19
Euclidean Embeddings

Distances are functions that can be defined on *any* set of objects. Euclidean distances, in contrast, are functions that can only be defined on sets that possess a particular structure. Given a set of dissimilarities, one can test whether these values are distances and, moreover, whether they can even be interpreted as Euclidean distances. More generally, one can ask the same questions allowing for particular transformations of the given dissimilarities such as adding a constant to each value. For ordinal transformations, the hypothesis that dissimilarities are Euclidean distances is trivially true. Hence, in ordinal MDS, we learn nothing from the fact that the dissimilarities can be represented in a Euclidean space. In interval MDS, in contrast, Euclidean embedding is not trivial. If the data can be mapped into Euclidean distances, one can ask how many dimensions at most are necessary for a perfect representation. A further question, related to classical MDS, is how to find an interval transformation that leads to approximate Euclidean distances, while keeping the dimensionality of the MDS space as low as possible.

19.1 Distances and Euclidean Distances

Given a matrix of distances, one can ask whether these distances can be interpreted as Euclidean distances. This is true only if they can be embedded into a Euclidean space. The answer is positive if the scalar product matrix $\mathbf{B}$ derived from these distances (see Section 7.9 or 18.4) can be de-

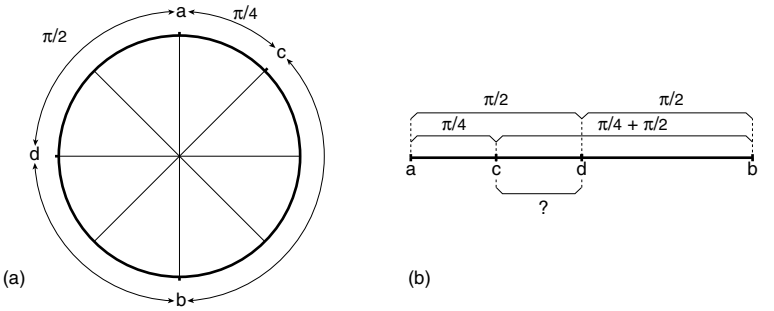

FIGURE 19.1. (a) Radian distances among points $a, \ldots, d$, and (b) their interpretation as Euclidean distances.

TABLE 19.1. Distances between points $a, \ldots, d$ on the circle in Figure 19.1a measured along the circle (radius=1).

Point	a	b	c	d
a	0.0000	3.1416	0.7854	1.5708
b	3.1416	0.0000	2.3562	1.5708
c	0.7854	2.3562	0.0000	2.3562
d	1.5708	1.5708	2.3562	0.0000

composed into $\mathbf{B} = \mathbf{X}\mathbf{X}'$, with real $\mathbf{X}$, or, equivalently, if $\mathbf{B}$'s eigenvalues are nonnegative (see Chapter 7). Conversely, if $\mathbf{B}$ has negative eigenvalues, the dissimilarities on which it is based can still be distances, albeit non-Euclidean distances. Consider an example.

Distances on a Circle

Figure 19.1a shows a configuration of four points on a circle. To determine their distances, we usually employ a straight ruler. This yields Euclidean distances. But here we measure the length of the shortest path ("geodesic") between points i and j on the circle. The circumference of a circle with radius 1 is equal to 2π. Thus, $d_{ab} = \pi$, $d_{ac} = \pi/4$, and so on, leading to the values in Table 19.1. These values are definitely distances: they are symmetric, they are nonnegative and exactly equal to 0 in the main diagonal, and the triangle inequality holds for all triples.

In fact, all triangle inequalities turn out to be equalities; for example, $d_{ab} = d_{ac} + d_{cb}$. In Euclidean geometry, this implies that a, b, and c lie on a straight line. Moreover, $d_{ab} = d_{ad} + d_{db}$ and, thus, the points a, b, and d must also lie on a straight line if the dissimilarities are interpreted as Euclidean distances. But in Euclidean geometry, there is just one line through the points a and b; hence, a, b, c, and d must all lie on it.

Figure 19.1b shows this line. The points c and d are positioned on it so that their distances satisfy the two triangle equalities above, and this implies that the distance between c and d should be $\pi/4$, which, however, is not in agreement with the value in Table 19.1.

Similarly, the scalar-product matrix derived from the distances in Table 19.1 using formula (18.17) yields the eigenvalues 5.61, 2.22, 0.00, and −1.21. Hence, this matrix is not positive semidefinite and so we are led to the same conclusion as before: the distances in Table 19.1 cannot be embedded into a Euclidean space.

Properties of Euclidean Distances

Because we did not arrive at the values in Table 19.1 by using a straight ruler, they cannot be Euclidean distances. Indeed, checking through them, we are led to contradictions if we assume that they were. Euclidean distances, therefore, have properties above and beyond those of general distances. The contradiction to which we were led in Figure 19.1b rests on the fact that for Euclidean distances there is just one geodesic path between any two points; that is, all points x that satisfy $d_{ab} = d_{ax} + d_{xb}$ must lie between a and b on the line through a and b.

This is not always true for other Minkowski distances. If points a and b lie on a line not parallel to the coordinate axes, then the city-block metric, for example, allows for infinitely many geodesics between a and b, so that the above triangle equality for x does not mean that x will be crossed if we move from a to b on a path of length d_{ab}. Hence, other Minkowski distances have special properties that require investigation.

Investigations of a mathematical structure typically begin by considering particular cases (such as, e.g., a plane with a straight-ruler distance measurement). One then attempts to describe the "essential" properties of these cases and to write them up in a simple list of axioms from which all of the theorems one has in mind may be proved. The axioms should be abstract in the sense that they do not rely on ad hoc features of the cases such as the dimensionality of the chosen geometry or a particular coordination for its points.

