

21

Three-Way Procrustean Models

In this chapter, we look at some varieties of generalized Procrustes analysis. The simplest task is to fit several given coordinate matrices \mathbf{X}_k ($k = 1, \dots, K$) to each other in such a way that uninformative differences are eliminated. We also consider generalizations of the Procrustean problem that first find an optimal average configuration for all \mathbf{X}_k and then attempt to explain each individual \mathbf{X}_k in turn by some simple transformation of the average configuration. One important case is to admit different weights on the dimensions of the average configuration. This case defines an interesting model for individual differences scaling: if the fit is good, then the perceptual space of individual k corresponds to the group's perceptual space, except that k weights the space's dimensions in his or her own idiosyncratic way.

21.1 Generalized Procrustean Analysis

We now begin by generalizing the Procrustes problem to the case of more than two configurations. To introduce the problem, assume that we had K proximity matrices and that each matrix was generated by one of K different individuals. Assume further that we had computed an MDS solution \mathbf{X}_k for each of these K individuals. What we would have, then, is a stack of \mathbf{X}_k s as depicted in Figure 21.1, a *three-way array* of coordinates x_{iak} ($i = 1, \dots, n; a = 1, \dots, m; k = 1, \dots, K$).

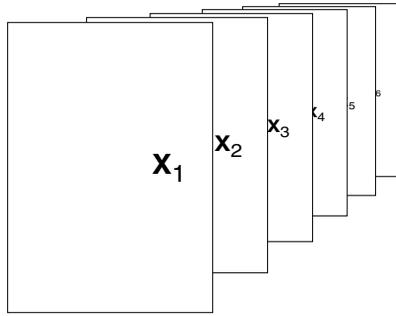


FIGURE 21.1. Schematic representation of a three-way data matrix, where the coordinate matrices \mathbf{X}_k of six subjects are stacked one after the other.

We now ask to what extent the K different \mathbf{X}_k “really” differ. We know from Chapter 20 that just looking at different \mathbf{X}_k s may be misleading, because one may notice differences that are uninformative in terms of the data. The task to visually separate uninformative from data-based differences becomes difficult or, indeed, unsolvable in the case of higher-dimensional spaces, but even in 2D it is at least helpful to first align different MDS solutions before comparing them.

Technically, given a set of K matrices \mathbf{X}_k , generalized Procrustean analysis is confronted with the task of optimally fitting these matrices to each other under a choice of rigid motions, dilations, and translations.

All of the above transformations are *admissible* ones, because they do not change the ratio of the distances and, hence, do not affect the way in which the various \mathbf{X}_k represent the corresponding proximity data. Generalized Procrustean fitting can, however, be generalized further by admitting *nonadmissible* free parameters to the transformations. For example, after fitting the K *individual configurations* \mathbf{X}_k to each other by similarity transformations, one may compute from them a *group configuration*,¹ \mathbf{Z} . We may then attempt to explain how the individuals differ from each other by considering certain simple transformations of \mathbf{Z} that allow one to approximate each \mathbf{X}_k in turn. The most important example is to compress and stretch \mathbf{Z} along its dimensions so that it best explains \mathbf{X}_k . The dimensional weights used in these deformations of \mathbf{Z} may be interpreted psychologically, for example, as expressions of the different importance that individual k attaches to the dimensions of the group space.

Gower and Dijksterhuis (2004) discuss the Procrustes problem and its three-way extensions in great mathematical depth. It is an excellent overview

¹This choice of terminology refers to the frequent case where each \mathbf{X}_k represents the proximity data from one individual k . The group configuration, then, is some kind of multidimensional average that represents the respective group of K individuals.

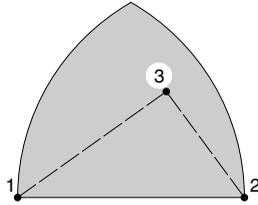


FIGURE 21.2. Data collection device used by Helm (1959).

of developments in this area. In the remainder of this chapter, we discuss a selection of three-way Procrustean models.

21.2 Helm's Color Data

We now consider an experiment by Helm (1959) that is used later on to illustrate various formal considerations. Different individuals were asked to judge the similarity of colors. The stimuli were 10 chips with different hues but constant brightness and saturation. The colors ranged over the entire spectrum from red to purple. With 10 stimuli, 120 different triples can be formed. For each triple, the subjects were asked first to identify the two colors that appeared most different. The respective chips were then placed onto points 1 and 2 in a schema like the one shown in Figure 21.2. The remaining chip of the triple was positioned somewhere in the shaded area so that the resulting distances would correspond to the perceived similarities among the colors. In this way, each subject generates more than the usual $n(n-1)/2$ distance judgments, because each stimulus pair is presented not just once but in combination with each of the remaining eight colors. The data were averaged to get more reliable estimates of the perceived distances than presenting each pair only once. The resulting values are reported in Table 21.1, where each column contains the $n(n-1)/2 = 45$ dissimilarity pairs of colors for one subject.

There were 14 different subjects, two of whom replicated the experiment after a four-week interval ($s_6^{[1]}, s_6^{[2]}$ and $s_{12}^{[1]}, s_{12}^{[2]}$) leading to a total of 16 replications. The resulting dissimilarity vectors of the two replications correlate with $r(s_6^{[1]}, s_6^{[2]}) = .96$ and $r(s_{12}^{[1]}, s_{12}^{[2]}) = .91$, which indicates high reliability for the judgments.

The subjects fall into two groups. Some of them have normal color vision, and others are deuteranopic (red-green deficient) in varying degrees. For a deuteranopic person, both red and green stimuli look gray. The subjects with deuteranopia are ordered, from least to most severe disability, as $s_{11} < s_{12} < s_{13} < s_{14}$.

Helm (1959, 1964) treats the data in Table 21.1 as direct distance estimates and applies classical scaling in maximal dimensionality, without any

TABLE 21.1. Distance estimates for color pairs (Helm, 1959) $s_6^{[1]}$ and $s_6^{[2]}$, and $s_{12}^{[1]}$ $s_{12}^{[2]}$ are replications for one subject each. Subjects s_1 to s_{10} are color-normals, and s_{11} to s_{14} are color-deficient.

