18

Scalar Products and Euclidean Distances

Scalar products are functions that are closely related to Euclidean distances. They are often used as an index for the similarity of a pair of vectors. A particularly well-known variant is the product-moment correlation for (deviation) scores. Scalar products have convenient mathematical properties and, thus, it seems natural to ask whether they can serve not only as indices but as models for judgments of similarity. Although there is no direct way to collect scalar product judgments, it seems possible to derive scalar products from "containment" questions such as "How much of A is contained in B?" Because distance judgments can be collected directly, but scalar products are easier to handle numerically, it is also interesting to study whether distances can be converted into scalar products.

18.1 The Scalar Product Function

The earliest papers on MDS paid more attention to scalar products than to distances. The reason was simply computational. Given a matrix of scalar products, it is easy to find a representing MDS configuration for them. In fact, this MDS problem can be solved analytically (see Chapter 7).

In the usual geometry, the scalar product b_{ij} of the points i and j is defined as the sum of the products of the coordinates of i and j:

$$b_{ij} = \sum_{a=1}^{m} x_{ia} x_{ja}. \tag{18.1}$$

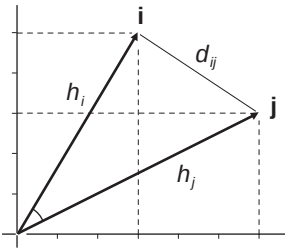

FIGURE 18.1. Illustration of two vectors (arrows) and of the Euclidean distance between their endpoints.

To illustrate, Figure 18.1 shows three distinct points in the $X-Y$-plane: the origin O with coordinates $(0,0)$, the point i with coordinates $(x_{i1}, x_{i2}) = (3,5)$, and the point $j = (6,3)$. The points i and j are depicted as endpoints of vectors ("arrows") emanating from O.

Once a particular point in a plane is chosen as the origin, then all points can be conceived as vectors bound to this origin, and vice versa. So one can alternate between the notions of point and vector whenever it seems useful to do so. Notationally, the origin to which the vectors are bound is not explicitly shown, so one simply writes a bold **j** for the vector from O to j.

For the scalar product of i and j in Figure 18.1, we find $b_{ij} = 3 \cdot 6 + 5 \cdot 3 = 33$. But formula (18.1) can also be used on each vector alone. For example, for **j**, one finds $b_{jj} = 6 \cdot 6 + 3 \cdot 3 = 45$. This corresponds to the length of **j**. For the length of j, one often writes h_j. So, $h_j = \sqrt{b_{jj}} = d_{Oj}$.[1]

Some of the relations between a scalar product, the lengths of its vectors, and the angle subtended by them may be seen by considering the triangle formed by the points O, i, and j in Figure 18.1. The lengths of its sides are easily found by using the Pythagorean theorem, which yields $h_i^2 = 6^2 + 3^2 = 45, h_j^2 = 5^2 + 3^2 = 34$, and $d_{ij}^2 = (6-3)^2 + (5-3)^2 = 13$. The angle α in Figure 18.1 is computed by using the cosine law for a triangle with sides a, b, and c, which says that $a^2 = b^2 + c^2 - 2bc\cos(\alpha)$, where α is the angle between the sides b and c. Thus, $d_{ij}^2 = h_i^2 + h_j^2 - 2h_i h_j \cos(\alpha)$, and solving this equation for $\cos(\alpha)$ yields

$$\cos(\alpha) = \frac{h_i^2 + h_j^2 - d_{ij}^2}{2h_i h_j}, \tag{18.2}$$

[1]The term h_j is the image of a scalar function on the vector argument **j**. The function is the Euclidean norm $\|\mathbf{j}\|$ (see Chapter 7). Norm functions on vectors have general properties that are similar to those of distances between points, but, unlike distances, they cannot be defined on just any set. Rather, they require sets that possess the properties of *vector spaces* so that operations such as + in (7.2) have a particular well-defined meaning (see Chapter 19).

TABLE 18.1. Scalar product matrix for three variables.

Variable	x	y	z
x	25	15	20
y	15	25	12
z	20	12	25

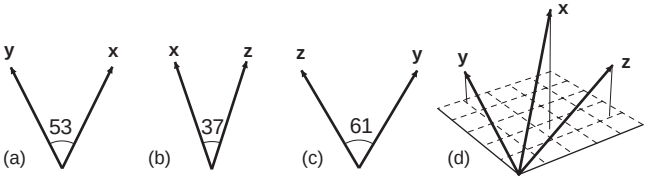

FIGURE 18.2. Vector configurations for scalar products in Table 18.1; panels (a), (b), and (c) show combinations of pairs of vectors; panel (d) results from combining panels (a), (b), and (c).