Euclidean distances are defined abstractly (coordinate-free and dimension-free) as the square root of the scalar product $b(\mathbf{i}-\mathbf{j}, \mathbf{i}-\mathbf{j})$, where $\mathbf{i}-\mathbf{j}$ is the difference vector of the vectors $\mathbf{i}$ and $\mathbf{j}$. Thus, Euclidean distances have properties related to those of scalar products. Two of these properties correspond to the axioms of (general) distances, namely, *symmetry* and *nonnegativity*. The remaining property, *linearity*, brings in the special properties of Euclidean distances: $b(s \circ \mathbf{u} + t \circ \mathbf{v}, \mathbf{w}) = s \cdot b(\mathbf{u}, \mathbf{w}) + t \cdot b(\mathbf{v}, \mathbf{w})$, for any vectors $\mathbf{u}$,$\mathbf{v}$,$\mathbf{w}$ and scalars s, t. The operation $\cdot$ denotes the usual multiplication of real numbers, whereas $\circ$ is different. It denotes that a vector is multiplied by a number (scalar). Also, + denotes addition of vectors,

not the usual addition of numbers.[1] As long as the rules that govern scalar multiplication and vector addition are not specified, the linearity axiom remains meaningless. But what are these rules?

The rules are collected in a system of axioms known as *Abelian vector spaces*. It comprises two structures, a *field* and a *group*. The field is usually the set of real numbers, with its two operations of addition and multiplication. A group is a set of elements with one operation that satisfies the following axioms.

g1 for any three of its elements, x, y, and z, $(x + y) + z = x + (y + z)$ ($+$ is *associative*);

g2 there exists a *zero element*, z, so that $x + z = x$, for any x;

g3 there exists an *inverse element* $x^{(i)}$ for any x, so that $x + x^{(i)} = z$;

g4 (for Abelian groups only) for any elements x, y, $x + y = y + x$ ($+$ is *commutative*).

A vector space ties together the field and the group (whose elements are now called vectors and written in this book in bold fonts) by an operation $\circ$ so that:

v1 $k \circ (\mathbf{x} + \mathbf{y}) = k \circ \mathbf{x} + c \circ \mathbf{y}$;

v2 $(s + t) \circ \mathbf{x} = s \circ \mathbf{x} + t \circ \mathbf{x}$;

v3 $s \circ (t \circ \mathbf{x}) = (s \cdot t) \circ \mathbf{x}$;

v4 $e \circ \mathbf{x} = \mathbf{x}$,

where s, t, e are scalars, e is the neutral element of the field, and $\mathbf{x}, \mathbf{y}$ are any elements of the group.

What does that tell us? It means that when we talk about Euclidean distances we are necessarily talking (at least by implication) about a rich mathematical structure. The notion of Euclidean distance is defined only in this system. It can be defined on a set of points u, v, w only if these points are first linked to corresponding elements $\mathbf{u}$,$\mathbf{v}$,$\mathbf{w}$ of a vector space ("embedding"). Distances in general need no such structural embeddings. The trivial distance, for example, defined as $d_{ij} = 1$ and $d_{ii} = 0$ for all i, j, exists on any set of elements i, j, whether they can be interpreted as vectors or not.

This also means that the properties of vector spaces cannot be tested for any finite set of vectors, because they must hold, for example, for any

[1] A different symbol (such as $\oplus$) might be better to denote vector addition. We do not use such particular notation here because we are almost always dealing with vectors that are n-tuples of real numbers in this book. In this case, addition of vectors is defined as the familiar addition of corresponding elements.

TABLE 19.2. (a) Dissimilarities for five objects; - denotes a missing value; (b) completing the proximity matrix by setting $\delta_{ii} = 0$ and $\delta_{ij} = \delta_{ji}$, for all i, j; (c) matrix after adding 4.8 to each element.

(a)	1	2	3	4	5
1	-	-	-	-	-
2	0.2	-	-	-	-
3	1.2	0.2	-	-	-
4	0.2	3.2	0.2	-	-
5	-1.8	-0.8	-1.8	-0.8	-

(b)	1	2	3	4	5
1	0.0	0.2	1.2	0.2	-1.8
2	0.2	0.0	0.2	3.2	-0.8
3	1.2	0.2	0.0	0.2	-1.8
4	0.2	3.2	0.2	0.0	-0.8
5	-1.8	-0.8	-1.8	-0.8	0.0

(c)	1	2	3	4	5
1	0	5	6	5	3
2	5	0	5	8	4
3	6	5	0	5	3
4	5	8	5	0	4
5	3	4	3	4	0

scalars s and t, and, therefore, involve *all* vectors of the space. This is why testing whether a given set of numbers are Euclidean distances is often called, more correctly, testing whether these numbers can be embedded into distances of a Euclidean space.

19.2 Mapping Dissimilarities into Distances

MDS models almost never assume that the given dissimilarities are distances. Rather, all models (except absolute MDS) admit some transformation on the dissimilarities such as, for example, a free choice of additive and multiplicative constants on the dissimilarities in interval MDS. We now study to what extent one can claim that some given dissimilarities can be embedded into a Euclidean space, given that some such transformation can be picked in an optimal way.

Allowing for a Multiplier on the Dissimilarities

Consider the dissimilarity matrix in Table 19.2a. This table is typical insofar as often only the δ_{ij}s for $i < j$ are collected. This immediately makes it impossible to test whether these values satisfy two of the properties of distances: $\delta_{ij} = \delta_{ji}$ and $\delta_{ii} = 0$, for all i, j. With no data to the contrary, we assume that these conditions are satisfied and complete the matrix as usual (Table 19.2b).

