MSc II: Further Statistical Analysis (Autumn Term)
**Advanced Multivariate Analysis**

# 3 Procrustes Rotation

Comparison of Two or More Graphical Representations

It is frequently the case that in the course of a multivariate investigation of the structure in a set of data, a number of configurations are obtained of the same set of sample individuals represented by coordinates in a low dimensional Euclidean space. Comparisons between the different pictures may be required.

For example, where a dissimilarity matrix has been calculated using a particular dissimilarity measure, a number of different such measures might have been calculated from these data. Moreover a different set, of say, quantitative measurements might have been made and inter-sample distances calculated from these. Each possible dissimilarity matrix might have been subjected to both a classical and an ordinal scaling analysis. It is not difficult to see how perhaps 6 or more coordinate representations might be obtained.

We need an analytical technique that will measure the extent to which two or more such graphical representations differ.

Comparison of two $n$-point configurations

Suppose that, for example, both classical and ordinal scaling have been carried out on the same dissimilarity matrix and the two corresponding $(p-1)$-dimensional configurations $X$ and $Y$ of the $n$ points have been obtained.

$$X \quad (n \times (p-1)) \qquad \text{and} \qquad Y \quad (n \times (p-1))$$

Since the internal relationships between points in any Euclidean representation remain the same under, translation, orthogonal rotation, reflection and dilation, we consider one configuration, say $X$, as fixed and modify $Y$ using these operations so as to match $X$ as closely as possible (in the least squares sense).

Since $X$ and $Y$ have origin at the centroid we do not need to consider translation. Also, if we extend $X$ and $Y$ to $p$ dimensions (adding a $p$'th coordinate 0) then we do not need to consider reflection explicitly either, since it may be achieved by a rotation in $p$ dimensions.

The 'problem' may be formulated as

$$\text{Minimize } R^2 = \text{trace}\{(X - Y^*)(X - Y^*)^T\} \quad \text{w.r.t. } Y^*$$

where $Y^*$ is obtained from $Y$ by orthogonal rotation and then dilation.

*Note*: this corresponds to a sensible least squares measure of the degree of coincidence between the two configurations $X$ and $Y^*$, via

$$\sum_{i=1}^{n}\left\{\sum_{j=1}^{p}(x_{ij}-y_{ij}^*)^2\right\} = \sum_{i=1}^{n}(\mathbf{x}_i-\mathbf{y}_i^*)^T(\mathbf{x}_i-\mathbf{y}_i^*) = \mathrm{tr}\left\{(X-Y^*)(X-Y^*)^T\right\}$$

<u>Solution</u>: The 'best' orthogonal rotation of $Y$ relative to $X$ is $YQ$ where $Q$ $(p \times p)$ is given by $Q = UV^T$. (This is <u>Procrustes Rotation</u>)

$U$ and $V$ are obtained from the singular value decomposition (SVD) of $Y^TX$, i.e. $Y^TX = U\Gamma V^T$ - (see Appendix, *details <u>not</u> examinable).*

Following rotation, the solution is completed by setting

$$Y^* = cYQ \qquad \text{with } c = \frac{\mathrm{tr}(\Gamma)}{\mathrm{tr}(YY^T)}$$

(This is <u>Procrustes rotation with scaling</u>).

$$R^2_{\min} = \mathrm{tr}(XX^T) + c^2\mathrm{tr}(YY^T) - 2c\mathrm{tr}(\Gamma)$$

<u>Note</u> the procedure is not symmetric with respect to $X$ and $Y$ unless we set

$$\mathrm{tr}(XX^T) = \mathrm{tr}(YY^T) \qquad (= 1, \text{ usual standardisation})$$

This gives $R^2_{\min} = 1 - \{\mathrm{tr}(\Gamma)\}^2$.

<u>Simultaneous Comparison of $g$ $n$-point Configurations (Generalized Procrustes Analysis)</u>
[See Krzanowski (Section 5.2)].

The idea of Procrustes Analysis can be generalized to allow a <u>single</u> analysis in which all $g$ configurations are simultaneously translated, rotated, reflected and scaled so that a goodness-of-fit criterion is optimized. It is then possible to investigate sources of difference between the $g$ configurations. The underlying theory and algebraic development are much more complicated than in the 2 configuration case and the optimization has to be carried out in a numerical iterative procedure.

