20
Procrustes Procedures

The Procrustes problem is concerned with fitting a configuration (testee) to another (target) as closely as possible. In the simplest case, both configurations have the same dimensionality and the same number of points, which can be brought into a 1–1 correspondence by substantive considerations. Under orthogonal transformations, the testee can be rotated and reflected arbitrarily in an effort to fit it to the target. In addition to such rigid motions, one may also allow for dilations and for shifts. In the oblique case, the testee can also be distorted linearly. Further generalizations include an incompletely specified target configuration, different dimensionalities of the configurations, and different numbers of points in both configurations.

20.1 The Problem

We now consider a problem that arose repeatedly throughout the text. In Figure 2.14, using rotations, reflections, and dilations, we found it possible to match two configurations almost perfectly. Without these transformations, it would have been difficult to see that ratio and ordinal MDS led to virtually the same configurations. If the dimensionality of two configurations is higher than 2D, such comparisons become even more difficult or, indeed, impossible. Therefore, one needs procedures that eliminate *meaningless* differences as much as possible by transforming one configuration (*testee*) by a set of *admissible* transformations so that it most closely approximates a given *target* configuration. Such fitting problems are known as *Procrustes problems* (Hurley & Cattell, 1962).[1]

In geometry, two figures (configurations) are called *similar* if they can be brought to a complete match by rigid motions and dilations. These transformations are admissible for all MDS solutions up to ratio MDS, so we can freely exploit similarity transformations to facilitate comparisons of different MDS configurations. Before considering similarity transformations, however, we first consider a restricted Procrustes problem, the orthogonal Procrustes. Once this problem is solved, it can be easily extended to cover the similarity case.

20.2 Solving the Orthogonal Procrustean Problem

Let $\mathbf{A}$ be the target configuration and $\mathbf{B}$ the corresponding testee. Assume that $\mathbf{A}$ and $\mathbf{B}$ are both of order $n \times m$. We now want to fit $\mathbf{B}$ to $\mathbf{A}$ by rigid motions. That is, we want $\mathbf{A} \approx \mathbf{BT}$ by picking a best-possible matrix $\mathbf{T}$ out of the set of all orthogonal $\mathbf{T}$. Geometrically, $\mathbf{T}$ therefore is restricted to rotations and reflections.

Without the restriction $\mathbf{TT}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$, $\mathbf{T}$ could be *any* matrix, which means, geometrically, that $\mathbf{T}$ is some *linear transformation*. Such transformations, however, do not, in general, preserve $\mathbf{B}$'s "shape". Rather, linear transformations can cause shears, stretch $\mathbf{B}$ differentially along some directions, or collapse its dimensionality (see, e.g., Green & Carroll, 1976). Such transformations are clearly inadmissible ones, because they generally change the ratios of the distances among $\mathbf{B}$'s points and, thus, affect the fit of these distances to the data. For the moment, we are not interested in such transformations.

As for the $\approx$ criterion, a reasonable definition would be to measure the distances between corresponding points, square these values, and add them to obtain the sum-of-squares criterion L. The transformation $\mathbf{T}$ should be chosen to minimize this L. Expressed in matrix notation, the differences of the coordinates of $\mathbf{A}$ and $\mathbf{BT}$ are given by $\mathbf{A} - \mathbf{BT}$. We want to minimize the sum of the squared error, that is,

$$L(\mathbf{T}) = \text{tr } (\mathbf{A} - \mathbf{BT})'(\mathbf{A} - \mathbf{BT}) \tag{20.1}$$

or, equivalently,

$$L(\mathbf{T}) = \text{tr } (\mathbf{A} - \mathbf{BT})(\mathbf{A} - \mathbf{BT})',$$

[1] Procrustes was an innkeeper in Greek mythology who "fitted" his guests to his beds by stretching them or by chopping off their legs. The terminology "Procrustes problem" is now standard, even though it is generally misleading, inasmuch we do *not* want to mutilate or distort the testee configuration.

as explained in Table 7.4. In other words, $L(\mathbf{T})$ measures the squared distances of the points of $\mathbf{A}$ and the corresponding points of $\mathbf{BT}$.

Expanding (20.1), we get

$$\begin{aligned} L(\mathbf{T}) &= \text{tr } (\mathbf{A} - \mathbf{BT})'(\mathbf{A} - \mathbf{BT}) \\ &= \text{tr } \mathbf{A}'\mathbf{A} + \text{tr } \mathbf{T}'\mathbf{B}'\mathbf{BT} - 2\text{tr } \mathbf{A}'\mathbf{BT} \\ &= \text{tr } \mathbf{A}'\mathbf{A} + \text{tr } \mathbf{B}'\mathbf{B} - 2\text{tr } \mathbf{A}'\mathbf{BT} \end{aligned}$$

over $\mathbf{T}$ subject to $\mathbf{T}'\mathbf{T} = \mathbf{TT}' = \mathbf{I}$. Note that the simplification tr $\mathbf{T}'\mathbf{B}'\mathbf{BT} =$ tr $\mathbf{B}'\mathbf{B}$ is obtained by using the property of invariance of the trace function under cyclic permutation (see Table 7.4, property 3), which implies tr $\mathbf{T}'\mathbf{B}'\mathbf{BT} =$ tr $\mathbf{B}'\mathbf{BTT}'$, and using $\mathbf{T}'\mathbf{T} = \mathbf{TT}' = \mathbf{I}$, so that tr $\mathbf{B}'\mathbf{BTT}' =$ tr $\mathbf{B}'\mathbf{B}$. Because tr $\mathbf{A}'\mathbf{A}$ and tr $\mathbf{B}'\mathbf{B}$ are not dependent on $\mathbf{T}$, minimizing $L(\mathbf{T})$ is equivalent to minimizing

$$L(\mathbf{T}) = c - 2\text{tr } \mathbf{A}'\mathbf{BT} \tag{20.2}$$

over $\mathbf{T}$ subject to $\mathbf{T}'\mathbf{T} = \mathbf{I}$, where c is a constant that is not dependent on $\mathbf{T}$.

Minimization of $L(\mathbf{T})$ can be accomplished by applying the concept of an attainable lower bound (Ten Berge, 1993).[2] Suppose that we can find an inequality that tells us that $L(\mathbf{T}) \geq h$ and also gives the condition under which $L(\mathbf{T}) = h$. Solving $L(\mathbf{T}) = h$ for $\mathbf{T}$ (subject to the appropriate constraints) automatically gives us the smallest possible value of $L(\mathbf{T})$ and hence the global minimum.