The resulting values violate the nonnegativity condition for distances. However, ratio MDS does not claim that the dissimilarities are distances but only that $k \cdot \delta_{ij} = d_{ij}, k \neq 0$. Hence, one can ask whether there exists a multiplier k such that the $k \cdot \delta_{ij}$ values satisfy all three distance axioms. For Table 19.2a, the answer is easily found: there is no such constant k for these data, because a negative k would make the positive values negative, and a positive one would not reverse the sign of the negative values. Hence, the hypothesis that the values in Table 19.2a are distances except for a

multiplicative constant k is wrong. Because they are not distances, they are not, a fortiori, Euclidean distances.

Generally, we note that the relation $k \cdot \delta_{ij} = d_{ij}$ (for some appropriately chosen k) is a hypothesis that may prove to be empirically wrong. Such hypotheses are called (empirically) *falsifiable*.

Allowing for an Interval Transformation on the Dissimilarities

More important than ratio MDS is interval MDS. Interval MDS also allows for an additive constant and, hence, claims that $k \cdot \delta_{ij} + c = d_{ij}$, for some $k \neq 0$ and c. Under this condition, we can transform all of the values in Table 19.2a into positive numbers. We simply add a number $c > 1.8$ to each δ_{ij} ($c = 1.9$, say), which transforms, for example, $\delta_{35} = -1.8$ into the new value $\delta^*_{35} = \delta_{35} + 1.9 = 0.1$.

This then leaves only the triangle inequality as a distance criterion. We find that it is violated for the δ^*_{ij}-values, because $\delta^*_{45} + \delta^*_{52} < \delta^*_{42}$. However, this inequality can be reversed by adding a larger constant c to all δ_{ij}s, because c appears twice on the left-hand side $\delta^*_{45} + \delta^*_{52} = \delta_{45} + c + \delta_{52} + c$ and only once in $\delta^*_{24} = \delta_{24} + c$. To find the smallest possible c that gives all triangle inequalities the desired sense, we check through all inequalities and find that $\delta_{45} + \delta_{52} = -1.6 \geq 3.2 = \delta_{42}$ is most violated; adding c to the dissimilarities, we should obtain $-1.6 + 2c > 3.2 + c$ or, at least, $-1.6+2c = 3.2+c$; hence, the minimal c is $c = 4.8$. If we turn this inequality around in the desired way by adding some $c \geq 4.8$ to all dissimilarities, then all of the other inequalities will also have the proper sense, because in each case c is added twice to the side that should be greater and only once to the other side. Taking $c = 4.8$ and setting all $\delta_{ii} = 0$, we arrive at Table 19.2c, which satisfies all distance axioms. We can conclude that the proposition that given dissimilarities are distances apart from an appropriate interval transformation is always true (tautological) if δ_{ij}s are given for only $i < j$.

Adding a positive additive constant will, in any case, transform any set of dissimilarities δ_{ij}, $i < j$, into distances, provided the constant is large enough. Yet, in the extreme case where $c \to \infty$, the distances thus generated approximate trivial distances.

If, on the other hand, a complete data matrix is given, it cannot be guaranteed that such constants exist. In fact, if just the δ_{ii}s are given, then the constants k and c must be chosen such that $k \cdot \delta_{ii} + c = d_{ii} = 0$. This restricts them so much that it is impossible to transform the dissimilarities into distances if $n \geq 3$.

Interval Transformed Dissimilarities and Euclidean Distances

We now go on and ask whether it is always possible to transform dissimilarities δ_{ij}, $i < j$, not only into distances, but into Euclidean distances by picking appropriate additive and multiplicative constants. The answer is

yes. Assume that some constant has already been added to the dissimilarities to make them all positive and that $\delta_{ii} = 0$, for all i, by definition. The factor k is irrelevant in the following and is set to $k = 1$. Substituting $\delta_{ij} + c$ for d_{ij} in (18.17) should yield a matrix of b_{ij}s that is positive semidefinite if an appropriate c is chosen. Setting $\delta_{ij} + c$ for d_{ij} (for $i \neq j$) and $d_{ii} = 0$ (for all i) in (18.17), or, more compactly, $d_{ij} = \delta_{ij} + (1 - \theta_{ij})c$, where $\theta_{ij} = 1$ (for $i = j$) and $\theta_{ij} = 0$ (for $i \neq j$), we obtain

$$\begin{aligned} b^*_{ij} &= \left[\tfrac{1}{2}(\delta^2_{i.} + \delta^2_{.j} - \delta^2_{..} - \delta^2_{ij})\right] \\ &\quad +2c\left[\tfrac{1}{2}(\delta_{i.} + \delta_{.j} - \delta_{..} - \delta_{ij})\right] + \frac{c^2}{2}\left[\theta_{ij} - \frac{1}{n}\right], \end{aligned} \tag{19.1}$$

where the point subscripts mean that the δs are averaged over the respective indices.

