22 Three-Way MDS Models

In the Procrustean context, the dimension-weighting model was used in order to better match a set of K given configurations $\mathbf{X}_k$ to each other. We now ask how a solution of the dimension-weighting model can be found directly from the set of K proximity matrices without first deriving individual MDS spaces $\mathbf{X}_k$ for each individual k. We discuss how dimension weighting can be incorporated into a framework for minimizing Stress. Another popular algorithm for solving this problem, Indscal, is considered in some detail. Then, some algebraic properties of dimension-weighting models are investigated. Finally, matrix-conditional and -unconditional approaches are distinguished, and some general comments on dimension-weighting models are made. Table 22.1 gives an overview of the (three-way) Procrustean models discussed so far and the three-way MDS models of this chapter.

22.1 The Model: Individual Weights on Fixed Dimensions

We now return to procedures that find a solution to dimension-weighting models directly. That is, given K proximity matrices, each of order $n \times n$, a group space and its associated subject space are computed without any intermediate analyses. This situation is depicted in Figure 3.10.

TABLE 22.1. Overview of models for three-way data in Chapters 21 and 22.

Distance	$\mathbf{X}_k$	Model	Chapter/ Section
Equal dimension weights	Given	Generalized Procrustes	21.3
Equal dimension weights	Derived from $\mathbf{P}_k$	Identity model Stress	8.6
Equal dimension weights	Derived from $\mathbf{P}_k$	Classical scaling	12
Weighted Euclidean	Given	PINDIS	21.4
Weighted Euclidean	Derived from $\mathbf{P}_k$	Three-way Stress	22.1
Weighted Euclidean	Derived from $\mathbf{P}_k$	INDSCAL	22.1
Generalized Euclidean	Given	Oblique Procrustes	21.4
Generalized Euclidean	Derived from $\mathbf{P}_k$	Three-way Stress	22.2
Generalized Euclidean	Derived from $\mathbf{P}_k$	IDIOSCAL	22.2

The Weighted Euclidean Model

The problem consists of representing the dissimilarity δ_{ijk} between objects i and j as seen by individual (or replication) k by the distance d_{ijk}:

$$\begin{aligned} d_{ijk}(\mathbf{GW}_k) &= \left[\sum_{a=1}^{m}(w_{aak}g_{ia} - w_{aak}g_{ja})^2\right]^{1/2} \\ &= \left[\sum_{a=1}^{m} w_{aak}^2(g_{ia} - g_{ja})^2\right]^{1/2}, \end{aligned} \tag{22.1}$$

where $i, j = 1, \ldots, n; k = 1, \ldots, K; a = 1, \ldots, m;$ $\mathbf{W}_k$ is an $m \times m$ diagonal matrix of nonnegative weights w_{aak} for every dimension a for individual k; and $\mathbf{G}$ is the matrix of coordinates of the *group stimulus space* $\mathbf{G}$. Note that $\mathbf{G}$ does not have subscript k: individual differences are possible only in the weights on the dimensions of $\mathbf{G}$. The group stimulus space is also called a *common space* (Heiser, 1988b). Equation (22.1) is called the *weighted Euclidean distance*, which we encountered before in (21.7).

In terms of an individual k, the weighted Euclidean model says that

$$\mathbf{X}_k = \mathbf{GW}_k, \tag{22.2}$$

where $\mathbf{X}_k$ is the individual configuration. Because distances do not change under translation, we may assume that $\mathbf{G}$ is column centered. $\mathbf{X}_k = \mathbf{GW}_k$ is similar to $\mathbf{ZSW}_k$ in (21.11), where $\mathbf{Z}$ was defined as the average configuration of N individual configurations $\mathbf{X}_k$ transformed to an optimal fit in the sense of the generalized Procrustean loss function in (21.1). However, in this chapter there are no individual configurations $\mathbf{X}_k$ to begin with, and thus $\mathbf{G}$ must be computed differently.

There is an inherent indeterminacy in the weighted Euclidean model: the dimension weights depend on the particular definition of the group space. Let $\mathbf{D}$ be any diagonal matrix with full rank. Then

$$\mathbf{X}_k = \mathbf{GW}_k = \mathbf{GDD}^{-1}\mathbf{W}_k = (\mathbf{GD})(\mathbf{D}^{-1}\mathbf{W}_k) = \mathbf{G}^*\mathbf{W}_k^*; \tag{22.3}$$

that is, if $\mathbf{G}$ is stretched by $\mathbf{D}$, and the weights in $\mathbf{W}_k$ are subjected to the inverse transformation, the product remains the same. For the group space $\mathbf{G}$, no restriction was defined yet, except for the irrelevant centering convention. Yet, in order to make $\mathbf{G}$ identifiable, it must be normed somehow. One such norming is to require that $\mathbf{GG}' = \mathbf{I}$. Although this norming is a purely formal requirement, it nevertheless affects the interpretation of the weights in each $\mathbf{W}_k$: they are conditional to $\mathbf{G}$, as 22.3 makes clear. Hence, care must be taken with claims that, for example, a person weights dimension X twice as much as dimension Y. This assertion is only true relative to the given group space $\mathbf{G}$. However, it is possible to compare the weights of different persons on each dimension in turn without restrictions.

The weighted Euclidean model can be implemented in several ways. First, we discuss a method that minimizes Stress to find a group space $\mathbf{G}$ and dimension weights $\mathbf{W}_k$ from K proximity matrices. Then, we discuss the popular INDSCAL algorithm, which finds $\mathbf{G}$ and the $\mathbf{W}_k$s from the scalar product matrices derived from the K proximity matrices.

Fitting the Dimension-Weighting Model via Stress

Dimension weights can be implemented fairly easily in the Stress framework by applying the constrained MDS theory (De Leeuw & Heiser, 1980) from Section 10.3. Let us assume that the proximities are dissimilarities. Then, the Stress that needs to be minimized equals

$$\sigma_r(\mathbf{X}_1, \ldots, \mathbf{X}_k) = \sum_{k=1}^{K} \sum_{i<j} \left(\delta_{ijk} - d_{ij}(\mathbf{X}_k)\right)^2, \tag{22.4}$$

subject to the constraints that $\mathbf{X}_k = \mathbf{GW}_k$ as required for the dimension-weighting model. This minimization can also be viewed as doing MDS on a $Kn \times Kn$ dissimilarity supermatrix $\mathbf{\Delta}^*$ (with the individual K dissimilarity matrices $\mathbf{\Delta}_k$ on the diagonal blocks, and other blocks missing) and a configuration supermatrix $\mathbf{X}^*$ (with the individual configuration matrices $\mathbf{X}_k$ stacked under each other); that is,

$$\mathbf{\Delta}^* = \begin{bmatrix} \mathbf{\Delta}_1 & & & \\ & \mathbf{\Delta}_2 & & \\ & & \ddots & \\ & & & \mathbf{\Delta}_K \end{bmatrix} \text{ and } \mathbf{X}^* = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_K \end{bmatrix}$$

subject to the constraints that $\mathbf{X}_k = \mathbf{GW}_k$. The theory of Section 10.3 says that every iteration of the majorization algorithm for confirmatory MDS consists of the following two steps.

1. Compute the unconstrained update $\overline{\mathbf{X}}^*$ by the Guttman transform (8.28).

2. Minimize tr $(\mathbf{X}-\overline{\mathbf{X}}^*)'\mathbf{V}^*(\mathbf{X}-\overline{\mathbf{X}}^*)$ over $\mathbf{X}$ subject to the constraints to obtain the update $\mathbf{X}^u$, where here $\mathbf{V}^*$ is a block-diagonal matrix with $n\mathbf{J}$ on the diagonal blocks and where $\mathbf{J} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}'$ is the centering matrix.