Pair	Color-Normals										Color-Deficients					
	s_1	s_2	s_3	s_4	s_5	$s_6^{[1]}$	$s_6^{[2]}$	s_7	s_8	s_9	s_{10}	s_{11}	$s_{12}^{[1]}$	$s_{12}^{[2]}$	s_{13}	s_{14}
AC	6.8	5.9	7.1	7.5	6.6	5.2	5.8	6.2	7.5	6.0	9.2	11.5	9.3	9.0	10.4	9.9
AE	12.5	11.1	10.2	10.3	10.5	9.4	10.5	10.8	9.1	9.4	10.8	13.1	10.7	10.0	12.4	13.2
AG	13.8	18.8	11.1	10.7	10.2	11.4	13.4	9.9	10.2	9.5	9.7	12.6	10.7	10.4	12.8	12.3
AI	14.2	17.3	12.5	11.6	9.6	13.3	14.0	11.1	12.1	9.5	10.1	10.6	11.9	10.0	13.7	11.1
AK	12.5	16.6	11.8	10.6	10.8	12.0	13.2	10.3	12.5	9.8	10.3	10.6	11.0	9.3	11.8	8.7
AM	11.0	16.5	9.9	9.7	9.7	12.3	11.7	8.8	9.7	8.7	9.7	10.8	9.8	8.6	4.3	5.6
AO	8.6	8.3	8.6	8.4	8.5	10.6	10.2	7.6	9.8	6.7	9.0	7.3	8.9	8.8	4.0	7.4
AQ	5.5	5.7	4.3	5.8	4.9	4.9	6.4	5.8	8.3	4.9	6.6	5.4	8.9	7.5	5.5	6.4
AS	3.5	4.2	2.9	3.6	3.5	3.5	3.5	3.0	6.7	4.1	4.6	5.0	5.1	5.8	4.1	5.8
CE	5.4	4.9	5.7	6.9	5.5	6.2	4.9	7.5	4.4	7.1	5.5	6.0	6.5	6.9	8.1	7.3
CG	8.3	10.6	11.5	8.5	9.6	11.2	12.2	8.9	7.9	9.5	8.2	7.9	8.0	8.9	10.8	7.9
CI	10.4	14.3	10.7	10.7	9.3	13.5	14.8	10.7	10.4	9.5	9.4	8.4	8.2	8.4	10.4	6.9
CK	11.6	16.6	11.8	11.1	9.9	12.9	14.6	10.8	11.2	9.9	10.1	9.4	8.9	8.3	4.6	6.8
CM	13.8	17.3	11.2	12.2	11.7	12.0	14.1	10.6	12.6	10.6	10.5	10.2	9.3	9.7	9.6	9.9
CO	14.3	14.5	12.5	10.8	11.6	11.5	13.4	10.4	11.4	10.6	10.8	11.3	10.7	11.1	12.3	13.1
CQ	11.8	9.5	9.2	9.9	10.3	8.2	9.7	9.0	11.3	8.5	11.2	11.5	10.1	10.6	14.2	12.7
CS	8.9	7.3	8.2	8.0	8.0	6.3	7.9	7.5	10.4	7.9	10.5	11.5	9.6	10.3	13.0	12.1
EG	5.2	4.8	6.7	4.9	7.2	5.6	4.6	6.3	5.7	7.6	4.6	6.2	4.4	6.0	3.5	4.5
EI	7.2	8.3	8.9	6.6	8.3	8.2	8.3	8.7	8.3	8.9	6.7	8.4	7.0	6.8	4.3	5.3
EK	9.5	13.2	9.4	8.7	9.3	9.6	10.7	9.6	10.2	9.8	9.8	9.9	10.8	8.2	7.9	9.7
EM	11.3	14.6	11.3	10.6	11.3	12.7	12.8	10.1	11.3	10.5	11.3	10.3	10.4	10.9	13.0	11.5
EO	13.5	16.1	12.5	11.7	11.9	13.7	14.1	10.8	12.2	10.7	11.9	12.7	11.8	11.6	13.8	13.7
EQ	14.6	14.0	11.9	11.1	11.8	13.4	12.9	11.7	11.9	9.7	11.5	12.9	11.6	9.6	14.8	14.1
ES	14.1	13.8	10.5	12.0	11.5	11.7	10.9	9.4	10.7	10.2	10.2	10.7	10.2	10.5	13.9	13.4
GI	3.7	3.6	3.7	3.5	4.7	4.0	3.5	3.9	3.9	3.8	3.7	5.2	4.6	4.2	3.5	5.3
GK	5.9	5.3	5.9	6.3	6.2	5.8	4.7	6.8	6.5	5.3	6.6	6.5	9.6	7.3	9.0	8.6
GM	10.1	8.2	10.3	7.8	8.9	6.8	8.8	9.4	8.7	7.3	8.7	8.8	10.8	10.1	12.3	12.5
GO	11.1	14.5	11.6	10.4	10.3	9.3	11.0	9.7	10.3	7.6	10.6	11.2	11.9	10.2	12.3	13.4
GQ	12.3	17.0	10.9	11.6	11.6	10.5	11.8	10.4	10.7	9.2	10.0	11.7	11.3	10.6	12.9	14.1
GS	12.5	17.3	11.5	11.3	10.2	12.2	11.7	9.7	12.6	10.1	7.7	10.2	10.9	10.3	14.5	13.1
IK	4.2	3.5	3.6	4.1	3.3	3.8	3.6	5.0	4.6	4.8	4.0	4.1	5.8	5.2	7.0	6.9
IM	6.9	6.8	8.2	6.5	6.3	5.4	6.9	8.3	7.8	6.2	7.5	7.0	8.0	7.6	13.1	9.0
IO	10.2	11.0	9.8	8.6	9.1	7.9	9.4	9.0	9.9	8.2	9.9	10.4	10.5	9.2	13.1	12.2
IQ	12.1	15.8	11.3	10.0	11.1	9.9	12.4	10.9	11.2	9.1	10.9	10.8	10.4	10.3	13.6	12.5
IS	11.2	15.8	11.1	10.8	10.4	13.2	13.7	9.6	11.6	9.7	10.6	10.6	10.7	10.3	14.1	13.4
KM	4.3	3.8	5.1	5.0	4.2	3.6	4.1	4.3	6.3	4.7	5.4	6.4	7.7	6.4	9.9	6.7
KO	6.8	7.4	8.1	7.4	8.9	5.6	6.9	7.3	9.6	6.7	9.3	9.9	9.6	9.5	11.3	9.7
KQ	9.9	13.8	10.2	9.1	9.4	9.0	10.6	9.0	10.6	8.8	9.9	9.4	10.6	10.0	13.6	11.3
KS	10.7	15.1	10.6	10.7	10.6	10.4	12.2	8.8	11.6	9.9	9.7	10.1	10.7	9.6	12.3	9.9
MO	4.8	5.7	4.9	5.9	6.6	4.2	4.1	4.9	4.8	4.5	5.6	4.2	7.4	7.0	3.9	5.5
MQ	7.4	10.9	8.7	8.7	8.9	8.2	10.0	7.2	6.8	7.2	8.2	8.4	9.0	7.9	5.3	7.4
MS	8.7	13.9	9.7	9.6	9.2	9.8	11.1	7.6	9.1	6.8	9.7	8.1	8.7	8.7	6.4	5.4
OQ	4.5	5.0	6.3	5.6	5.8	5.1	4.1	4.7	4.6	4.0	5.3	4.5	4.5	4.8	4.7	4.2
OS	6.1	6.0	7.5	6.7	7.3	6.8	6.9	5.6	7.4	5.3	6.3	6.4	7.0	6.7	3.2	4.0
QS	3.6	3.5	3.0	3.5	2.9	3.8	3.4	3.5	5.2	3.4	3.4	3.0	4.5	4.3	2.4	4.3

TABLE 21.2. Eigenvalues obtained by classical scaling on each of the 16 individual dissimilarity matrices of the Helm (1959) data; color-normal subjects in upper table, color-deficient subjects in lower table; “Average” shows eigenvalues for averaged data.

	s_1	s_2	s_3	s_4	s_5	$s_6^{[1]}$	$s_6^{[2]}$	s_7	s_8	s_9	s_{10}	Average
1	260.0	449.2	191.6	179.7	166.8	233.8	276.4	147.7	182.0	126.6	164.8	204.4
2	178.3	276.5	143.6	125.8	127.1	165.1	190.1	110.4	150.0	105.9	102.9	160.2
3	28.6	32.0	44.1	22.8	29.5	23.3	27.5	32.7	28.5	28.4	48.6	12.1
4	17.9	15.8	18.4	18.6	26.8	16.4	8.9	21.6	24.9	18.2	32.3	6.7
5	4.8	5.2	11.3	13.3	15.8	6.6	7.0	11.2	14.7	12.9	10.7	6.5
6	4.3	.0	5.2	8.9	7.3	1.3	2.3	7.8	10.7	7.2	6.7	4.8
7	.0	-12.6	2.8	1.0	5.6	.0	.0	6.7	7.9	3.0	6.0	2.5
8	-9.5	-17.2	.0	.0	.0	-6.2	-6.2	.0	.0	.0	.0	1.3
9	-18.2	-40.6	-5.4	-3.4	-2.1	-10.5	-17.3	-.4	-5.9	-2.4	-3.1	.0
10	-30.5	-71.8	-17.1	-8.6	-15.3	-35.8	-26.5	-8.0	-11.3	-3.0	-14.1	-.4

	s_{11}	$s_{12}^{[1]}$	$s_{12}^{[2]}$	s_{14}	s_{14}	Average
1	213.2	175.2	154.0	347.7	296.0	232.0
2	80.7	92.5	72.5	98.7	56.9	59.9
3	48.4	47.5	51.7	34.2	38.3	51.5
4	36.0	32.7	31.3	25.0	26.3	22.0
5	14.9	28.3	19.9	9.8	21.6	15.8
6	10.9	14.8	13.2	.0	13.1	11.6
7	.8	7.0	6.8	-1.8	.4	8.6
8	.0	.0	3.3	-4.0	.0	2.6
9	-3.0	-2.5	.0	-6.2	-3.1	.0
10	-13.0	-5.0	-3.0	-17.7	-13.1	-1.3

prior transformations. The eigenvalues of classical scaling for each subject are reported in Table 21.2. One notes some negative eigenvalues, because the dissimilarities are not exact distances. The negative eigenvalues are, however, relatively small and can be explained by the Messick–Abelson model (Section 19.4). On the average, the color-normal subjects have two rather large eigenvalues, with the remaining eight eigenvalues close to zero. For the deuteranopic subjects, on the other hand, we find essentially only one large eigenvalue.

If a configuration is sought that is most representative for all color-normal subjects, the simplest answer is to derive it from the scores averaged over all respective data sets. This leads to the eigenvalues shown in the column “Average” of Table 21.2. Their distribution suggests that the color-normal subjects have a true 2D MDS configuration and that further dimensions are due to errors in perception and judgment. This interpretation is buttressed by the fact that the plane spanned by the first two eigenvectors shows the expected color circle as shown in Figure 21.3a. Classical scaling on the average data of the color-deficient subjects leads to Figure 21.3b. For

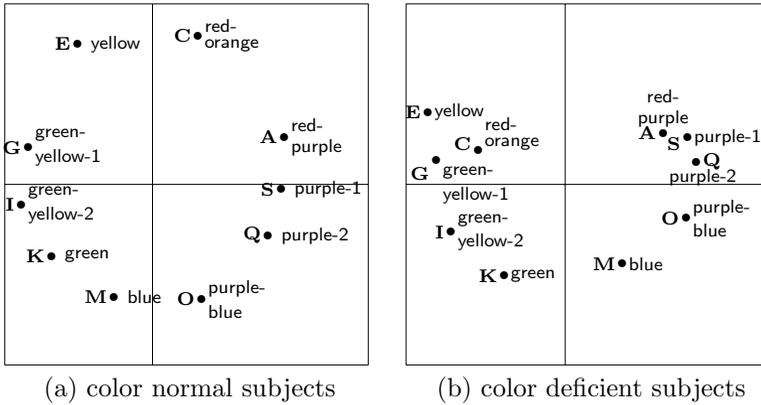


FIGURE 21.3. The 2D configuration obtained by classical scaling on the average of (a) the color-normal subjects, (b) the color-deficient subjects.

these subjects, one notes that the second principal axis of the color circle is clearly less pronounced.