If $c = 0$, then (19.1) is equal to (18.17). Otherwise, there are two additional terms. If we store the bracketed terms in (19.1) in the ij cells of the matrices $\mathbf{B}$, $\mathbf{B}_r$, and $\mathbf{J}$, respectively, then (19.1) reads in matrix notation

$$\mathbf{B}^* = \mathbf{B} + 2c\mathbf{B}_r + \frac{c^2}{2}\mathbf{J}. \tag{19.2}$$

Note that $\mathbf{B}$ is the usual scalar-product matrix associated with the δ_{ij}s, and $\mathbf{B}_r$ is the scalar-product matrix associated with the square roots of the dissimilarities. $\mathbf{J}$, finally, is the centering matrix used in (12.2). Our task is to choose c such that $\mathbf{B}^*$ is positive semidefinite. There are several equivalent ways to state this condition. So far, we have seen two closely related tests: $\mathbf{B}^*$ has nonnegative eigenvalues; $\mathbf{B}^*$ can be factored into $\mathbf{X}\mathbf{X}'$, with real $\mathbf{X}$. A third way to state positive semidefiniteness is that $\mathbf{x}'\mathbf{B}^*\mathbf{x} \geq 0$, for all $\mathbf{x}$. That is, the number resulting from premultiplying $\mathbf{B}^*$ by any (real) vector $\mathbf{x}'$ and then postmultiplying $\mathbf{x}'\mathbf{B}^*$ by $\mathbf{x}$ must be nonnegative (see Chapter 7).

The condition $\mathbf{x}'\mathbf{B}^*\mathbf{x} \geq 0$ is trivially true if $\mathbf{x}$ is the zero-vector: then we have $\mathbf{x}'\mathbf{B}^*\mathbf{x} = 0$. If $\mathbf{x}$ is any other vector, this product should also be nonnegative. This condition is generally not as convenient as the eigenvalue test, but sometimes it leads to insights. The condition requires that

$$\begin{aligned} \mathbf{x}'\mathbf{B}^*\mathbf{x} &= \mathbf{x}'\left[\mathbf{B} + 2c\mathbf{B}_r + \frac{c^2}{2}\mathbf{J}\right]\mathbf{x} \\ &= \mathbf{x}'\mathbf{B}\mathbf{x} + 2c\mathbf{x}'\mathbf{B}_r\mathbf{x} + \frac{c^2}{2}\mathbf{x}'\mathbf{J}\mathbf{x} \\ &= k_1 + c \cdot k_2 + c^2 \cdot k_3 \geq 0. \end{aligned} \tag{19.3}$$

We find that $k_3 > 0$, because $\mathbf{x}'\mathbf{J}\mathbf{x}$ is positive for any $\mathbf{x} \neq \mathbf{0}$. ($\mathbf{x}'\mathbf{J}\mathbf{x}$ simply says $\sum_i (x_i - \bar{x})^2$ in summation notation.) The term k_3 is multiplied by c^2, but k_2 is multiplied by c only, and k_1 does not change as a function of c at all. Thus, if c is chosen ever larger, then $c^2 \cdot k_3$ will eventually dominate

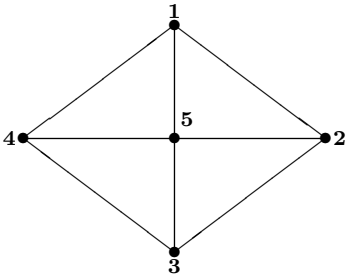

FIGURE 19.2. Five-point configuration, with distances among points as in Table 19.2c.

the sum of the other two terms and make $\mathbf{x}'\mathbf{B}^*\mathbf{x}$ positive semidefinite. It is therefore always possible to find an additive constant c that turns dissimilarities δ_{ij} $(i < j)$ into Euclidean distances.

19.3 Maximal Dimensionality for Perfect Interval MDS

We now know that dissimilarities $\delta_{ij}, i < j$, can always be mapped into Euclidean distances by an interval transformation and by setting $\delta_{ij} = \delta_{ji}$ and $\delta_{ii} = 0$, for all i, j. With respect to the additive constant c, any sufficiently large value will do. There are reasons, however, to choose the smallest possible value for c. For the values in Table 19.2a, we saw that they can be transformed into distances by adding $c_1 = 4.8$. This value turns the triangle inequality that was most violated into an equality. The resulting distances in Table 19.2c are Euclidean distances, because, by applying straight-ruler measurements, we obtain the configuration in Figure 19.2. Adding some $c_2 > c_1 = 4.8$ also leads to values that satisfy the triangle inequalities, but wherever we had a triangle equality for c_1 we will have a triangle inequality for c_2. Geometrically, adding some segment of length $c_2 - c_1$ to each line segment in Figure 19.2 will force point 5 out of the plane of the paper, so that our 5-point configuration will form a pyramid, and a space of three dimensions will be required to represent the data.

Because this makes the representation unnecessarily inaccessible for interpretation, it should be avoided. Of course, there is nothing in the data that would allow us to decide whether the pyramid or the square-with-midpoint configuration from Figure 19.2 is the true configuration, but, in the absence of any further knowledge or hypotheses, there is no reason not to assume that point 5 lies in the middle of the shortest path from 1 to 3.

We show how many dimensions are needed at most for a geometric embedding of an $n \times n$ matrix of Euclidean distances. In equation (12.2), $\mathbf{D}^{(2)}$ is double-centered by $\mathbf{J}$. This makes the rows/columns of $\mathbf{B}$ linearly depen-

TABLE 19.3. Matrix for finding the minimal additive constant c for data in Table 19.1 using formula (19.4); $c = 1.291$, the largest real eigenvalue of this matrix.