To apply this notion to the problem in (20.2), let us first consider a lower bound inequality derived by Kristof (1970). If $\mathbf{Y}$ is a *diagonal* matrix with nonnegative entries, and $\mathbf{R}$ is orthogonal, Kristof's inequality states that

$$-\text{tr } \mathbf{RY} \geq -\text{tr } \mathbf{Y}, \tag{20.3}$$

with equality if and only if $\mathbf{R} = \mathbf{I}$.

To prove this theorem, note that because $\mathbf{Y}$ is diagonal, we may express (20.3) as

$$-\text{tr } \mathbf{RY} = -\sum_i r_{ii} y_{ii} \geq -\sum_i y_{ii}.$$

Now, because $\mathbf{RR}' = \mathbf{R}'\mathbf{R} = \mathbf{I}$, it holds for each column j of $\mathbf{R}$ that $\mathbf{r}_j'\mathbf{r}_j = \sum_i r_{ij}^2 = 1$, so that $-1 \leq r_{ii} \leq 1$. Thus, $-r_{ii}y_{ii} \geq -y_{ii}$. Obviously, only if $r_{ii} = 1$ or, in matrix terms, only if $\mathbf{R} = \mathbf{I}$, then inequality (20.3) is an equality.

---

[2] The orthogonal Procrustes problem was first solved by Green (1952) and later simultaneously by Cliff (1966) and Schönemann (1966). Their solutions are, however, somewhat less easy to understand and to compute.

We can use this theorem to find $L(\mathbf{T})$ as follows. Let $\mathbf{P\Phi Q}'$ be the singular value decomposition of $\mathbf{A}'\mathbf{B}$, where $\mathbf{P}'\mathbf{P} = \mathbf{I}$, $\mathbf{Q}'\mathbf{Q} = \mathbf{I}$, and $\mathbf{\Phi}$ is the *diagonal* matrix with the singular values. Using the invariance of the trace function under cyclic permutation (see Table 8.3)

$$\begin{aligned} L(\mathbf{T}) &= c - 2\text{tr } \mathbf{A}'\mathbf{BT} = c - 2\text{tr } \mathbf{P\Phi Q}'\mathbf{T} \\ &= c - 2\text{tr } \mathbf{Q}'\mathbf{TP\Phi} \\ &\geq c - 2\text{tr } \mathbf{\Phi}. \end{aligned}$$

Because $\mathbf{T}$ is orthonormal, so is $\mathbf{Q}'\mathbf{TP}$. Now the minimization of $L(\mathbf{T})$ is written in the form of (20.3) with $\mathbf{R} = \mathbf{Q}'\mathbf{TP}$ and $\mathbf{Y} = \mathbf{\Phi}$. We know that $L(\mathbf{T})$ is minimal if $\mathbf{R} = \mathbf{I}$ or, equivalently, $\mathbf{Q}'\mathbf{TP} = \mathbf{I}$. Hence, we have to choose $\mathbf{T}$ as

$$\mathbf{T} = \mathbf{QP}', \tag{20.4}$$

because substitution of (20.4) in $\mathbf{Q}'\mathbf{TP}$ yields $\mathbf{Q}'\mathbf{QP}'\mathbf{P} = \mathbf{I}$, so that $L(\mathbf{T}) = c - 2\text{tr } \mathbf{\Phi}$.

20.3 Examples for Orthogonal Procrustean Transformations

We now consider a simple artificial case where $\mathbf{T}$ can be computed by hand. In Figure 20.1, two vector configurations, $\mathbf{A}$ and $\mathbf{B}$, are shown. Their points are connected to form rectangles. If panels 1 and 2 of Figure 20.1 are superimposed (panel 3), then $L(\mathbf{T})$ is equal to the sum of the squared lengths of the dashed-line segments that connect corresponding points of $\mathbf{A}$ and $\mathbf{B}$. Computing $\mathbf{T}$ as discussed above, we find

$$\mathbf{T} = \begin{pmatrix} -.866 & -.500 \\ -.500 & .866 \end{pmatrix}.$$

What does $\mathbf{T}$ do to $\mathbf{B}$? From Figure 20.1, we see that $\mathbf{T}$ should first reflect $\mathbf{B}$ along the horizontal axis (or, reflect it on the vertical axis) and then rotate it by 30° counterclockwise. The reflection matrix is thus

$$\mathbf{U}_1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

and the rotation matrix by 30° is

$$\mathbf{R}_1 = \begin{pmatrix} \cos 30^\circ & \sin 30^\circ \\ -\sin 30^\circ & \cos 30^\circ \end{pmatrix} = \begin{pmatrix} .866 & .500 \\ -.500 & .866 \end{pmatrix}.$$

Applying $\mathbf{U}_1$ first and $\mathbf{R}_1$ afterwards yields $\mathbf{U}_1\mathbf{R}_1 = \mathbf{T}$ and $\mathbf{BT} = \mathbf{BU}_1\mathbf{R}_1$. But the decomposition of $\mathbf{T}$ into $\mathbf{U}_1$ and $\mathbf{R}_1$ is not unique. This may be

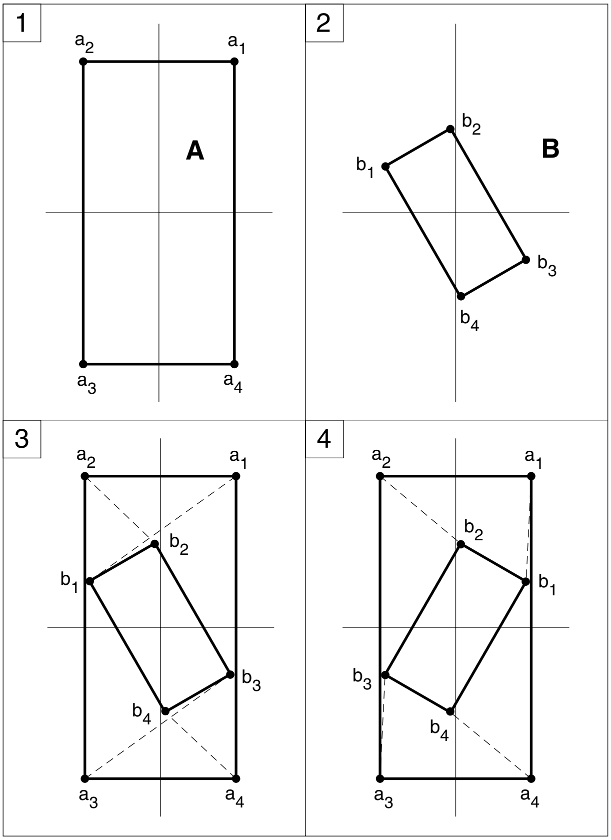

FIGURE 20.1. Illustration of some steps involved in fitting **B** to **A** by an orthogonal transformation.