or, in more detail,

$$\cos(\alpha) = \frac{\sum_a x_{ia}^2 + \sum_a x_{ja}^2 - \sum_a (x_{ia} - x_{ja})^2}{2(\sum_a x_{ia}^2)^{1/2}(\sum_a x_{ja}^2)^{1/2}}, \tag{18.3}$$

which simplifies to

$$\cos(\alpha) = \frac{\sum_a x_{ia}x_{ja}}{(\sum_a x_{ia}^2 \sum_a x_{ja}^2)^{1/2}} = \frac{b_{ij}}{h_i h_j}. \tag{18.4}$$

(This is the formula for the product-moment correlation coefficient for deviation scores, which can therefore be interpreted as an angle function of the data vectors **i** and **j**.) The scalar product b_{ij} is thus

$$b_{ij} = h_i h_j \cos(\alpha). \tag{18.5}$$

One notes that the value of b_{ij} depends on three arguments: the length of the vector **i**; the length of **j**; and the angle α subtended by **i** and **j**. If $\alpha = 0$, then $\cos(\alpha) = 1$, and the scalar product is equivalent to the squared Euclidean distance between the origin O and the endpoint of vector **i**.

If all scalar products are given for a set of vectors, then it is possible to construct the corresponding vector configuration from these values. This was shown algebraically in Section 7.9, but it is also easy to understand geometrically. Consider an example. Assume that Table 18.1 is a matrix whose entries are scalar products.[2] Then, the values in the main diagonal

[2]The matrix does not obviously violate this assumption. It is symmetric and its main diagonal elements are nonnegative. Symmetry must be satisfied by all scalar product

tell us that the vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ all have the same length, $h_x = h_y = h_z = \sqrt{25} = 5$. To construct the vector configuration, we need to know the angles between each pair of vectors. The angle between $\mathbf{x}$ and $\mathbf{y}$, say, is found from $b_{xy} = 15$. By formula (18.5), $b_{xy} = 5 \cdot 5 \cdot \cos(\alpha) = 15$ or $\cos(\alpha) = 3/5$, and this yields $\alpha = 53.13°$. Figure 18.2a shows the resulting configuration of the two vectors $\mathbf{x}$ and $\mathbf{y}$. Proceeding in the same way for the other vector pairs, we arrive at Figures 18.2b and 18.2c. If everything is put together, we find the configuration of all three vectors, which requires a 3D space (Figure 18.2d).

18.2 Collecting Scalar Products Empirically

A scalar product is a more complex measure than a distance. In terms of points in a space, it not only involves the endpoints i and j but also a third point that serves as the origin. Does such a complicated function serve any other purpose than a purely mathematical one? Is it possible to translate all or some of the properties of scalar products into real questions on the similarity of two objects, i and j?

Building Scalar Products from Empirical Judgments

It would be futile, of course, to ask a subject to directly rate the "scalar product" of two stimuli i and j, although we could certainly ask him or her to rate their "distance". A scalar product is a notion that has no intuitive meaning. Ekman (1963), therefore, suggested an indirect approach, asking for two particular judgments, which are then combined to form a scalar product. Consider Figure 18.3. Let vectors $\mathbf{i}$ and $\mathbf{j}$ in panel (b) represent two stimuli such as the colors blue and red. We could ask the subject to assess the ratio c_{ij}/h_j, that is, judge the length of the projection of $\mathbf{i}$ onto $\mathbf{j}$ relative to the length of $\mathbf{j}$. Concretely, this could be operationalized in a question like, "How much of this blue is contained in this red?" to which the subject may answer by giving a percentage judgment (such as "80%"). The question is then inverted to, "How much of this red is contained in this blue?" and the subject's answer is taken as an assessment of the ratio c_{ji}/h_i.

---

matrices, because it follows from (18.1) that $b_{ij} = b_{ji}$. Moreover, all elements in the main diagonal must be nonnegative, because, for $i = j$, formula (18.1) is but a sum of squared numbers. These conditions do not guarantee, however, that the matrix is a scalar product matrix. A matrix $\mathbf{B}$ is a (Euclidean) scalar product matrix only if it can be decomposed into the matrix product $\mathbf{XX'}$, with real $\mathbf{X}$ (see Chapter 19). If this is possible, then each element of $\mathbf{B}$ satisfies formula (18.1).

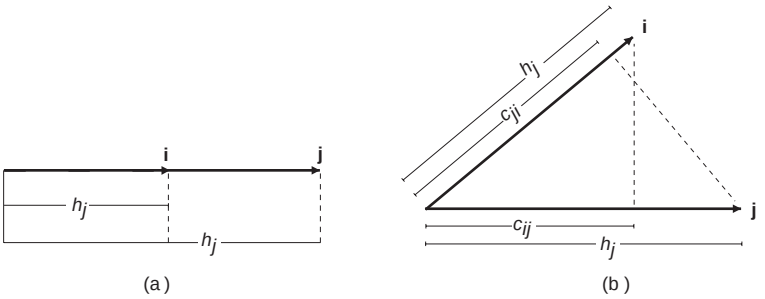

FIGURE 18.3. Vector representations of two stimuli, i and j.