$$\left[\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 3.16 & -5.48 & 2.24 & 0.08 \\ 0 & 0 & 0 & 0 & -5.48 & 5.63 & -1.47 & 1.31 \\ 0 & 0 & 0 & 0 & 2.24 & -1.47 & 2.55 & -3.32 \\ 0 & 0 & 0 & 0 & 0.08 & 1.31 & -3.32 & 1.93 \\ \hline -1 & 0 & 0 & 0 & -2.55 & 2.95 & -0.98 & 0.59 \\ 0 & -1 & 0 & 0 & 2.95 & -4.12 & 1.37 & -0.20 \\ 0 & 0 & -1 & 0 & -0.98 & 1.37 & -2.55 & 2.16 \\ 0 & 0 & 0 & -1 & 0.59 & -0.20 & 2.16 & -2.55 \end{array}\right]$$

dent so that $\text{rank}(\mathbf{B}) < n$: the centering matrix $\mathbf{J}$ generates deviation scores in the matrix it operates on, and, thus, the rows or columns, respectively, of the matrix product sum to the null vector $\mathbf{0}$. Hence, $\text{rank}(\mathbf{B}) \leq n-1$, so that the maximum dimensionality of a Euclidean distance matrix is $n-1$. But, as we saw above in Figure 19.2, there may be a c that reduces the dimensionality further. Cailliez (1983) presents a solution for c which guarantees distances that can be represented in at most $n-2$ dimensions. The minimal additive constant c is given by

$$c = \text{largest (real) eigenvalue of} \quad = \quad \left[\begin{array}{cc} \mathbf{0} & 2\mathbf{B} \\ -\mathbf{I} & -4\mathbf{B}_r \end{array}\right]. \qquad (19.4)$$

The matrix in (19.4) is set up by collecting the matrices $2\mathbf{B}$, $4\mathbf{B}_r$, the null matrix $\mathbf{0}$, and the identity matrix $\mathbf{I}$ into one supermatrix. All four matrices have the order $n \times n$; hence, the supermatrix has the order $2n \times 2n$. For the values in Table 19.1, we find by formula (19.4) that $c \approx 1.29$. Adding 1.29 to all numbers in Table 19.1 leads (almost precisely) to a positive semidefinite $\mathbf{B}^*$ with two zero eigenvalues or $\text{rank}(\mathbf{B}^*) = n - 2 = 2$.

If we deal with an ordinal MDS problem, we are not restricted to interval transformations for mapping dissimilarities into Euclidean distances. However, it seems that this does not allow one to reduce the maximal dimensionality of the MDS space below $n-2$. Lingoes (1971), in an earlier paper, describes a simple monotonic transformation on the dissimilarities that guarantees Euclidean distances but does not reduce the dimensionality below $n-2$.

19.4 Mapping Fallible Dissimilarities into Euclidean Distances

In the preceding sections, we ignored the issue of measurement error. But now that we understand how error-free dissimilarities are related to dis-

tances and Euclidean distances under various transformations, some statistical considerations should be made. For fallible data, the transformation problem becomes $k \cdot p_{ij} + c = d_{ij} + e_{ij}$, where d_{ij} is the true distance and e_{ij} is an error component. The task, then, is to find an additive constant c such that the transformed dissimilarities are distances except for a random component. In other words, the shifted data values may violate the critical triangle inequality condition only to such an extent that the violations can be attributed to error. This requires an error theory and results in a much more complicated problem than those considered above. We may require, in addition, that the d_{ij}s be Euclidean distances and that their representation space be as small as possible. This represents a difficult problem, which is subject to different interpretations. We consider the formulation of Messick and Abelson (1956), which, in combination with the double-centering conversion in formula (12.3), is known as *classical MDS*.

The Minimum Statistical Additive Constant

For error-free Euclidean distances, the eigenvalues of the associated scalar-product matrix $\mathbf{B}$ are all positive or zero. The number of positive eigenvalues is equal to the rank of $\mathbf{B}$. Thus, an additive constant c should be chosen such that (a) $\mathbf{B}$ becomes positive semidefinite and (b) the number of zero eigenvalues is maximal.

For error-affected Euclidean distances, this c would be too large. Because of error, the distance estimates cannot be expected to be Euclidean distances so that $\mathbf{B}$ has, in general, some negative eigenvalues. But the distribution of the eigenvalues should have a peculiar form. If the error component is small, there should be some large eigenvalues and some small ones. The large eigenvalues represent the true structure, and the small ones are due to the random over- and under-estimation of the distances. Moreover, "... with fallible data ... the small roots will probably not equal zero but will vary positively and negatively around zero" (Messick & Abelson, 1956, p. 7). If this assumption is made, the sum of the small eigenvalues should be equal to zero, and c should be chosen accordingly.

We start with equation (12.2) and see what can be derived from this assumption about the eigenvalue distribution. Messick and Abelson (1956) use a theorem from matrix algebra which says that the trace of a symmetric matrix $\mathbf{B}$ is equal to the sum of its eigenvalues. That is, if $\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$ is the eigendecomposition of $\mathbf{B}$, then

$$\text{tr } \mathbf{B} = \text{tr } \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}' = \text{tr } \mathbf{\Lambda}\mathbf{Q}'\mathbf{Q} = \text{tr } \mathbf{\Lambda},$$

which uses $\mathbf{Q}'\mathbf{Q} = \mathbf{I}$ and the invariance of the trace function under cyclic permutation (property 3 of Table 7.4). Assume that the eigendecomposition of $\mathbf{B}^*$—which, of course, cannot be computed before c is defined—yields the eigenvalues $\lambda_1, \ldots, \lambda_n$ and the corresponding eigenvectors $\mathbf{q}_1, \ldots, \mathbf{q}_n$.