more evident geometrically: in order to transform $\mathbf{B}$ into $\mathbf{BT}$, it would also be possible to first rotate $\mathbf{B}$ by $-30°$ (i.e., clockwise by $30°$) and then reflect it horizontally. This reverses the order of rotation and reflection but leads to the same result. Another possibility would be to reflect $\mathbf{B}$ vertically and then turn it by $210°$. To see that this produces the same effect, we simply find the corresponding reflection and rotation matrices,

$$\mathbf{U}_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

$$\mathbf{R}_2 = \begin{pmatrix} \cos 210° & \sin 210° \\ -\sin 210° & \cos 210° \end{pmatrix} = \begin{pmatrix} -.866 & -.500 \\ .500 & -.866 \end{pmatrix},$$

which yield $\mathbf{T} = \mathbf{U}_2\mathbf{R}_2 = \mathbf{U}_1\mathbf{R}_1$. Thus, $\mathbf{T}$ can be interpreted in different ways.

20.4 Procrustean Similarity Transformations

We now return to our original Procrustean problem of fitting one MDS configuration (testee) to another (target) MDS configuration. Because the overall size and the origin of MDS configurations are irrelevant, we now attempt to optimally exploit these additional transformations in fitting the testee matrix to the target. That is, we now extend the rotation/reflection task by finding an optimal dilation factor and an optimal translation for $\mathbf{B}$ (Schönemann & Carroll, 1970). In the context of Figure 20.1, this means that $\mathbf{BT}$ should also be scaled to the size of $\mathbf{A}$, so that the corresponding points are *incident*, i.e., lie on top of each other. The translation generalizes the fitting problem so that it can be used for distance representations where there is no fixed origin.

Consider now the example in Figure 20.2, where $\mathbf{Y}$ is derived from $\mathbf{X}$ by reflecting it horizontally, then rotating it by $30°$, shrinking it by $s = 1/2$, and finally shifting it by the translation vector $\mathbf{t}' = (1.00, 2.00)$. Formally, $\mathbf{Y} = s\mathbf{XT} + \mathbf{1t}'$, where $\mathbf{T}$ is the rotation/reflection matrix and $\mathbf{1}$ is a vector of 1s. Given the coordinate matrices

$$\mathbf{X} = \begin{pmatrix} 1 & 2 \\ -1 & 2 \\ -1 & -2 \\ 1 & -2 \end{pmatrix} \text{ and } \mathbf{Y} = \begin{pmatrix} 0.07 & 2.62 \\ 0.93 & 3.12 \\ 1.93 & 1.38 \\ 1.07 & 0.88 \end{pmatrix},$$

we want to find s, $\mathbf{T}$, and $\mathbf{t}$ that transform $\mathbf{Y}$ back to $\mathbf{X}$. In this case, we know the solutions: because $\mathbf{Y} = s\mathbf{XT} + \mathbf{1t}'$, we subtract first $\mathbf{1t}'$ on both sides, which yields $\mathbf{Y} - \mathbf{1t}' = s\mathbf{XT}$; then, premultiplying by $1/s$ and postmultiplying by $\mathbf{T}^{-1} = \mathbf{T}'$ gives $(1/s)(\mathbf{Y} - \mathbf{1t}')\mathbf{T}' = \mathbf{X}$, which is $(1/s)\mathbf{YT}' - (1/s)\mathbf{1t}'\mathbf{T}' = \mathbf{X}$. In words: we first multiply $\mathbf{Y}$ by $1/s$, then

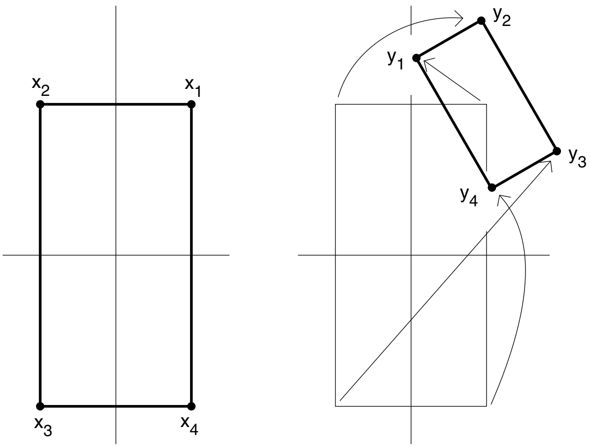

FIGURE 20.2. Illustration of fitting $\mathbf{Y}$ to $\mathbf{X}$ by a similarity transformation.

rotate it clockwise by 30° and reflect it horizontally, and then subtract the translation vector $(1/s)\mathbf{1t}'\mathbf{T}'$ from it. Because the $\mathbf{T}$ matrix is the same as the one discussed in the last section, and $1/s = 2$ and $\mathbf{t}' = (1,2)$ are also known, the transformations involved in mapping $\mathbf{Y}$ back into $\mathbf{X}$ can be computed easily.

In general, of course, only $\mathbf{X}$ and $\mathbf{Y}$ are given, and we have to find optimal s, $\mathbf{T}$, and $\mathbf{t}$. The loss function $L(s,\mathbf{t},\mathbf{T})$ is therefore

$$L(s,\mathbf{t},\mathbf{T}) = \text{tr } [\mathbf{X} - (s\mathbf{YT} + \mathbf{1t}')]'[\mathbf{X} - (s\mathbf{YT} + \mathbf{1t}')], \tag{20.5}$$

subject to $\mathbf{T}'\mathbf{T} = \mathbf{I}$. An optimal translation vector $\mathbf{t}$ is obtained by setting the derivative of $L(s,\mathbf{t},\mathbf{T})$ with respect to $\mathbf{t}$ equal to zero and solving for $\mathbf{t}$, i.e.,

$$\partial L(s,\mathbf{t},\mathbf{T})/\partial \mathbf{t} = 2n\mathbf{t} - 2\mathbf{X}'\mathbf{1} + 2s\mathbf{T}'\mathbf{Y}'\mathbf{1} = \mathbf{0}, \tag{20.6}$$