Minimizing tr $(\mathbf{X}-\overline{\mathbf{X}}^*)'\mathbf{V}^*(\mathbf{X}-\overline{\mathbf{X}}^*)$ in the second step is equal to minimizing

$$\sum_k \text{tr } n(\mathbf{X}_k - \overline{\mathbf{X}}_k)'\mathbf{J}(\mathbf{X} - \overline{\mathbf{X}}_k) =$$
$$\sum_k \text{tr } n(\mathbf{G}\mathbf{W}_k - \overline{\mathbf{X}}_k)'(\mathbf{G}\mathbf{W}_k - \overline{\mathbf{X}}_k) \tag{22.5}$$

over $\mathbf{G}$ and $\mathbf{W}_k$. The centering matrix $n\mathbf{J}$ may be removed from (22.5), because $\overline{\mathbf{X}}_k$ is already column centered. De Leeuw and Heiser (1980) give a solution that is based on dimensionwise solving (22.5). Let $\overline{\mathbf{X}}_a$ denote the $n \times K$ matrix with column a of each unconstrained update $\overline{\mathbf{X}}_k$ stacked next to each other. Let $\mathbf{g}_a$ be column a of $\mathbf{G}$ and $\mathbf{w}_a$ the $K \times 1$ vector of the dimension weight w_{aak} for individual k in dimension a. Then, minimizing (22.5) is the same as minimizing

$$\sum_a \text{tr } (\mathbf{g}_a\mathbf{w}_a' - \overline{\mathbf{X}}_a)'(\mathbf{g}_a\mathbf{w}_a' - \overline{\mathbf{X}}_a). \tag{22.6}$$

This problem can be solved for each dimension separately by an alternating least squares algorithm, where in each iteration (22.6) is minimized over $\mathbf{g}_a$, keeping $\mathbf{w}_a$ fixed, followed by the minimization over $\mathbf{w}_a$, keeping $\mathbf{g}_a$ fixed. Alternatively, the analytic minimum is obtained by computing the singular value decomposition of $\overline{\mathbf{X}}_a = \mathbf{P}\mathbf{\Phi}\mathbf{Q}'$ and setting $\mathbf{g}_a = \mathbf{p}_1$ and $\mathbf{w}_a = \phi_1\mathbf{q}_1$. The PROXSCAL program implements the dimension-weighting model for Stress with more options (such as fixing coordinates and allowing for missing proximities). For the detailed mathematics of that approach, we refer to Heiser (1988b) and Commandeur and Heiser (1993). A different algorithm for dimension weighting with constrained dimensions is given by Winsberg and De Soete (1997).

If all weights $\mathbf{w}_a$ are constrained to be equal, we get the *identity* model for three-way proximities (Commandeur & Heiser, 1993). Then, the only thing that needs to be estimated is the group stimulus space $\mathbf{G}$. This allows (22.4) to be written as

$$\begin{aligned}\sigma_r(\mathbf{G}) &= \sum_k^K \sum_{i<j} (\delta_{ijk} - d_{ij}(\mathbf{G}))^2 \\ &= K \sum_{i<j} \left(\bar{\delta}_{ij} - d_{ij}(\mathbf{G})\right)^2 + \sum_{i<j} \sum_k^K \left(\bar{\delta}_{ij} - \delta_{ijk}\right)^2,\end{aligned}$$

where $\overline{\delta}_{ij} = K^{-1}\sum_k^K \delta_{ijk}$. The first term of $\sigma_r(\mathbf{G})$ amounts to simple MDS of the average dissimilarity matrix, and the second term measures the difference of the individual dissimilarity matrices to their average.

Heiser (1989b) discusses the minimization of the weighted Euclidean model for Stress with city-block distances. The minimization can be done by a combinatorial approach (similar to combinatorial methods used for unidimensional scaling) combined with a majorizing approach that accommodates negative disparities, or by majorization of city-block distances (Groenen et al., 1995).

The INDSCAL *Algorithm*

A popular algorithm for solving the dimension-weighting model is based on the scalar-product matrix, similar to classical scaling. Let $\mathbf{B}_{\mathbf{\Delta}_k} = -\frac{1}{2}\mathbf{J}\mathbf{\Delta}_k^{(2)}\mathbf{J}$ be the $n \times n$ scalar-product matrix for individual k derived from the distances via (12.2). Classical scaling for individual k minimizes

$$\frac{1}{4}||\mathbf{J}[\Delta_k^{(2)} - \mathbf{D}^{(2)}(\mathbf{X})]\mathbf{J}||^2 = \frac{1}{4}4||\mathbf{B}_{\mathbf{\Delta}_k} - \mathbf{X}\mathbf{X}'||^2.$$

This is extended by including dimension weights in the INDSCAL loss function; that is,

$$L_{IND}(\mathbf{G}, \mathbf{W}_1, \ldots, \mathbf{W}_K) = \sum_k^K ||\mathbf{B}_{\mathbf{\Delta}_k} - \mathbf{G}\mathbf{W}_k^2\mathbf{G}'||^2 \quad (22.7)$$

$$= \sum_{k=1}^K \sum_{i,j} \left(b_{ijk} - \sum_{a=1}^m g_{ia}g_{ja}w_{aak}^2 \right)^2. \quad (22.8)$$

It is assumed that the scalar-product matrices $\mathbf{B}_{\mathbf{\Delta}_k}, k = 1, \ldots, K$, are given. In the case of interval-scale proximities, an additive constant that leads to Euclidean distances must be computed, and scalar products are then derived from these distances. If only ordinal proximities (possibly even with missing data values) are given as data, one often proceeds as in PINDIS, that is, by first computing the individual configurations $\mathbf{X}_k, k = 1, \ldots, K$, via ordinal MDS, and then from these deriving the needed scalar products (e.g., Krantz & Tversky, 1975). We now describe a solution for (22.7).

The INDSCAL procedure (Carroll & Chang, 1970) proceeds as follows. The INDSCAL loss function L_{IND} has to be solved over two sets of parameters, $\mathbf{G}$ and the $\mathbf{W}_k$s. Unfortunately, this loss function does not have an analytical solution, except in the error-free case (Schönemann, 1972). INDSCAL uses the alternating update strategy in which an update of $\mathbf{G}$ for fixed $\mathbf{W}_k$s is followed by an update of the $\mathbf{W}_k$s for fixed $\mathbf{G}$. These updates are iterated until convergence. The two steps are computed as follows.

1. The update for the $\mathbf{W}_k$s for fixed $\mathbf{G}$ is found by standard regression. However, L_{IND} has to be rewritten. First, we string out each $\mathbf{B}_{\mathbf{\Delta}_k}$

into one column vector with n^2 elements and then form an $n^2 \times K$ matrix $\mathbf{B}^*$ by stacking these column vectors next to each other. In a similar fashion, we then stack the diagonals of the K weight matrices $\mathbf{W}_k^2$ in an $m \times K$ matrix $\mathbf{W}$. Finally, we compute the products $\mathbf{g}_a\mathbf{g}_a'$, string out its elements into one column vector, and place them for each dimension $a = 1, \ldots, m$ next to each other in the $n^2 \times m$ matrix $\mathbf{V}$. This leads to a compact way of writing L_{IND} as

$$L_{IND} = \text{tr } (\mathbf{B}^* - \mathbf{VW})'(\mathbf{B}^* - \mathbf{VW}). \tag{22.9}$$

The update for $\mathbf{W}$ is found by differentiating (22.9) with respect to $\mathbf{W}$ and setting the result equal to the null matrix $\mathbf{0}$, which yields

$$\mathbf{W} = (\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}'\mathbf{B}^*. \tag{22.10}$$

The columns of $\mathbf{W}$ are the diagonals of the individual weight matrices $\mathbf{W}_k^2$. Note, however, that some elements of $\mathbf{W}$ may be negative, so that the corresponding dimension weight is not a real number. This problem of negative squared dimension weights in the INDSCAL algorithm could be avoided by minimizing (22.9) over $\mathbf{W}$ under the constraint that $\mathbf{W} \geq \mathbf{0}$, as suggested by Ten Berge, Kiers, and Krijnen (1993), who used nonnegative least-squares (Lawson & Hanson, 1974). De Soete, Carroll, and Chaturvedi (1993) imposed these constraints using the alternating least-squares method discussed in Section 9.6.