We show what can be done with such data. To simplify the notation, let

$$v_{ij} = c_{ij}/h_j, \tag{18.6}$$
$$v_{ji} = c_{ji}/h_i. \tag{18.7}$$

Note that what one observes are the v-values; the expressions on the right-hand side of the equations are how these data are explained by the vector model. Note also that one can assume that the v-values are nonnegative because a score of zero is the least possible containment. In terms of the model, the c-terms are projections,

$$c_{ij} = h_i \cdot \cos(\alpha), \tag{18.8}$$
$$c_{ji} = h_j \cdot \cos(\alpha). \tag{18.9}$$

Thus,

$$\cos(\alpha) = (v_{ij} \cdot v_{ji})^{1/2}, \tag{18.10}$$
$$h_i/h_j = (v_{ij}/v_{ji})^{1/2}. \tag{18.11}$$

If v_{ij} and v_{ji} can be estimated empirically, we can derive from them (a) the angle α between the vectors **i** and **j** via (18.10), and (b) the ratio of the lengths of these vectors via (18.11). If one of the vectors is fixed arbitrarily (say, by setting $h_i = 1$), then a unit for scaling is given, and the vector configuration can be constructed.

Consider some data (Ekman, 1963). Six monochromatic lights of equal brightness served as stimuli. Their wavelengths were 593, 600, 610, 628, 651, and 674 nm; that is, the lights were in the red-yellow range. All possible pairs of lights were projected onto a screen, and 10 subjects were asked for "contained-in" judgments on each pair. The averages of the observed values are presented in Table 18.2. This data matrix is denoted as $\mathbf{V}$. From $\mathbf{V}$ we can derive a matrix $\mathbf{H}^{(2)}$ that contains the quotients v_{ij}/v_{ji} as its elements (Table 18.3). These values are, by formula (18.11), the quotients of the squared lengths of our six vectors. For example, the second element in the first row is $v_{12}/v_{21} = h_1^2/h_2^2 = .94/.95 = .99$. Summing over all elements

TABLE 18.2. Averaged v-data for colors with wavelengths 593, ... , 674 nm (Ekman, 1963).

nm	593	600	610	628	651	674
593	1.00	.94	.67	.22	.08	.04
600	.95	1.00	.80	.31	.16	.06
610	.63	.75	1.00	.78	.56	.38
628	.21	.37	.78	1.00	.81	.72
651	.14	.23	.61	.85	1.00	.86
674	.07	.13	.40	.80	.90	1.00

TABLE 18.3. $\mathbf{H}^{(2)}$ matrix, based on v-values in Table 18.2; $\mathbf{H}^{(2)} = (v_{ij}/v_{ji})$.

nm	593	600	610	628	651	674	First Est. h_i^2	Sec. Est. h_i^2
593	1.00	.99	1.07	1.05	.56	.62	5.29	4.11
600	1.01	1.00	1.07	.85	.70	.47	5.10	3.93
610	.94	.93	1.00	1.00	.91	.94	5.72	3.87
628	.95	1.18	1.00	1.00	.95	.89	5.97	4.13
651	1.79	1.44	1.09	1.05	1.00	.96	7.33	4.33
674	1.61	2.14	1.06	1.11	1.05	1.00	7.97	4.58

of row i of $\mathbf{H}^{(2)}$ yields, symbolically, $\sum_{j=1}^{n} h_i^2/h_j^2 = h_i^2 \sum_{j=1}^{n}(1/h_j^2)$. Hence, this sum always involves a constant term, $\sum_{j=1}^{n}(1/h_j^2)$. Ekman (1963) suggested simply setting this term equal to 1, thus introducing a scaling norm for the vectors. With $\sum_{j=1}^{n}(1/h_i^2) = 1$, we get $h_1^2 = 5.29$, $h_2^2 = 5.10$, and so on, as shown in Table 18.3.

Further Considerations on Constructing Scalar Products

Selecting a scaling norm for the vectors is a trivial matter in the case of error-free data. The simplest choice would be to arbitrarily select one vector and take its length as the norming length for all vectors. With real data, however, Ekman's suggestion for norming by setting $\sum_i(1/h_i^2) = 1$ seems better, because the vector lengths are derived from all data, not just a subset. This should lead to more robust vector-length estimates for fallible data.

Computational accuracy remains a problem in any case, because our estimates rely very much on divisions and multiplications. Such operations are numerically unstable. Consider, for example, the v-values for 593 and 674. Their quotient forms two entries in the $\mathbf{H}^{(2)}$ matrix. For example, we should find the value for the element in row 593 and column 674 of $\mathbf{H}^{(2)}$ from 0.04/0.07. But 0.04/0.07 = 0.57, and not 0.62, as we find in the table. The discrepancy is a consequence of the fact that Table 18.3 reports only

two decimal places. Other small changes in the v-values also render quite different h^2-values.

