the points along this curve can be used to seriate the data. (See D. G. Kendall, 1971.)

14.2 Classical Solution

14.2.1 Some theoretical results

Definition *A distance matrix* $\mathbf{D}$ *is called* Euclidean *if there exists a configuration of points in some Euclidean space whose interpoint distances are given by* $\mathbf{D}$*; that is, if for some* p*, there exists points* $\mathbf{x}_1, \ldots, \mathbf{x}_n \in R^p$ *such that*

$$d_{rs}^2 = (\mathbf{x}_r - \mathbf{x}_s)'(\mathbf{x}_r - \mathbf{x}_s). \tag{14.2.1}$$

The following theorem enables us to tell whether $\mathbf{D}$ is Euclidean, and, if so, how to find a corresponding configuration of points. First we need some notation. For any distance matrix $\mathbf{D}$, let

$$\mathbf{A} = (a_{rs}), \qquad a_{rs} = -\tfrac{1}{2} d_{rs}^2 \tag{14.2.2}$$

and set

$$\mathbf{B} = \mathbf{HAH}, \tag{14.2.3}$$

where $\mathbf{H} = \mathbf{I} - n^{-1}\mathbf{11}'$ is the $(n \times n)$ centring matrix.

Theorem 14.2.1 *Let* $\mathbf{D}$ *be a distance matrix and define* $\mathbf{B}$ *by* (14.2.3). *Then* $\mathbf{D}$ *is Euclidean if and only if* $\mathbf{B}$ *is p.s.d. In particular, the following results hold:*

(a) *If* $\mathbf{D}$ *is the matrix of Euclidean interpoint distances for a configuration* $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)'$*, then*

$$b_{rs} = (\mathbf{z}_r - \bar{\mathbf{z}})'(\mathbf{z}_s - \bar{\mathbf{z}}), \qquad r, s = 1, \ldots, n. \tag{14.2.4}$$

In matrix form (14.2.4) *becomes* $\mathbf{B} = (\mathbf{HZ})(\mathbf{HZ})'$ *so* $\mathbf{B} \geq 0$. *Note that* $\mathbf{B}$ *can be interpreted as the "centred inner product matrix" for the configuration* $\mathbf{Z}$.

(b) *Conversely, if* $\mathbf{B}$ *is p.s.d. of rank* p *then a configuration corresponding to* $\mathbf{B}$ *can be constructed as follows. Let* $\lambda_1 > \cdots > \lambda_p$ *denote the positive eigenvalues of* $\mathbf{B}$ *with corresponding eigenvectors* $\mathbf{X} = (\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(p)})$ *normalized by*

$$\mathbf{x}_{(i)}'\mathbf{x}_{(i)} = \lambda_i, \qquad i = 1, \ldots, p. \tag{14.2.5}$$

Then the points P_r *in* R^p *with coordinates* $\mathbf{x}_r = (x_{r1}, \ldots, x_{rp})'$ *(so* $\mathbf{x}_r$ *is the* r*th row of* $\mathbf{X}$*) have interpoint distances given by* $\mathbf{D}$*. Further, this* *configuration has centre of gravity* $\bar{\mathbf{x}}=\mathbf{0}$, *and* $\mathbf{B}$ *represents the inner product matrix for this configuration.*

Proof We first prove (a). Suppose

$$d_{rs}^2=-2a_{rs}=(\mathbf{z}_r-\mathbf{z}_s)'(\mathbf{z}_r-\mathbf{z}_s). \tag{14.2.6}$$

We can write

$$\mathbf{B}=\mathbf{HAH}=\mathbf{A}-n^{-1}\mathbf{AJ}-n^{-1}\mathbf{JA}+n^{-2}\mathbf{JAJ}, \tag{14.2.7}$$

where $\mathbf{J}=\mathbf{11}'$. Now

$$\frac{1}{n}\mathbf{AJ}=\begin{bmatrix}\bar{a}_{1\cdot}&\dots&\bar{a}_{1\cdot}\\ \cdot&&\cdot\\ \cdot&&\cdot\\ \cdot&&\cdot\\ \bar{a}_{n\cdot}&\dots&\bar{a}_{n\cdot}\end{bmatrix},\qquad \frac{1}{n}\mathbf{JA}=\begin{bmatrix}\bar{a}_{\cdot1}&\dots&\bar{a}_{\cdot n}\\ \cdot&&\cdot\\ \cdot&&\cdot\\ \cdot&&\cdot\\ \bar{a}_{\cdot1}&\dots&\bar{a}_{\cdot n}\end{bmatrix},\qquad \frac{1}{n^2}\mathbf{JAJ}=\begin{bmatrix}\bar{a}_{\cdot\cdot}&\dots&\bar{a}_{\cdot\cdot}\\ \cdot&&\cdot\\ \cdot&&\cdot\\ \cdot&&\cdot\\ \bar{a}_{\cdot\cdot}&\dots&\bar{a}_{\cdot\cdot}\end{bmatrix},$$

where

$$\bar{a}_{r\cdot}=\frac{1}{n}\sum_{s=1}^{n}a_{rs},\qquad \bar{a}_{\cdot s}=\frac{1}{n}\sum_{r=1}^{n}a_{rs},\qquad \bar{a}_{\cdot\cdot}=\frac{1}{n^2}\sum_{r,s=1}^{n}a_{rs}. \tag{14.2.8}$$