2. A better $\mathbf{G}$, relative to the given $\mathbf{W}_k^2$s, is computed by INDSCAL as follows. With fixed $\mathbf{W}_k$s, we minimize

$$\begin{aligned} L_{IND}(\mathbf{G}, \mathbf{H}) &= \sum_k^K ||\mathbf{B}_{\mathbf{\Delta}_k} - \mathbf{H}\mathbf{W}_k^2\mathbf{G}'||^2 \\ &= \sum_k^K \text{tr } \mathbf{B}_{\mathbf{\Delta}_k}^2 + \text{tr } \mathbf{G} \left[\sum_k \mathbf{W}_k^2\mathbf{H}'\mathbf{H}\mathbf{W}_k^2\right] \mathbf{G})' \\ &\quad -2\text{tr } \mathbf{G} \left[\sum_k \mathbf{W}_k^2\mathbf{H}'\mathbf{B}_{\mathbf{\Delta}_k}\right] \end{aligned} \tag{22.11}$$

over both $\mathbf{G}$ and $\mathbf{H}$, the so-called CANDECOMP algorithm (Carroll & Chang, 1970). After convergence, it turns out that $\mathbf{G}$ and $\mathbf{H}$ are equal or can be made equal. Differentiating (22.11) with respect to $\mathbf{G}$ and setting the result equal to $\mathbf{0}$ gives the update for $\mathbf{G}$; that is,

$$\mathbf{G} = \left(\sum_k \mathbf{B}_{\mathbf{\Delta}_k}\mathbf{H}\mathbf{W}_k^2\right)\left(\sum_k \mathbf{W}_k^2\mathbf{H}'\mathbf{H}\mathbf{W}_k^2\right)^{-1}.$$

$\mathbf{H}$ is updated with the same update formula by reversing the roles of $\mathbf{G}$ and $\mathbf{H}$.

These two steps are repeated until the process converges to a final solution for $\mathbf{W}$ and $\mathbf{G}$, which "almost always" is the global optimum, according to Carroll and Wish (1974a, p. 90) and Ten Berge and Kiers (1991).

22.2 The Generalized Euclidean Model

The weighted Euclidean distance can be extended by the *generalized Euclidean distance*, where the individual space is defined as $\mathbf{X}_k = \mathbf{G}\mathbf{T}_k$, with $\mathbf{T}_k$ an $m \times m$ (real-valued) matrix that need not be diagonal.

Interpreting the Generalized Euclidean Model

The generalized Euclidean model can be interpreted as follows. Consider the singular value decomposition of $\mathbf{T}_k$, $\mathbf{T}_k = \mathbf{P}\mathbf{\Lambda}\mathbf{Q}'$. Then, the transformation $\mathbf{G}\mathbf{T}_k = \mathbf{G}\mathbf{P}\mathbf{\Phi}\mathbf{Q}'$ can be interpreted as: take group space $\mathbf{G}$, rotate it by $\mathbf{P}$, and stretch it along its dimensions by $\mathbf{\Phi}$. Because we are concerned with the distances of $\mathbf{G}\mathbf{T}_k$, the final rotation by $\mathbf{Q}'$ is irrelevant. This shows that in the generalized Euclidean model every individual k transforms the group space first by a rotation and/or a reflection, and then by stretching. In contrast to the weighted Euclidean model, each individual may weight a different set of dimensions of the group space. Therefore, this model is somewhat less restrictive than the dimension-weighting model.

Other interpretations are possible. For example, Tucker (1972) and Harshman (1972) proposed decomposing $\mathbf{T}_k = \mathbf{D}_k\mathbf{M}_k$, where the diagonal matrix $\mathbf{D}_k$ contains the standard deviation of the column elements of $\mathbf{T}_k$ [so that $\text{diag}(\mathbf{D}_k^2) = \text{diag}(\mathbf{T}_k'\mathbf{T}_k)$]. Thus, $\mathbf{M}_k'\mathbf{M}_k$ has diagonal elements 1 and can be seen as a correlation matrix or as a matrix of cosines of angles among oblique dimensions. The interpretation of the generalized Euclidean model using this decomposition is that the individual space $\mathbf{X}_k$ can be obtained from the group space $\mathbf{G}$ by first stretching its dimensions by $\mathbf{D}_k$ and then applying an oblique rotation by $\mathbf{M}_k$. Harshman and Lundy (1984) proposed a model with only one $\mathbf{M}$ that is common to all individuals. However, this model is not equivalent to the generalized Euclidean model.

Whether or not there are applications for the generalized Euclidean model, there is nothing that rules it out formally. Indeed, even more exotic interpretations derived from other decompositions (Carroll & Wish, 1974a, 1974b) are possible.

If generalized Euclidean models are interpreted as a distance model in $\mathbf{G}$,

$$\begin{aligned} d_{ijk}^2(\mathbf{G}) &= (\mathbf{g}_i\mathbf{T}_k - \mathbf{g}_j\mathbf{T}_k)'(\mathbf{g}_i\mathbf{T}_k - \mathbf{g}_j\mathbf{T}_k) \\ &= (\mathbf{g}_i - \mathbf{g}_j)'\mathbf{C}_k(\mathbf{g}_i - \mathbf{g}_j), \end{aligned}$$

then $\mathbf{C}_k$ must be positive definite, not just positive semidefinite, as Carroll and Wish (1974b) declare, because otherwise one may obtain $d_{ijk}(\mathbf{G}) = 0$ even though $i \neq j$. That is, if we want to interpret the model in such a way that each individual k picks his or her own particular distance function from the family of weighted Euclidean distances or, as mathematicians sometimes call it, from the family of *elliptical distances* (Pease, 1965, p. 219) on the group space $\mathbf{G}$, then all dimension weights must be positive. If some of these weights are zero, then this interpretation has to be changed slightly to one in which individual k first reduces $\mathbf{G}$ to a subspace and then computes distances in this, possibly further transformed, subspace of $\mathbf{G}$. The first model has been called a *subjective metrics model* (Schönemann & Borg, 1981a), and the latter, due to Schulz (1972, 1975, 1980), may be called a *subjective transformations model*. From a practical point of view, however, these distinctions are irrelevant because, in the subjective metrics model, $\mathbf{G}$ may be almost reduced to a lower rank by choosing extremely small weights for some of its dimensions.

Fitting the Generalized Euclidean Model via Stress

The method for minimizing Stress with the generalized Euclidean model is the same as for the weighted Euclidean model via Stress, except that $\mathbf{X}_k$ is restricted as $\mathbf{X}_k = \mathbf{G}\mathbf{T}_k$, where $\mathbf{T}_k$ may be *any* real-valued $m \times m$ matrix. In the second step of the algorithm, the restrictions are imposed by minimizing

$$\sum_k \text{tr } (\mathbf{G}\mathbf{T}_k - \overline{\mathbf{X}}_k)'(\mathbf{G}\mathbf{T}_k - \overline{\mathbf{X}}_k) \tag{22.12}$$

over $\mathbf{G}$ and the $\mathbf{T}_k$s. Let the $n \times mK$ matrix $\overline{\mathbf{X}}$ contain the $\overline{\mathbf{X}}_k$s stacked next to each other, and the $m \times mK$ matrix $\mathbf{T}$ the $\mathbf{T}_k$s stacked next to each other. Then, (22.12) is equal to

$$\text{tr } (\mathbf{G}\mathbf{T} - \overline{\mathbf{X}})'(\mathbf{G}\mathbf{T} - \overline{\mathbf{X}}),$$

which is solved analytically (De Leeuw & Heiser, 1980) by taking the singular value decomposition of $\overline{\mathbf{X}} = \mathbf{P}\mathbf{\Phi}\mathbf{Q}'$ and setting $\mathbf{G} = \mathbf{P}_m$ and $\mathbf{T} = \mathbf{\Phi}_m\mathbf{Q}'_m$, where the subscript m implies taking only the first m singular values and vectors.