The reliability of the v-data could also be taken into account. One may argue that very small v-data should be less reliable because they express that the subject felt that the two objects had essentially nothing in common. A subject should, therefore, be relatively indifferent whether, say, one object is said to contain 4% or 3% of the other. Forming ratios of such small contained-in percentages is, however, very much affected by such small changes in the magnitude of the contained-in data. Ekman (1963) therefore suggested skipping such small v-values in estimating the vector lengths. In Table 18.2, he decided to ignore the values in the upper right-hand and the lower left-hand corners of the matrix, because these corners contain relatively many small v-values. The vector lengths were estimated only from the values in the upper 4×4 left-hand and the lower 3×3 right-hand submatrices of $\mathbf{H}^{(2)}$ in Table 18.3. That means, for example, that h^2_{593} results from adding the first four elements in row 593. For h^2_{651}, the last three elements of row 651 are added, and, because this sum involves one element less than before, the sum is rescaled by the adjustment factor 1.45 to yield $h^2_{651} = 4.33$. The factor 1.45 is computed from row 628, where the first four elements add up to 4.13 and the last three to 2.84, giving the ratio $4.13/2.84 = 1.45$. If one compares the results of this estimation (see column "Second Estimates" in Table 18.3) with those obtained before, one notes (apart from the irrelevant differences in the magnitudes of the h^2_i-values) that their proportions are quite different. Whether these values are indeed better estimates is, of course, impossible to say without any further evidence. We note, however, that the estimation approach is obviously not very robust.

The angles for each vector pair could be found in a similar way using (18.10). This would then yield scalar products by (18.5). However, there is a more direct estimation approach that also provides a test. It follows from (18.8) that $b_{ij} = h^2_j \cdot v_{ij}$. Because we also obtain b_{ji} from $h^2_i \cdot v_{ji}$, nothing guarantees that $\mathbf{B}$ is symmetric so that $b_{ij} = b_{ji}$, for all i, j. This provides a test for assessing to what extent the properties derived from the vector model are consistent with the data. We find that there are indeed some asymmetries, for example, $b(593,600) = (3.93)(0.94) = 3.69$, but $b(600,593) = (4.12)(0.95) = 3.91$ (Table 18.4). However, these asymmetries are quite small, as Figure 18.4 makes clear, and can be assumed to lie within the error range. Thus, finally, a scalar product matrix is obtained by averaging these b-values.

Different Vector Lengths and v-Data

One notes that the $\mathbf{V}$-matrix in Table 18.2 is not symmetric; that is, $v_{ij} \neq v_{ji}$ for most i, j. The asymmetries are, however, minor in their magnitudes. Empirically, one also finds cases where asymmetries are substantial. One

TABLE 18.4. Preliminary (nonsymmetrized) **B** matrix for values in Table 18.2.

nm	593	600	610	628	651	674
593	4.12	3.68	2.59	.92	.35	.20
600	3.89	3.93	3.09	1.28	.71	.27
610	2.58	2.92	3.87	3.21	2.42	1.71
628	.87	1.44	3.01	4.13	3.51	3.27
651	.59	.92	2.37	3.51	4.33	3.89
674	.29	.50	1.54	3.30	3.88	4.55

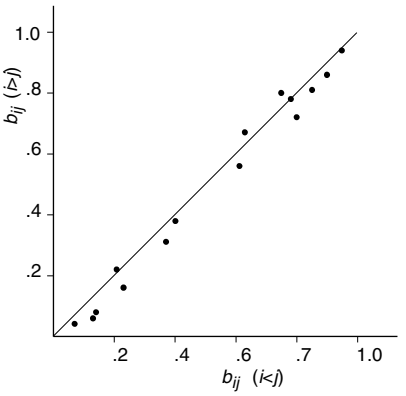

FIGURE 18.4. Scatter plot of scalar products in Table 18.4; points' coordinates on X-axis (Y-axis) are values in upper (lower) half of Table 18.4.

example is a study by Sixtl (1967) on the similarity of different emotional experiences or feelings. He reports that his subjects felt that "wrath" has 83% in common with "aggressiveness", but "aggressiveness" overlaps with "wrath" only to 60%. Within the vector model, such asymmetries of the containment judgments imply different lengths for the representing vectors. This can be seen from Figure 18.3b, where **i** does have more in common with **j** than vice versa. The reason is that **j** is longer and, thus, its perpendicular projection is also longer.

This example also shows that, in the model, symmetric contained-in judgments are not necessary for symmetric scalar products: $b_{ij} = b_{ji}$ implies that $h_j^2 v_{ij} = h_i^2 v_{ji}$. Hence, asymmetries of v_{ij} and v_{ji} judgments can be compensated by the different lengths of **i** and **j**.

Vastly different vector lengths may, on the other hand, lead to a serious problem for the contained-in judgments. Consider Figure 18.3a. Here, $\mathbf{j} = 2 \cdot \mathbf{i}$ and, so, $v_{ji} = 2$. This yields consistent equations: if one sets $h_j = 1$ (units), then $b_{ij} = h_j^2 v_{ij} = (1)^2(1/2) = 0.5$ and $b_{ji} = h_i^2 v_{ji} = (1/2)^2(2) = 0.5$. However, the operationalization, "How much of j is contained in i?" does not work anymore for this case, because it seems impossible that a containment judgment is smaller than 0% or larger than 100%.

