Procrustes Alignment in N dimensions

Consider two corresponding sets of N coordinates ordered as the columns of the matrices $\mathbf{X}$ and $\mathbf{Y}$. In a 2-dimensional space, these sets are written as –

$$\mathbf{X}=\begin{bmatrix} x_1 & x_2 & \cdots & \cdots & x_N \\ y_1 & y_2 & \cdots & \cdots & y_N \end{bmatrix} \quad and \quad \mathbf{Y}=\begin{bmatrix} x'_1 & x'_2 & \cdots & \cdots & x'_N \\ y'_1 & y'_2 & \cdots & \cdots & y'_N \end{bmatrix} \qquad [1]$$

If we generalise to a P-dimensional space with axes $x_1, x_2, \cdots x_P$, the P x N coordinate matrices may be written as –

$$\mathbf{X}=\begin{bmatrix} x_1\;1 & x_1\;2 & \cdots & x_1\;N-1 & x_1\;N \\ x_2\;1 & x_2\;2 & \cdots & x_2\;N-1 & x_2\;N \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ x_{P-1}\;1 & x_{P-1}\;2 & \cdots & x_{P-1}\;N-1 & x_{P-1}\;N \\ x_P\;1 & x_P\;2 & \cdots & x_P\;N-1 & x_P\;N \end{bmatrix} = \begin{bmatrix} \overrightarrow{\mathbf{x}_1} & \overrightarrow{\mathbf{x}_2} & \cdots & \overrightarrow{\mathbf{x}_{N-1}} & \overrightarrow{\mathbf{x}_N} \end{bmatrix}$$

and

$$\mathbf{Y}=\begin{bmatrix} x'_1\;1 & x'_1\;2 & \cdots & x'_1\;N-1 & x'_1\;N \\ x'_2\;1 & x'_2\;2 & \cdots & x'_2\;N-1 & x'_2\;N \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ x'_{P-1}\;1 & x'_{P-1}\;2 & \cdots & x'_{P-1}\;N-1 & x'_{P-1}\;N \\ x'_P\;1 & x'_P\;(2) & \cdots & x'_P\;(N-1) & x'_P\;(N) \end{bmatrix} = \begin{bmatrix} \overrightarrow{\mathbf{x}'_1} & \overrightarrow{\mathbf{x}'_2} & \cdots & \overrightarrow{\mathbf{x}'_{N-1}} & \overrightarrow{\mathbf{x}'_N} \end{bmatrix}$$

[2]

where $\overrightarrow{\mathbf{x}_i}$ and $\overrightarrow{\mathbf{x}'_i}$ are the ith points (and thus ith column vectors) of matrices **X** and **Y** respectively.

The aim is to transform (i.e. align) the <u>input</u> coordinates **X** to the <u>reference</u> coordinates **Y** so as to minimise <u>the total sum of the squared Euclidean distances between the corresponding points</u>. The three transformations of translation, scaling and rotation are succesively applied, each independently satisfying this criterion.

Step 1 – Translation

The aim is to find that global translation of the coordinates in matrix $\mathbf{X}$ which minimises the total sum of the squared Euclidean distances between the translated coordinates and their corresponding values in the reference matrix $\mathbf{Y}$.

Mathematically, the criterion is thus to minimise a cost function Q defined as –

$$Q = \sum_{i=1}^{N} \left[\vec{\mathbf{x}}_i + \vec{\mathbf{t}} - \vec{\mathbf{x}}'_i \right]^{\mathbf{T}} \left[\vec{\mathbf{x}}_i + \vec{\mathbf{t}} - \vec{\mathbf{x}}'_i \right] \quad [3]$$

where $\vec{\mathbf{t}}$ is the translation vector applied.