The IDIOSCAL *Model*

The generalized Euclidean model gained its popularity in the framework of scalar products. This idiosyncratic weighting model is also called the IDIOSCAL *model* (Carroll & Wish, 1974a, 1974b; Schulz, 1980). In scalar product notation, this model minimizes

$$\begin{aligned} L_{IDIO}(\mathbf{G}, \mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_K) &= \sum_{k=1}^{K} ||\mathbf{B}_k - (\mathbf{G}\mathbf{T}_k)(\mathbf{G}\mathbf{T}_k)'||^2 \\ &= \sum_{k=1}^{K} ||\mathbf{B}_k - \mathbf{G}\mathbf{T}_k\mathbf{T}_k'\mathbf{G}'||^2 \\ &= \sum_{k=1}^{K} ||\mathbf{B}_k - \mathbf{G}\mathbf{C}_k\mathbf{G}'||^2, \end{aligned} \tag{22.13}$$

where $\mathbf{T}_k\mathbf{T}_k' = \mathbf{C}_k$ and $\mathbf{C}_k$ is positive semidefinite.

Apart from the many possibilities for factoring $\mathbf{C}_k$, it is of interest to ask whether there is only one $\mathbf{C}_k$ and one $\mathbf{G}$ that solve (22.13). This is not so, because

$$\begin{aligned} \mathbf{B}_k &= \mathbf{G}\mathbf{C}_k\mathbf{G}' \\ &= \mathbf{G}(\mathbf{A}\mathbf{A}^{-1})\mathbf{C}_k(\mathbf{A}\mathbf{A}^{-1})'\mathbf{G}' \\ &= (\mathbf{G}\mathbf{A})[\mathbf{A}^{-1}\mathbf{C}_k(\mathbf{A}')^{-1}](\mathbf{G}\mathbf{A})' \\ &= \mathbf{G}^*\mathbf{C}_k^*(\mathbf{G}^*)', \end{aligned} \tag{22.14}$$

where $\mathbf{A}$ is an arbitrary $m \times m$ matrix with full rank. In comparison to (22.3), the more general IDIOSCAL model is less unique. This has the practical implication that if this model is applied to a set of data matrices, many quite different group spaces $\mathbf{G}$ can be derived, and it is impossible to say which one is the *true* common structure. Schönemann (1972) proposed imposing the restriction that the $\mathbf{C}_k$s average to $\mathbf{I}$; that is,

$$\frac{1}{K}\sum_{k=1}^{K} \mathbf{C}_k = \mathbf{I}. \tag{22.15}$$

Given a set of $k = 1, \ldots, K$ arbitrary $\mathbf{C}_k^*$ as in (22.14), property (22.15) can be imposed by choosing a transformation matrix $\mathbf{A}$ such that

$$\begin{aligned} \mathbf{I} &= K^{-1}[\mathbf{A}^{-1}\mathbf{C}_1^*(\mathbf{A}')^{-1} + \ldots + \mathbf{A}^{-1}\mathbf{C}_K^*(\mathbf{A}')^{-1}] \\ &= K^{-1}\mathbf{A}^{-1}(\mathbf{C}_1^* + \ldots + \mathbf{C}_K^*)(\mathbf{A}')^{-1}, \end{aligned}$$

whence $K\mathbf{A}\mathbf{A}' = \mathbf{C}_1^* + \ldots + \mathbf{C}_K^*$. Because each $\mathbf{C}_k^*$ is symmetric by (22.13), $\mathbf{C}_1^* + \ldots + \mathbf{C}_K^*$ is also symmetric, so $\mathbf{A}$ is found by factoring the average of all $\mathbf{C}_k$s into $\mathbf{A}\mathbf{A}'$. If (22.15) holds, then

$$\frac{1}{K}\sum_{k=1}^{K} \mathbf{B}_k = \frac{1}{K}\sum \mathbf{G}\mathbf{C}_k\mathbf{G} = \frac{1}{K}\mathbf{G}(\mathbf{C}_1 + \ldots + \mathbf{C}_K)\mathbf{G}' = \mathbf{G}\mathbf{G}'. \tag{22.16}$$

For error-free data, this equation can be solved immediately (by classical scaling) to yield the group space $\mathbf{G}$, or, more properly, one possible $\mathbf{G}$

because each such $\mathbf{G}$ can be arbitrarily rotated and/or reflected and would still satisfy (22.16).

To find each individual $\mathbf{C}_k$ is also simple. We just solve the following equation for $\mathbf{C}_k$,

$$\begin{aligned} \mathbf{B}_k &= \mathbf{G}\mathbf{C}_k\mathbf{G}', && (22.17)\\ \mathbf{G}'\mathbf{B}_k\mathbf{G} &= \mathbf{G}'\mathbf{G}\mathbf{C}_k\mathbf{G}'\mathbf{G}, && (22.18)\\ (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{B}_k\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1} &= \mathbf{C}_k. && (22.19) \end{aligned}$$

Note that the pre- and postmultiplications in (22.18) serve the purpose of generating the matrix $\mathbf{G}'\mathbf{G}$, which, assuming that $\text{rank}(\mathbf{G}) = m$, is invertible, whereas $\mathbf{G}$ generally is not. Thus, for error-free $\mathbf{B}_k$s, the IDIOSCAL loss function (22.13) can be solved analytically (Schönemann, 1972).

Chaturvedi and Carroll (1994) imposed the additional restriction on $\mathbf{G}$ that every row only contains a single 1 and the rest 0, which makes $\mathbf{G}$ an indicator matrix. Thus, $\mathbf{G}$ classifies each stimulus i to one of M clusters. This model, called INDCLUS, falls somewhere between clustering and MDS.

22.3 Overview of Three-Way Models in MDS

To develop some geometric feeling for the various three-way models discussed in this chapter, let us demonstrate with the help of a simple example how they relate a common space to the individual space of each subject.[1]. An overview of these models is given in Figure 22.1 The *identity* model is trivial boundary case: every subject space should be equal to the common space, that is, $\mathbf{X}_k = \mathbf{G}$. This model is equivalent to computing the average dissimilarity and doing an ordinary MDS (see Section 22.1). Note that the weight plot shows dimension weights of one for all subjects on all dimensions.

However, we also know that only the relative distances between the points in a configuration are of importance, not the absolute distances. Therefore, instead of the identity model, it is better to fit the *dilation* model that allows for a dilation factor for each subject; that is, $\mathbf{X}_k = w_k\mathbf{G}$. This model is shown in the second row of Figure 22.1. Inserting the dilation factors ensures that the size of the individual configuration reflects the fit (see Section 11.1). As the weights do not differ per dimension, the points for individuals in the weights plot are on a line.

The third row in Figure 22.1 shows the weighted Euclidean model that allows each subject to weight the fixed dimensions of the common space, that is, $\mathbf{X}_k = \mathbf{G}\mathbf{W}_k$. In this example, the weights are $w_{111} = 1.5$ and $w_{221} =$

[1]Note that we are not discussing models with idiosyncratic origins or with vector weightings, as discussed in Chapter 21. Rather, the models considered here are all within the dimension-weighting family for group spaces centered at the origin.