But are such cases really impossible? The projection of **j** onto **i** should be longer than **i** itself if **j** contains more of **i** than **i** itself. This case is not quite as paradoxical as it may appear at first sight. A conceivable instance of this situation involves the colors **j** = bright red and **i** = pale reddish, where the latter is but a "pale" instance of the prototypical color. If this situation seems likely, conventional contained-in judgments do not appear to be sufficient to measure scalar products. It would be desirable, for example, to somehow assess the vector lengths independently of any notions of similarity or containment. A set-theoretic approach where objects are equated with feature sets might be a possibility (see, e.g., Restle, 1959).

18.3 Scalar Products and Euclidean Distances: Formal Relations

In an exploratory context, the mapping of v-values into a vector configuration does not have to pass major tests that would allow one to conclude that the model is inappropriate. The only such test is the required rough symmetry of the preliminary **B**-matrix. If this matrix is grossly asymmetric, the model should be dropped as inappropriate.

One could devise further tests though, for example, constraints on the dimensionality of the vector configuration or predictions as to how the vectors should be positioned relative to each other. The more such tests there are, the more can be learned about the data. Ekman, Engen, Künnapas, and Lindman (1964) suggested collecting further measures besides the

contained-in judgments (v-data), and then checking whether everything fits together. Given two stimuli i and j as in Figure 18.1, we would expect that the dissimilarity judgments on i and j could be mapped into the distance d_{ij}, and the v-data would mirror the discussed projection-to-length ratios.

Converting Scalar Products into Euclidean Distances, and Vice Versa

Scalar products and Euclidean distances are closely related; for example, for vectors of constant length, they stand in an inverse monotonic relation to each other, so that if d_{ij} grows, b_{ij} gets smaller, and vice versa. But there is a major difference between scalar products and distances: if the origin of the coordinate system is shifted in space, then the scalar products will also change, whereas the distances remain the same. Expressed in terms of the formulas, we have $b_{ij} = \sum_a x_{ia}x_{ja}$ and $d_{ij}^2 = \sum_a (x_{ia} - x_{ja})^2$ for the old coordinate system. Shifting the coordinate system by the translation vector $(t_1, \ldots, t_m)$, one obtains $b_{ij(t)} = \sum_a (x_{ia} + t_a)(x_{ja} + t_a) \neq b_{ij}$, unless $t_1 = 0, \ldots, t_m = 0$. For distances, on the other hand, one gets $d_{ij(t)}^2 = \sum_a [(x_{ia} + t_a) - (x_{ja} + t_a)]^2 = d_{ij}^2$. However, once some point has been chosen to serve as the origin, we can compute scalar products from distances and vice versa (see also Chapter 12). To see this, consider Figure 18.5. Let point k be the origin. Then, by the cosine theorem,

$$d_{ij}^2 = d_{kj}^2 + d_{ki}^2 - 2d_{kj}d_{ki}\cos(\alpha), \tag{18.12}$$

where α is the angle between the vectors from point k to j and from k to i, respectively. Rearranging (18.12), we find

$$d_{kj}d_{ki}\cos(\alpha) = \tfrac{1}{2}(d_{kj}^2 + d_{ki}^2 - d_{ij}^2), \tag{18.13}$$

which is, by (18.5),

$$b_{ij} = \tfrac{1}{2}(d_{kj}^2 + d_{ki}^2 - d_{ij}^2), \tag{18.14}$$

because d_{kj} and d_{ki} are just the lengths of the vectors $\mathbf{j}$ and $\mathbf{i}$. Thus, we find the scalar product b_{ij} from three distances. Conversely, we find the distance d_{ij} from three scalar products: observing that $d_{kj}^2 = b_{jj}$ and $d_{ki}^2 = b_{ii}$, we have $d_{ij}^2 = b_{ii} + b_{jj} - 2b_{ij}$. Note that the origin k always enters into these conversions.

Typically, one chooses the centroid as the origin, because this point is supposedly more reliable than any point representing a single variable (Torgerson, 1958). This choice should therefore lead to more robust scalar-product estimates. The centroid is the point z with coordinates

$$(z_1, \ldots, z_m) = \left(\frac{1}{n}\sum_{i=1}^{n} x_{i1}, \ldots, \frac{1}{n}\sum_{i=1}^{n} x_{im} \right). \tag{18.15}$$

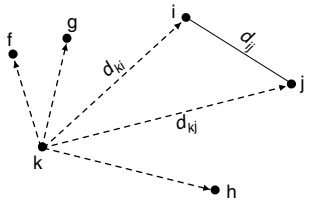

FIGURE 18.5. Defining a vector configuration on the points $f, \ldots, k$ by choosing one point, k, as an origin.