21.3 Generalized Procrustean Analysis

Using average data is a rather crude approach. A possible alternative is to map all 11 data sets simultaneously into one configuration. Another possibility is *generalized Procrustes analysis (GPA)*, which transforms all K individual configurations, $\mathbf{X}_1, \dots, \mathbf{X}_K$, at the same time so that each configuration matches all others as closely as possible. The admissible transformations consist of rotations, reflections, dilations, and translations, as in Procrustean similarity transformations.

Expressed in terms of a loss function, generalized Procrustes analysis amounts to minimizing

$$GPA = \sum_{k < l}^K \text{tr} (\tilde{\mathbf{X}}_k - \tilde{\mathbf{X}}_l)' (\tilde{\mathbf{X}}_k - \tilde{\mathbf{X}}_l), \tag{21.1}$$

where $\tilde{\mathbf{X}}_k = s_k \mathbf{X}_k \mathbf{T}_k + \mathbf{1} \mathbf{t}'_k$ and $\mathbf{T}'_k \mathbf{T}_k = \mathbf{I}$. The function (21.1) is to be minimized through a proper choice of K scale factors s_k , K orthonormal matrices \mathbf{T}_k , and K translation vectors \mathbf{t}_k . The trivial solution where $s_k = 0$ must be avoided by imposing additional restrictions. For example, Commandeur (1991) proposes to require $\sum_k^K s_k^2 \text{tr} \tilde{\mathbf{X}}'_k \tilde{\mathbf{X}}_k = \sum_k^K \text{tr} \mathbf{X}'_k \mathbf{X}_k$, which we assume implicitly whenever needed.

The GPA loss function (21.1) has to be minimized with an iterative algorithm, because no direct analytical solution is known. We describe three

methods for minimizing *GPA*. The first method consists of cyclically updating one configuration while keeping the others fixed. Thus, each iteration consists of first updating $\tilde{\mathbf{X}}_1$ while keeping the remaining configurations fixed, then updating $\tilde{\mathbf{X}}_2$ while keeping the remaining configurations fixed, and so on. Writing only the terms of the *GPA* function dependent on $\tilde{\mathbf{X}}_k$ gives

$$\begin{aligned} GPA_k(\tilde{\mathbf{X}}_k) &= (K-1)\text{tr } \tilde{\mathbf{X}}_k' \tilde{\mathbf{X}}_k - 2\text{tr } \tilde{\mathbf{X}}_k' \sum_{l \neq k} \tilde{\mathbf{X}}_l + c \\ &= (K-1)(\text{tr } \tilde{\mathbf{X}}_k' \tilde{\mathbf{X}}_k - 2\text{tr } \tilde{\mathbf{X}}_k' \mathbf{Y}) + c, \end{aligned}$$

where $\mathbf{Y} = (K-1)^{-1} \sum_{l \neq k} \tilde{\mathbf{X}}_l$, and c contains terms that are not dependent on $\tilde{\mathbf{X}}_k$. The minimum of $GPA_k(\tilde{\mathbf{X}}_k)$ can be found by the Procrustean similarity transformation procedure outlined in Section 20.4. This procedure is iteratively repeated over all k s until *GPA* no longer drops. The proposed algorithm must converge, because (21.1) can never become greater as a consequence of any individual Procrustean fitting and because (21.1) has a lower bound of 0. Usually, very few iterations are required to reach convergence. The current procedure is used by Kristof and Wingersky (1971) and Ten Berge (1977).

A second procedure for solving the *GPA* problem is described by Gower (1975). Differentiating (21.1) with respect to \mathbf{t}_k , he first finds that all configurations must be translated so that their respective centroids are all incident with the origin. Hence, all \mathbf{X}_k s must be centered so that their columns sum to 0. This solves the translation problem directly. The rotation/reflection problems associated with \mathbf{T}_k can then be solved in the iterative manner described above. Finally, a direct solution exists for the scale factors s_k (Ten Berge, 1977). Let \mathbf{B} be the $K \times K$ matrix with elements $b_{kl} = \text{tr } \mathbf{X}_k' \mathbf{X}_l$ and $\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$ the eigendecomposition of \mathbf{B} . Then, the scale factors should be chosen as $s_k = (\sum_l \mathbf{X}_k' \mathbf{X}_l / \text{tr } \mathbf{X}_k' \mathbf{X}_k)^{1/2} q_{k1}$, where q_{k1} is element k of the largest eigenvector of \mathbf{B} .

A third method for minimizing *GPA* uses the centroid configuration \mathbf{Z} of all $\tilde{\mathbf{X}}_k$ s, $\mathbf{Z} = (1/K) \sum_k \tilde{\mathbf{X}}_k$. The function *GPA* in (21.1) is equivalent to

$$GPA = K \sum_{k=1}^K \text{tr } (\tilde{\mathbf{X}}_k - \mathbf{Z})' (\tilde{\mathbf{X}}_k - \mathbf{Z}), \quad (21.2)$$

which is minimized by updating the $\tilde{\mathbf{X}}_k$ s and the centroid configuration \mathbf{Z} one at a time while keeping the others fixed. Commandeur (1991) notes, however, that this procedure has slower convergence properties than the method described above.

To see that (21.1) is the same as (21.2), consider the following.

$$GPA = \sum_{k < l}^K \text{tr } (\tilde{\mathbf{X}}_k - \tilde{\mathbf{X}}_l)' (\tilde{\mathbf{X}}_k - \tilde{\mathbf{X}}_l) \quad (21.3)$$

$$= \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \text{tr} (\tilde{\mathbf{X}}_k - \tilde{\mathbf{X}}_l)' (\tilde{\mathbf{X}}_k - \tilde{\mathbf{X}}_l) \quad (21.4)$$

$$= \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \text{tr} \tilde{\mathbf{X}}_k' \tilde{\mathbf{X}}_k + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \text{tr} \tilde{\mathbf{X}}_l' \tilde{\mathbf{X}}_l \\ - \sum_{k=1}^K \sum_{l=1}^K \text{tr} \tilde{\mathbf{X}}_k' \tilde{\mathbf{X}}_l. \quad (21.5)$$

Summing the first two terms of (21.5) yields $K \sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' \tilde{\mathbf{X}}_k$. Because $\tilde{\mathbf{X}}_k$ in the last term of (21.5) does not depend on l , this term can be written as $K \sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' (K^{-1} \sum_{l=1}^K \tilde{\mathbf{X}}_l)$, so that

$$GPA = K \sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' \tilde{\mathbf{X}}_k - K \sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' (K^{-1} \sum_{l=1}^K \tilde{\mathbf{X}}_l) \\ = K \left(\sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' \tilde{\mathbf{X}}_k - \sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' \mathbf{Z} \right).$$

Using this result, the derivation continues as

$$GPA = K \left(\sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' \tilde{\mathbf{X}}_k + \sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' \mathbf{Z} - 2 \sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' \mathbf{Z} \right) \\ = K \left(\sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' \tilde{\mathbf{X}}_k + K(K^{-1} \sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k)' \mathbf{Z} - 2 \sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' \mathbf{Z} \right) \\ = K \left(\sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' \tilde{\mathbf{X}}_k + K(\text{tr} \mathbf{Z}' \mathbf{Z}) - 2 \sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' \mathbf{Z} \right) \\ = K \left(\sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' \tilde{\mathbf{X}}_k + \sum_{k=1}^K \text{tr} \mathbf{Z}' \mathbf{Z} - 2 \sum_{k=1}^K \text{tr} \tilde{\mathbf{X}}_k' \mathbf{Z} \right) \\ = K \sum_{k=1}^K \text{tr} (\tilde{\mathbf{X}}_k - \mathbf{Z})' (\tilde{\mathbf{X}}_k - \mathbf{Z}). \quad (21.6)$$

This shows that *GPA* minimizes the squared differences of all $\tilde{\mathbf{X}}_k$ to the centroid configuration \mathbf{Z} . Therefore, \mathbf{Z} can be used as the configuration that summarizes all of the optimally transformed \mathbf{X}_k s. Dijksterhuis and Gower (1991) go one step further. They provide a much more detailed analysis-of-variance-like decomposition of the error of the *GPA* loss function, so that several sources of misfit can be attributed.

Geometrically, each of \mathbf{Z} 's points is the centroid of the corresponding points from the fitted individual configurations. Thus, if (21.1) is small, these centroids lie somewhere in the middle of a tight cluster of K points, where each single point belongs to a different $\tilde{\mathbf{X}}_k$.