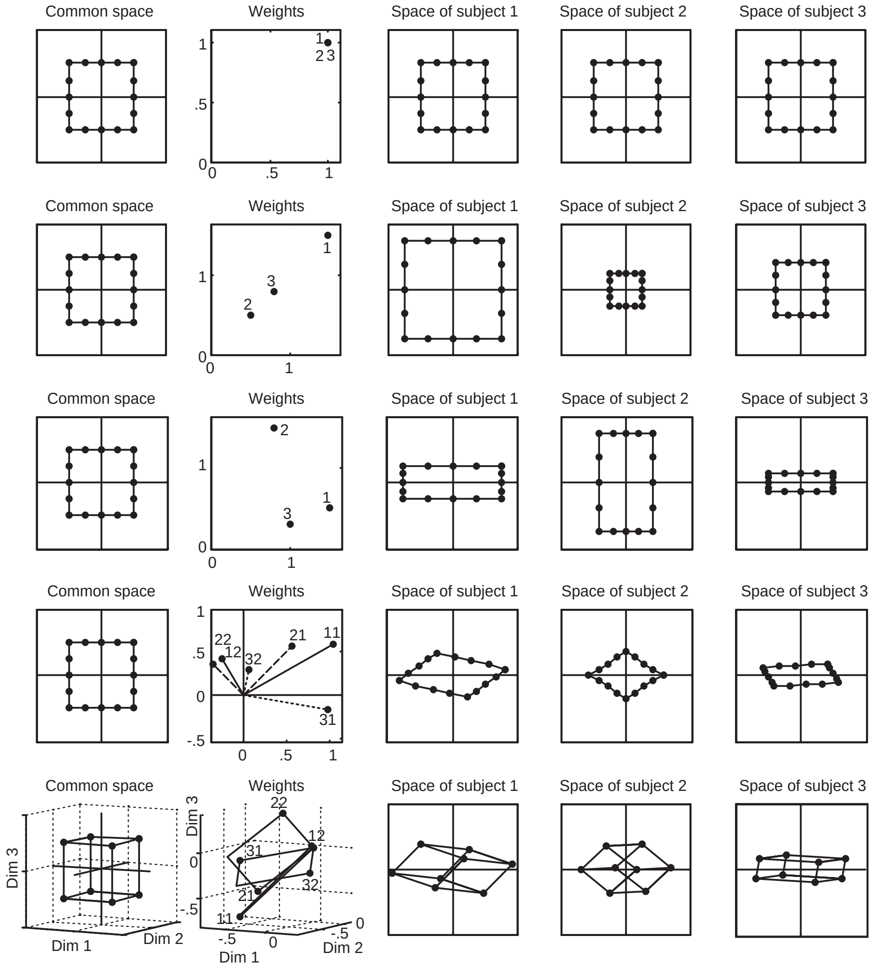

FIGURE 22.1. Overview of five three-way models for MDS. For each model the common space, the weights, and three individual spaces are given. The first row considers the identity model, the second row the dilation model, the third row the weighted Euclidean distance model, the fourth row the generalized Euclidean distance model, and the fifth row the reduced rank model.

.5 for Subject 1, $w_{112} = .8$ and $w_{222} = 1.5$ for Subject 2, and $w_{113} = 1$ and $w_{223} = .3$ for Subject 3. The weights plot shows each subject as a point with its dimension weights as coordinates. The weighted Euclidean model generalizes the dilation model by allowing for each subject to have unequal weights per dimension. In this example, we see that the Subjects 1 and 3 emphasize the first dimension in their individual spaces and Subject 2 the second dimension.

The generalized Euclidean model is given in the fourth row of Figure 22.1. In this model, the common space is first rotated to each individual's *principal directions*, as Young (1984) calls this orientation, and subsequently weighted to obtain the individual space, that is, $\mathbf{X}_k = \mathbf{G}\mathbf{T}_k\mathbf{W}_k$, where $\mathbf{T}_k$ is a rotation matrix and $\mathbf{W}_k$ is again a diagonal matrix with weights.

In our example, we choose

$$\mathbf{T}_1 = \begin{bmatrix} .866 & -.500 \\ .500 & .866 \end{bmatrix}, \quad \mathbf{W}_1 = \begin{bmatrix} 1.2 & .0 \\ .0 & .5 \end{bmatrix},$$

$$\mathbf{T}_2 = \begin{bmatrix} .707 & -.707 \\ .707 & .707 \end{bmatrix}, \quad \mathbf{W}_2 = \begin{bmatrix} .8 & .0 \\ .0 & .5 \end{bmatrix},$$

$$\mathbf{T}_3 = \begin{bmatrix} .985 & .174 \\ -.174 & .985 \end{bmatrix}, \quad \mathbf{W}_3 = \begin{bmatrix} 1.0 & .0 \\ .0 & .3 \end{bmatrix},$$

where the rotation matrices $\mathbf{T}_1, \mathbf{T}_2$, and $\mathbf{T}_3$ correspond to rotation by $30°, 45°$, and $-10°$. The weight plot is different than before. It shows how to obtain the subject space from the common space. For example, the solid vectors 11 and 12 show that the space of subject 1 is obtained by rotating the common space by 30° and to obtain dimension 1, stretch in the direction of vector 11 by a factor 1.2 (i.e., the length of vector 11) and for dimension 2, shrink in the direction of vector 12 by a factor 0.5 (the length of vector 12). Thus, the weight vectors belonging to each subject always have an angle of 90°. These vectors are obtained by $\mathbf{T}'_k\mathbf{W}_k$.

The last row of Figure 22.1 displays the *reduced rank* model. In this case, the individual spaces are allowed to have a lower rank than the common space, that is, $\mathbf{X}_k = \mathbf{G}\mathbf{T}_k\mathbf{W}_k$, where $\mathbf{G}$ is $n \times m$, $\mathbf{T}_k$ is an $m \times q$ rotation projection matrix with $q < m$, and $\mathbf{W}_k$ a diagonal $q \times q$ matrix with dimension weights. Thus, the dimensionality m of the common space is reduced to q for each subject space by the $\mathbf{T}_k$s. The example shows the common space of a cube in 3D and the individual spaces in 2D (thus $m = 3$ and $q = 2$). Then, the weights plot is interpreted in the same way as for the generalized Euclidean model. For example, the vectors 21 and 22 for subject 2 are connected to form a rectangle so that it is easy to see which 2D plane is associated with the space of Subject 2. Again, the length of each vector indicates the weighting factor for stretching or shrinking along its direction to obtain the dimension for this subject. Unless the common space and the subject spaces are very structured, it may be hard to interpret the reduced rank model in empirical applications.

22.4 Some Algebra of Dimension-Weighting Models

The fit measure provided by INDSCAL is the correlation between the original scalar products, given in the K matrices $\mathbf{B}_k$, and the reproduced scalar products, computed by $\widehat{\mathbf{B}}_k = \mathbf{G}\mathbf{W}_k^2\mathbf{G}'$. These correlations are usually extremely high, even if the model is not adequate. This was shown by MacCallum (1976). He generated synthetic data from a group space that was stretched not only differentially for each $k = 1, \ldots, K$, but also, in violation of the INDSCAL model, along different directions for each k. He observed fit coefficients that were not lower than $r = .97$ and commented that "one must wonder whether this index provides in any sense a measure of the appropriateness of the INDSCAL model to a given set of data" (pp. 181–182).

Finer fit indices can, however, be derived from an algebraic analysis of the model. Such an analysis starts out by assuming the ideal case, where data are given that perfectly satisfy the model. Of course, this is unrealistic, because data always have error components. Hence, one can never expect to satisfy a deterministic model strictly, except in trivial cases. Nevertheless, by studying the ideal case, one can derive certain properties that data must possess if they are to be accounted for by a particular model. Real data should then also satisfy these conditions "more or less". They may also violate the model conditions systematically, and this provides potentially very informative insights into the structure of the data.

The Common-Space Condition

What does the dimension-weighting model imply for the data? That is, what properties must the data possess so that they can be explained by such models? For the subjective metrics interpretation, it is necessary that $\text{rank}(\mathbf{B}_k) = m$, for all k, because $\text{rank}(\mathbf{C}_k) = m$ and $\text{rank}(\mathbf{G}) = m$. Moreover, because $\mathbf{C}_k = \mathbf{T}_k\mathbf{T}_k'$, $\text{rank}(\mathbf{T}_k) = m$. Even if $\text{rank}(\mathbf{C}_k) < m$ (as may be the case in the subjective transformations model), it must hold that each individual space, $\mathbf{X}_k = \mathbf{G}\mathbf{T}_k$, lies in the column space of $\mathbf{G}$; that is, the columns of $\mathbf{X}_k$ must be linear combinations of the columns of $\mathbf{G}$. For example, column 1 of $\mathbf{X}_k$ should result from adding the columns of $\mathbf{G}$ with the weights of column 1 of $\mathbf{T}_k$.

