multidimensional Scaling
Mark L. Davison

CHAPTER 4 BF39.2 M85 D38
Lockwood

1992

Torgerson's Metric Group Method

Greatly assisted by Ledyard Tucker and building on work by Young and Householder (1938, 1941), Torgerson (1952, 1958) proposed one of the first MDS algorithms. Gower (1966, 1982) discusses and extends Torgerson's (1952) results. The assumptions of Torgerson's (1952) algorithm are much more restrictive than those of recent methods. Consequently, his approach is seldom used in its original form. Various features of his method have been incorporated into the algorithms described in Chapters 5 and 6 and, therefore, the Torgerson method will be described in some detail.

At this point, a few words are in order about the first example in this chapter. Table 4.1 shows a hypothetical dissimilarity matrix for six sports. The dissimilarity matrix was constructed to reflect two dimensions along which these sports differ: the speed of the games and the degree of contact between players. Hockey and football are two fast contact sports. Tennis and basketball are two fast noncontact sports. Golf and croquet are slow noncontact sports. The data in Table 4.1 will be used to illustrate how Torgerson's method can be employed to recover the two-dimensional configuration of stimuli underlying the data matrix in Table 4.1. The illustration will constitute a dimensional application of MDS.

TORGERSON'S MODEL

In Torgerson's model, the dissimilarity estimates comprising the data are assumed equal to distances in a Euclidean multidimensional space. Again let δ_{ij} be the dissimilarity between objects i and j. Let x_{ik} and x_{jk} ($i = 1,\ldots, I$; $j = 1,\ldots, J$; $I = J$; $k = 1,\ldots, K$) be the coordinates of stimuli i and j

Table 4.1. The Dissimilarity Matrix Δ for Five Sports and the Scalar Product Matrix Δ* Derived from It.

Dissimilarity Matrix Δ

	H	F	B	T	G	C	$\delta_{i.}^2$
Hockey	0.00	0.71	1.41	1.73	2.00	2.00	2.25
Football	0.71	0.00	1.41	1.73	2.00	2.00	2.25
Basketball	1.41	1.41	0.00	1.00	1.41	1.41	1.50
Tennis	1.73	1.73	1.00	0.00	1.00	1.00	1.50
Golf	2.00	2.00	1.41	1.00	0.00	0.71	1.92
Croquet	2.00	2.00	1.41	1.00	0.71	0.00	1.92
$\delta_{.j}^2$	2.25	2.25	1.50	1.50	1.92	1.92	

$\delta_{..}^2 = 1.89$ ← dot product

Scalar Product Matrix Δ*

	H	F	B	T	G	C
Hockey	1.31	1.06	−0.07	−0.57	−0.86	−0.86
Football	1.06	1.31	−0.07	−0.57	−0.86	−0.86
Basketball	−0.07	−0.07	0.56	0.06	−0.24	−0.24
Tennis	−0.57	−0.57	0.06	0.56	0.27	0.27
Golf	−0.86	−0.86	−0.24	0.27	0.98	0.73
Croquet	−0.86	−0.86	−0.24	0.27	0.73	0.98

along dimension k. Note that the number of rows in the dissimilarity matrix I will equal the number of columns J because the rows and columns correspond to the same stimuli. Torgerson's fundamental assumption is:

$$\delta_{ij} = d_{ij} = \left[\sum_k (x_{ik} - x_{jk})^2\right]^{1/2}. \tag{4.1}$$

Without loss of generality, it can be assumed that the mean coordinate along each stimulus dimension equals zero:

$$\sum_i x_{ik} = \sum_j x_{jk} = 0.0. \tag{4.2}$$

Torgerson began by constructing a double-centered matrix Δ* with elements δ_{ij}^* computed directly from the data matrix. A double-centered matrix is one in which the mean of the elements in each row and the mean of the elements in each column equals 0.0. Each element of the new matrix Δ^* is of the form:

$$\delta^*_{ij} = -\tfrac{1}{2}\left(\delta^2_{ij} - \delta^2_{i.} - \delta^2_{.j} + \delta^2_{..}\right). \tag{4.3}$$

Here $\delta^2_{i.}$, $\delta^2_{.j}$, and $\delta^2_{..}$ are defined as follows:

$$\delta^2_{i.} = \frac{1}{J}\sum_j \delta^2_{ij}$$

$$\delta^2_{.j} = \frac{1}{I}\sum_i \delta^2_{ij} \tag{4.4}$$

$$\delta^2_{..} = \frac{1}{IJ}\sum_i \sum_j \delta^2_{ij}.$$

The matrix Δ^* computed from matrix Δ is also shown in Table 4.1. The element in row 2 and column 3 of Δ^*, $\delta^*_{23} = -0.07$, was obtained by taking $\delta^2_{23} = (1.41)^2 = 2.00$, subtracting the mean of squared elements in row 2, $\delta^2_{2.} = 2.25$, subtracting the mean of squared elements in column 3, $\delta^2_{.3} = 1.50$, adding the grand mean squared element, $\delta^2_{..} = 1.89$, and multiplying the result by -0.50:

$$\begin{aligned}\delta^*_{23} &= -\tfrac{1}{2}\left(\delta^2_{23} - \delta^2_{2.} - \delta^2_{.3} + \delta^2_{..}\right)\\ &= -\tfrac{1}{2}(2.00 - 2.25 - 1.50 + 1.89) = -0.07.\end{aligned} \tag{4.5}$$

The element in row 4 column 5 of Δ^* was computed as follows:

$$\begin{aligned}\delta^*_{45} &= -\tfrac{1}{2}\left(\delta^2_{45} - \delta^2_{4.} - \delta^2_{.5} + \delta^2_{..}\right)\\ &= -\tfrac{1}{2}(1.00 - 1.50 - 1.92 + 1.89) = 0.27,\end{aligned} \tag{4.6}$$

and so forth.