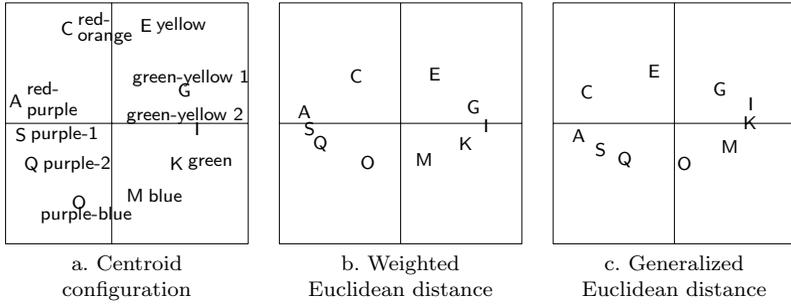


FIGURE 21.4. (a) Centroid configuration for 11 color-normal subjects in Table 21.1. (b) Example of an individual space using the weighted Euclidean distance for a hypothetical subject with dimension weights 1.0 (dim 1) and .5 (dim 2). (c) Example of an individual space using the generalized Euclidean distance for a hypothetical subject with dimension weights 1.0 (dim 1), .5 (dim 2), and idiosyncratic anticlockwise rotation of 30°.

21.4 Individual Differences Models: Dimension Weights

The generalized Procrustean transformation problem is of limited interest in practice. A more interesting question to ask is whether each individual configuration can be accounted for by stretching the centroid configuration appropriately along the dimensions. This idea for explaining individual differences was introduced by Horan (1969) and Bloxom (1968) and developed by Carroll and Chang (1970) in the INDSCAL procedure (see Section 22.1).

To illustrate the model, let us look at Figure 21.4a which shows the centroid configuration obtained by *GPA*. Every individual is allowed his or her own weights for every column of \mathbf{Z} . For example, a mildly color-deficient subject k could weight the first dimension by 1.0 and the second dimension by 0.5, showing that the second dimension accounts for less variance in his or her data than the first dimension. Figure 21.4a shows the centroid configuration obtained by *GPA*. The weighted centroid configuration in Figure 21.4b shows clearly that for this subject the first dimension of \mathbf{Z} is more important than the second dimension.

The weighted centroid configuration for subject k can be expressed as $\mathbf{Z}\mathbf{W}_k$, where \mathbf{W}_k is an $m \times m$ diagonal matrix of nonzero dimension weights. Hence, the corresponding distance between points i and j is

$$\begin{aligned}
 d_{ijk}(\mathbf{Z}\mathbf{W}_k) &= \left[\sum_{a=1}^m (w_{aak}z_{ia} - w_{aak}z_{ja})^2 \right]^{1/2} \\
 &= [(\mathbf{z}_i - \mathbf{z}_j)' \mathbf{W}_k^2 (\mathbf{z}_i - \mathbf{z}_j)]^{1/2}, \tag{21.7}
 \end{aligned}$$

where \mathbf{z}'_i is row i of \mathbf{Z} . Equation (21.7) is called the *weighted Euclidean distance*.² Note w_{aa_k} may be positive or not: a negative dimension weight simply reflects the corresponding axis but does not change the distances.

An extension of dimension weighting allows for *idiosyncratic rotations* as well. Before applying dimension weights, an individual would first orient \mathbf{Z} in his or her particular way. The transformed centroid configuration for individual k becomes $\mathbf{Z}\mathbf{S}_k\mathbf{W}_k$, where \mathbf{S}_k is a rotation matrix with $\mathbf{S}'_k\mathbf{S}_k = \mathbf{S}_k\mathbf{S}'_k = \mathbf{I}$. The *generalized Euclidean distance* is

$$d_{ijk}(\mathbf{Z}\mathbf{S}_k\mathbf{W}_k) = [(\mathbf{z}_i - \mathbf{z}_j)'\mathbf{S}_k\mathbf{W}_k^2\mathbf{S}'_k(\mathbf{z}_i - \mathbf{z}_j)]^{1/2}. \quad (21.8)$$

The use of this type of distance was popularized by Carroll and Wish (1974a) in the IDIOSCAL model (see Section 22.2). In Figure 21.4c, the perceptual space of a hypothetical individual is shown. This individual first rotates the centroid configuration of Figure 21.4a by 30° anticlockwise and then weights the newly obtained axes by $w_{11k} = 1.0$ and $w_{22k} = 0.5$.

Helm's Color Data and the Subject Space

Consider our color perception example. Figure 21.4a shows the centroid configuration \mathbf{Z} derived from the 11 MDS configurations of the color-normal subjects by minimizing (21.1). \mathbf{Z} matches each of the 11 individual configurations exceedingly well, as can be seen from the $r^2(\tilde{\mathbf{X}}_k, \mathbf{Z})$ values in Table 21.3. None of the fit values shows an agreement of less than 96%; hence, \mathbf{Z} is truly representative for these subjects. In Figure 21.4a, the coordinate axes are rotated so that the vertical dimension intersects the color circle at the points red-purple and green-blue. But these are just the colors that the deuteranopic subjects cannot reliably discriminate, although they have no problems distinguishing yellow from blue. Thus, their color circles should be squeezed together in the red-green direction, because the point distances represent the perceived dissimilarities. Figure 21.4b shows what distance structures would be expected for a mildly deuteranopic subject. In this case, the transformation can be represented in terms of point coordinates as $\tilde{\mathbf{X}}_k \approx \mathbf{Z}\mathbf{W}_k$, where $\mathbf{Z}\mathbf{W}_k$ is the approximated MDS configuration of individual k , and \mathbf{W}_k is a 2×2 diagonal matrix consisting of the weights $w_{11} = 1.00$ and $w_{22} = 0.5$, which has the effect of shrinking all coordinates in \mathbf{Z} 's second column to a half of their original magnitude.

In general, this fitting problem can be expressed as

$$\text{tr}(\mathbf{Z}\mathbf{W}_k - \tilde{\mathbf{X}}_k)'(\mathbf{Z}\mathbf{W}_k - \tilde{\mathbf{X}}_k) = \min, \quad (21.9)$$

²Actually, d_{ijk} is simply a Euclidean distance on a "weighted" MDS space, $\mathbf{X}_k = \mathbf{Z}\mathbf{W}_k$. The term "weighted Euclidean distance", therefore, characterizes formula (21.7) but does not imply that we are dealing with a special type of distance.

TABLE 21.3. Fit measures for simple and weighted Procrustean analyses of MDS configurations of Helm data, split by color-normal subjects and color-deficient subjects.

Color-Normals	$r^2(\tilde{\mathbf{X}}_k, \mathbf{Z})$	$r^2(\tilde{\mathbf{X}}_k, \mathbf{Z}\mathbf{W}_k)$	Color-Deficients	$r^2(\tilde{\mathbf{X}}_k, \mathbf{Z})$	$r^2(\tilde{\mathbf{X}}_k, \mathbf{Z}\mathbf{W}_k)$
s_1	0.98	0.98	s_{11}	0.89	0.92
s_2	0.98	0.98	$s_{12}^{[1]}$	0.91	0.92
s_3	0.99	0.99	$s_{12}^{[2]}$	0.94	0.97
s_4	0.99	0.99	s_{13}	0.50	0.75
s_5	0.99	0.99	s_{14}	0.44	0.82
$s_6^{[1]}$	0.97	0.97	Average	0.74	0.88
$s_6^{[2]}$	0.98	0.99			
s_7	0.98	0.98			
s_8	0.97	0.97			
s_9	0.98	0.98			
s_{10}	0.96	0.97			
Average	0.98	0.98			

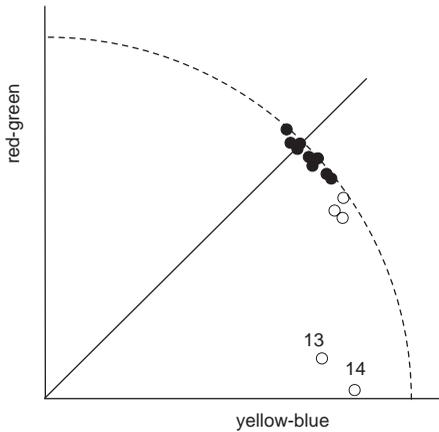


FIGURE 21.5. Subject space for centroid configuration in Fig. 21.4a and data in Table 21.1; solid points represent color-normal and open points color-deficient subjects.

where \mathbf{W}_k is the unknown diagonal matrix of weights, and $\tilde{\mathbf{X}}_k$ is the i th individual configuration optimally fitted to \mathbf{Z} by similarity transformations. Because \mathbf{W}_k is diagonal, finding the best \mathbf{W}_k amounts to solving a set of simple regression problems. To see this, let \mathbf{z}_a and \mathbf{x}_a be the column vectors of \mathbf{Z} and $\tilde{\mathbf{X}}_a$, respectively, and z_{ia} and x_{ia} the i th elements in these vectors. Then, find the weight w_a (the a th diagonal element of \mathbf{W}_k) for \mathbf{z}_a such that $\sum_{i=1}^n (w_a z_{ia} - x_{ia})^2$ is minimal. Differentiating and setting the derivative equal to 0 leads to $w_a = (\sum_i x_{ia} z_{ia}) / \sum_i z_{ia}^2$, the formula for the regression coefficient. With such weights in each \mathbf{W}_k , the agreement of the $\mathbf{Z}\mathbf{W}_k$ s and the individual configurations of the deuteranopic subjects goes up substantially relative to the unweighted case (Table 21.3). The size of the increments mirrors the degree of deuteranopia.