Setting the derivative of Q with respect to $\vec{\mathbf{t}}$ to zero, we obtain –

$$\frac{\partial Q}{\partial \vec{\mathbf{t}}} = 2\sum_{i=1}^{N} \left[\vec{\mathbf{x}}_i + \vec{\mathbf{t}} - \vec{\mathbf{x}}'_i \right] = 0 \quad \Rightarrow \quad \vec{\mathbf{t}} = \frac{1}{N}\sum_{i=1}^{N} \left[\vec{\mathbf{x}}'_i - \vec{\mathbf{x}}_i \right]$$
$$\vec{\mathbf{t}} = \langle \vec{\mathbf{y}} \rangle - \langle \vec{\mathbf{x}} \rangle \quad [4]$$

where $\langle \vec{\mathbf{x}} \rangle$ and $\langle \vec{\mathbf{y}} \rangle$ are the average (centroid) coordinate vectors of the respective sets of N points.

Normally, we refer the reference coordinates (matrix **Y**) to the origin in which case $\langle \vec{\mathbf{y}} \rangle = 0$. With this condition, the first step in the Procrustes alignment simply subtracts the sample mean value from each of the coordinates and the translated coordinates are described by $\vec{\mathbf{x}}_i - \langle \vec{\mathbf{x}} \rangle$.

Step 2 – Scaling

A uniform scaling of all the ordinates in **X** can be achieved by a *diagonal* matrix $\mathbf{S} = s\mathbf{I}$ where s is the scaling parameter and **I** is the identity matrix. The aim of the scaling is to identify that value of the parameter s which scales the coordinates in matrix $\mathbf{X}$ so as to minimise the total sum of the squared Euclidean distances between the scaled coordinates and their corresponding values in the reference matrix $\mathbf{Y}$.

We must minimise the least-squares cost function -

$$Q = \sum_{i=1}^{N} \left[\mathbf{S}\vec{\mathbf{x}}_i - \vec{\mathbf{x}}'_i \right]^{\mathbf{T}} \left[\mathbf{S}\vec{\mathbf{x}}_i - \vec{\mathbf{x}}'_i \right] \quad [5]$$

where $\mathbf{S} = s\mathbf{I}$.

Differentiating with respect to the parameter s and setting to zero, we have -

$$\frac{\partial Q}{\partial s} = \frac{\partial}{\partial s}\left\{\sum_{i=1}^{N} \overrightarrow{\mathbf{x}_i}^T \mathbf{S}^T \mathbf{S}\overrightarrow{\mathbf{x}_i} - \overrightarrow{\mathbf{x}_i}^T \mathbf{S}^T \overrightarrow{\mathbf{x}_i'} - \overrightarrow{\mathbf{x}_i'}^T \mathbf{S}\overrightarrow{\mathbf{x}_i} + \overrightarrow{\mathbf{x}_i'}^T \overrightarrow{\mathbf{x}_i'}\right\} = 0$$

$$\frac{\partial Q}{\partial s} = \frac{\partial}{\partial s}\left\{\sum_{i=1}^{N} s^2 \overrightarrow{\mathbf{x}_i}^T \overrightarrow{\mathbf{x}_i} - s\overrightarrow{\mathbf{x}_i}^T \overrightarrow{\mathbf{x}_i'} - s\overrightarrow{\mathbf{x}_i'}^T \overrightarrow{\mathbf{x}_i} + \overrightarrow{\mathbf{x}_i'}^T \overrightarrow{\mathbf{x}_i'}\right\} = 0$$

$$\frac{\partial Q}{\partial s} = \sum_{i=1}^{N} 2s\overrightarrow{\mathbf{x}_i}^T \overrightarrow{\mathbf{x}_i} - \overrightarrow{\mathbf{x}_i}^T \overrightarrow{\mathbf{x}_i'} - \overrightarrow{\mathbf{x}_i'}^T \overrightarrow{\mathbf{x}_i} = 0$$

and since $\overrightarrow{\mathbf{x}_i}^T \overrightarrow{\mathbf{x}_i'} = \overrightarrow{\mathbf{x}_i'}^T \overrightarrow{\mathbf{x}_i}$ [1], we obtain –

$$s = \frac{\sum_{i=1}^{N} \overrightarrow{\mathbf{x}_i'}^T \overrightarrow{\mathbf{x}_i}}{\sum_{i=1}^{N} \overrightarrow{\mathbf{x}_i}^T \overrightarrow{\mathbf{x}_i}} \qquad [6]$$