Torgerson showed that if the data satisfy Eq. (4.1), then each element in the new matrix Δ^* would be of the form

$$\delta^*_{ij} = \sum_k x_{ik}x_{jk}. \tag{4.7}$$

For the interested reader, a proof is given later in this chapter. Equation (4.7) is the fundamental theorem around which Torgerson designed his algorithm. Matrix Δ^* is often called a *scalar product matrix* because, as Eq. (4.7) shows, each of its elements is the sum of products between scalars x_{ik} and x_{jk}.

The matrix form of Eq. (4.7) is as follows:

$$\Delta^* = \mathbf{XX}' \tag{4.8}$$

where $\mathbf{X}$ is the $(I \times K)$ matrix of stimulus coordinates. As described in Chapter 2, a principal components factor program can be used to find a matrix $\mathbf{X}$ satisfying Eq. (4.8) so long as such a matrix exists. Interested readers can consult a text on matrix algebra (Green, 1978; Hohn, 1973) for a description of methods for computing $\mathbf{X}$ and of the conditions under which such a matrix $\mathbf{X}$ will exist.

Figure 4.1 shows the matrix $\mathbf{X}$ obtained by extracting the first two principal components from the scalar product matrix Δ^* in Table 4.1.

ROTATION

The matrix $\mathbf{X}$ obtained from the principal components analysis is a solution of Eq. (4.8) but not the only solution. To see why it is not the only solution, imagine a $(K \times K)$ orthogonal transformation matrix $\mathbf{T}$. If $\mathbf{X}$ satisfies Eq. (4.8), then any matrix $\mathbf{X}^* = \mathbf{XT}$ will also satisfy Eq. (4.8). That is, if

$$\Delta^* = \mathbf{XX}', \tag{4.9}$$

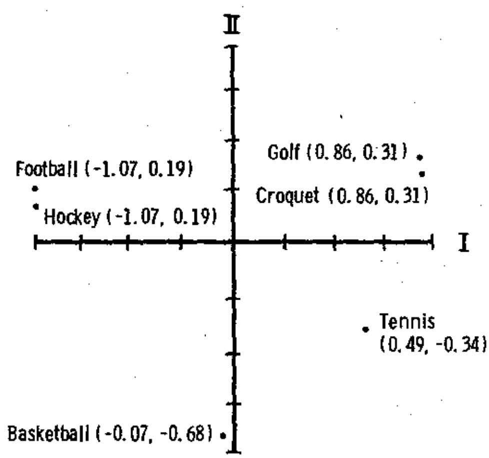

Figure 4.1. Unrotated metric scaling of sports data.

then

$$\Delta^* = \mathbf{X}^*\mathbf{X}^{*\prime}. \tag{4.10}$$

Since **T** is orthogonal, $\mathbf{TT}' = \mathbf{I}$ (see the definition of orthogonal transformation matrix in Chapter 2). Hence

$$\mathbf{X}^*\mathbf{X}^{*\prime} = (\mathbf{XT})(\mathbf{XT})' \tag{4.11}$$

Theorem 2 of Chapter 2 tells us that $(\mathbf{XT})'$ in Eq. (4.11) equals $\mathbf{T}'\mathbf{X}'$. Inserting this result into Eq. (4.11) yields

$$\begin{aligned} \mathbf{X}^*\mathbf{X}^{*\prime} &= (\mathbf{XT})(\mathbf{T}'\mathbf{X}') \\ &= \mathbf{X}(\mathbf{TT}')\mathbf{X}' \\ &= \mathbf{XIX}' \\ &= \mathbf{XX}' \\ &= \Delta^*. \end{aligned} \tag{4.12}$$

As the proof in Eq. (4.12) shows, if **X** is a solution to Eq. (4.8), then so is any matrix $\mathbf{X}^*$. If there are several rotations of **X** that can reproduce Δ^* equally well, then which rotation should one prefer?

The above question is usually a moot issue in data reduction or configural applications of MDS as long as K does not exceed two dimensions. In such a small number of dimensions, the important features of the configuration should be visually recognizable irrespective of rotation. In dimensional applications, however, it is not a moot issue. If the dimensions are not rotated to a suitable orientation, the coordinates will not correspond to meaningful stimulus attributes, and it will be difficult to interpret the dimensions. The phrase "meaningful stimulus attributes" will be explained below.

The configuration in Figure 4.1 illustrates the kind of interpretational problem that can occur. Dimension I has the slow noncontact games at one end, the fast contact games at the other, and fast noncontact sports in the middle. The scale represented by this dimension cannot, therefore, be interpreted as representing the speed of each game or the degree of contact. It is a confound of the two. Dimension II is also a confound of the two stimulus attributes, speed and contact. It has fast noncontact sports at the positive end. Slow noncontact and fast contact sports are found at the other end.

In deciding on a rotation for the solution, there are three basic options. If the unrotated solution is interpretable, one need not rotate the solution at all. Such is not the case for the data in Figure 4.1. Consequently, either an objective rotation or a hand rotation must be employed to obtain an interpretable representation of the configuration. It is to these latter options, objective and hand rotations, to which our attention now turns.