If the weights are normalized appropriately (see below), they can be displayed as in Figure 21.5. The diagram, known as a *subject space* (Carroll & Chang, 1970), shows the color-normal persons represented by solid points and the color-deficient subjects by open points. The coordinates of these points correspond to the weights assigned to the dimensions by the respective individual. The color-normal persons weight the centroid configuration on both the red-green (X -axis) and the yellow-blue (Y -axis) dimensions about equally, because the points cluster tightly around the bisector. The color-deficient persons, on the other hand, weight the red-green dimension of \mathbf{Z} less than the yellow-blue dimension. Moreover, the open point closest to the yellow-blue axis represents s_{14} , the individual with the most severe case of deuteranopia, and the open point next to it stands for s_{13} , who is next in color deficiency. For these individuals, the red-green dimension is practically irrelevant for their dissimilarity judgments, as predicted.

For the subject space, the dimension weights w_{aak} were normed so that their sum-of-squares is equal to $r^2(\tilde{\mathbf{X}}_k, \mathbf{Z}\mathbf{W}_k)$. This equality holds if the normed weights, \bar{w}_{aak} , satisfy $\bar{w}_{aak} = w_{aak} / (\sum_i z_{ia}^2)^{1/2}$, which follows from writing out the squared correlation $r^2(\tilde{\mathbf{X}}_k, \mathbf{Z}\mathbf{W}_k)$ in scalar notation (Borg, 1977a). Thus, the distance of the points in the subject space from the origin corresponds to the communality of the weighted average configuration and an individual configuration. We have $\tilde{\mathbf{X}}_k = \mathbf{Z}\mathbf{W}_k$ if the weight point of individual k lies on the circle with radius 1 around the origin.

Common Misinterpretations of the Subject Space

We should note here that the subject space depends on how the group space \mathbf{Z} is defined. In the above example, \mathbf{Z} was the centroid configuration of the color-normal persons only. Alternatively, it would also be possible to derive \mathbf{Z} from, say, the configurations of all subjects. But then \mathbf{Z} would have a different shape: it would be more elliptical, and this would entail that all points and stars in the subject space be rotated towards the red-green dimension, so that the color-normals would not be distributed around

the bisector. Hence, it is not possible to infer from a subject's point in the subject space that he or she weights dimension a more than dimension b , because these ratios change if we change the group space. The weights, thus, are only relative to those of other subjects.

Another misinterpretation can occur if we take the name "subject space" in the sense of a Euclidean point space. The meaning of distances computed by the Euclidean distance formula in the subject space is completely obscure, because each of its points is placed so that the distance from the origin denotes the communality of the respective configuration $\tilde{\mathbf{X}}_k$ with \mathbf{Z} , and the direction of the ray on which the point lies represents the weights in \mathbf{W}_k . Moreover, as we mentioned above, the distances in the subject space are conditional on how \mathbf{Z} is defined: if we choose different \mathbf{Z} s, different subject spaces result. To avoid the distance interpretation between points in the subject space, one can project the points on the dimensions and interpret the weights for each dimension separately.

Dimension Weighting with Idiosyncratic Rotations

The dimensional weighting of \mathbf{Z} in Figure 21.4 was done along the given dimensions red-green and yellow-blue. Different ellipses would have resulted in Figures 21.4b and 21.4c if \mathbf{Z} were squeezed together in other *directions*. For example, we could squeeze \mathbf{Z} in the direction purple 1/green-yellow 2, which would bring the points S and I in close proximity. Expressed more technically, if the dimensional system in Figure 21.4a were rotated by \mathbf{S} , then weighting \mathbf{ZS} by \mathbf{W}_k would lead to a different result than weighting \mathbf{Z} by \mathbf{W}_k , in general. Thus, for each \mathbf{S} , we obtain the loss function

$$\text{tr}(\mathbf{ZS}\mathbf{W}_k - \tilde{\mathbf{X}}_k)'(\mathbf{ZS}\mathbf{W}_k - \tilde{\mathbf{X}}_k), \quad (21.10)$$

where $\mathbf{S}'\mathbf{S} = \mathbf{I}$, and $\tilde{\mathbf{X}}_k$ is an individual configuration fitted optimally to \mathbf{ZS} . Note that \mathbf{S} does not have a subscript here, so that \mathbf{ZS} is the group space for all individuals.

How do we minimize (21.10) over all $k = 1, \dots, K$ individuals? To do this, we need to find the best \mathbf{S} in

$$L = \sum_{k=1}^K \text{tr}(\mathbf{ZS}\mathbf{W}_k - \mathbf{X}_k\mathbf{T}_k)'(\mathbf{ZS}\mathbf{W}_k - \mathbf{X}_k\mathbf{T}_k), \quad (21.11)$$

where we write $\mathbf{X}_k\mathbf{T}_k$ (with $\mathbf{T}_k'\mathbf{T}_k = \mathbf{I}$) for $\tilde{\mathbf{X}}_k$, because it turns out that the optimal translation of an individual configuration is always to center it, and because the optimal scaling factor becomes irrelevant when correlations are used as similarity measures (Lingoes & Borg, 1978). Equation (21.11) involves the unknown \mathbf{S} , and K unknown weight matrices \mathbf{W}_k and rotations \mathbf{T}_k . To find the optimal matrices is a difficult problem, and it is useful to consider a simpler case first.

Let \mathbf{S}_k be an *idiosyncratic* rotation, that is, a different rotation matrix \mathbf{S}_k for every subject k . Find \mathbf{S}_k and \mathbf{W}_k in

$$\text{tr}(\mathbf{ZS}_k\mathbf{W}_k - \mathbf{X}_k\mathbf{T}_k)'(\mathbf{ZS}_k\mathbf{W}_k - \mathbf{X}_k\mathbf{T}_k). \quad (21.12)$$

In terms of our color perception example, we want to find a rotation \mathbf{S}_k and a set of dimensional weights \mathbf{W}_k that distort the color circle in Figure 21.4a such that individual k -th's configuration, rotated appropriately by \mathbf{T}_k , is approximated as closely as possible. A direct solution for this problem is known only for the 2D case (Lingoes & Borg, 1978; Mooijaart & Commandeur, 1990). This solution can be used for each plane of the space in turn, and then one can repeat the fittings iteratively, because every $m \times m$ rotation matrix can be expressed as the product of $m(m-1)/2$ planar rotation matrices (see also Section 7.10). The average of all \mathbf{ZS}_k s is then used as a target matrix to solve for the \mathbf{ZS} of (21.11).

We obtain a group space that is uniquely rotated; that is, L in (21.11) is minimal for one particular \mathbf{S} and increases for any other rotation, except in special cases. This will be so for any \mathbf{Z} , whether it represents empirical data or whether it has been defined completely at random. Hence, to conclude that "this method automatically gives psychologically meaningful axes, if they exist" (Indow, 1974, p.497) appears too optimistic. It may just be the case that no dimensional theory is psychologically relevant and that the dimension-weighting model, which leads to the unique dimensions, yields nothing but substantively empty formal relations. But even when the model is adequate, we should keep in mind that the rotational uniqueness is an algebraic property. For real data, which always have error components, it may just be that the resulting dimensions are a consequence of the particular error distribution. In any case, we usually find that the rotational uniqueness is statistically weak. Thus, for formal as well as substantive reasons, we recommend rotating \mathbf{Z} such that the dimensions reflect some substantive theory (as in Figure 21.4) rather than leaving the finding of the coordinate axes to a blind procedure.