Step 3 – Rotation

The aim of the rotation is to identify an orthogonal matrix **R** which acts on the coordinate matrix $\mathbf{X}$ so as to minimise the total sum of the squared Euclidean distances between the scaled coordinates and their corresponding values in the reference matrix $\mathbf{Y}$.

To define this criterion mathematically, we thus define an error matrix $\mathbf{E} = \mathbf{Y} - \mathbf{RX}$ and seek that matrix **R** which minimises the cost Q given by –

$$Q = Tr\{\mathbf{E}^T\mathbf{E}\} = Tr\ (\mathbf{Y} - \mathbf{RX})^{\mathbf{T}}(\mathbf{Y} - \mathbf{RX}) \qquad [7]$$

where Tr denotes the trace operator (the sum of the diagonal elements of a square matrix).

(That eq. [7] does indeed define the sum of the squared Euclidean distances between the points can be seen by writing the reference and rotated coordinates explicitly in the tableau form of eq.2 and considering the diagonal elements of the matrix product.)

Since $Tr\ \mathbf{A} + \mathbf{B}\ = Tr\ \mathbf{A}\ + Tr\ \mathbf{B}$, we expand eq. [7] to obtain –

$$\begin{aligned} Q &= Tr\ \mathbf{Y}^T\mathbf{Y} - \mathbf{X}^T\mathbf{R}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{RX} + \mathbf{X}^T\mathbf{R}^T\mathbf{RX} \\ &= Tr\ \mathbf{Y}^T\mathbf{Y}\ - Tr\ \mathbf{X}^T\mathbf{R}^T\mathbf{Y}\ - Tr\ \mathbf{Y}^T\mathbf{RX}\ + Tr\ \mathbf{X}^T\mathbf{R}^T\mathbf{RX} \end{aligned} \qquad [8]$$

Since $\mathbf{X}^T\mathbf{R}^T\mathbf{Y} = \left[\mathbf{Y}^T\mathbf{RX}\right]^T$ [2], the trace of these two matrices is equal and since **R** is an orthogonal matrix $\mathbf{R}^T\mathbf{R} = \mathbf{RR}^T = \mathbf{I}$ the identity matrix, eq. [8] simplifies to –

[1] It is a <u>scalar</u> quantity and the transpose of a scalar must equal itself.

$$Q = Tr(\mathbf{Y}^T\mathbf{Y}) + Tr(\mathbf{X}^T\mathbf{X}) - 2Tr(\mathbf{X}^T\mathbf{R}^T\mathbf{Y}) \qquad [9]$$

Recall that we must choose matrix **R** in eq. [8] so as to *minimise* Q. Since the first two terms do not depend on **R**, we equivalently seek that orthogonal matrix **R** that **maximises** the quantity $Tr(\mathbf{X}^T\mathbf{R}^T\mathbf{Y})$ (since this will automatically minimise Q). Thus –

$$\min\{Q\} = \max\{Tr(\mathbf{X}^T\mathbf{R}^T\mathbf{Y})\} \qquad [10]$$

The cyclic property of the trace operator states that –

$$Q = Tr(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_k) = Tr(\mathbf{A}_k\mathbf{A}_2\cdots\mathbf{A}_1) = Tr(\mathbf{A}_2\cdots\mathbf{A}_k\mathbf{A}_1) \qquad [11]$$

Applying the cyclic property of trace and the matrix identity $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ to eq. [10], we may write –

$$Tr(\mathbf{X}^T\mathbf{R}^T\mathbf{Y}) = Tr(\mathbf{YX}^T\mathbf{R}^T) = Tr(\mathbf{RXY}^T)$$