Objective Rotations

An objective rotation is a mathematical algorithm for finding an interpretable rotation of a solution. Since objective rotations were designed primarily for use in factor analysis and are used only occasionally in MDS, they will be mentioned briefly in this book. Interested readers can consult a text on factor analysis (Harman, 1976) for a detailed description of such rotations. The objective rotations commonly available were designed for rotating a factor analytic configuration of tests so the configuration conforms as closely as possible to the criterion of simple structure (Thurstone, 1947; Tucker, 1967). Stated simply, a MDS solution would satisfy the criterion of simple structure if each stimulus had a nonzero scale value on one or, at most, a few stimulus attributes. Such rotations are seldom used in MDS because, to this author's knowledge, there is no reason to believe that naturally occurring stimuli would satisfy this criterion. Nor is there any reason to believe that rotation to such a criterion will yield a more interpretable solution. In some applications, an objective rotation to simple structure, such as varimax (Kaiser, 1958) or equimax (Saunders, 1960), will yield a highly interpretable solution. Users should not, however, assume that such an objective rotational algorithm will automatically produce the most interpretable possible solution.

Figure 4.2 gives the varimax rotation of the original solution matrix in Figure 4.1. The varimax solution is little different from the unrotated solution, and hence shares its interpretational problems. For this particular example, the varimax dimensions are no easier to interpret than are the unrotated dimensions in Figure 4.1.

Hand Rotations

Computers perform objective rotations. People perform hand rotations. A hand rotation is one performed by the experimenter and chosen on the basis of her or his visual inspection of the unrotated configuration. In practice, one can sometimes see by inspection what rotation of the solution would yield interpretable dimensions. If dimension I in Figure 4.1 were rotated 45°, as shown in Figure 4.3, then all of the fast sports would fall at the

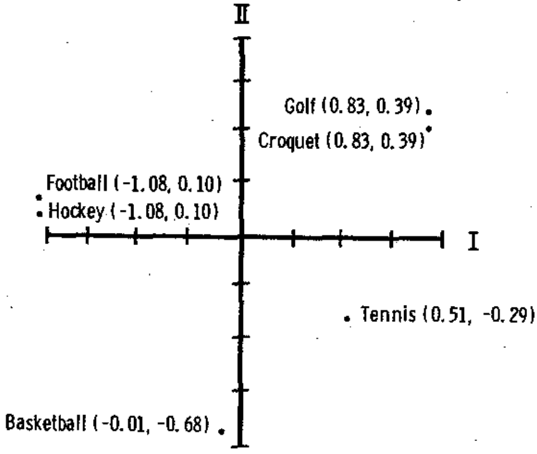

Figure 4.2. Varimax rotation of metric sports dimensions.

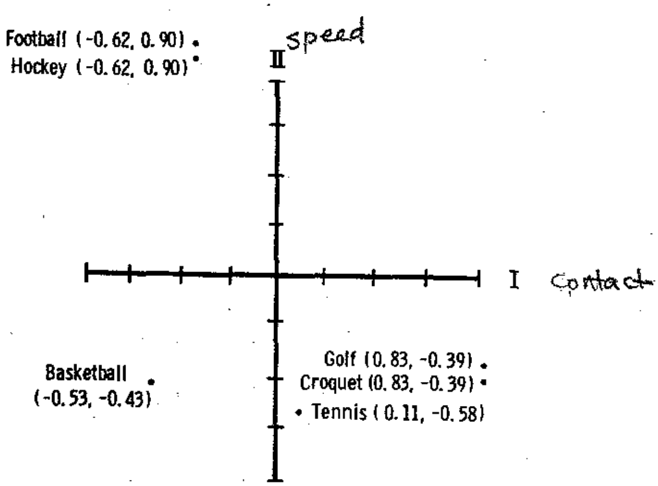

Figure 4.3. Hand rotation of metric sports data.

positive end. Slow sports would fall at the negative end. Scale values along the resulting dimension could be said to reflect the speed of the various sports. Rotating dimension II 45°, as shown in Figure 4.3, would yield an axis with contact sports at one end and noncontact sports at the other. The resulting dimension could be said to reflect the degree of contact in each game.

One can quickly compute the angles between dimensions I and II in Figure 4.2 and dimensions 1 and 2 in Figure 4.3. The angle between dimensions I and 1 would be 45°; the angle between dimensions II and 1 would be 315°; the angle between dimensions I and 2 would be 135°; the angle between dimensions II and 2 would be 45°. The corresponding cosines would be .71, .71, −.71, and .71. Assembling these cosines yields the orthogonal transformation matrix

$$\mathbf{T} = \begin{array}{c} \\ \mathrm{I} \\ \mathrm{II} \end{array} \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{bmatrix} .71 & -.71 \\ .71 & .71 \end{bmatrix} \end{array} \tag{4.13}$$

Postmultiplying the unrotated scale values $\mathbf{X}$ in Figure 4.1 by $\mathbf{T}$ yields the rotated coordinate matrix plotted in Figure 4.3. The resulting dimensions

$$\mathbf{X}^* = \begin{bmatrix} -0.62 & 0.90 \\ -0.62 & 0.90 \\ -0.53 & -0.43 \\ 0.11 & -0.58 \\ 0.83 & -0.39 \\ 0.83 & -0.39 \end{bmatrix} \begin{array}{l} \text{Hockey} \\ \text{Football} \\ \text{Basketball} \\ \text{Tennis} \\ \text{Golf} \\ \text{Croquet} \end{array} \tag{4.14}$$

can be interpreted as reflecting the degree of contact in and the speed of each sport.

In dimensional applications, finding an interpretable rotation is an important step in the MDS process. Several approaches are possible. If the original solution is interpretable, then the user need not rotate at all. If the unrotated solution is not easily interpreted, then an objective rotation such as varimax (Kaiser, 1958) or equimax (Saunders, 1960) can be tried. If neither the unrotated solution nor an objectively rotated solution yield interpretable dimensions, the user can try a hand rotation.