Thus

$$b_{rs}=a_{rs}-\bar{a}_{r\cdot}-\bar{a}_{\cdot s}+\bar{a}_{\cdot\cdot}\,. \tag{14.2.9}$$

After substituting for a_{rs} from (14.2.6) and using (14.2.8), this formula simplifies to

$$b_{rs}=(\mathbf{z}_r-\bar{\mathbf{z}})'(\mathbf{z}_s-\bar{\mathbf{z}}). \tag{14.2.10}$$

(See Exercise 14.2.1 for further details.) Thus (a) is proved.

Conversely, to prove (b) suppose $\mathbf{B}\geq 0$ and consider the configuration given in the thorem. Let $\mathbf{\Lambda}=\operatorname{diag}(\lambda_1,\dots,\lambda_p)$ and let $\mathbf{\Gamma}=\mathbf{X\Lambda}^{-1/2}$, so that the columns of $\mathbf{\Gamma}$, $\boldsymbol{\gamma}_{(i)}=\lambda_i^{-1/2}\mathbf{x}_{(i)}$ are *standardized* eigenvectors of $\mathbf{B}$. Then by the spectral decomposition theorem (Remark 4 after Theorem A.6.4),

$$\mathbf{B}=\mathbf{\Gamma\Lambda\Gamma}'=\mathbf{XX}';$$

that is, $b_{rs} = \mathbf{x}_r'\mathbf{x}_s$, so $\mathbf{B}$ represents the inner product matrix for this configuration.

We must now show that $\mathbf{D}$ represents the matrix of interpoint distances for this configuration. Using (14.2.9) to write $\mathbf{B}$ in terms of $\mathbf{A}$, we get

$$\begin{aligned}(\mathbf{x}_r-\mathbf{x}_s)'(\mathbf{x}_r-\mathbf{x}_s) &= \mathbf{x}_r'\mathbf{x}_r - 2\mathbf{x}_r'\mathbf{x}_s + \mathbf{x}_s'\mathbf{x}_s \\ &= b_{rr} - 2b_{rs} + b_{ss} \\ &= a_{rr} - 2a_{rs} + a_{ss} \\ &= -2a_{rs} = d_{rs}^2 \qquad (14.2.11)\end{aligned}$$

because $a_{rr} = -\frac{1}{2}d_{rr}^2 = 0$ and $-2a_{rs} = d_{rs}^2$.

Finally, note that $\mathbf{B1} = \mathbf{HAH1} = \mathbf{0}$, so that $\mathbf{1}$ is an eigenvector of $\mathbf{B}$ corresponding to the eigenvalue 0. Thus $\mathbf{1}$ is orthogonal to the columns of $\mathbf{X}$, $\mathbf{x}_{(i)}'\mathbf{1} = 0$, $i = 1, \ldots, p$. Hence

$$n\bar{\mathbf{x}} = \sum_{r=1}^{n} \mathbf{x}_i = \mathbf{X}'\mathbf{1} = (\mathbf{x}_{(1)}'\mathbf{1}, \ldots, \mathbf{x}_{(p)}'\mathbf{1})' = \mathbf{0}$$

so that the centre of gravity of this configuration lies at the origin. ■

Remarks (1) The matrix $\mathbf{X}$ can be visualized in the following way in terms of the eigenvectors of $\mathbf{B}$ and the corresponding points:

$$\begin{array}{llcl}
 & & \overbrace{\lambda_1 \;\; \lambda_2 \;\;\;\;\;\; \lambda_p}^{\text{Eigenvalues}} & \text{Vector notation} \\
\text{Points} & \left\{\begin{array}{l} P_1 \\ P_2 \\ \vdots \\ P_n \end{array}\right. & \begin{array}{l} x_{11}\; x_{12} \ldots x_{1p} \\ x_{21}\; x_{22} \ldots x_{2p} \\ \;\vdots \quad\; \vdots \qquad \vdots \\ x_{n1}\; x_{n2} \ldots x_{np} \end{array} & \begin{array}{c} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{array} \\
\text{Vector notation.} & & \mathbf{x}_{(1)}\mathbf{x}_{(2)} \ldots \mathbf{x}_{(p)} &
\end{array}$$

Centre of gravity:

$$\bar{x}_1 = 0, \bar{x}_2 = 0, \ldots, \bar{x}_p = 0, \qquad \bar{\mathbf{x}} = \frac{1}{n}\sum \mathbf{x}_r = \mathbf{0}.$$

In short, the rth *row* of $\mathbf{X}$ contains the coordinates of the rth point, whereas the ith *column* of $\mathbf{X}$ contains the eigenvector corresponding to λ_i.