We may take the singular value decomposition (SVD) of the matrix $\mathbf{XY}^T = \mathbf{USV}^T$ where **U** and **V** are orthogonal matrices of eigenvectors and **S** is a diagonal matrix of singular values. Replacing $\mathbf{XY}^T$ by its SVD and again applying the cyclic property of the trace operator, we obtain –

$$Tr(\mathbf{R}(\mathbf{XY}^T)) = Tr(\mathbf{RUSV}^T) = Tr(\mathbf{SV}^T\mathbf{RU}) = Tr(\mathbf{SH}) \qquad [12]$$

where we note that $\mathbf{H} = \mathbf{V}^T\mathbf{RU}$, a product of orthogonal matrices, must therefore itself be orthogonal.

Thus, we seek

$$\min\{Q\} = \max\{Tr(\mathbf{SH})\} \qquad [13]$$

Writing the matrix product in tableau form –

---

[2] Matrix identity $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

$$\mathbf{SH}=\begin{bmatrix} s_1 & 0 & \cdots & 0 & 0 \\ 0 & s_1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & s_{P-1} & 0 \\ 0 & 0 & \cdots & 0 & s_P \end{bmatrix}\begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1P-1} & h_{1P} \\ h_{21} & h_{22} & \cdots & h_{2P-1} & h_{2P} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ h_{P-11} & h_{P-12} & \cdots & h_{P-1P-1} & h_{11} \\ h_{P1} & h_{P2} & \cdots & h_{PP-1} & h_{PP} \end{bmatrix} \qquad [14]$$

it is readily apparent that

$$\min\ Q = \max\ Tr\ \mathbf{SH} = \max\left\{\sum_{k=1}^{P} h_{kk}s_k\right\} \qquad [15]$$

Now, since **H** is an orthogonal matrix and the Euclidean length of the ith column vectors is required to be unity ($\sum_{k=1}^{P} h_{ik}{}^2 = 1$), we obtain a maximum in eq.[15] if $h_{kk} = 1$ for all k – i.e. if $\mathbf{H}=\mathbf{I}$.

Thus the condition for Q to be minimised as required is that –

$$\mathbf{H}=\mathbf{V}^T\mathbf{RU}=\mathbf{I} \qquad \Rightarrow \qquad \mathbf{R}=\mathbf{VU}^T \qquad [16]$$

where **U** and **V** are obtained from the SVD of $\mathbf{XY}^T$ - ie from $\mathbf{XY}^T = \mathbf{USV}^T$.

In summary, the solution steps for all dimensionalities of the problem are –

- **Translation:** Place the origin at the centroid of your reference coordinates (that set of points to which you wish to align) and translate the input coordinates (the points you wish to align) to this origin by subtracting the centroids.
- **Scaling:** Scale the input coordinates using eq. [6].
  Rotation:
- We form the product of the coordinate matrices $\mathbf{XY}^T$.
- We calculate its singular value decomposition (SVD) as $\mathbf{XY}^T = \mathbf{USV}^T$
- The matrix which minimises the total squared Euclidean error is $\mathbf{R}=\mathbf{VU}^T$

Finally, note that we can obtain the solution for the scaling part of Procrustes using the same mathematical approach of *minimising the trace of a cost matrix*. In this case,

Define an error matrix $\mathbf{E}=\mathbf{Y}-\mathbf{SX}$ and seek that matrix $\mathbf{S}=s\mathbf{I}$ which minimises the cost Q given by –

$$Q = Tr\ \mathbf{E}^T\mathbf{E} = Tr\ \ \mathbf{Y}-\mathbf{SX}\ ^T\ \mathbf{Y}-\mathbf{SX}$$

The solution follows in an analogous way to the analysis above as –

$$s = \frac{Tr\ \mathbf{X}^T\mathbf{Y}}{Tr\ \mathbf{X}^T\mathbf{X}} \qquad [17]$$