In practical applications of the model, we typically would not use the $\mathbf{G}$ resulting from (22.16) as the group space but only its first few dimensions. But, with only a subspace of the complete $\mathbf{G}$, each $\mathbf{B}_k$ can only be approximated by the model, because the model requires $\text{rank}(\mathbf{B}_k) = \text{rank}(\mathbf{G}) = m$. However, if the dropped dimensions represent just error,

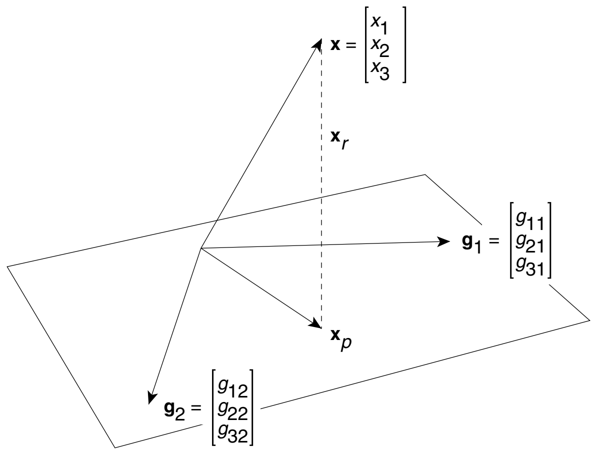

FIGURE 22.2. Geometric view of projecting a vector $\mathbf{x}$ onto the column space of a matrix with column vectors $\mathbf{g}_1$ and $\mathbf{g}_2$.

then the low-dimensional $\mathbf{G}$ should account for most of the variance of each $\mathbf{B}_k$ or, at least, for more variance than could be expected by chance.

This is easier to understand geometrically. In Figure 22.2, the column vectors $\mathbf{g}_1$ and $\mathbf{g}_2$ span a plane onto which the vector $\mathbf{x}$ is projected. The vectors $\mathbf{g}_1$ and $\mathbf{g}_2$ together form the 3×2 matrix $\mathbf{G}$. The projection of $\mathbf{x}$ onto the $\mathbf{G}$-plane, $\mathbf{x}_p$, is equal to some linear combination $w_1 \cdot \mathbf{g}_1 + w_2 \cdot \mathbf{g}_2$ or $\mathbf{x}_p = \mathbf{Gw}$, where $\mathbf{w}' = (w_1, w_2)$ is the weight or coordinate vector of $\mathbf{x}_p$. The residual vector (i.e., the component of $\mathbf{x}$ not contained in $\mathbf{G}$) is $\mathbf{x}_r = \mathbf{x} - \mathbf{x}_p = \mathbf{x} - \mathbf{Gw}$. As Figure 22.2 shows, $\mathbf{x}_r$ is orthogonal to $\mathbf{x}_p$. Thus, $\mathbf{x}_p'\mathbf{x}_r = 0$ or $(\mathbf{Gw})'(\mathbf{x} - \mathbf{Gw}) = 0$ or $\mathbf{w}'(\mathbf{G}'\mathbf{x} - \mathbf{G}'\mathbf{Gw}) = 0$. Because $\mathbf{w} \neq \mathbf{0}$, in general, we have $\mathbf{G}'\mathbf{x} - \mathbf{G}'\mathbf{Gw} = \mathbf{0}$ and $\mathbf{w} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{x}$. With this weight vector, we obtain $\mathbf{x}_p = \mathbf{Gw} = \mathbf{G}[(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{x}] = [\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}']\mathbf{x} = \mathbf{P}_G\mathbf{x}$, where $\mathbf{P}_G$ denotes the matrix that effects the projection of $\mathbf{x}$ onto the column space of $\mathbf{G}$.

Now, the common-space index is constructed as follows. The portion of $\mathbf{B}_k$ that can be reproduced from $\mathbf{P}_G\mathbf{X}_k$ is $\widehat{\mathbf{B}}_k = (\mathbf{P}_G\mathbf{X}_k)(\mathbf{P}_G\mathbf{X}_k)' = \mathbf{P}_G\mathbf{B}_k\mathbf{P}_G'$. The sum-of-squares of its elements can be expressed as $\text{tr}\ (\mathbf{P}_G\mathbf{B}_k\mathbf{P}_G)^2$, and the sum-of-squares of $\mathbf{B}_k$s elements is $\text{tr}\ \mathbf{B}_k^2$. The ratio of these two sums-of-squares is a possible measure for how well the common-space condition is satisfied empirically:

$$v_k = \frac{\text{tr}\ (\mathbf{P}_G\mathbf{B}_k\mathbf{P}_G)^2}{\text{tr}\ \mathbf{B}_k^2}, \tag{22.20}$$

the common-space index for individual k (Schönemann, James, & Carter, 1979). We would require, of course, that this index be close to 1 or "high" before any of the models in the IDIOSCAL family could be considered seriously as an explanation for an individual's data.

The Diagonality Condition

The common-space condition is a rather weak criterion the data must satisfy so that they can be represented by a model of the IDIOSCAL type. This weakness is simply a consequence of the generality of the models, which, without many more additional constraints on $\mathbf{G}$ and/or $\mathbf{C}_k$, are not likely to lead to much scientific insight. Thus, we now go on to the more restrictive dimension-weighting model $\mathbf{B}_k = \mathbf{G}\mathbf{W}_k^2\mathbf{G}'$ and investigate what further properties must hold in the $\mathbf{B}_k$s for such a representation to be possible.

We first impose a condition similar to the one in (22.15),

$$\frac{1}{K}\sum_{k=1}^{K}\mathbf{W}_k = \mathbf{I}, \tag{22.21}$$

which leads to

$$\frac{1}{K}\sum_{k}\mathbf{B}_k = \mathbf{G}\mathbf{G}' \tag{22.22}$$

and thus to a direct solution[2] for $\mathbf{G}$. To compute $\mathbf{W}_k$ is somewhat more demanding than to find $\mathbf{C}_k$ in (22.19) because $\mathbf{W}_k$ must be diagonal. Thus, we first find $\mathbf{C}_k$ and then try to "diagonalize" it. This is done as follows. We note again that $\mathbf{G}$ is determined only up to an orthogonal matrix $\mathbf{S}$, because $\mathbf{G}^*(\mathbf{G}^*)' = (\mathbf{GS})(\mathbf{GS})' = \mathbf{GSS}'\mathbf{G}' = \mathbf{GG}'$. Hence, we want to find that $\mathbf{S}$ which diagonalizes $\mathbf{C}_k$; that is,

$$\mathbf{B}_k = (\mathbf{GS})\mathbf{C}_k(\mathbf{GS})' = \mathbf{GSC}_k\mathbf{S}'\mathbf{G}' = \mathbf{G}(\mathbf{SC}_k\mathbf{S}')\mathbf{G}', \tag{22.23}$$

so that

$$\mathbf{SC}_k\mathbf{S}' = \mathbf{W}_k^2. \tag{22.24}$$

If we write

$$\mathbf{C}_k = \mathbf{S}'\mathbf{W}_k^2\mathbf{S}, \tag{22.25}$$

we see that $\mathbf{S}$ and $\mathbf{W}_k^2$ are the eigenvector and eigenvalue matrices of $\mathbf{C}_k$. Because $\mathbf{C}_k$ is symmetric and positive definite, $\mathbf{S}$ is orthogonal or can be so constructed, and $\mathbf{W}_k^2$ is positive definite. Note, however, that $\mathbf{S}$ does not have a subscript, and thus (22.25) cannot be guaranteed to hold for every set of $\mathbf{B}_k$s. Rather, these $\mathbf{B}_k$s must have a common set of eigenvectors. Otherwise, the data cannot be explained by the dimension-weighting model. Geometrically, the reason for this condition is apparent and simply expresses that the model requires one fixed dimension system for all individuals.