The main reference is

Gower, J. (1975) Generalzed Procrustes Analysis. Psychometrika, **40**, 33-51.

# Appendix: The Singular Value Decomposition (SVD) of a Rectangular Matrix

Let $X$ be an arbitrary $m \times n$ matrix of rank $r$. Then, $X$ can be written as

$$X = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \ldots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T \tag{1}$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$. The numbers $\sigma_i$ (the singular values) are real and positive, the $m \times 1$ column vectors $\mathbf{u}_i$ are orthonormal and the $n \times 1$ column vectors $\mathbf{v}_i$ are orthonormal. The orthonormal sets $\mathbf{u}_1, \ldots, \mathbf{u}_r$ and $\mathbf{v}_1, \ldots, \mathbf{v}_r$ can be completed to sets:

$$\mathbf{u}_1, \ldots, \mathbf{u}_r, \ldots, \mathbf{u}_m \quad \text{and,} \quad \mathbf{v}_1, \ldots, \mathbf{v}_r, \ldots, \mathbf{v}_n$$

although this completion is not unique if $r + 2 \leq \max(m, n)$. Let $k = \min(m, n)$. Then a complete singular value decomposition of $X$ is given by

$$X = \sum_{j=1}^{k} \sigma_j \mathbf{u}_j \mathbf{v}_j^T$$

where $\sigma_j = 0$ for $j = r + 1, \ldots, k$ if $r < k$.

Thus if we set

$$U = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m)$$
$$V = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$$

and,

$$\Gamma = (\sigma_{ij}) \quad \text{where} \quad \begin{cases} \sigma_{jj} = \sigma_j, & \text{j=1, ..., k;} \\ \sigma_{ij} = 0, & \text{i}\neq\text{j.} \end{cases}$$

Then,

$$X = U \Gamma V^T \tag{2}$$

where,

$$U^T U = U U^T = \mathbf{I} \tag{3}$$
$$V^T V = V V^T = \mathbf{I} \tag{4}$$

(i.e. $U$ and $V$ are orthogonal matrices).

**Note 1**: Expressing $X$ in the form

$$\begin{aligned} X &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \ldots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T \\ &= (\mathbf{u}_1, \ldots, \mathbf{u}_k)\ diag(\sigma_i) \begin{pmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_k^T \end{pmatrix} \\ &= U^*\ diag(\sigma_i) V^{*T} \end{aligned} \tag{5}$$

we have,

$$\begin{aligned} XX^T &= U^*\ diag(\sigma_i) V^{*T} V^*\ diag(\sigma_i) U^{*T} \\ &= U^*\ diag(\sigma_i^2) U^{*T} \end{aligned} \tag{6}$$

And, similarly

$$X^T X = V^*\ diag(\sigma_i^2) V^{*T} \tag{7}$$

That is, it is the case that the $\mathbf{u}_1, \ldots, \mathbf{u}_r$ are the eigenvectors of $XX^T$ corresponding to the $r$ positive eigenvalues $\sigma_1^2 \geq \ldots \geq \sigma_r^2 > 0$, and the $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are the eigenvectors of $X^T X$ corresponding to the same eigenvalues.

Clearly, $\mathbf{u}_i \propto X\mathbf{v}_i$ and is easily shown that $\mathbf{u}_i^T \mathbf{u}_i = 1$ requires that $\mathbf{u}_i = \dfrac{1}{\sigma_i} X \mathbf{v}_i$.

Similarly, $\mathbf{v}_i = \dfrac{1}{\sigma_i} X^T \mathbf{u}_i$.

**Note 2**: <u>The Eckart-Young Theorem</u>

If $X$ is of rank $r$, having the SVD expressed in the form (1), i.e.

$$X = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \ldots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

then the matrix $Y$ of rank $k < r$ that approximates $X$ most closely in the least squares sense is

$$Y = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \ldots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T \tag{8}$$

That is, (8) minimises

$$S(Y) = \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - y_{ij})^2 = \operatorname{tr}\left\{(X - Y)(X - Y)^T\right\}$$

**Note 3**: If $X$ is a $p \times p$ square symmetric matrix, then $\mathbf{u}_i = \mathbf{v}_i$, $i = 1, \ldots, p$ and are the eigenvectors of $X$ while the $\sigma_i, i = 1, \ldots, p$ are the corresponding eigenvalues. We then have the Spectral Decomposition Theorem.