$$\mathbf{t} = n^{-1}(\mathbf{X} - s\mathbf{YT})'\mathbf{1}. \tag{20.7}$$

Inserting the optimal $\mathbf{t}$ (20.7) into (20.5) gives

$$\begin{aligned} &L(s,\mathbf{T}) \\ &\quad = \text{tr } [(\mathbf{X} - s\mathbf{YT}) - \frac{\mathbf{11}'}{n}(\mathbf{X} - s\mathbf{YT})]'[(\mathbf{X} - s\mathbf{YT}) - \frac{\mathbf{11}'}{n}(\mathbf{X} - s\mathbf{YT})] \\ &\quad = \text{tr } [(\mathbf{I} - \frac{\mathbf{11}'}{n})(\mathbf{X} - s\mathbf{YT})]'[(\mathbf{I} - \frac{\mathbf{11}'}{n})(\mathbf{X} - s\mathbf{YT})] \\ &\quad = \text{tr } [\mathbf{JX} - s\mathbf{JYT}]'[\mathbf{JX} - s\mathbf{JYT}], \end{aligned}$$

with $\mathbf{J}$ the centering matrix $\mathbf{I} - n^{-1}\mathbf{11}'$. Similarly, setting the partial derivative of $L(s,\mathbf{T})$ to s equal to zero and solving for s yields

$$\partial L(s,\mathbf{T})/\partial s = 2s(\text{tr } \mathbf{Y}'\mathbf{JY}) - 2\text{tr } \mathbf{X}'\mathbf{JYT} = 0, \tag{20.8}$$

$$s = (\text{tr } \mathbf{X}'\mathbf{JYT})/(\text{tr } \mathbf{Y}'\mathbf{JY}). \tag{20.9}$$

Inserting the optimal s into $L(s, \mathbf{T})$ gives

$$\begin{aligned} L(s, \mathbf{T}) &= \text{tr } [\mathbf{JX} - \frac{\text{tr } \mathbf{X'JYT}}{\text{tr } \mathbf{Y'JY}}\mathbf{JYT}]'[\mathbf{JX} - \frac{\text{tr } \mathbf{X'JYT}}{\text{tr } \mathbf{Y'JY}}\mathbf{JYT}] \\ &= \text{tr } \mathbf{X'JX} + \frac{(\text{tr } \mathbf{X'JYT})^2}{\text{tr } \mathbf{Y'JY}} - 2\frac{(\text{tr } \mathbf{X'JYT})^2}{\text{tr } \mathbf{Y'JY}} \\ &= \text{tr } \mathbf{X'JX} - \frac{(\text{tr } \mathbf{X'JYT})^2}{\text{tr } \mathbf{Y'JY}}. \end{aligned} \tag{20.10}$$

Minimizing (20.10) over $\mathbf{T}$ ($\mathbf{TT'}=\mathbf{I}$) is equal to minimizing $-$tr $\mathbf{X'JYT}$ over $\mathbf{T}$ because $\mathbf{T}$ may always be chosen such that tr $\mathbf{X'JYT}$ is nonnegative. Therefore, we can apply the results from the previous section to find the optimal $\mathbf{T}$. This also explains why maximizing the correlation $r(\mathbf{A}, \mathbf{BT})$ (see Section 20.6) or (20.1) yields the same $\mathbf{T}$ as the Procrustes problem (20.1).

The steps to compute the Procrustean similarity transformation are:

1. Compute $\mathbf{C} = \mathbf{X'JY}$.
2. Compute the SVD of $\mathbf{C}$; that is, $\mathbf{C} = \mathbf{P\Phi Q'}$.
3. The optimal rotation matrix is $\mathbf{T} = \mathbf{QP'}$.
4. The optimal dilation factor is $s = (\text{tr } \mathbf{X'JYT})/(\text{tr } \mathbf{Y'JY})$.
5. The optimal translation vector is $\mathbf{t} = n^{-1}(\mathbf{X} - s\mathbf{YT})'\mathbf{1}$.

20.5 An Example of Procrustean Similarity Transformations

We now return to Figure 20.2. To transform $\mathbf{Y}$ back to $\mathbf{X}$, the original transformations that led to $\mathbf{Y}$ have to be undone. According to our computation scheme of the previous section, what has to be found first is the orthogonal matrix $\mathbf{T}$, then the dilation factor s, and finally $\mathbf{t}$.

$\mathbf{C} = \mathbf{X'JY}$ turns out to be simply $\mathbf{C} = \mathbf{X'Y}$ in the present case, because $\mathbf{J} = \mathbf{I} - \mathbf{11'}/n$ can be seen to center the rows of $\mathbf{X'}$ or the columns of $\mathbf{Y}$. But the columns of $\mathbf{X}$ are centered already (i.e., the values in the columns of $\mathbf{X}$ sum to 0); thus $\mathbf{J}$ is not needed here. For $\mathbf{C} = \mathbf{X'Y}$ we obtain

$$\mathbf{C} = \begin{pmatrix} -1.72 & -1.00 \\ -4.00 & 6.96 \end{pmatrix}.$$

The singular value decomposition of $\mathbf{C}$, $\mathbf{C} = \mathbf{P\Phi Q'}$, is

$$\mathbf{C} = \begin{pmatrix} .00 & -1.00 \\ 1.00 & .00 \end{pmatrix} \begin{pmatrix} 8.03 & .00 \\ .00 & 1.99 \end{pmatrix} \begin{pmatrix} -.50 & .87 \\ .87 & .50 \end{pmatrix}.$$

Thus, $\mathbf{T}$ is given by

$$\mathbf{T} = \mathbf{QP}' = \begin{pmatrix} -.87 & -.50 \\ -.50 & .87 \end{pmatrix}.$$

It is easier to see what $\mathbf{T}$ does when it is decomposed into a rotation and a subsequent reflection:

$$\mathbf{T} = \mathbf{RU} = \begin{pmatrix} .87 & -.50 \\ .50 & .87 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

In $\mathbf{R}$, we have $\cos(\alpha) = .87$ and hence[3] $\alpha = 30°$. Also, $\sin(\alpha) = .50$; thus, $\alpha = 30°$. Hence, $\mathbf{R}$ rotates $\mathbf{Y}$ by 30° to the right or clockwise, which aligns the sides of $\mathbf{Y}$ in Figure 20.2 with the coordinate axes. $\mathbf{U}$ then undoes the previous reflection along the horizontal axis, because all of the coordinates in the first column of $\mathbf{YR}$ are reflected by -1.

The transformations s and $\mathbf{t}$ are also easy to compute. For s we compute $s = (\text{tr } \mathbf{X}'\mathbf{JYT})/(\text{tr } \mathbf{Y}'\mathbf{JY}) = 10.02/5.01 = 2$, which is just the inverse of the dilation factor from above. Finally, we find $\mathbf{t}' = (3.73, -2.47)$. It is harder to understand why such a translation is obtained, and not just $(-1, -2)$. At the beginning of the previous section, it was shown algebraically that to undo the translation $\mathbf{t}$ it is generally not correct to set $-\mathbf{t}$. This is so because other transformations are also done at the same time; thus, what has to be back-translated is not $\mathbf{Y}$, but $\mathbf{Y}$ after it has been back-rotated, -reflected, and -dilated. If we check what these transformations do to $\mathbf{Y}$ in Figure 20.2, we can see that $\mathbf{t} = (3.73, -2.47)$ must result. (Note, in particular, that $\mathbf{R}$ rotates $\mathbf{Y}$ about the origin, not about the centroid of $\mathbf{Y}$.)