DIMENSIONALITY

To this point, the discussion of Torgerson's method has proceeded as if the number of dimensions K were known. In practice, however, it is not known and must be estimated in the analysis. In most MDS methods, the user must obtain several solutions in different dimensionalities and choose between them on the basis of three criteria: interpretability, fit to the data, and reproducibility. Torgerson's algorithm minimizes the following measure of fit: $F = \sum_{(i,j)}(\delta_{ij}^* - \sum_k \hat{x}_{ik}\hat{x}_{jk})^2$, where $\hat{x}_{ik}$ and $\hat{x}_{jk}$ are the estimates of coordinates for stimuli i and j along dimension k. That is, the algorithm minimizes the sum of squared discrepancies between the predicted, $\hat{\delta}_{ij}^* = \sum_k \hat{x}_{ik}\hat{x}_{jk}$, and actual scalar products, δ_{ij}^*.

Torgerson's method is one of two methods discussed in this book in which the fit measure plays little or no role in deciding how many dimensions are required to adequately reproduce the data. There is, however, a series of eigenvalues (also called characteristic roots or eigen roots) that do play a role in the dimensionality decision. Each eigenvalue is associated with one dimension in the solution. For our purposes, the eigenvalue associated with a given dimension is simply the sum of squared stimulus scale values along that dimension. That is, if we let $\hat{x}_{ik}$ refer to the estimated scale value for stimulus i along dimension k, then the kth eigenvalue is

$$\lambda_k = \sum_i \hat{x}_{ik}^2. \tag{4.15}$$

If one takes stimulus 1 to be hockey, stimulus 2 to be football, 3 to be basketball, 4 to be tennis, 5 to be golf, and 6 to be croquet, then in the solution of Figure 4.1, the second eigenvalue is simply

$$\begin{aligned} \lambda_2 = \sum \hat{x}_{i2}^2 &= \hat{x}_{12}^2 + \hat{x}_{22}^2 + \hat{x}_{32}^2 + \hat{x}_{42}^2 + \hat{x}_{52}^2 + \hat{x}_{62}^2 \\ &= (.19)^2 + (.19)^2 + (-.68)^2 + (-.34)^2 + (.31)^2 \qquad (4.16) \\ &+ (.31)^2 = .84. \end{aligned}$$

A plot like the one in Figure 4.4 can be useful in determining dimensionality. The vertical axis represents eigenvalues for the *unrotated* solution, and the horizontal axis corresponds to dimensions. The graph is constructed by plotting one point for each dimension at a height corresponding to the eigenvalue associated with that dimension. For instance, the point corresponding to the second dimension indicates that the eigenvalue associated with the second dimension of the unrotated solution was 0.84.

If the data conform exactly to the model of Eq. (4.1), then the plot should level off at exactly $(K + 1)$ dimensions, just as the plot in Figure 4.4 levels off at 3 dimensions. In other words, there should be an "elbow" in the graph one dimension beyond K, the correct number of dimensions. In real data that do not conform exactly to the model or in which there is a large amount of measurement and sampling error, an elbow may be difficult to discern. Indeed, the elbow is difficult to discern in Figure 4.4. In such cases, a plot of the eigenvalues may not suffice to determine the correct number of

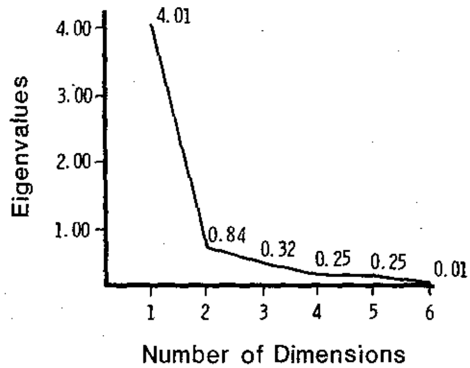

Figure 4.4. Eigenvalues plotted against dimensions for the unrotated scaling of sports.

dimensions. Interpretation and reproducibility of dimensions must also be considered.

Reproducibility can be used as a criterion only when there are two or more subsamples. The basic idea is that one should retain as many dimensions in the final solution as emerge consistently in the separate subsamples. If one derives separate solutions for each subsample, and there are *K* dimensions that appear consistently in all of the subsamples, then the final solution should contain exactly *K* dimensions. Each of the subsamples should come from the same population.

Interpretability as a criterion requires some subjective judgment on the user's part. The basic idea, however, is that a higher dimensional solution is preferred over a lower dimensional solution if there are important stimulus features that appear in the higher dimensional solution but fail to appear in the lower dimensional solution. Conversely, the lower dimensional solution is preferred if there are no important stimulus features that fail to appear in the lower dimensional solution.

In our example, neither stimulus feature, degree of contact, nor speed of game can be distinctly discerned in the one-dimensional solution, which consists solely of the first dimension plotted in Figure 4.1. Both features are confounded in the lone dimension of the one-dimensional solution. Only after the extraction of two dimensions can the solution be rotated so that each stimulus feature corresponds to a unique dimension as in Figure 4.3. Because it is more readily interpreted, the two-dimensional solution is preferred.

INTERPRETATION

Interpretability was discussed above in the section on deciding dimensionality, but a few more comments need to be made about interpreting solutions.

Particularly, the phrase "meaningful stimulus features" from the prior discussion needs clarification. Such features are typically either orderings or groupings of stimuli.

A substantively meaningful grouping of stimuli is a set of stimuli that cluster together in a region of the multidimensional solution space and that possess some common attribute. For instance, in a study of occupations, sales jobs may cluster together to form a meaningful grouping. Women's magazines (*MS.*, *Ladies Home Journal*, *Vogue*, etc.) may group together in a study of popular periodicals.