TABLE 18.5. Numerical example for the relation (18.17).

Point	Coordinates		Centered Coordinates		Squared Distances				Scalar Products			
	1	2	1	2	1	2	3	4	1	2	3	4
1	1	2	-1.5	0.75	0	1	8	10	2.81	1.31	-1.69	-2.44
2	2	2	-0.5	0.75	1	0	5	5	1.31	0.81	-1.19	-0.94
3	3	0	0.5	-1.25	8	5	0	2	-1.69	-1.19	1.81	1.06
4	4	1	1.5	-0.25	10	5	2	0	-2.44	-0.94	1.06	2.31
Sum	2.5	1.25	0.0	0.0	19	11	15	17				

With the centroid of all points as the origin, we obtain the scalar product

$$b_{ij} = \sum_a (x_{ia} - z_a)(x_{ja} - z_a), \tag{18.16}$$

because each coordinate is now expressed as a deviation score from the origin z. This expression is transformed into a formula with only distances appearing on the right-hand side, as in (18.14). Such a formula allows one to convert distances—for which empirical estimates are assumed to be given—into scalar products relative to the centroid. Inserting z values into (18.16), one obtains, after some rearrangements of terms,

$$b_{ij} = -\tfrac{1}{2}\left(d_{ij}^2 - \frac{1}{n}\sum_i d_{ij}^2 - \frac{1}{n}\sum_j d_{ij}^2 + \frac{1}{n^2}\sum_i\sum_j d_{ij}^2\right). \tag{18.17}$$

Table 18.5 shows an example for this conversion. Given the squared distances, b_{ij} values are found by first subtracting from each d_{ij}^2 value the mean of row i and column j, then adding to it the mean of all squared distances, and finally multiplying all values by $-\frac{1}{2}$. For example, for points 1 and 2 we get $d_{12}^2 = (1-2)^2 + (2-2)^2 = 1$ from the coordinates. The scalar product relative to the centroid is $b_{12} = -\frac{1}{2}(1 - 19/4 - 11/4 + 62/16) = 1.31$. But this should be the same as computing the scalar product directly by formula (18.5) from the centered coordinates. Indeed, we find $b_{12} = (-1.5)(-0.5) + (0.75)(0.75) = 1.31$.

Table 18.5 also demonstrates that the scalar product between **i** and **j** depends on the origin, whereas the distance of **i** and **j** does not. For example, the scalar product for points 1 and 2 is 5.00 for the raw coordinates but, as we saw, 1.31 for the centered coordinates. On the other hand, the distance $d_{12} = 1.00$ for any origin. Thus, scalar products are in a sense stronger or richer in information than distances, because they depend on the n points of a configuration and, in addition, on an origin. This issue is unrelated to the scale level of the data. Even for absolute proximities, any point of the MDS configuration can be chosen to serve as an origin. Hence, in distance scaling, the origin is, by itself, meaningless, although meaning may be brought in from elsewhere, as, for example, in the radexes in Chapter 5. For scalar-product data, in contrast, the origin necessarily has an empirical meaning. In Figure 18.7, it represents the color gray, and the fact that all points have the same distance from it reflects the equal saturation of the six colors used in the experiment. Consequently, Ekman (1963) interprets the different directions of the color vectors as due to their *qualitative* differences, whereas different vector lengths represent their *quantitative* differences.

18.4 Scalar Products and Euclidean Distances: Empirical Relations

The formal relations between scalar products and distances may be used in empirical research. If one collects both contained-in data (v-data) and also asks the subjects to directly assess the global similarity (s-data) of the objects of interest, it becomes possible to test whether the subjects' proximity judgments can be accounted for by their scalar products. If so, our confidence in the empirical validity of the geometrical models should be increased. The converse, however, is not possible, because one cannot uniquely derive scalar products from given distances due to the arbitrary choice of origin.

A number of researchers have studied whether there exist *empirical* relationships between distance and scalar-product data that allow such two-way conversions. Let us first consider the special case where $h_i = h_j$, for all i, j. Under this equal-length condition, Ekman proposed that the relation $s_{ij} = \cos(\alpha)/\cos(\alpha/2)$ could be shown to hold very well empirically, where $\cos(\alpha) = \sqrt{v_{ij}v_{ji}}$ from equations (18.6)–(18.10). (Both the s- and the v-data were collected on percentage scales, in which 100 meant "identity" for proximity judgments, and "completely contained in" for contained-in judgments.) If such a relation would indeed hold, then we could arrive at a natural origin by converting proximity data into scalar products. This has the advantage that all we need are proximity data, which are much easier to collect. Of course, this should work only if $h_i = h_j$ holds for all i and j, a condition that supposedly is guaranteed by proper instruction

of the subjects. Therefore, in an experiment on the similarity of different emotions, Ekman et al. (1964) asked their subjects "to consider emotions of equal intensity" (p. 532) and "to disregard possible quantitative differences in the intensity of the (emotions) and base their judgments on qualitative characteristics" (p. 533).