20.6 Configurational Similarity and Correlation Coefficients

So far, we have considered Procrustean procedures primarily for transforming a configuration so that it becomes easier, in one sense or another, to look at. We now discuss a measure that assesses the degree of similarity between the transformed configuration and its target. One obvious choice for such a measure is the product-moment correlation coefficient computed over the corresponding coordinates of $\mathbf{X}$ and $\mathbf{YT}$.

Consider the three data matrices in Table 20.1, taken from a study by Andrews and Inglehart (1979). The matrices show the product-moment

[3] Note that $\mathbf{R}' = \mathbf{R}^{-1}$ rotates a configuration to the left or counterclockwise. See (7.31), which shows a rotation matrix for the plane that moves the points counterclockwise.

TABLE 20.1. Intercorrelations of items in three studies on well-being in the U.S.A., Italy, and Denmark, respectively. Items are: (1) housing, (2) neighborhood, (3) income, (4) standard of living, (5) job, (6) spare time activities, (7) transportation, (8) health, (9) amount of spare time, (10) treatment by others, (11) getting along with others. Decimal points omitted.

	Italy																					
	1	2	3	4	5	6	7	8	9	10	11	1	2	3	4	5	6	7	8	9	10	11
1	–	33	44	52	38	37	26	23	20	27	23	–										
2	38	–	23	19	29	23	28	18	21	29	31	46	–									
3	30	21	–	77	57	43	38	25	23	17	19	38	29	–								
4	42	30	66	–	56	52	38	28	29	22	2	46	35	64	–							
5	10	18	34	23	–	49	35	23	23	32	29	29	30	42	47	–						
6	27	28	33	36	31	–	28	28	39	25	32	27	29	28	37	41	–					
7	14	19	28	29	26	26	–	18	24	18	22	16	16	22	24	20	18	–				
8	15	11	23	21	17	15	22	–	27	17	21	08	13	14	12	24	15	17	–			
9	17	15	18	26	25	29	18	08	–	20	28	22	21	19	26	24	37	15	14	–		
10	23	30	27	30	38	26	26	20	26	–	69	27	33	26	30	32	26	17	14	23	–	
11	18	18	14	24	23	21	21	29	14	36	–	31	39	29	37	34	38	22	13	28	53	-
	U.S.A.											Denmark										

TABLE 20.2. Similarity coefficients of three attitude structures on well-being. Lower half: squared correlations over coordinates. Upper half: squared congruence coefficients of distances.

	U.S.A.	Italy	Denmark
U.S.A.	1.000	0.883	0.859
Italy	0.347	1.000	0.857
Denmark	0.521	0.515	1.000

correlations of 11 questions on subjective well-being asked in the U.S.A., Italy, and Denmark, respectively. The questions were always phrased according to the schema “How satisfied are you with [X]?”. The interviewees responded by giving a score on a rating scale. The scores were correlated over persons. Data from representative samples in each of nine different Western societies were available. The general hypothesis was that the attitude structures on well-being in these countries would be very similar.

Andrews and Inglehart (1979) represented each of these correlation matrices in a 3D MDS space. For the matrices in Table 20.1, this leads to the Stress values .09, .08, and .04, respectively. It was then asked how similar each pair of these configurations is, and Procrustean transformations were used to “remove inconsequential differences in the original locations, orientations, and sizes of the configurations” (Andrews & Inglehart, 1979, p.78). For our three data sets, this leads to the indices in Table 20.2 (lower half). (Note that we report squared correlations here, which is in agreement with the notion of common variance in statistics.) On the basis of such measures, Andrews and Inglehart conclude that “there seems to be a basic similarity in structures among all nine of these Western societies” (p.83).

Such an evaluation assumes that the observed similarity indices are greater than can be expected by chance alone. For two configurations, $\mathbf{X}$ and $\mathbf{Y}$, both chosen completely at random, $r(\mathbf{X}, s\mathbf{YT} + \mathbf{1t}') = r(\mathbf{X}, \mathbf{Y}^*)$ would probably not be zero but should be positive. The fewer points there are in $\mathbf{X}$ and $\mathbf{Y}$, the greater the correlation should be. The Procrustean transformations are designed to maximize $r(\mathbf{X}, \mathbf{Y}^*)$; the fewer points there are, the greater the effect of these transformations, in general. Langeheine (1980b, 1982) has studied by extensive computer simulations what r-values could be expected for different numbers of points (n) and different dimensionalities (m). He finds virtually the same results for different error models (such as sampling the points from multidimensional rectangular or normal distributions). For $n = 10$ and $m = 3$, the parameters relevant for the present 3D MDS configurations with ten points, he reports $0.072 \leq r^2(\mathbf{X}, \mathbf{Y}^*) \leq 0.522$ and $\bar{r}^2(\mathbf{X}, \mathbf{Y}^*) = 0.260$. Furthermore, only 5% of the observed coefficients were greater than 0.457. We should therefore conclude that the degree of similarity observed for these structures is hardly impressive.

20.7 Configurational Similarity and Congruence Coefficients

It is possible to skip the Procrustean transformations altogether and still arrive at a measure of similarity for each pair of configurations. This can be done by directly comparing the distances of $\mathbf{X}$ and $\mathbf{Y}$, because their ratios remain the same under any transformations where $\mathbf{T}'\mathbf{T} = \mathbf{I}$. Thus, Shepard (1966) computes the product-moment correlation coefficient over the corresponding distances of $\mathbf{X}$ and $\mathbf{Y}$, and Poor and Wherry (1976) report extensive simulations on the behavior of such correlations in randomly chosen configurations. Yet, the usual correlation is an inadmissible and misleading index when used on distances. To see why, consider the following example. Assume that $\mathbf{X}$ and $\mathbf{Y}$ consist of three points each. Let the distances in $\mathbf{X}$ be $d_{12}(\mathbf{X}) = 1, d_{23}(\mathbf{X}) = 2, d_{13}(\mathbf{X}) = 3$ and the distances in $\mathbf{Y}$, $d_{12}(\mathbf{Y}) = 2, d_{23}(\mathbf{Y}) = 3, d_{13}(\mathbf{Y}) = 4$. The correlation of these distances is $r = 1$, indicating perfect similarity of $\mathbf{X}$ and $\mathbf{Y}$. But this is false; $\mathbf{X}$ and $\mathbf{Y}$ do not have the same shape: $\mathbf{Y}$ forms a triangle, but $\mathbf{X}$'s points lie on a straight line because they satisfy the equation $d_{12}(\mathbf{X}) + d_{23}(\mathbf{X}) = d_{13}(\mathbf{X})$. If a constant s is subtracted from each distance in this equation, the inequality $d_{12}(\mathbf{X}) - k + d_{23}(\mathbf{X}) - k \neq d_{13}(\mathbf{X}) - k$ results. The translated values $v_{ij} = d_{ij}(\mathbf{X}) - k$ are therefore not distances of three collinear points. Thus, pulling out any nonzero constant from the distances implies that the new values are either distances of a configuration different from the one we wanted to assess, or are not even distances at all, i.e., they correspond to no geometric configuration whatsoever. Hence, correlating distances does not properly assess the similarity of geometric figures (configurations).