A meaningful ordering of stimuli is an ordering that corresponds to the arrangement of stimuli along an important stimulus attribute. For instance, the ordering of the stimuli along dimension I in Figure 4.3 corresponds to their ordering by degree of contact. The ordering along dimension II corresponds to the ordering of the games by speed. Both of these dimensions represent meaningful orderings because they correspond to important attributes of games, speed, and contact. Ideally, the dimensions will be rotated so that each represents one of the meaningful orderings.

Interpreting a solution involves identifying the important groupings and orderings of stimuli. For groupings, one must identify the feature or features that the members of each cluster share in common. For orderings, one must identify the attribute corresponding to the ordering. One way to interpret a solution is by simple inspection of the configuration. More formal methods are considered in Chapter 8.

EXAMPLE

Smith and Siegel (1967) use MDS to derive dimensions of job tasks for the position Office of Civil Defense (OCD) director. In three successive stages, they identify 34 job functions which, in their opinion, are representative of the total tasks performed by OCD directors. Thirty-five supervisory level OCD personnel then rate the dissimilarity of each pair of tasks on an 11-point scale. For each job task pair, the dissimilarity judgments of the 35 subjects are pooled to obtain a dissimilarity matrix.

Torgerson's algorithm was used to obtain a four-dimensional solution. Smith and Siegel use an objective rotation called equimax (Saunders, 1960). For each of the four dimensions, Table 4.2 shows the functions that have the highest positive scale values and the lowest negative scale values. The authors summarize their interpretations of the dimensions in the labels they assign to each: internal versus external system maintenance (dimension I), routine versus emergency programming (dimension II), resource use versus resource evaluation (dimension III), and emergency system integration (dimension IV).

Table 4.2. Dimensions of Civil Defense Director Job Tasks.

Scale Value	Job Task
	Dimension I: Internal vs. External System Maintenance
3.43	Issuing necessary emergency orders and instructions to the public
4.21	Prescribing information channels
4.70	Relaying information from higher-level CD organizations
5.66	Informing public and private groups of CD activities
−3.19	Assisting in legal actions arising from CD activities
−3.96	Accounting for CD funds and property
−4.94	Assuring the preservation of essential CD records
	Dimension II: Routine vs. Emergency Programming
3.58	Administering the protected facilities program
4.46	Preparing and presenting a CD budget
4.82	Inspecting and reporting on installations assigned to or related to CD
6.39	Accounting for CD funds and property
−2.16	Prescribing mobilization procedures
−2.24	Evaluating potential emergencies
−3.00	Programming the continuity of government
−4.88	Establishing the order of succession within the CD system
	Dimension III: Resource Use vs. Resource Evaluation
2.81	Accounting for CD funds and property
4.27	Advising on needed CD legislation
6.33	Assisting on legal actions arising from CD activities
−3.20	Conducting required research
−3.68	Determining the availability of human and material resources
−3.94	Assuring the proficiency of CD workers
−4.65	Conducting and evaluating CD tests
	Dimension IV: Emergency System Integration
3.79	Advising on needed CD legislation
4.14	Preparing and presenting a CD budget
4.26	Appointing CD technical advisory committees
4.91	Conducting required research
−2.11	Prescribing information channels
−2.49	Relaying information from higher-level CD organizations
−3.26	Maintaining liaison with federal and state military groups
−3.35	Issuing necessary emergency orders and instructions to the public
−4.90	Alerting and mobilizing the CD system

Source: Smith and Siegel (1967). Copyright 1967 by the American Psychological Association. Adapted by permission of the publisher and author.

The authors suggest that the obtained dimensions could be used as a basis for developing unidimensional, employee evaluation scales. For instance, since resource evaluation tasks emerged at one pole of dimension III, one may wish to develop a corresponding employee evaluation scale. The authors go on to suggest that employee selection measures and training programs might be planned around the obtained dimensions.

PROOF: TORGERSON'S FUNDAMENTAL THEOREM†

Torgerson's method rests on his proof that if one starts from a data matrix $\boldsymbol{\Delta}$ with elements of the form

$$\delta_{ij} = d_{ij} = \left[\sum_k (x_{ik} - x_{jk})^2 \right]^{1/2}, \tag{4.17}$$

and one applies the following transformation to those elements

$$\delta_{ij}^* = -\tfrac{1}{2}\left(\delta_{ij}^2 - \delta_{i.}^2 - \delta_{.j}^2 + \delta_{..}^2\right), \tag{4.18}$$

then one obtains quantities of the form

$$\delta_{ij}^* = \sum_k x_{ik} x_{jk}. \tag{4.19}$$

Although Torgerson did not do so, one can assume without loss of generality that

$$\sum_i x_{ik} = \sum_j x_{jk} = 0 \qquad \text{for all } k. \tag{4.20}$$

As preliminary steps in the proof, it is necessary to derive expressions for $\delta_{i.}^2$, $\delta_{.j}^2$, and $\delta_{..}^2$ in terms of the stimulus coordinates x_{ik} and x_{jk}. The re-expression of $\delta_{i.}^2$ in terms of stimulus coordinates will be sought first.

Squaring and expanding the term on the right side of Eq. (4.17) yields

$$\delta_{ij}^2 = \sum_k x_{ik}^2 + \sum_k x_{jk}^2 - 2\sum_k x_{ik} x_{jk}. \tag{4.21}$$

Squaring and taking the average over j of the quantity on the left side of Eq.

†Throughout this book, an asterisk designates a more technical section containing a proof or a description of an algorithm. Readers can omit such sections without loss of continuity.