21.5 An Application of the Dimension-Weighting Model

Consider another example. Green and Rao (1972) asked 41 individuals to evaluate the 15 breakfast objects described in Table 14.1 with respect to their pairwise similarities. (The 41 proximity matrices are presented in Green & Rao's book.) By an ordinal MDS of each of the 41 15×15 proximity matrices, 2D representations are obtained. Using the program PINDIS (Lingoes & Borg, 1977), the centroid configuration \mathbf{Z} shown in Figure 21.6 is derived. The representation shows \mathbf{Z} in its optimal rotation with respect to the loss function (21.12), so that we already have \mathbf{ZS} . The

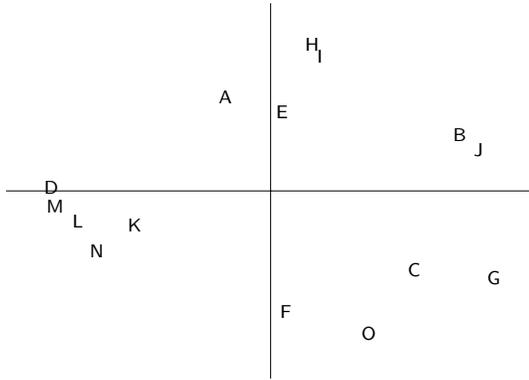


FIGURE 21.6. The PINDIS centroid configuration derived from 41 individual configurations obtained by MDS of the 15 breakfast items reported in Table 14.1 (after Borg & Lingoes, 1977).

fit measures (squared correlations, common variances, or *communalities*) between this group space and each individual configuration are given in Table 21.4 in the column $r^2(\tilde{\mathbf{X}}_k, \mathbf{Z})$. These values vary substantially over the individuals. Although $\tilde{\mathbf{X}}_{22}$ and \mathbf{Z} share some 92% of their variance, $\tilde{\mathbf{X}}_{38}$ and \mathbf{Z} have almost nothing in common.

There is one typical misinterpretation of such results. Table 21.4 shows that, for example, $\tilde{\mathbf{X}}_{25}$ and $\tilde{\mathbf{X}}_{35}$ correlate with \mathbf{Z} with about $r^2 = .50$. Yet we cannot infer from this value that $\tilde{\mathbf{X}}_{25}$ and $\tilde{\mathbf{X}}_{35}$ have anything in common: a simple, pairwise Procrustean analysis could show that $r^2(\tilde{\mathbf{X}}_{25}, \tilde{\mathbf{X}}_{35}) = 0$. It is easy to see why this is so. $\tilde{\mathbf{X}}_{25}$ and $\tilde{\mathbf{X}}_{35}$ each share some 50% variance with \mathbf{Z} , but these variance proportions may be complementary, so that \mathbf{Z} shares with $\tilde{\mathbf{X}}_{25}$ one-half of its variance, and with $\tilde{\mathbf{X}}_{35}$ the remaining half. To see how similar $\tilde{\mathbf{X}}_{25}$ and $\tilde{\mathbf{X}}_{35}$ are, we would have to do a pairwise Procrustean analysis.

We now use dimensional weightings with and without idiosyncratic rotations. This is a purely formal exercise here because, in contrast to the color perception data considered above, there is no reason why the individuals should perceive the similarity of these breakfast items dimensionally, and also no reason why they should differ with respect to the importance of dimensions. This lack of a substantive theory is, in fact, evidenced by the very fact that we use such blindly optimizing rotations in the first place. Not surprisingly, it turns out that both dimension-weighting models do not account for much additional variance relative to the model using unit weights. Table 21.4 shows the respective fit values in columns $r^2(\tilde{\mathbf{X}}_k, \mathbf{ZSW}_k)$ and $r^2(\tilde{\mathbf{X}}_k, \mathbf{ZS}_k\mathbf{W}_k)$. On the average, the fit increments are just 2.6% and 4.5%, and in no individual case is there an increment of the magnitude found in Table 21.3 for the severely color-deficient persons.

TABLE 21.4. Communalities of individual configurations fitted to different transformations of centroid configuration \mathbf{Z} .

Subject	$r^2(\tilde{\mathbf{X}}_k, \mathbf{Z})$	$r^2(\tilde{\mathbf{X}}_k, \mathbf{ZS}_k)$	$r^2(\tilde{\mathbf{X}}_k, \mathbf{ZS}_k \mathbf{W}_k)$	$r^2(\tilde{\mathbf{X}}_k, \mathbf{V}_k \mathbf{Z})$
1	0.7999	0.8005	0.8008	0.8335
2	0.8520	0.8713	0.8714	0.8926
3	0.9088	0.9112	0.9159	0.9519
4	0.9194	0.9222	0.9283	0.9488
5	0.5669	0.6811	0.7352	0.8410
6	0.8939	0.9056	0.9065	0.9376
7	0.8376	0.8469	0.8480	0.8692
8	0.8365	0.9014	0.9032	0.8900
9	0.8518	0.8520	0.8849	0.8977
10	0.7369	0.7404	0.7439	0.8407
11	0.7765	0.7958	0.8803	0.8631
12	0.7044	0.7188	0.7649	0.8099
13	0.7833	0.9152	0.9161	0.9154
14	0.7772	0.7814	0.8020	0.9219
15	0.8982	0.9175	0.9175	0.9339
16	0.6199	0.6698	0.6747	0.7751
17	0.6871	0.7765	0.7805	0.8015
18	0.7881	0.8174	0.8324	0.8821
19	0.8050	0.8469	0.8493	0.8540
20	0.1118	0.1307	0.1307	0.7782
21	0.6179	0.6310	0.6631	0.6754
22	0.9222	0.9429	0.9470	0.9588
23	0.8770	0.8794	0.9005	0.9184
24	0.8721	0.8777	0.8785	0.8866
25	0.5101	0.5135	0.5419	0.8383
26	0.6827	0.6884	0.6895	0.7731
27	0.8251	0.8280	0.8385	0.8712
28	0.7198	0.7268	0.7644	0.8292
29	0.8493	0.8931	0.8936	0.9199
30	0.8593	0.8978	0.9068	0.9289
31	0.3929	0.4067	0.4695	0.8143
32	0.2728	0.2994	0.3642	0.6585
33	0.8498	0.8700	0.8776	0.9448
34	0.7973	0.8299	0.8451	0.8860
35	0.5126	0.6170	0.6623	0.6976
36	0.6076	0.6212	0.6894	0.6806
37	0.8137	0.8192	0.8322	0.8786
38	0.0192	0.0262	0.0331	0.4690
39	0.7077	0.7426	0.7478	0.8903
40	0.7824	0.8004	0.8019	0.8306
41	0.8559	0.8581	0.9178	0.9178
Mean	0.7196	0.7456	0.7647	0.8465

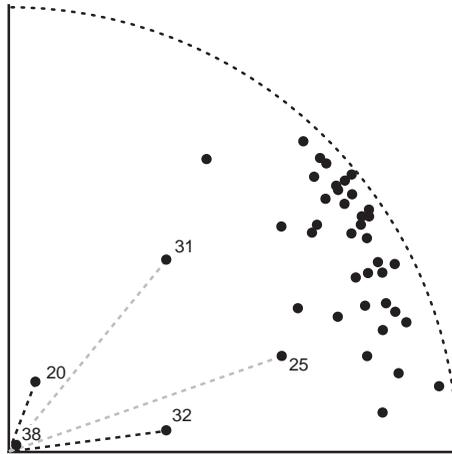


FIGURE 21.7. The subject space for the dimension-weighting model of the Green–Rao data (after Borg & Lingoës, 1977).

Most programs (e.g., INDSICAL; see Chapter 22) designed specifically to represent data in the dimension-weighting model skip the step with the unit weights, which yields the fit values in column $r^2(\bar{\mathbf{X}}_k, \mathbf{Z})$ of Table 21.3. In other words, they analyze the entire $41 \times 15 \times 15$ data block at once in the sense of the dimension-weighting models, which yields a group space and its related subject space of weights. This information alone is difficult to interpret, however. Consider the subject space for the dimension-weighting model of the present data (Figure 21.7). We observe that: (1) the subject points scatter considerably in their distance from the origin, which expresses the different communalities of \mathbf{ZS} and the individual configurations; (2) the subject points also scatter in terms of their North-West directions, and this, as we saw, indicates that the subjects weight the dimensions differently. But, as Table 21.4 demonstrates, this second scatter really does not mean much, because if all of the points were forced onto the bisector, the model would be reduced to the unit-weighting case, whose average communality is just 2.6% lower. Thus, it would be risky to infer that because the subject points scatter so much in terms of direction, the differential weights (and with them, the particular dimensions) are meaningful or even descriptively important. The scatter simply reflects the fact that no restrictions were placed on the Procrustean procedure, so that whatever reduces the loss function most is chosen for a weight.

21.6 Vector Weightings

Because the dimension-weighting models proved ineffective in explaining the interindividual differences in the Green–Rao breakfast data, we might

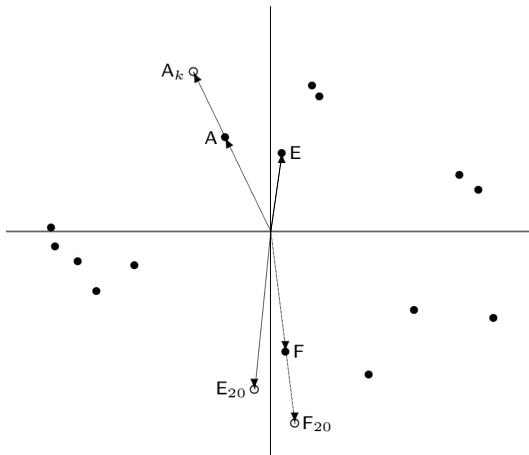


FIGURE 21.8. Illustration of some vector weights for centroid configuration and MDS configuration of subject 20 in Green–Rao study.

seek other, more successful models. Dimensional weightings are constrained by the property that the neighborhood relations of the points in \mathbf{Z} are preserved in a certain way when transforming \mathbf{Z} into $\mathbf{Z}\mathbf{W}_k$. Namely, those points that are close together in \mathbf{Z} are also close together in $\mathbf{Z}\mathbf{W}_k$.