Thus, $\mathbf{B}^*\mathbf{q}_i = \lambda_i\mathbf{q}_i$ or $\mathbf{q}_i'\mathbf{B}^*\mathbf{q}_i = \mathbf{q}_i'\lambda_i\mathbf{q}_i = \lambda_i\mathbf{q}_i'\mathbf{q}_i = \lambda_i$, because $\mathbf{q}_i'\mathbf{q}_i = 1$, by convention. Now, let the first r eigenvalues be large and the remaining $n - r$ small, as discussed above. Then, r is the dimensionality of the true distances and their scalar products. The sum of the first r eigenvalues is $\sum_{i=1}^{r}\lambda_i = \sum_{i=1}^{r}\mathbf{q}_i'\mathbf{B}^*\mathbf{q}_i$. Hence, by the trace-eigenvalue theorem (see also Section 7.4), we find

$$\sum_{i=1}^{n} b_{ii}^* = \sum_{i=1}^{r}\lambda_i, \tag{19.5}$$

or

$$\text{tr } \mathbf{B}^* = \sum_{i=1}^{r}\mathbf{q}_i'\mathbf{B}^*\mathbf{q}_i. \tag{19.6}$$

Substituting $\mathbf{B} + 2c\mathbf{B}_r + (c^2/2)\mathbf{J}$ for $\mathbf{B}^*$ leads to

$$\text{tr } \left[\mathbf{B} + 2c\mathbf{B}_r + \frac{c^2}{2}\mathbf{J}\right] = \sum_{i=1}^{r}\mathbf{q}_i'\left[\mathbf{B} + 2c\mathbf{B}_r + \frac{c^2}{2}\mathbf{J}\right]\mathbf{q}_i, \tag{19.7}$$

a quadratic equation with the unknown c. The derivation hinges on (19.5): the sum of the first r eigenvalues of $\mathbf{B}^*$ is equal to the trace of $\mathbf{B}^*$ only if the sum of the remaining $n - r$ eigenvalues is equal to zero. This means that the $n - r$ smallest eigenvalues are either all equal to zero or they are distributed symmetrically about zero, as assumed.

Equation (19.7) involves two unknowns, r and c. However, even if we assume for a moment that r has been estimated in some way, we note that it still is not possible to solve the equation for c, because the eigenvectors $\mathbf{q}_i$ are computed from $\mathbf{B}^*$ and thus also depend on c. Solving the problem may therefore be attempted in the usual iterative fashion. First, choose some value for $c^{[0]}$, compute the eigenvalues for $\mathbf{B}^*$, and solve (19.7) for a new c, $c^{[1]}$. This $c^{[1]}$ leads to a new $\mathbf{B}^*$, new eigenvalues, and a new c, $c^{[2]}$, and so on. We show that it is better to choose $c^{[0]}$ too large than too small. A good choice for $c^{[0]}$ would be the additive constant that strictly satisfies all triangle inequalities.

An Illustration for Finding the Statistical Additive Constant

It is peculiar that Messick and Abelson (1956) illustrate their method by an example in which there is no error at all in the distances, that is, a case where we do not really have to estimate the additive constant c but can simply compute it. We nevertheless present this example here because it is transparent and instructive. We start by defining the configuration in Figure 19.3, which yields the true Euclidean distances. As before, only the values in one-half of the distance matrix are considered. Assume that subtracting 1 from these distances generates the dissimilarities that we observe; for example, $d_{AB} = 1$ and hence $\delta_{AB} = d_{AB} - 1 = 1 - 1 = 0$.

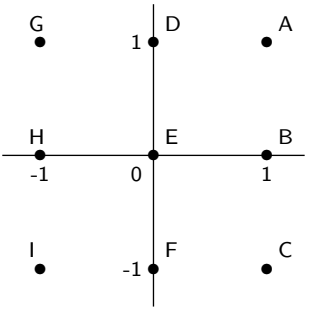

FIGURE 19.3. Configuration used in Messick and Abelson (1956) study.

Because $\delta_{AC} = 1$ and $\delta_{CB} = 0$, the triangle inequality $\delta_{AC} \leq \delta_{AB} + \delta_{BC}$ is violated for the dissimilarities.

To find the true additive constant c in the sense of Messick and Abelson (1956) (which here is $c = 1$ because there is no error in the dissimilarities) a starting value $c^{[0]}$ has to be chosen so that $\mathbf{B}^*$ is defined and its eigenvectors can be computed. Table 19.4 shows the effect of different $c^{[0]}$-values on the eigenvalues and eigenvectors of $\mathbf{B}^*$. All values equal to or greater than 1 transform the dissimilarities into Euclidean distances. For $c^{[0]} = 1$, the true additive constant, only two nonzero eigenvalues result. (One eigenvalue is equal to 0 in all cases due to the centering of $\mathbf{B}^*$.) For $c^{[0]} < 1$, negative eigenvalues arise, because the triangle inequalities remain violated under this condition. Moreover, for $c^{[0]} = 0$, the first two eigenvectors define a configuration very similar to the one in Figure 19.3, but this is not the case for $c^{[0]} = -1$ and $= -2$. Messick and Abelson (1956) claim that, in these latter cases, it is the eighth and ninth eigenvectors whose coordinates define a configuration similar to the one in Figure 19.3. However, such similarities are more apparent than real, because negative eigenvalues correspond to negative distances, and it is quite unclear what this means geometrically. What is definite, in contrast, is that choosing a "small" value for $c^{[0]}$ may lead to problems, because it may result in using the "wrong" r eigenvectors in (19.7). We also note that, for larger initial c-values, two eigenvalues are definitely dominant, which enables us to make a decision on the true dimensionality r.