(4.17) yields

$$\delta_{i.}^2 = \frac{1}{J}\sum_j \delta_{ij}^2. \tag{4.22}$$

Substituting the quantity on the right side of Eq. (4.21) for the expression on the right side of Eq. (4.22) yields

$$\begin{aligned}\delta_{i.}^2 &= \frac{1}{J}\sum_j\left(\sum_k x_{ik}^2 + \sum_k x_{jk}^2 - 2\sum_k x_{ik}x_{jk}\right)\\ &= \frac{1}{J}\sum_j\sum_k x_{ik}^2 + \frac{1}{J}\sum_j\sum_k x_{jk}^2 - 2\frac{1}{J}\sum_j\sum_k x_{ik}x_{jk}.\end{aligned} \tag{4.23}$$

Consider the third term on the right side of Eq. (4.23), $-2(1/J)\Sigma_j\Sigma_k x_{ik}x_{jk}$. Since x_{ik} does not depend on j, this term can be rewritten as $-2(1/J)\Sigma_k x_{ik}(\Sigma_j x_{jk})$. According to Eq. (4.20), $\Sigma_j x_{jk} = 0$, and hence $-2(1/J)\Sigma_k x_{ik}(\Sigma_j x_{jk}) = 0$. Consequently, Eq. (4.23) can be rewritten as

$$\delta_{i.}^2 = \sum_k x_{ik}^2 + \sum_k x_{.k}^2, \tag{4.24}$$

where $x_{.k}^2 = (1/J)\Sigma_j x_{jk}^2$. Equation (4.24) states that $\delta_{i.}^2$ can be expressed as the sum of squared coordinates for stimulus i plus the sum of the average squared coordinates. Both of these sums are taken across the K dimensions. Equation (4.24) provides the desired re-expression of $\delta_{i.}^2$ in terms of stimulus coordinates. Let's turn our attention now to a similar re-expression of $\delta_{.j}^2$.

Squaring and taking the average over i of the quantity on the left side of Eq. (4.17) yields

$$\delta_{.j}^2 = \frac{1}{I}\sum_i \delta_{ij}^2. \tag{4.25}$$

Substituting the quantity on the right side of Eq. (4.21) for the expression on the right side of Eq. (4.25) yields

$$\begin{aligned}\delta_{.j}^2 &= \frac{1}{I}\sum_i\left(\sum_k x_{ik}^2 + \sum_k x_{jk}^2 - 2\sum_k x_{ik}x_{jk}\right)\\ &= \frac{1}{I}\sum_i\sum_k x_{ik}^2 + \frac{1}{I}\sum_i\sum_k x_{jk}^2\\ &\quad -2\frac{1}{I}\sum_i\sum_k x_{ik}x_{jk}.\end{aligned} \tag{4.26}$$

Consider the third term on the right side of Eq. (4.26). Since x_{jk} does not depend on i, the term can be rewritten as $-2(1/I)\Sigma_k x_{jk}(\Sigma_i x_{ik})$. According to Eq. (4.20), $\Sigma_i x_{ik} = 0$, and hence $-2(1/I)\Sigma_k x_{jk}(\Sigma_i x_{ik}) = 0$. Consequently, Eq. (4.26) can be written as

$$\delta_{.j}^2 = \sum_k x_{.k}^2 + \sum_k x_{jk}^2, \tag{4.27}$$

where

$$x_{.k}^2 = \frac{1}{I}\sum_i x_{ik}^2 = \frac{1}{J}\sum_j x_{jk}^2.$$

Equation (4.27) expresses $\delta_{.j}^2$ in a form directly analogous to the expression for $\delta_{i.}^2$ in Eq. (4.24), and it provides the desired re-expression of $\delta_{.j}^2$ in terms of stimulus coordinates. Before beginning the proof of Torgerson's fundamental theorem, Eq. (4.19), we need only re-express $\delta_{..}^2$ in terms of stimulus coordinates.

Squaring and taking the average over j and i of the quantity on the left side of Eq. (4.17) yields

$$\delta_{..}^2 = \frac{1}{IJ}\sum_i \sum_j \delta_{ij}^2. \tag{4.28}$$

Substituting the quantity on the right side of Eq. (4.21) for the expression on the right side of Eq. (4.28) yields

$$\begin{aligned}\delta_{..}^2 &= \frac{1}{IJ}\sum_i \sum_j \left(\sum_k x_{ik}^2 + \sum_k x_{jk}^2 - 2\sum_k x_{ik}x_{jk}\right)\\ &= \frac{1}{IJ}\sum_i \sum_j \sum_k x_{ik}^2 + \frac{1}{IJ}\sum_i \sum_j \sum_k x_{jk}^2 \\ &\quad -2\frac{1}{IJ}\sum_i \sum_j \sum_k x_{ik}x_{jk}.\end{aligned} \tag{4.29}$$

Consider the third term on the right side of Eq. (4.29). According to Eq. (4.20), $\Sigma_i x_{ik} = \Sigma_j x_{jk} = 0$, and hence $-2(1/IJ)\Sigma_i\Sigma_j\Sigma_k x_{ik}x_{jk} = -2(1/IJ)\Sigma_k(\Sigma_i x_{ik})(\Sigma_j x_{jk}) = 0$. Consequently, Eq. (4.29) can be rewritten as

$$\delta_{..}^2 = \sum_k x_{.k}^2 + \sum_k x_{.k}^2 = 2\sum_k x_{.k}^2. \tag{4.30}$$

To prove Torgerson's theorem in Eq. (4.19), we need only combine the results in Eqs. (4.24), (4.27), and (4.30) with the expression in Eqs. (4.18) and (4.21).