The problem is easily resolved, however, if we do not extract the mean from the distances and compute a correlation about the origin, not the centroid. The resulting *congruence coefficient* is

$$c(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{i<j} w_{ij} d_{ij}(\mathbf{X}) d_{ij}(\mathbf{Y})}{[\sum_{i<j} w_{ij} d_{ij}^2(\mathbf{X})]^{1/2} [\sum_{i<j} w_{ij} d_{ij}^2(\mathbf{Y})]^{1/2}} \quad ,$$

with w_{ij} nonnegative weights. Because distances are nonnegative and, by the Cauchy–Schwarz inequality, $c(\mathbf{X}, \mathbf{Y})$ ranges from 0 to 1, we have $c(\mathbf{X}, \mathbf{Y}) = 1$ if $\mathbf{X}$ and $\mathbf{Y}$ are perfectly similar [i.e., if $r(\mathbf{X}, s\mathbf{YT} + \mathbf{1t}') = 1$], and $c = 0$ if $r = 0$. But apart from these boundary cases, there is no easy relation of r and c and it seems impossible to convert a given r-value directly into the corresponding c-value, and vice versa.

Computing the congruence coefficients for the MDS configurations of the data in Table 20.1 yields the values in the upper half of Table 20.2. In comparison with the Procrustean correlation values in the lower half of the matrix, these measures lead to a different interpretation of the similarity pattern: the similarity of the Italian and the U.S.A. configurations is lowest in terms of r but highest in terms of c. Indeed, the order of the similarities among countries is exactly the opposite for both coefficients. Thus, using two equally admissible indices leads to different conclusions. Why this is so is not difficult to see: each coefficient must condense a great deal of information on the similarity of two configurations into a single number, and this can be done by weighting this or that piece of information more or less. Furthermore, the distinction between geometric and correlational similarity should be noted in problems of this sort.

The question that remains is whether r and c typically yield different answers in practical applications. In particular, by comparing r with its statistical norms (Langeheine, 1982) and c with analogous norms (Leutner & Borg, 1983), are we likely to conclude in one case that the configuration pair is significantly similar and in the other that it is not? In simulation studies, Borg and Leutner (1985) showed that, for randomly chosen configurations with different numbers of points and dimensionalities, r and c led to the same statistical judgment in not more than 60% of the cases. Hence, if we claim that two configurations are more similar than can reasonably be expected by chance alone, both the r and c values should be well above their respective statistical norms.

The problems associated with such similarity coefficients are ultimately due to the fact that these measures are extrinsic to substantive problems. It would be an illusion to believe that somehow a better coefficient could be constructed, because any such coefficient must condense the given complex information into a single number. It is evident that the way this should be done depends on the substantive question being studied. Moreover, it seems that, in a case like the Andrews–Inglehart study on attitude structures, the question of how close corresponding points can be brought to each other is much too parametric. The formal reason is that with 10 points in a 3D space, the MDS configurations are not strongly determined by Stress; that is, many other configurations exist (some more similar, some less similar among themselves) that represent the data almost equally well. This was shown by Borg and Bergermaier (1981). The deeper scientific reason is that there is actually no basis for expecting such a point-by-point matching of different attitude structures in the first place. In Section 5.3, the similarity question was therefore asked quite differently: can two (or more) MDS configurations both be partitioned into regions by facets from the *same* facet design [see also Shye (1981)]. Because this could be done, it was concluded that the structures were indeed similar in the sense that they all reflected the same facets. In other contexts, the pointwise matching of two configurations may be more meaningful, but this has to be checked in each single case. For the psychophysical example discussed in Section 17.4, for example, such indices are adequate in Figure 17.8 to assess the fit of the design configuration (transformed in a theoretically meaningful way) and the respective MDS representations. It is a widespread fallacy, however, to believe that such indices are somehow "harder" and "more meaningful" than the pictures themselves. Rather, the indices play only a supplementary role, because the pictures show in detail where the configurations match and where they do not.

20.8 Artificial Target Matrices in Procrustean Analysis

Procrustean procedures were first introduced in factor analysis because it frequently deals with relatively high-dimensional vector configurations which would otherwise be hard to compare. Moreover, with very few exceptions [e.g., the radex of Guttman (1954) or the positive manifold hypothesis of Thurstone (1935)], factor analysts have been interested only in dimensions, whose similarity can be seen directly from comparing $\mathbf{X}$ and $\mathbf{YT}$. In addition, it was soon noted that the Procrustean methods can also be used in a confirmatory manner, where $\mathbf{X}$ does not relate to a configuration of empirical data but is a matrix constructed to express a substantive hypothesis; for example, $\mathbf{X}$ could contain the point coordinates for an expected configuration in a psychophysical study such as the one on rectangles in Section 17.4. Here, we might simply take the solid grid configuration in Figure 17.7 as a target for the MDS configuration in Figure 17.8 (assuming, for the sake of argument, that the MDS configuration was generated under the Euclidean metric, because otherwise no rotations are admissible). Of course, in the plane, such rotations are more cosmetic (to obtain better aligned plots, e.g.) and we can easily do without them. However, imagine that the stimuli had been boxes rather than rectangles. A theory for the similarity judgments on such stimuli would certainly ask for a 3D representation, but the common principal-component orientation routinely used by most MDS procedures may not give us the desired orientation. Hence, even though using the design configuration without, say, any logarithmic rescalings of its dimensions may be primitive, it may lead to a more interpretable orientation of the MDS configuration.