Assume now that $c^{[0]} = 4$ was chosen. This defines $\mathbf{B}^*$ in (19.6), which can then be factored. Studying the resulting eigenvalue distribution suggests setting $r = 2$. This defines (19.7) and yields as the solutions for its unknown $c_1 = 0.997$ and $c_2 = -0.55$. The value -0.55 is evidently not the desired additive constant, because it does not eliminate violations of the triangle inequalities. Hence, 0.997 must be the solution. We know that the true $c = 1$, so $c_1 = 0.997$ is quite close. The Messick–Abelson procedure has, thus, after just one iteration, almost recovered the true value. But why is c_1 not exactly equal to 1? The reason is that $c^{[0]} = 4$ was too large a value.

TABLE 19.4. First two eigenvectors (fitted to correspond to configuration in Fig. 19.3) and all eigenvalues for different choices of $c^{[0]}$; eigenvalues with star correspond to shown eigenvectors; after Messick and Abelson (1956).

	$c^{[0]}=4$		3		2		1		0		−1		−2	
	$\mathbf{q}_1$	$\mathbf{q}_2$	$\mathbf{q}_1$	$\mathbf{q}_2$	$\mathbf{q}_1$	$\mathbf{q}_2$	$\mathbf{q}_1$	$\mathbf{q}_2$	$\mathbf{q}_1$	$\mathbf{q}_2$	$\mathbf{q}_1$	$\mathbf{q}_2$	$\mathbf{q}_1$	$\mathbf{q}_2$
A	.97	.97	.98	.98	.99	.99	1.00	1.00	1.05	1.05	.72	.72	.90	.90
B	1.05	.00	1.04	.00	1.03	.00	1.00	.00	.88	.00	1.31	.00	1.15	.00
C	.97	−.97	.98	−.98	.99	−.99	1.00	−1.00	1.05	−1.05	.72	−.72	.90	−.90
D	.00	1.05	.00	1.04	.00	1.03	.00	1.00	.00	.88	.00	1.31	.00	1.15
E	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
F	.00	−1.05	.00	−1.04	.00	−1.03	.00	−1.00	.00	−.88	.00	−1.31	.00	−1.15
G	−.97	.97	−.98	.98	−.99	.99	−1.00	1.00	−1.05	1.05	−.72	.72	−.90	.90
H	−1.05	.00	−1.00	.00	−1.03	.00	−1.00	.00	−.88	.00	−1.31	.00	−1.15	.00
I	−.97	−.97	−.98	−.98	−.99	−.99	−1.00	−1.00	−1.05	−1.05	−.72	−.72	−.90	−.90
λ_1	23.12*		16.41*		1.70*		6.00*		2.31*		1.03		3.05	
λ_2	23.12*		16.41*		1.70*		6.00*		2.31*		1.03		3.05	
λ_3	8.02		4.34		1.67		.00		.02		.86		2.69	
λ_4	7.32		3.88		1.44		.00		.00		.11		2.01	
λ_5	6.93		3.66		1.33		.00		−.14		.00		1.68	
λ_6	6.36		3.24		1.12		.00		−.14		− .34		.99	
λ_7	6.36		3.24		1.12		.00		.33		− .34		.00	
λ_8	5.95		2.97		.98		.00		−.33		−.52*		−2.17*	
λ_9	.00		.00		.00		.00		−.44		−.52*		−2.17*	

On the other hand, we see from Table 19.3 that the first two coordinate vectors (which are the eigenvectors rotated to match the true coordinate vectors of Figure 19.3 as closely as possible) are very similar across different values for $c \geq 1$. Thus, it hardly matters which eigenvectors are used in (19.7). For this reason, c_1 is found to be so close to the true value after just one iteration. If, on the other hand, too small a value had been chosen for $c^{[0]}$, negative eigenvalues would have resulted for $\mathbf{B}^*$. In this case, one should start all over again using a larger constant.

Geometric Effects of Nonminimal Additive Constants

Table 19.4 shows that choosing any value other than the true additive constant has a distorting effect on the recovered configuration.[2] The true underlying configuration in Figure 19.3 is a pattern of squares in which the points lie on a network of straight lines. If we plot the point coordinates for $c = 4$ in Table 19.4, we find that the resulting configuration is very similar to Figure 19.3, but the grid is bent convexly outwards from the origin. For example, point B is shifted away from the origin on the Y-axis, whereas A and C stay put. The analogous situation is true for D, F, and H. Moreover, in the 3D MDS space, the plane that best represents Figure 19.3 is warped to form a peculiar saddle shape: A and I are pulled upwards, but G and

[2]Similar distorting effects can be observed when a metric is chosen in MDS that does not correspond to the metric used to generate the distance estimates in the true underlying space. See Section 17.3.

C are pushed downwards, with all other points in a horizontal plane. In contrast, if $c = 0$, the points on the coordinate axes of the 2D space are, relative to the other points, shifted towards the origin, resulting in a convex distortion of the grid. Hence, choosing an inappropriate additive constant results not merely in higher dimensionality but in a *systematic distortion* of the configuration.

Once the data are transformed into distances, statistically or strictly speaking, any further additive constant will change the distance function and thus affect the geometric representation (in a metric context). This is important because dissimilarity data may already be distances, without any transformation, and so adding a constant to them has direct effects on their geometry. In practice, one finds, for example, that ratings of dissimilarity typically require an additive constant that is negative. Such data satisfy the properties of distances so that adding a constant merely serves the purpose of transforming them into Euclidean distances of low dimensionality or into distances with particular segmental additivity properties (see Chapter 17). In that case, an alternative and possibly more fruitful way to proceed would be to consider alternative geometries in which the given distances can be embedded as they are.