For the four terms on the right side of Eq. (4.18): δ_{ij}^2, $\delta_{i.}^2$, $\delta_{.j}^2$, and $\delta_{..}^2$, it is necessary to substitute the corresponding expressions on the right sides of Eqs. (4.21), (4.24), (4.27), and (4.30). These substitutions yield

$$\delta_{ij}^* = -\tfrac{1}{2}\Bigg[\Big(\sum_k x_{ik}^2 + \sum_k x_{jk}^2 - 2\sum_k x_{ik}x_{jk}\Big) - \Big(\sum_k x_{ik}^2 + \sum_k x_{.k}^2\Big) - \Big(\sum_k x_{.k}^2 + \sum_k x_{jk}^2\Big) + 2\sum_k x_{.k}^2\Bigg]. \tag{4.31}$$

Combining terms and multiplying through by $(-\frac{1}{2})$ yields the desired result.

$$\delta_{ij}^* = \sum_k x_{ik}x_{jk}. \tag{4.32}$$

Thus one arrives at Torgerson's fundamental theorem expressed originally in this chapter by Eq. (4.7).

OTHER METRIC MODELS

Torgerson's (1952, 1958) algorithm makes very restrictive assumptions. A slightly less restrictive model is the following:

$$\delta_{ij} = d_{ij} + c. \tag{4.33}$$

where c is an additive constant. One way to analyze such proximity data would be to first estimate c, subtract the estimate of c from each proximity δ_{ij}, and then analyze the new data points $(\delta_{ij} - \hat{c})$ via Torgerson's algorithm. The problem of first estimating c is often called the *additive constant problem* in MDS (Cooper, 1972).

One can set the estimate of c equal to the following:

$$\hat{c} = (-1)\max_{(h,i,j)}\big(\delta_{hj} - \delta_{hi} - \delta_{ij}\big). \tag{4.34}$$

If one generates new proximities γ_{ij} such that

$$\gamma_{ij} = 0 \qquad \text{if } i = j \qquad (4.35)$$

$$\gamma_{ij} = \delta_{ij} - c \qquad \text{if } i \neq j,$$

then the proximities γ_{ij} will satisfy the triangular inequality of Eq. (1.4). The additive constant estimate $\hat{c}$ is the smallest value one can subtract from each proximity δ_{ij} $(i \neq j)$ that will ensure that the transformed data satisfy the triangular inequality (Carroll and Wish, 1974a).

Ramsey (1978, 1980) developed maximum likelihood estimates for the stimulus coordinates. He proposed two models. The first assumes that each proximity δ_{ij} is a normally distributed random variable with unknown mean $\mu_{ij} = d_{ij}$ and variance σ^2. The other model assumes that the natural logarithms of the data $\ln(\delta_{ij})$ are normally distributed random variables with unknown means $\mu_{ij} = d_{ij}$ and variance σ^2. The maximum likelihood theory on which these algorithms are based makes it possible to develop a fit measure that is approximately distributed as a chi square variable under the null hypothesis represented by the scaling model. Early versions of Ramsey's algorithm required so much computer time that they were practical only for small data sets. If the computational problems can be overcome, the maximum likelihood approach may enable researchers to examine the fit of the model to their data more rigorously than has been possible with other approaches.

SUMMARY

Torgerson (1952) assumed that dissimilarities were equal to distances in a Euclidean space. From this assumption, he derived one of the first multidimensional scaling algorithms. Using data that satisfy Torgerson's metric assumption, one can solve for the coordinate dimensions by applying a principal components analysis to the scalar product matrix Δ^*.

One can decide upon the number of dimensions by considering the replicability of dimensions across subsamples, the interpretability of solutions with varying numbers of dimensions, and a dimensions-by-eigenvalues plot. The solution can be left unrotated, it can be rotated by hand, or it can be rotated by some objective algorithm such as varimax (Kaiser, 1958) or equimax (Saunders, 1960). Of these three rotation options, the one that gives the most interpretable orientation of the axes is preferred. Interpreting the solution involves identifying groupings of stimuli or orderings of stimuli that correspond to meaningful stimulus attributes.

Smith and Siegel (1967) use Torgerson's algorithm to derive dimensions of job task performance. They conclude that the dimensions could be used as a basis for developing unidimensional employee evaluation scales and for planning employee training programs.

PROBLEMS

1. Imagine that matrix Δ below contains dissimilarity data for all possible pairs of eight countries. Compute the scalar product matrix Δ^* from the matrix of dissimilarities.

$\Delta =$

	An	Ar	Au	Ch	Cu	J	US	Z
Angola	0.00	1.41	1.00	1.00	1.41	1.41	1.73	0.71
Argentina	1.41	0.00	1.00	1.73	1.41	1.41	1.00	1.41
Australia	1.00	1.00	0.00	1.41	1.73	1.00	1.41	1.00
China	1.00	1.73	1.41	0.00	1.00	1.00	1.41	1.00
Cuba	1.41	1.41	1.73	1.00	0.00	1.41	1.00	1.41
Japan	1.41	1.41	1.00	1.00	1.41	0.00	1.00	1.41
United States	1.73	1.00	1.41	1.41	1.00	1.00	0.00	1.73
Zimbabwe	0.71	1.41	1.00	1.00	1.41	1.41	1.73	0.00

2. Use a principal components analysis to extract three components from the scalar product matrix computed in Problem 1. Be sure to have all eight eigenvalues printed. Also, print both the unrotated and varimax rotated solutions in three dimensions. Then answer each of the following questions.
   a. What are the scale values along the first three unrotated dimensions? Can you interpret these dimensions?
   b. What are the scale values along the first three varimax rotated dimensions? Can you interpret these dimensions?
   c. What are the eight eigenvalues?