Radial Point Shifts by Vector Weightings

For the Green–Rao breakfast data, it may be possible that most individuals do indeed perceive the breakfast objects just as the *average person* \mathbf{Z} does but that they see some neighborhood relations differently. We consider a particularly simple transformation, whose loss function is expressed by

$$\sum_{k=1}^K \text{tr} (\mathbf{V}_k \mathbf{Z} - \tilde{\mathbf{X}}_k)' (\mathbf{V}_k \mathbf{Z} - \tilde{\mathbf{X}}_k), \tag{21.13}$$

where \mathbf{V}_k is an $n \times n$ diagonal matrix of unknown weights and $\tilde{\mathbf{X}}_k$ is the individual configuration \mathbf{X}_k optimally fitted to $\mathbf{V}_k \mathbf{Z}$. The elements in \mathbf{V}_k act on the points and are called *vector weights*. Formally, (21.13) differs from (21.9) only insofar as \mathbf{Z} is now weighted by a diagonal matrix from the left, not the right. The solution of (21.13) is simple in the 2D case, but to find all transformations (\mathbf{V}_k and all those on \mathbf{X}_k) in higher-dimensional spaces simultaneously appears intractable. Hence, to minimize (21.13) we have to iterate over all planes of the space (see Lingoes & Borg, 1978). We continue with an example of the vector-weighting model, and then discuss several ways to interpret and apply vector weights.

Analyzing \mathbf{X}_{20} and the \mathbf{Z} configuration from Figure 21.6 with the vector-weighting model yields 15 weights, one for each point. Most of these weights

are very close to +1, except those for points A , E , and F , where the procedure finds 1.7, 1.6, and -1.4 , respectively. Premultiplying \mathbf{Z} by \mathbf{V}_{20} has the effect shown in Figure 21.8. Point A is shifted away from the origin in the direction and sense of the vector associated with it. In other words, the vector with endpoint A is simply stretched by the factor 1.7. The analogous movement is true for point F . For point E , in contrast, the weight -1.4 not only stretches the respective vector but also flips it over, or *reverses its sense*. From Table 21.4, column $r^2(\tilde{\mathbf{X}}_k, \mathbf{V}_k \mathbf{Z})$, we see that these movements lead to a communality increment of almost 70% relative to the unweighted Procrustean fitting. Thus, moving the points A , E , and F into different neighborhoods in the described way seems to capture an important characteristic in which person 20 differs from most others.

Evaluating the Fit in Vector Weighting

Table 21.4 shows that the vector weightings allow a much better approximation of the individual configurations than the dimension-weighting models. However, the dimension-weighting models use only 2 (the dimensional weights) or 3 (the idiosyncratic rotation angle, in addition) parameters, but the vector weightings use up to 15 parameters. Of course, the sheer number of free parameters cannot be compared directly if the models are restricted in different ways, but simulation studies (Langeheine, 1980a, 1982) show that the vector-weighting model can be expected to fit 2D random configurations considerably better than dimensional weightings. For $K = 41$, $m = 2$, and $n = 15$, the average fit value of 0.169 was found for the unweighted Procrustean fitting, 0.186 for the dimension-weighting model, 0.200 for the dimension-weighting model with idiosyncratic orientation, and 0.699 for the vector-weighting model. If we evaluate the observed fit values against these expectations for random configurations, the performance of the vector weighting is less impressive in this example.

As the dimensionality m goes up, the dimension-weighting model offers increasingly more fitting parameters, whereas their number remains constant in the vector-weighting model. This partly explains why, when m goes up and n remains constant, the communalities for random configurations grow substantially for unweighted Procrustean fittings and dimensional weightings but do not increase much for the vector-weighting models. Naturally, this is also a consequence of how these parameters are used in the analyses. Because there are various complicated interdependencies, it becomes difficult to say what should be expected for random configurations in general, but fortunately Langeheine (1980a) provides extensive tables.

Interpreting Vector Weights

In contrast to dimensional weighting, there is no convincing interpretation of vector weighting as a psychological model. For dimension weights,

such terms as *relative importance* or *saliency* may be appropriate. No such interpretations can be given to the vector weights, except when their values are constrained to be nonnegative (see below: the perspective model). However, vector weighting may provide valuable index information. If we find that an optimal fitting of \mathbf{Z} to each individual configuration can be done only with weights varying considerably around +1, then it makes little sense to consider the centroid configuration \mathbf{Z} as a structure *common* to all individuals. To see this, consider the distribution of voters in France. This distribution is almost perfectly bimodal over the political left-right continuum. But it would be foolish to say that “the” French voter is politically “in the middle”. In fact, no one really is. Similarly, in respect to centroid configurations, it may just be the case that this configuration does not really represent anybody. But there may be groups of individuals with very similar perceptions. Whether this is so may be seen from studying the distributions of the vector weights.

Because we can also arrive at such conclusions by directly studying the data, we return to the question of whether there is an interpretation of vector weighting as a psychological model. The answer is yes, but only under some restrictions on (21.13). One possibility would be to carry out the vector weightings with respect not to the centroid of \mathbf{Z} but to a substantively meaningful origin. In a radex (see Chapter 5), for example, the centroid is extrinsic to the scientific problem under investigation, but the point chosen as the radex origin is not. If several such radexes were given, we could first translate them all such that their origins lie at these points. If we then fit each individual configuration to an average configuration derived from them all, the vector weights would express the different relative centrality of the points.

A different interpretation is given by Feger (1980) which clarifies why the vector-weighting transformations are called the *perspective model* under certain conditions. Feger asked nine subjects to rate pairs of 10 attitude objects (the six major political parties in West Germany; the trade unions; the Church; the employers’ association; the subject him- or herself) with respect to the criterion “closeness”. The ratings were replicated 12 times in intervals of 3 weeks. The data led to 108 2D MDS configurations. A centroid configuration was computed, and all configurations were translated such that their origins were at the points representing the object *self*. Using vector weightings on \mathbf{Z} , it was possible to explain most of the intra- and interindividual differences. It seemed as if the individuals perceived the attitude objects from this perspective in space, sometimes pulling some objects closer to themselves, sometimes pushing them farther away. Feger (1980) goes on to interpret the *self* point as the ideal point of the subjects (just as in the unfolding models) and the distances from the ideal point to the (possibly shifted) points of the nine other objects as indicators of the strength of preference of the respective person for these objects. In this interpretation, the point shiftings assessed by the vector weights are

due to changes of preferences over time and individuals. This complex but interesting interpretation implies a dependency between similarity judgments and preferences but becomes intractable if negative vector weights are observed.

Adding an Idiosyncratic Origin to Vector Weighting

Finally, as in the idiosyncratic rotations in the dimension-weighting model, we can generalize the vector-weighting transformations to a model with an idiosyncratic origin. In other words, rather than fixing the perspective origin externally either at the centroid or at some other more meaningful point, it is also possible to leave it to the model to find an origin that maximizes the correspondence of an individual configuration and a transformed \mathbf{Z} . The relevant loss function becomes rather complex:

$$\sum_{k=1}^K \text{tr } \mathbf{E}'_k \mathbf{E}_k, \text{ with } \mathbf{E}_k = \mathbf{V}_k(\mathbf{Z} - \mathbf{1}\mathbf{t}'_k) - s_k(\mathbf{X}_k - \mathbf{1}\mathbf{u}'_k)\mathbf{T}_k, \quad (21.14)$$

where \mathbf{t}_k and \mathbf{u}_k are translation vectors for \mathbf{Z} and \mathbf{X}_k , respectively, $\mathbf{T}'_k \mathbf{T}_k = \mathbf{I}$, \mathbf{V}_k is diagonal, and s_k is a scalar. Differentiating (21.14) with respect to the unknowns \mathbf{V}_k , \mathbf{t}_k , \mathbf{u}_k , s_k , and \mathbf{T}_k shows that none of these unknowns is redundant and that they are interrelated in a complicated way (Lingoes & Borg, 1978). However, minimizing (21.14) is uninteresting in itself, because what was true for idiosyncratic rotations holds here: an origin chosen by substantive considerations is always preferable to one found by blind optimization. The latter may, at best, serve an exploratory purpose as an index.

21.7 PINDIS, a Collection of Procrustean Models

The transformations of \mathbf{Z} for individual k discussed so far are the dimension weights \mathbf{W}_k , the idiosyncratic rotations \mathbf{S}_k , and the vector weightings \mathbf{V}_k . Table 21.5 summarizes all combinations of these transformations and the resulting models. Note that the models with the idiosyncratic rotation \mathbf{S}_k but without the dimension weights \mathbf{W}_k are equivalent to the same model without \mathbf{S}_k . In this case, the idiosyncratic rotation appears both in \mathbf{Z} and in $\tilde{\mathbf{X}}_k$, so that one of them can be omitted. The vector dimension-weighting model, $\mathbf{V}_k \mathbf{Z} \mathbf{W}_k$, and the full model, $\mathbf{V}_k \mathbf{Z} \mathbf{S}_k \mathbf{W}_k$, are difficult to interpret. This explains why the other models are more popular.