Sometimes the target matrix $\mathbf{X}$ and testee matrix $\mathbf{Y}$ do not to have the same dimensionality. For example, in the rectangle study from above, we might have various other variables associated with the rectangles (such as different colorings and patterns). A higher-dimensional MDS space is then probably necessary to represent their similarity scores. Nevertheless, the 2D design lattice can still serve to derive a *partial* target matrix $\mathbf{X}$, which can serve as a partial hypothesis structure for the higher-dimensional MDS configuration. Technically, what needs to be done in this case to guarantee that the necessary matrix computations can be carried out is to append columns of zeros on $\mathbf{X}$ until the column orders of the augmented $\mathbf{X}$ and the $\mathbf{Y}$ matrix match. A reference for procedures on missing dimensionalities in Procrustean analysis is Peay (1988).

A further generalization of the Procrustean procedures allows partial specification of $\mathbf{X}$ by leaving some of its elements undefined (Browne, 1972a; Ten Berge, Kiers, & Commandeur, 1993). This possibility is needed when only partial hypotheses exist. A typical application is the case in which some points represent previously investigated variables and the remaining variables are "new" ones. We might then use the configuration from a previous study as a partial target for the present data in order to check how well this structure has been replicated. A different strategy was pursued by Commandeur (1991) in the Matchals program where entire rows can be undefined.

Another case of an incomplete formal match of $\mathbf{X}$ and $\mathbf{Y}$ is one in which the configurations contain a different number of points. Consider a study of Levy (1976) concerned with the psychology of well-being. Using a facet-theoretical approach, Levy used items based on two facets: A = $\{$*state, resource*$\}$ and B = *life area* with eight elements. Respondents were asked how satisfied they were with the content of an item on a 9-point rating scale. For example, the respondent had to indicate how satisfied he or she was with "the city as place to live" on a scale of -4 for "very dissatisfied" to $+4$ for "very satisfied". The data were taken from two studies carried out in 1971, one in the U.S. and one in Israel. Similarity coefficients were derived from the items (correlations for the U.S. study and μ_2 for the Israel study), followed by an MDS analysis for each country. The resulting coordinate matrices for the configurations are given in Table 20.3. There

TABLE 20.3. Generating comparable matrices $\mathbf{X}_c$ and $\mathbf{Y}_c$ by averaging the coordinates of points in $\mathbf{X}$ and $\mathbf{Y}$ that have equivalent structuples, dropping rows that do not have matching structuples, and permuting the rows of the resulting matrices into a common order of structuples. Bold-face structuples are common to both studies.

	U.S. Study		Structuple	Structuple	$\mathbf{X}_c$	
1	82.878	-42.163	**23**	23	83.014	-41.638
2	88.993	-60.939	**23**	17	-4.189	-31.551
3	60.183	-46.662	**23**	14	3.004	-8.451
4	100.000	-16.787	**23**	26	-100.000	-28.496
5	-13.325	-87.959	21	22	19.631	-46.593
6	-19.009	-100.000	21	18	8.226	-15.692
7	-4.189	-31.551	**17**			
8	3.004	-8.451	**14**			
9	-100.000	-28.496	**26**			
10	27.065	-38.147	12			
11	19.631	-46.593	**22**			
12	41.695	20.110	29			
13	-7.944	40.172	25			
14	7.994	15.670	15			
15	8.226	-15.692	**18**			
	Israel Study		Structuple	Structuple	$\mathbf{Y}_c$	
1	55.109	-38.601	**22**	23	100.000	-87.625
2	100.000	-87.625	**23**	17	-20.501	45.374
3	-100.000	-59.374	**26**	14	9.139	9.563
4	-89.866	-100.000	**26**	26	-94.933	-79.687
5	-50.625	-60.229	16	22	55.109	-38.601
6	3.523	-48.208	**18**	18	-12.976	-39.149
7	-20.501	45.374	**17**			
8	-31.567	49.099	27			
9	-29.474	-30.089	**18**			
10	9.139	-9.563	14			

are 15 points in the U.S. representation, but only 10 in the Israeli solution. However, most of these points are associated with structuples that are common across the two studies. Hence, we can proceed as indicated in Table 20.3: (1) in each configuration, average the coordinates of all points that have common structuples; (2) set up matrices $\mathbf{X}_c$ and $\mathbf{Y}_c$ consisting of the average coordinate vectors in such a way that the rows of $\mathbf{X}_c$ and $\mathbf{Y}_c$ correspond substantively (i.e., in terms of their structuples); centroids without a partner in the other configuration are dropped; (3) with $\mathbf{X}_c$ and $\mathbf{Y}_c$ proceed as in a usual Procrustean problem; (4) finally, use the transformations computed in (3) to transform the original matrices (Borg, 1977b, 1978a). Provided there are enough different common structuples, this procedure does what can be done to make the configurations easier to compare.

20.9 Other Generalizations of Procrustean Analysis

Here, we consider variants of the Procrustes problem. In particular, we discuss the so-called oblique Procrustean rotation, rotation to optimal congruence, and robust Procrustean rotation.

The problem of *oblique* Procrustean rotation has been encountered previously in this book under different names. It consists of rotating each coordinate axis independently of the others in such a way that $\mathbf{BT}$ approximates $\mathbf{A}$ as closely as possible. Thus, we want to minimize (20.1) without any constraint on $\mathbf{T}$. Such a solution can be readily found by multiple regression for each dimension separately.

Oblique Procrustes rotation can be interpreted as follows. Let $\mathbf{T}$ be decomposed by a singular value decomposition; then $\mathbf{T} = \mathbf{P\Phi Q}'$. It follows what $\mathbf{T}$ does: first, $\mathbf{B}$ is rotated by $\mathbf{P}$; then $\mathbf{\Phi}$ multiplies the coordinate vectors of $\mathbf{BP}$ with different weights, thus geometrically stretching $\mathbf{BP}$ differentially along the axes, and finally $\mathbf{BP\Phi}$ is rotated by $\mathbf{Q}'$. Hence, only if $\mathbf{\Phi} = \mathbf{I}$ is $\mathbf{T} = \mathbf{PQ}'$ an orthonormal matrix. The transformation problem turns out to be the same as the one encountered in Section 4.3, where *external* scales had to be optimally placed into a given configuration. In factor analysis, certain additional constraints are often placed on $\mathbf{T}$, so that the oblique Procrustes problem is not always equivalent to the linear fitting. However, these additional constraints are not relevant in the MDS context (see, e.g., Browne, 1969, 1972b; Browne & Kristof, 1969; Mulaik, 1972). Applying oblique Procrustes rotation of MDS solutions has to be done with caution, because the transformed solution has different distances.