19.5 Fitting Dissimilarities into a Euclidean Space

We have seen that the additive constant problem for interval-scaled dissimilarities $\delta_{ij}, i < j$, has a simple solution if it is formulated in an algebraic or error-free way. A statistical model, in which the unknown additive constant is not computed but estimated, is more demanding. The Messick–Abelson solution is complicated, however, and its underlying model is not entirely clear. It suggests, perhaps, that we should not insist on an additive constant strictly satisfying the requirement that the transformed dissimilarities be Euclidean distances. Yet, it seems that in most applications we could drop the parameter r from those that have to be estimated and simply set it to some value that appears theoretically appropriate. With a fixed r, and with the requirement that the distances should be approximately mapped into Euclidean distances, we end up with a familiar problem: interval MDS.

In this context, the transformation question gets a positive answer if the resulting value for the loss criterion is sufficiently small, so that the required conditions are more or less satisfied. What should be considered sufficiently small depends on the context. Among the earliest proposals for treating the additive constant problem in this way are those of Cooper (1972) and Roskam (1972). These authors use the algebraic solution for c as a starting value; that is, $c^{[0]} = \max[\delta_{ij} - (\delta_{ik} + \delta_{kj})]$, over all i, j, k. The resulting $\mathbf{B}^*$ is decomposed into $\mathbf{X}\mathbf{X}'$, and the first r columns of $\mathbf{X}$ are used as the starting configuration. With these starting parameters, a flip-flop procedure

for minimizing $L = \sum[d_{ij} - (k \cdot \delta_{ij} + c)]^2 / \sum d_{ij}^2$ is entered. As we have seen, however, this procedure may not produce the best possible solution for c. Nevertheless, the method works in practice, and we can always check the optimality of the solution by trying other starting configurations. In any case, it is important to distinguish the *optimization* approach conceptually from the *algebraic* and the *statistical* viewpoints taken above. In the first case, c is optimized, in the second it is computed, and in the third it is estimated.

The so-called rational starting configurations for ordinal MDS are constructed by using the optimization method of interval MDS. Often, ranking-numbers are first substituted for the given dissimilarities: if the data are dissimilarities, the smallest δ_{ij} is set equal to 1, the second-smallest to $2, \ldots,$ and the largest to $\binom{n}{2}$; for similarities, the largest δ_{ij} is set equal to 1, the second largest to 2, and so on. We can also use the δ_{ij} values as they are. In either case, there are several options for proceeding. One possibility would be to add the algebraic additive constant, find the associated $\mathbf{B}$, decompose this into $\mathbf{XX}'$, and use the first r dimensions as an initial configuration. Another possibility would be to use the data or ranking-number matrix without adding any constant c and check whether the resulting $\mathbf{X}$ has some small imaginary dimensions. If so, we keep the first r and proceed with ordinal optimization. If not, a constant c can be added to the dissimilarities repeatedly until this situation results: if there are no negative eigenvalues for $\mathbf{B}^*$, then we choose $c < 0$; otherwise, we set $c > 0$.

19.6 Exercises

Exercise 19.1 Consider the similarities in Table 4.1 on p. 65. For this exercise you need software that can do matrix algebra.

(a) Transform the similarities into dissimilarities.

(b) Then, find the smallest possible additive constant that turns these values into Euclidean distances.

(c) Use classical scaling on the transformed dissimilarities. Compare the solution to the one obtained in Exercise 12.1 and in Figure 4.1. What do you conclude?

(d) Instead of the distances being Euclidean, find the smallest possible additive constant that turns dissimilarities into distances (not necessarily Euclidean). Is this constant the same as the one for Euclidean distance?

(d) Using a large additive constant, the dissimilarities are turned into distances. Are they also turned into Euclidean distances? Try a few cases numerically.

Exercise 19.2 Consider the data matrix below (Torgerson, 1958). It shows "absolute distances" based on 84 judgments of closeness for all possible triads of nine colors. The colors were all of the same red hue (=5R in Munsell notation) but differed from each other in brightness (value) and saturation (chroma). The conversion of the triadic closeness judgments into the values shown below involved a series of conversions aimed at adding the best additive constant.

			Number of Stimulus								
No	Value	Chroma	1	2	3	4	5	6	7	8	9
1	7	4	–	1.23	3.48	2.98	3.83	5.16	4.69	5.62	5.83
2	6	6	1.23	–	2.59	1.67	2.70	4.40	3.13	4.65	4.38
3	6	10	3.48	2.59	–	4.30	2.28	2.93	4.67	4.30	6.22
4	5	4	2.98	1.67	4.30	–	2.82	4.85	1.85	3.88	2.88
5	5	8	3.83	2.70	2.28	2.82	–	2.58	2.37	1.95	4.09
6	5	12	5.16	4.40	2.93	4.85	2.58	–	4.17	2.93	5.48
7	4	6	4.69	3.13	4.67	1.85	2.37	4.17	–	2.42	2.30
8	4	10	5.62	4.65	4.30	3.88	1.95	2.93	2.42	–	4.02
9	3	4	5.83	4.38	6.22	2.88	4.09	5.48	2.30	4.02	–

(a) Check whether these data violate any distance axioms.

(b) Determine the minimum additive constant that turns these values into distances, if possible. (If this constant exists, it may be equal to zero. When?)

(c) Same as (b), but now replace "distances" by "Euclidean distances".

(d) Use classical scaling to check to which extent the data mirror their physical design. (The design is given by the Munsell values for value and chroma; hue is constant.)

(e) Enforce an MDS structure that mirrors the physical design except for possible monotonic transformations along the coordinate axes value and chroma.