3. Construct a dimensions-by-eigenvalues plot. How many dimensions does this plot suggest should be retained?

4. Apply the transformation below to the varimax rotated dimensions? What are the obtained scale values? Interpret each of the dimensions.

$$\mathbf{T} = \begin{bmatrix} .63 & .53 & -.63 \\ .55 & -.63 & .59 \\ -.59 & -.63 & -.55 \end{bmatrix}$$

Answers

1.

$$\Delta = \begin{array}{c|cccccccc} & \text{An} & \text{Ar} & \text{Au} & \text{Ch} & \text{Cu} & \text{J} & \text{US} & \text{Z} \\ \hline \text{An} & .70 & -.21 & .17 & .17 & -.21 & -.34 & -.71 & .45 \\ \text{Ar} & -.21 & .89 & .26 & -.74 & -.12 & -.24 & .39 & -.21 \\ \text{Au} & .17 & .26 & .64 & -.37 & -.74 & .14 & -.24 & .17 \\ \text{Ch} & .17 & -.74 & -.37 & .64 & .26 & .14 & -.24 & .17 \\ \text{Cu} & -.21 & -.12 & -.74 & .26 & .89 & -.24 & .39 & -.21 \\ \text{J} & -.34 & -.24 & .14 & .14 & -.24 & .64 & .26 & -.34 \\ \text{US} & -.71 & .39 & -.24 & -.24 & .39 & .26 & .89 & -.71 \\ \text{Z} & .45 & -.21 & .17 & .17 & -.21 & -.34 & -.71 & .70 \end{array}$$

2. a.

$$\mathbf{X} = \begin{array}{c|ccc} & \text{I} & \text{II} & \text{III} \\ \hline \text{An} & -.73 & .00 & -.21 \\ \text{Ar} & .42 & .71 & -.45 \\ \text{Au} & -.28 & .71 & .26 \\ \text{Ch} & -.28 & -.71 & .26 \\ \text{Cu} & .42 & -.71 & -.45 \\ \text{J} & .24 & .00 & .77 \\ \text{US} & .95 & .00 & .04 \\ \text{Z} & -.73 & .00 & -.21 \end{array}$$

These dimensions are difficult, if not impossible, to interpret.

b.

$$\mathbf{X} = \begin{array}{c|ccc} & \text{I} & \text{II} & \text{II} \\ \hline \text{An} & -.75 & .10 & .10 \\ \text{Ar} & .18 & -.92 & .08 \\ \text{Au} & -.14 & -.25 & .75 \\ \text{Ch} & -.14 & .75 & -.25 \\ \text{Cu} & .18 & .08 & -.92 \\ \text{J} & .56 & .41 & .41 \\ \text{US} & .87 & -.27 & -.27 \\ \text{Z} & -.75 & .10 & .10 \end{array}$$

These dimensions are difficult, if not impossible, to interpret.

c.

Dimension	I	II	III	IV	V	VI	VII	VIII
Eigenvalue	2.55	2.01	1.22	.25	.02	.01	−.00	−.07

3. Figure 4.5 suggests that $K = 3$ because there is an elbow above the fourth dimension.

4.

$$\mathbf{X} = \begin{array}{lrrr} & \text{I} & \text{II} & \text{III} \\ \text{An} & -.48 & -.53 & .48 \\ \text{Ar} & -.45 & .63 & -.70 \\ \text{Au} & -.67 & -.39 & -.48 \\ \text{Ch} & .48 & -.39 & .67 \\ \text{Cu} & .70 & .63 & .45 \\ \text{J} & .34 & -.22 & -.34 \\ \text{US} & .56 & .80 & -.56 \\ \text{Z} & -.48 & -.53 & .48 \end{array}$$

Along the first dimension, stimuli are roughly arrayed along an axis running from north to south. The countries in the southern hemisphere—Angola, Argentina, Australia, and Zimbabwe—fall at the negative end of this axis. Countries in the northern hemisphere—China, Cuba, Japan, and the United States—fall at the positive end of this dimension. The arrangement from north to south is not perfect, how-

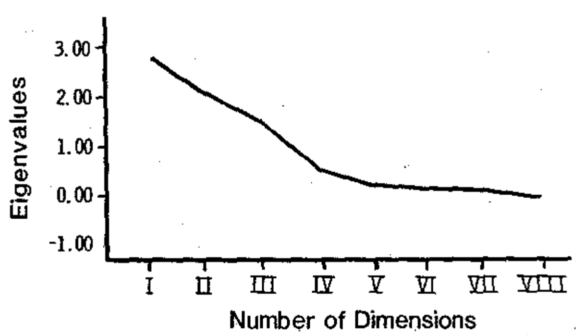

Figure 4.5. Dimensions-by-eigenvalues plot for metric scaling of countries' data.

ever; note, for instance, that Cuba has a higher scale value than the United States, even though Cuba is farther south. Nevertheless, dimension I can be interpreted as roughly reflecting the locations of the countries on a north–south axis.

Dimension II seems to be an east–west axis. The countries in the western hemisphere—Argentina, Cuba, and the United States—are located at the positive end of this dimension. Countries in the eastern hemisphere—Angola, Australia, China, Japan, and Zimbabwe—are found at the negative end.

Dimension III appears to be a Marxist–capitalist dimension. Countries headed by Marxist governments (in 1980)—Angola, China, Cuba, and Zimbabwe—appear at the positive end of this dimension. At the negative end, one will find the capitalist countries—Argentina, Australia, Japan, and the United States.