A different fit criterion for Procrustes rotation is based on the congruence between the columns of $\mathbf{A}$ and $\mathbf{BT}$. Brokken (1983) proposed a rotation method where the congruence between corresponding columns is optimized. If $\mathbf{A}$ and $\mathbf{B}$ have column means of zero, then rotation to optimal congruence can be interpreted as Procrustes rotation while assuming that each column of $\mathbf{A}$ and $\mathbf{B}$ is measured on an interval level. Kiers and Groenen (1996) developed a majorizing algorithm to optimize this criterion. This type of analysis is particularly suitable if the columns of the matrices are interval variables measured on the same scale.

In some cases, all but a few of the points of two configurations may be similar (after rotation). Verboon (1994) discusses the following artificial example. Let $\mathbf{A}$ contain the coordinates of cornerpoints of a centered square and $\mathbf{B}$ the same coordinates but rotated by 45°. An "outlier" in $\mathbf{B}$ is created by multiplying the second coordinate of point 1 by -10. This outlier has a different orientation (180°) and is located much farther from the origin than the other points. Ordinary Procrustes analysis yields a rotation matrix with an angle of $-18°$, a deviation of more than 60°.

Verboon and Heiser (1992) and Verboon (1994) propose a *robust* form of Procrustes analysis that is less sensitive to outliers. They start by decomposing the misfit into the contribution of error of each object to the total error; that is,

$$\begin{aligned} L(\mathbf{T}) &= \text{tr } (\mathbf{A} - \mathbf{BT})'(\mathbf{A} - \mathbf{BT}) \\ &= \sum_{i=1}^{n} (\mathbf{a}_i - \mathbf{T}'\mathbf{b}_i)'(\mathbf{a}_i - \mathbf{T}'\mathbf{b}_i) = \sum_{i=1}^{n} r_i^2, \end{aligned}$$

where $\mathbf{a}_i$ denotes row i of $\mathbf{A}$. The basic idea is to downweight large residuals so that outliers have less influence on the final rotation. This can be achieved by minimizing

$$L_r(\mathbf{T}) = \sum_{i=1}^{n} f(r_i),$$

with $f(x)$ a robust function. Some often-used robust functions are $|x|$, the Huber function (Huber, 1964), and the biweight function (Beaton & Tukey, 1974), all of which have as a main characteristic that $f(r_i) < r_i^2$ for large values of r_i. Clearly, choosing $f(x) = x^2$ reduces $L_r(\mathbf{T})$ to $L(\mathbf{T})$. Algorithms for minimizing $L_r(\mathbf{T})$ for different robust functions $f(x)$ based on iterative majorization can be found in Verboon (1994).

20.10 Exercises

Exercise 20.1 Consider the three correlation matrices in Table 20.1 on p. 438. Scale each data matrix individually via MDS. Then use Procrustean transformations to eliminate irrelevant differences among the MDS solutions. How do the three solutions differ from one another?

Exercise 20.2 It looks as if the plane spanned by dimension 1 and dimension 2 in Figure 4.3 corresponds closely to the 2D configuration in Figure 4.1.

(a) Replicate the scalings and then fit the 3D solution to the 2D solution by Procrustean methods.

(b) Compute indices that indicate the similarity of the 2D MDS solution and the fitted plane of the 3D solution. Use two different measures of similarity.

Exercise 20.3 Use the data matrix of Table 4.1 on p. 65 and represent it in a 3D MDS space. Then use an artificial target matrix to swing the MDS solution into a plane that shows a color circle.

Exercise 20.4 The matrices below show the point coordinates of two configurations in three dimensions.

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & -.5 \\ -1 & 2 & .5 \\ -1 & 0 & -.5 \\ 1 & 0 & .5 \\ -1 & -2 & .5 \\ 1 & -2 & -.5 \end{bmatrix} \text{ and } \mathbf{Y} = \begin{bmatrix} 1.2449 & -0.8589 & -1.7202 \\ 0.4572 & 1.1834 & -1.7793 \\ -0.9626 & 0.5000 & -0.5321 \\ 0.9086 & -0.4459 & 0.3364 \\ -1.1118 & 0.8645 & 1.8433 \\ -0.5361 & -1.2432 & 1.8519 \end{bmatrix}.$$

(a) Find the rotation that optimally fits $\mathbf{Y}$ to $\mathbf{X}$.

(b) Assess the fit of the fitted $\mathbf{Y}$ to the target $\mathbf{X}$.

Exercise 20.5 The matrix below shows the coordinates of four points in 4D. Transform this configuration so that it optimally fits into a 2D plane. (Hint: Procrustean transformations may not be the best method to solve this problem.)

$$\mathbf{M} = \begin{bmatrix} 1.4944 & -0.2109 & -1.5806 & -0.4718 \\ 0.2397 & 0.4019 & -1.9928 & 0.8993 \\ -1.4944 & 0.2109 & 1.5806 & 0.4718 \\ -0.2397 & -0.4019 & 1.9928 & -0.8993 \end{bmatrix}$$

Exercise 20.6 Use the coordinate matrices $\mathbf{X}$ and $\mathbf{Y}$ from Section 20.4.

(a) Augment matrix $\mathbf{Y}$ with a vector of random error so that $\mathbf{Y}$ becomes three-dimensional. Repeat the Procrustean transformations and assess the fit to the target configuration.

(b) Repeat the above with different amounts of random error. How does this error affect the fittings?

Exercise 20.7 Assume we drop the constraint that $\mathbf{TT}' = \mathbf{I}$ in Section 20.2 and admit any linear transformation $\mathbf{T}$ to solve the loss function in formula 20.1.

(a) Show that $\mathbf{T} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}$ minimizes the loss function in this case. (Hint: Expand the expression and use the rules developed in Section 8.3.)

(b) Apply the result to two simple 2D configurations, $\mathbf{A}$ and $\mathbf{B}$, that are both centered.

(c) Study geometrically in which way $\mathbf{T}$ affects $\mathbf{B}$ in fitting it to $\mathbf{A}$.

(d) Analyze what $\mathbf{T}$ does to $\mathbf{B}$ in terms of its singular value decomposition. (Hint: Note the the SVD decomposes $\mathbf{T}$ into a rotation/reflection, followed by a stretching along the dimensions, followed by another rotation/reflection.)

(e) What properties of a configuration **B** are generally left unchanged when using a linear transformation **T**? (Hint: Check points that are on a straight line in **B**. Where do they end up in **BT**? Also, consider the dashed grid in Figure 17.9 and how it is related to its design grid in Figure 17.7.)

(f) Repeat fitting **B** to **A**, but now make sure that neither **A** nor **B** is centered. Compare the shape of **BT** in this case to the shape of **BT** in the centered case above.