It seems hard to evaluate what the results of such experiments mean. Because they involve complicated formal relations and equally complicated instructions, it is impossible to see where things break down. Nevertheless, it is interesting to consider the more general principles from which Ekman derived such relations. He started by studying models for the subjective similarity of stimuli differing on one attribute only. For example, Ekman, Goude, and Waern (1961) report an experiment in which subjects had to assess all possible pairs of different grays (a) with respect to their global similarity on a 10-point scale, and (b) relative to their darkness ratios. The resulting proximity values were divided by 10, and a simple function was found closely describing the relation between ratio and similarity data:

$$s_{ij} = \frac{2h_i}{h_i + h_j}, \quad h_i \leq h_j, \tag{18.18}$$

where h_i and h_j are the values of stimuli i and j on the darkness scale, and s_{ij} is the (rescaled) distance estimate for i and j. In terms of the actual data collection procedure, (18.18) can be written as $s_{ij} = 2/(1 + h_j/h_i)$, with h_j/h_i being the empirical ratio judgment. Because the different gray stimuli on which the relation (18.18) is based do not differ qualitatively, the situation can be best understood by considering Figure 18.3a, where **i** is completely contained in **j**. Also, **i** is, of course, completely contained in itself. Hence, one can interpret the term $2h_i$ in formula (18.18) as an expression for what i and j have in common (e.g., in the sense of their stimulation). The term $h_i + h_j$, on the other hand, expresses what i and j comprise together. What equation (18.18) says, thus, is that the subjective dissimilarity of i and j is given as the ratio[3] of the *communality* and the *totality* of i and j,

$$s_{ij} = \frac{\text{communality of } i \text{ and } j}{\text{totality of } i \text{ and } j} = \frac{K_{ij}}{T_{ij}}. \tag{18.19}$$

We now want to drop the model constraint that **i** and **j** are both collinear (as in Figure 18.3a) and generalize the notions of communality and totality to the higher-dimensional case. One possibility is to set $K_{ij} = c_{ij} + c_{ji}$, because c_{ij} is just that component that **i** shares with **j**, and the converse is

[3]This is similar to feature-set models of stimuli, where communality is equated with the intersection of the object's feature sets, and totality with the union of these sets. The 2 in the numerator of (18.18) could be interpreted as a scaling factor on the similarity judgments.

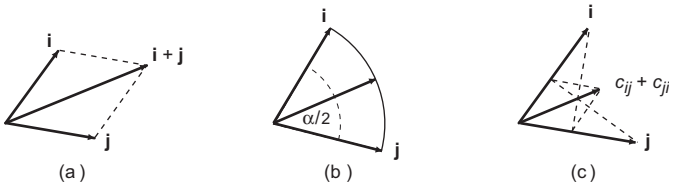

FIGURE 18.6. Illustrations of some notions of communality and totality.

true for c_{ji}. With $T_{ij} = h_i + h_j$ as before, this gives

$$s_{ij} = \frac{c_{ij} + c_{ji}}{h_i + h_j}. \tag{18.20}$$

But $c_{ij} = h_i \cos(\alpha)$ and $c_{ji} = h_j \cos(\alpha)$, so (18.20) is equal to

$$s_{ij} = \cos(\alpha), \tag{18.21}$$

assuming that $c_{ij} \leq h_j$ and $c_{ji} \leq h_i$, that is, that the projections of any vector onto another vector should not be longer than the vector itself (see Section 18.2). This equation means, in terms of the observations, that $s_{ij} = \sqrt{v_{ij} v_{ji}}$.

Unfortunately, this simple hypothesis on the relation between s- and v-values was found to describe empirical correspondences rather poorly. Hence, other proposals were made. Ekman et al. (1964) tested a version of (18.19) in which the totality of i and j was modeled by the vector sum of **i** and **j**, as shown in Figure 18.6a. If $h_i = h_j$, we obtain

$$s_{ij} = \frac{\cos(\alpha)}{\cos(\alpha/2)}. \tag{18.22}$$

But (18.22) can also be interpreted in a different way. The projections of **i** and **j** onto the stimulus vector that lies ("qualitatively") halfway between **i** and **j** (see Figure 18.6b) are $h_i \cos(\alpha/2)$ and $h_j \cos(\alpha/2)$, respectively. If $h_i = h_j$, then the these projections sum to $2h_i \cos(\alpha)$. If this term is used for T_{ij}, one also obtains (18.22). One could also reason that $K_{ij} = [h_i \cos(\alpha) + h_j \cos(\alpha)]/2 = h_i \cos(\alpha)$ and $T_{ij} = [h_i \cos(\alpha/2) + h_j \cos(\alpha/2)]/2 = h_i \cos(\alpha/2)$, which again implies (18.22).