[2]This $\mathbf{G}$ is taken as a rational starting configuration in the dimension-weighting option of ALSCAL (Takane et al., 1977). For ALSCAL, see Appendix A.

With an $\mathbf{S}$ computed from one particular $\mathbf{C}_k$ or from the average of all $\mathbf{C}_k$s, we can check how well it does in generating a diagonal matrix $\mathbf{W}_k^2$ from $\mathbf{S}\mathbf{C}_k\mathbf{S}'$. An index for how much the data violate this *diagonality condition* is provided by the sum-of-squares of the nondiagonal elements of all $\mathbf{W}_k^2$s, computed with this one $\mathbf{S}$, appropriately normed to make the index independent of the size of $\mathbf{G}$. Schönemann et al. (1979) define a diagonality index

$$\delta_k = \frac{\text{tr }[\tilde{\mathbf{W}}_k^2 - \mathbf{I}]^2}{(m-1)m}, \tag{22.26}$$

where $\tilde{\mathbf{W}}_k^2$ is a normalized[3] form of $\mathbf{W}_k^2$. If $\mathbf{W}_k$ is diagonal, then $\delta_k = 0$. Otherwise, $\delta_k > 0$, and we then must decide whether it is still acceptably small.

An Empirical Application: Helm's Color Similarities

To illustrate, we scale the Helm color data from Table 21.1 with COSPA (Schönemann, James, & Carter, 1978), a program that also computes a common-space and a diagonality index for each k. Table 22.2 shows these indices. If the model were strictly adequate, we should have $v_k = 1$ and $\delta_k = 0$ for all individuals. Even though this is not true, it holds that all v_ks are high and most δ_ks are small. Moreover, the v_k-indices of the color-deficient subjects are generally lower than those of the color-normal subjects. This could be expected from the results in Table 22.2, because the former persons have relatively much more variance accounted for by the "small" dimensions, possibly due to a greater error variance in their data. Also, s_{13} has the worst δ_k value, which mirrors this person's relatively low communality values from Table 21.3.

Schönemann et al. (1979) report some statistical norms for these indices, derived by computer simulations under various error conditions. In the least restrictive or—relative to the model—the "nullest" case, each individual scalar-product matrix $\mathbf{B}_k$ is generated by forming the product $\mathbf{X}_k\mathbf{X}_k'$ with a random $\mathbf{X}_k$. For $m = 2, n = 10$, and $N = 16$, it is found that 90% of the v_k-values are less than 0.40. Hence, common-space values of the magnitude of those in Table 22.2 are extremely unlikely if the null-hypothetical situation is true. For the diagonality index, 90% of the values obtained were greater than 0.04. Some of the δ_ks in Table 22.2 are greater than this value, and, if taken by themselves, would not lead to a rejection of the null hypothesis. But if all of the diagonality indices are taken together, then a value distribution like the one observed for the Helm data is highly improbable under this random condition. These tests provide just rough guidelines, because it is not clear when we should assume such a

[3]The normalization of $\mathbf{W}_k^2$ is achieved by pre- and postmultiplying it by $\text{diag}[(\mathbf{W}_k^2)'\mathbf{W}_k^2]^{-1}$.

TABLE 22.2. Model test, indices for Helm data of Table 21.1.

Subject	Common Space (v_k)	Diagonality (δ_k)
s_1	0.95	0.04
s_2	0.94	0.02
s_3	0.94	0.01
s_4	0.96	0.02
s_5	0.93	0.04
$s_6^{[1]}$	0.91	0.09
$s_6^{[2]}$	0.96	0.08
s_7	0.92	0.02
s_8	0.93	0.12
s_9	0.92	0.07
s_{10}	0.87	0.08
s_{11}	0.86	0.06
$s_{12}^{[1]}$	0.84	0.04
$s_{12}^{[2]}$	0.85	0.01
s_{13}	0.86	0.45
s_{14}	0.93	0.17

null hypothesis. In color perception, it is certainly not the incumbent hypothesis, which the null hypothesis should be (Guttman, 1977). Moreover, we are not really interested in "some" dimension-weighting model but in a model where a particular group space (i.e., the color circle) is expected, and where this configuration is individually transformed by weighting a particular dimension, not just any one. Because everything comes out as predicted (except that, for some individuals, there is some residual unspecified variance) it would be foolish to reject the model altogether, just because some formal norms are too high. Rather, it seems more fruitful to take this result as a reasonable approximation, modify and/or supplement the theory somewhat, and test it in further empirical studies.

22.5 Conditional and Unconditional Approaches

The dimension-weighting model $\mathbf{B}_k = \mathbf{G}\mathbf{W}_k^2\mathbf{G}$ comes in two variants. In one case, the individual scalar-product matrices are processed as they are; in the other, they first are normed so that their sum-of-squares is equal to 1 for each k. Some authors call the first case the *Horan model* and the latter the INDSCAL model (Schönemann et al., 1979). A more gripping distinction calls the first approach *unconditional* and the latter *matrix-conditional* (Takane et al., 1977). This reveals the similarity to the situation in unfolding, where we did not want to compare data values across the rows of the data matrix and so used a split-by-rows or row-conditional approach. Analogously, a matrix-conditional or split-by-matrices treatment

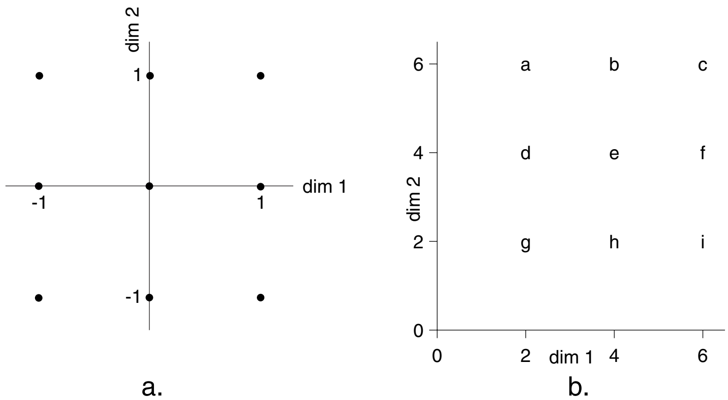

FIGURE 22.3. Synthetic (a) group space $\mathbf{G}$ and (b) subject space weights w_{11k} and w_{22k} for subjects a, b, ..., i used in the MacCallum (1977) study.

of the data implies that we do not want to compare the values of $\mathbf{B}_k$ and $\mathbf{B}_l$, for any $k \neq l$. The approach to be chosen depends on the particular data under investigation. For the Helm color data, for example, it seems that we should opt for the unconditional approach, because the data collection procedure suggests that all individuals used the same ratio scale for their proximity judgments. For the Green–Rao breakfast data (Table 14.1), on the other hand, the data were just ordinal, so ordinal MDS was used to arrive at ratio-scaled values. These values are the MDS distances, and they can be uniformly dilated or shrunk, of course, so that in this case we should prefer the matrix-conditional approach.

If the data are unconditionally comparable over individuals, then to norm all of the $\mathbf{B}_k$s in the same way leads to a loss of empirical information. This is apparent from the following demonstration due to MacCallum (1977). [Similar examples are given by Möbus (1975) and Schulz and Pittner (1978).] Figure 22.3 shows a group space $\mathbf{G}$ (panel a) and the associated subject space (panel b) that determines the weight matrices $\mathbf{W}_a, \mathbf{W}_b, \ldots, \mathbf{W}_i$. The nine $\mathbf{B}_k$s that can be derived from these figures as $\mathbf{B}_k = \mathbf{G}\mathbf{W}_k^2\mathbf{G}'$ differ in their sum-of-squares: for example, $\mathbf{B}_c$'s values are all much larger than the corresponding values in $\mathbf{B}_g$. Now, scaling the $\mathbf{B}_k$s with INDSCAL (which is always matrix-conditional) or with the matrix-conditional option of ALSCAL yields the subject space in Figure 22.4 (panel a). The different "size" of each individual's private perceptual space $\mathbf{X}_k$ is not represented. But because, for example, $\mathbf{X}_c$, $\mathbf{X}_e$, and $\mathbf{X}_g$ are perfectly similar and differ only in their sizes, the norming has the effect of projecting c, e, and g onto the same point in the subject space, as shown in Figure 22.4 (panel b). If the unconditional approach is used, the subject space is recovered perfectly.