Most of the Procrustean transformations discussed above are carried out by the program PINDIS³ (Lingoes & Borg, 1977). The program needs as in-

³A good overview of the least-squares estimation of the PINDIS models can be found in Commandeur (1991). Moreover, his MATCHALS algorithm can handle entire rows of miss-

TABLE 21.5. Overview of transformations of \mathbf{Z} in Procrustes models. For each model a ‘+’ indicates the presence of the factor, a ‘-’ the absence. The factors are: dimension weights \mathbf{W}_k , idiosyncratic rotations \mathbf{S}_k , vector weights \mathbf{V}_k . \mathbf{E}_k denotes the error of the model, and $\text{tr } \mathbf{E}'_k \mathbf{E}_k$ is the loss function minimized.

\mathbf{W}_k	\mathbf{S}_k	\mathbf{V}_k	Model	\mathbf{E}_k
-	-	-	<i>GPA</i>	$\mathbf{Z} - \tilde{\mathbf{X}}_k$
+	-	-	Dimension weighting	$\mathbf{Z}\mathbf{W}_k - \tilde{\mathbf{X}}_k$
-	+	-	<i>GPA</i>	$\mathbf{Z}\mathbf{S}_k - \tilde{\mathbf{X}}_k \Leftrightarrow \mathbf{Z} - \tilde{\mathbf{X}}_k$
-	-	+	Vector weighting	$\mathbf{V}_k\mathbf{Z} - \tilde{\mathbf{X}}_k$
+	+	-	Idiosyncratic rotations	$\mathbf{Z}\mathbf{S}_k\mathbf{W}_k - \tilde{\mathbf{X}}_k$
+	-	+	Vector and dimension weighting	$\mathbf{V}_k\mathbf{Z}\mathbf{W}_k - \tilde{\mathbf{X}}_k$
-	+	+	Vector weighting	$\mathbf{V}_k\mathbf{Z}\mathbf{S}_k - \tilde{\mathbf{X}}_k \Leftrightarrow \mathbf{V}_k\mathbf{Z} - \tilde{\mathbf{X}}_k$
+	+	+	Full	$\mathbf{V}_k\mathbf{Z}\mathbf{S}_k\mathbf{W}_k - \tilde{\mathbf{X}}_k$

put K individual configurations (\mathbf{X}_k). From these, it computes a centroid configuration \mathbf{Z} via the generalized Procrustean fitting in (21.1). Alternatively, we can input some configuration \mathbf{Z} derived externally from, say, substantive considerations. \mathbf{Z} can also be based on an empirical configuration that is fixed in some desirable rotation and/or translated to some meaningful origin.

The various models and transformations discussed above are summarized in Table 21.6. $\tilde{\mathbf{X}}_k$ always denotes an \mathbf{X}_k optimally rotated and translated relative to the (weighted, rotated, translated) \mathbf{Z} that tries to account for it. For the Procrustean transformations involving rigid motions and dilations only, all of the parameters chosen to maximize $r^2(\tilde{\mathbf{X}}_k, \mathbf{Z})$ are *admissible* and, thus, uninformative because they leave the distance ratios of \mathbf{X}_k and \mathbf{Z} invariant. *Informative* are those parameters on \mathbf{Z} that change the ratios of its distances directly or in combination with other transformations that are applied at the same time. The dimension-weighting model uses, in general, different weights for each of the m dimensions of \mathbf{Z} , which has the effect of changing the distance ratios of \mathbf{Z} . Hence, these weight parameters are informative about simple ways in which \mathbf{Z} relates to \mathbf{X}_k . In contrast to admissible fitting parameters that cannot possibly be interpreted in a substantive sense, informative parameters are potential candidates for interpretations: the dimensional weights, for example, might be viewed as *dimensional saliences*. Similarly, the rotation angles in the dimension-weighting model with idiosyncratic rotation lead to the interpretation that this subject uses dimensions differently from those chosen by the average person. Because there is one such angle for each of the $m(m - 1)/2$ planes, this model has $m(m - 1)/2$ additional inadmissible parameters. The perspec-

ing values in the coordinate matrices. Ten Berge et al. (1993) discuss a *GPA* algorithm in which only the missing values themselves are discarded.

TABLE 21.6. Overview of the transformations in PINDIS. \mathbf{Z}^r = optimally rotated \mathbf{Z} in model 2; \mathbf{Z}_k^r = idiosyncratically optimal \mathbf{Z} in model 3; \mathbf{W}_k^r is \mathbf{W}_k relative to \mathbf{Z}_k^r . Similarly for \mathbf{Z}^t , \mathbf{V}^t , and \mathbf{Z}_k^t , where t denotes an optimal translation.

Model	Number of Informative Fitting Parameters	Fit Index
(1) Similarity transformation (unit weighting)	0	$r^2(\tilde{\mathbf{X}}_k, \mathbf{Z})$
(2) Dimension weighting (dimensional salience)	m	$r^2(\tilde{\mathbf{X}}_k, \mathbf{Z}^r \mathbf{W}_k)$
(3) Dimension weighting with Idiosyncratic orientation	$m + m(m - 1)/2$	$r^2(\tilde{\mathbf{X}}_k, \mathbf{Z}^r \mathbf{W}_k^r)$
(4) Perspective model with fixed origin (vector weighting)	n	$r^2(\tilde{\mathbf{X}}_k, \mathbf{V}_k \mathbf{Z})$
(5) Perspective model with idiosyncratic origin	$n + m$	$r^2(\tilde{\mathbf{X}}_k, \mathbf{V}_k^t \mathbf{Z}_k^t)$

tive model involves n inadmissible fitting parameters, one for each point. In its generalized version with an idiosyncratic origin, there are m additional parameters corresponding to the m coordinates of the freely chosen origin of \mathbf{Z} .

The PINDIS transformations form two genuine hierarchies: the models denoted as 1, 2, and 3 in Table 21.6 establish one such set of nested approaches, and models 1, 4, and 5 the other. Moreover, in practical applications, n is usually much greater than m , so that $0 < m < m + m(m - 1)/2 < n < n + m$ results. Thus, in terms of complexity, the models in PINDIS are linearly ordered. The order of the various communality values typically mirrors the order of complexity.

21.8 Exercises

Exercise 21.1 Consider the data matrix from Exercise 1.6 on p. 16.

- (a) Compute Euclidean distances for its columns. Then scale these distances in a two-dimensional MDS space (ordinal MDS).
- (b) Next, randomly eliminate 20, 30, and 40 percent of the Euclidean distances and consider these distances as missing data. With the remaining distance values run three ordinal MDS analyses in 2D.
- (c) Then compare the four configurations using Procrustean methods. (Hint: Would you choose one fixed target, or would you rather do one generalized Procrustean fitting?)

- (d) Discuss the similarity of the MDS solutions in terms of robustness of MDS.

Exercise 21.2 The table below contains the coordinates of four configurations, \mathbf{X}_1 to \mathbf{X}_4 .

Point	\mathbf{X}_1		\mathbf{X}_2		\mathbf{X}_3		\mathbf{X}_4	
1	1	2	2	1	1	1	2.64	0.50
2	-1	2	-2	1	-1	1	-1.00	1.32
3	-1	-2	-2	-1	-1	-1	-2.64	-0.50
4	1	-2	2	-1	1	-1	1.00	-1.32

- (a) Find, by geometric means, the centroid configuration \mathbf{Z} of the configurations \mathbf{X}_1 to \mathbf{X}_3 . (Hint: Plot the configurations in one chart, then determine \mathbf{Z} 's points.)
- (b) Find the dimension weights that turn \mathbf{Z} into \mathbf{X}_1 , \mathbf{X}_2 , and \mathbf{X}_3 , respectively.
- (c) Find the dimension weights that turn (a possibly rotated) \mathbf{Z} into \mathbf{X}_4 .
- (d) Characterize in which way rotating \mathbf{Z} and then weighting its dimensions affects the resulting configuration?

Exercise 21.3 Consider the group space in Figure 21.6 on p. 463.

- (a) Interpret this configuration and its four clusters.
- (b) Interpret what weighting the X - and the Y -axis of this configuration means in terms of the perceived similarity of the breakfast items.
- (c) Assume that one particular person assigns a vector weight of -1 to item B, and weights of about $+1$ to all other items. What do you conclude from this?
- (d) Assume that we decided to adopt an interpretation for this group space with two orthogonal dimensions whose ends are defined by the characteristics of the items in the four clusters in Figure 21.6. In that case, we may also decide to translate the group space to a more meaningful origin. Sketch in Figure 21.6 what seems the most reasonable origin to you. Discuss what implications such a shift of origin has for dimensional- and for vector-weighting models.