\}^{\frac{1}{2}} + \{\lambda_2(\lambda_2-a)\}^{\frac{1}{2}}] .$$

For $\lambda_1, \lambda_2 >> a$, we have

$$m^2 \simeq \frac{a^2}{4}\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right) . \qquad (3.2)$$

Hence a small value of a does not alter the configuration much. Further, if a is distributed as $N(0,\sigma^2)$ with σ very small so that $d^{*2} = d^2 - 2a \geq 0$ with probability near 1, we have

$$4m^2 \sim \sigma^2\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right)\chi_1^2 .$$

This result can be easily extended.

In the general case, note that if the non-negative eigenvalues of $\underset{\sim}{B}$ under $a=0$ are $\lambda_1, \ldots, \lambda_p$, with the remaining $n-p$ zero, the corresponding (unstandardised) eigenvectors being $\underset{\sim}{v}_k$, $k=1,\ldots,p$ and $\underset{\sim}{w}_k$, $k=p+1,\ldots,n-1$, then under (3.1) the eigenvalues of $\underset{\sim}{B}$ are $\lambda_k - a$, $k=1,\ldots,p$, $-a$ repeated $n-p-1$ times, and zero, with corresponding eigenvectors $\underset{\sim}{v}_k$, $k=1,\ldots,p$, $\underset{\sim}{w}_k$, $k=p+1,\ldots,n-1$ and $\underset{\sim}{w}_n = \underset{\sim}{1}$. Thus the noise can produce negative eigenvalues.

One possible way to choose a is now described. Let $b_{ij}(a)$ be the $(i,j)^{th}$ inner product with the transformation (3.1). The residual after fitting r dimensions is

$$\Sigma\Sigma\{b_{ij}(a) - b^{*}_{ij}(a)\}^2 = \sum_{k=r+1}^{n-1} (\lambda_k - a)^2 , \qquad (3.3)$$

where the eigenvalues of $\underset{\sim}{B}$ under $a=0$ are $\lambda_1 \geq \ldots \geq \lambda_{n-1}$ and $\lambda_n = 0$ corresponds to the eigenvector $\underset{\sim}{1}$. Then (3.3) is minimised by

$$\hat{a}_r = \sum_{k=r+1}^{n-1} \lambda_k/(n-r-1) \qquad (3.4)$$

provided that

$$\hat{a}_r \leq \min_{i \neq j} d^2_{ij} : \qquad (3.5)$$

Let $\hat{r}$ be the smallest r for which (3.4) satisfies (3.5) and set $a = \hat{a}_{\hat{r}}$.

We applied this procedure to the colour data of Ekman (1954). First the data is transformed from similarity s to distance d^2 by $d^2 = 2(1-s)$. The modified transformation by the above method is $d^2 = 2(1-s-a)$ where $a=0.133$.

4. MISSING VALUES

We now consider the case when some elements of $\underset{\sim}{D}$ are missing. Let us assume that d_{12} is missing but configurations $P_1, P_3, \ldots, P_n$ and $Q_2, \ldots, Q_n$ can be found in p dimensions after deleting the second row and column, and the first row and first column of $\underset{\sim}{D}$ respectively. Then a rotation can be found sending Q_i to $Q_i' = P_i$, $i=3,..,n$, and $P_1, Q_2', P_3, \ldots, P_n$ is a configuration as desired. We only illustrate the method for $n>p+2$ where d_{12} can be recovered uniquely. (For $n \leq p+2$, d_{12} is not unique but can be made so by making d_{12} take its largest possible value.)

Explicitly, let $\underset{\sim}{x}_1, \underset{\sim}{x}_3, \ldots, \underset{\sim}{x}_n$ and $\underset{\sim}{y}_2, \underset{\sim}{y}_3, \ldots, \underset{\sim}{y}_n$ be the position vectors of $P_1, P_3, \ldots, P_n$ and $Q_2, Q_3, \ldots, Q_n$ respectively. Let $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ be the $(n-1)\times p$ matrices with rows $\underset{\sim}{x}_i + \frac{1}{n-2}\underset{\sim}{x}_1$, $i=3,..,n$ and $\underset{\sim}{y}_i + \frac{1}{n-2}\underset{\sim}{y}_2$, $i=3,..,n$ respectively. Then it can be shown that the required rotation $\underset{\sim}{P}$ for $\underset{\sim}{X} = \underset{\sim}{Y}\underset{\sim}{P}$ is given by

$$\underset{\sim}{P} = (\underset{\sim}{Y}'\underset{\sim}{Y})^{-1}\underset{\sim}{Y}'\underset{\sim}{X} .$$

The required configuration is then

$$\underset{\sim}{x}_1, \frac{1}{n-2}\{(n-1)\underset{\sim}{y}_2\underset{\sim}{P}-\underset{\sim}{x}_1\}, \underset{\sim}{x}_3,\ldots,\underset{\sim}{x}_n .$$

Generally, there will not be an exact representation in p-dimensions from $\underset{\sim}{D}$. In this case, one should match $P_3,\ldots,P_n$ and $Q_3,\ldots,Q_n$ from the Torgerson solution by procrustes rotation

$$\underset{\sim}{P} = (\underset{\sim}{Y}'\underset{\sim}{X})(\underset{\sim}{X}'\underset{\sim}{Y}\underset{\sim}{Y}'\underset{\sim}{X})^{-\frac{1}{2}} .$$

This work can easily be extended when more than one value is missing under certain conditions.

BIBLIOGRAPHY

Eckart, C. & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1, 211-218.

Ekman, G. (1954). Dimensions of colour vision. *J.Psychol.*, 38, 467-474.

Gower, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53, 325-338.

Gower, J.C. (1971). Statistical methods of comparing different multivariate analyses of the same data. *Mathematics in the Archaeological and Historical Sciences*. Eds: Hodson, F.R., Kendall, D.G. and Tantu, P. Edinburgh Univ.Press, 138-149.

Lingoes, J.C. (1971). Some boundary conditions for a monotone analysis of symmetric matrices. *Psychometrika*, 36, 195-203.

Schoenberg, I.J. (1935). Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espace distanciés vectoriellement applicable sur l'espace de Hilbert". *Ann.Maths.*, 36, 724-732.

Schönemann, P.H. & Carroll, R.M. (1970). Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika*, 35, 245-255.

Torgerson, W.S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, 17, 401-419.

Torgerson, W.S. (1958). *Theory and Methods of Scaling.* Wiley, New York.

Young, G. & Householder, A.S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika,* 3, 19-22.

Received July, 1978.

Refereed anonymously.