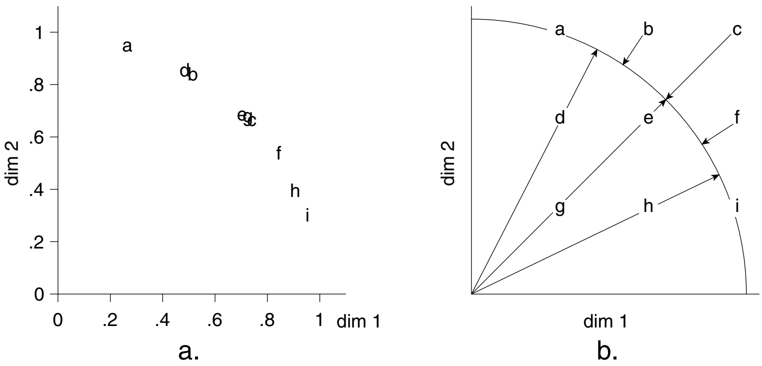

FIGURE 22.4. (a) INDSCAL reconstruction of subject space in Fig. 22.3b, and (b) visualization of norming effect.

22.6 On the Dimension-Weighting Models

The dimension-weighting models have received considerable attention in the literature. In their restrictive versions with fixed dimensions, they have been used in many applications because they promised to yield dimensions with a unique orientation while offering an intuitively appealing explanation for interindividual differences. The only other popular MDS model that accounts for interindividual differences is unfolding, but unfolding is for dominance data, not for similarity data. Unfolding assumes that the perceptual space is the same for all persons. It models different preferences by different ideal points in this space. The weighted Euclidean model allows for different perceptual spaces, related to each other by different weights attached to a set of fixed dimensions. Both models can be combined into one, an unfolding model with different dimensional weights for each person (see Chapter 16).

Some more recent developments should also be mentioned. They are motivated by practical and applied problems such as analyzing data sets where n is very large. The usual computer programs cannot handle such cases, or, more important, the subject space tends to be cluttered. "Marketing research suppliers often collect samples from thousands of consumers, and the ability of MDS procedures to fully portray the structure in such volumes of data is indeed limited. The resulting joint spaces or individual weight spaces become saturated with points/vectors, often rendering interpretation impossible . . . Yet, marketeers are rarely interested in the particular responses of consumers at the individual level . . . marketeers are more concerned with identifying and targeting market segments—homogeneous groups of consumers who share some designated set of characteristics (e.g., demographics, psychographics, consumption patterns, etc.) ..." (DeSarbo, Manrai, & Manrai, 1994, p. 191). In order to identify such segments, models were invented that combine a fuzzy form of cluster analysis with MDS. In essence, what one wants is an MDS solution where the subject space does not represent individual persons but types of persons. One procedure for that purpose is Clascal by Winsberg and De Soete (1993), a *latent class MDS model* (LCMDS). If the number of types of persons, S, is equal to 1, Clascal is but normal MDS. If $S = K$, Clascal corresponds to the Indscal model. For $1 < S < K$, Clascal estimates the probability that each person belongs to class S and, furthermore, computes an Indscal-like MDS solution for each class separately.

The dimension-weighting model therefore continues to be of interest. One may ask, however, whether it has led to noticeable substantive insights or to the establishment of scientific laws. In this regard, the model seems to have been much less successful, in contrast, for example, to the numerous regional laws established in the context of facet theoretical analyses of "normal" MDS data representations (see Chapter 5). Why is this so, even though the model certainly seems to be a plausible one? The answer may be found in the problems that we encountered with dimensional models in Chapter 17: if one takes a close look at dimensional models in the sense that the distance formula explains how dissimilarity judgments are generated from meaningful psychological dimensions, they are found to be less convincing, even in the case of stimuli as simple as rectangles. Adding interindividual differences to such models does not change things for the better. One should, therefore, be careful not to be misled by the dimension-weighting models: the dimensions they identify are not automatically meaningful ones, even though they may be rotationally unique.

22.7 Exercises

Exercise 22.1 Consider the three correlation matrices in Table 20.1 at p. 438.

(a) Without going into much theory, represent these data in the dimensionally weighted (DW) MDS model by using, for example, the Proxscal program in SPSS. How do you evaluate the outcome of this scaling effort?

(b) Scale each data matrix individually via MDS and then compare the configurations (by using Procrustean methods) and its Stress values with the DW solution.

(c) Use the DW configuration as a common starting configuration for an MDS scaling of each correlation matrix. How does this approach affect the MDS solutions?

Exercise 22.2 Consider Figure 17.7 at p. 373.

(a) Use the configuration of the 16 points on the *solid* grid to construct two different configurations, one by stretching this grid by factor 2 along the horizontal dimension (width), the other by stretching the grid by 2 along the vertical dimension (height). Compute the distances for the two resulting configurations.

(b) Use the two sets of distance as data in dimensional-weighting individual differences scaling. Check whether you succeed in recovering both the underlying configurations and the weights used in (a) to generate these distances.

(c) Add error to the distances and repeat the MDS analyses.

(d) Interpret the above weightings of the dimensions' width and height in substantive terms in the context of the perception of rectangles.

(e) Assume that you would generate more sets of distance matrices. This time, the configuration of points on the dashed grid in Figure 17.7 is stretched (or shrunk) along the dimensions width and height. Would these data lead to the same MDS configurations as the data generated above in (a)?

(f) Again assume that you would generate more sets of distances, this time by differentially stretching the configuration of points on the *dashed* grid in Figure 17.7 along a width-by-height coordinate system *rotated counterclockwise by 45 degrees*. Discuss what this would mean in terms of rectangle perception.

(g) Would you be able to discriminate persons using a weighted width-by-height model as in (a) and (e) from those using the rotated system in (f) by using INDSCAL or by using IDIOSCAL?

Exercise 22.3 Young (1987) reports the following hypothetical coordinates for four food stimuli and the dimension weights for five persons.

Food	I	II
Potato	-2	1
Spinach	-1	4
Lettuce	1	3
Tuna	4	-1

Person	I	II
1	.0	.9
2	.2	.8
3	.6	.6
4	.4	.4
5	.8	.2

.

(a) Interpret these data in the context of a dimensional salience model.

(b) Use the matrix equations of this model to compute the distance matrix for each person.

(c) Use a suitable MDS program to reconstruct the underlying configuration and weights from the set of distance matrices.

(d) Add an idiosyncratic rotation for each person, and repeat the above analyses with an MDS program that fits this model.

Exercise 22.4 The table below (Dunn-Rankin, Knezek, Wallace, & Zhang, 2004) shows ratings of five persons on the similarity of four handicaps: Learning Disability (LD), Mental Retardation (MR), Deafness (D), and Blindness (B).

Person	Handicap	LD	MR	D	B
1	LD	–			
1	MR	4	–		
1	D	4	5	–	
1	B	4	2	5	–
2	LD	–			
2	MR	6	–		
2	D	3	8	–	
2	B	2	2	4	–
3	LD	–			
3	MR	5	–		
3	D	4	6	–	
3	B	4	3	4	–

Person	Handicap	LD	MR	D	B
4	LD	–			
4	MR	2	–		
4	D	2	4	–	
4	B	6	2	5	–
5	LD	–			
5	MR	2	–		
5	D	6	7	–	
5	B	6	4	5	–

(a) Analyze these data with a dimensional salience model. Assess its fit.

(b) Interpret the dimensions of the solution space.

(c) Interpret the subject space (its meaning and how well it explains each person).