Of course, many more possibilities for K_{ij} and T_{ij} offer themselves if the special constraint $h_i = h_j$ is dropped. Sjöberg (1975) presents the partial overview shown in Table 18.6. It is not surprising that none of these hypotheses has been found to be universally superior. But this leads us back to the question raised at the beginning of this section, and we can now conclude that there is no empirical correspondence between proximities and scalar-product estimates that allows one to derive the latter from the former.

TABLE 18.6. Some formulations for communality K_{ij} and totality T_{ij} of two stimuli. $\bar{K}_{ij}$ in model 4 (Goude, 1972) denotes what is not common to i and j, so that $s_{ij} = 1 - \bar{K}_{ij}/T_{ij}$. The function $\min(a, b)$ selects the smaller of a and b. If $h_i = h_j$, model 2 (Ekman et al., 1964) is equal to formula (18.22). K_{ij} in model 3 (Ekehammar, 1972) is the vector sum of c_{ij} and c_{ji} in Figure 18.6c. Model 1 is by Ekman and Lindman (1961), and model 5 (the *content model*) is by Eisler and Roskam (1977) and Eisler and Lindman (1990).

Model	Communality K_{ij}	Totality T_{ij}
1	$(h_i + h_j)\cos(\alpha)$	$h_i + h_j$
2	$\min[h_j, h_i\cos(\alpha)] + \min[h_i, h_j\cos(\alpha)]$	$[h_i^2 + h_j^2 + 2h_ih_j\cos(\alpha)]^{1/2}$
3	$\cos(\alpha)[h_i^2 + h_j^2 + 2h_ih_j\cos(\alpha)]^{1/2}$	$h_i + h_j$
4	$[h_i^2 + h_j^2 - 2h_ih_j\cos(\alpha)]^{1/2} = \bar{K}_{ij} = d_{ij}$	$[h_i^2 + h_j^2 + 2h_ih_j\cos(\alpha)]^{1/2}$
5	$2 \cdot \min(h_i, h_j)\cos(\alpha)$	$h_i + h_j$

18.5 MDS of Scalar Products

Given a matrix of scalar products, we can compute—by solving $\mathbf{B} = \mathbf{X}\mathbf{X}'$ for $\mathbf{X}$—a configuration $\mathbf{X}$ that represents or approximates the scalar products (see Chapter 7). Because $\mathbf{B} = \mathbf{X}\mathbf{X}' = (\mathbf{X}\mathbf{T})(\mathbf{X}\mathbf{T})' = \mathbf{X}\mathbf{T}\mathbf{T}'\mathbf{X}' = \mathbf{X}\mathbf{X}'$ for $\mathbf{T}\mathbf{T}' = \mathbf{I}$, $\mathbf{X}$ is unique up to an orthogonal transformation $\mathbf{T}$. That is, $\mathbf{X}$ can be rotated and/or reflected freely without affecting the quality of the solution.

An Application on the Color Data

For the symmetrized matrix of Table 18.4, $\mathbf{B} = (\mathbf{B} + \mathbf{B}')/2$, Ekman (1963) reports the point coordinates in Table 18.7. The column $\widehat{h}_i^2$ shows the squared length of vector $\mathbf{i}$ in the MDS space; the hat denotes that this length is a reconstruction of the vector length computed directly from the data. For example, using (18.5) with $i = j$, we find for $i = 3$: $(1.29)(1.29) + (-1.47)(-1.47) + (0.15)(0.15) = 3.85$. In Table 18.3, we had concluded that this color's vector should have a length of 3.87, so the 3D MDS configuration comes very close to representing this value accurately. The five other vectors also represent their colors well.

Table 18.7 shows that the vectors are distributed primarily around the first principal axis: the sum of the squared projections onto this dimension is 14.87, and only 7.72 and 1.54 for the second and third principal axes, respectively. Thus, the vector configuration is essentially two-dimensional. The dimensions of this space are principal axes. Therefore, we know that the plane spanned by the first two axes is the best possible approximation to the 3D vector configuration $\mathbf{X}$.

Any coordinate system can be picked to coordinate this plane. Ekman (1963), for example, rotated the principal axes to a *simple structure ori-*

TABLE 18.7. Coordinates of vector configuration for symmetrized scalar products of Table 18.4; PAs are principal axes of 3D representation; D1 and D2 are dimensions of the 2D plane spanned by PA_1 and PA_2 after rotation to simple structure; SS is the sum-of-squares of the column elements; $\widehat{h}_i^2$ is the squared length of the vector in space.

nm	PA_1	PA_2	PA_3	$\widehat{h}_i^2$	D_1	D_2	$\widehat{h}_i^2$
593	1.14	-1.58	0.51	4.04	0.00	1.95	3.79
600	1.29	-1.47	0.15	3.84	0.18	1.95	3.81
610	1.75	-0.59	-0.64	3.81	1.07	1.50	3.40
628	1.82	0.60	-0.55	3.97	1.82	0.58	3.67
651	1.74	0.99	0.14	4.03	1.99	0.22	4.01
674	1.59	1.20	0.70	4.45	1.99	-0.05	3.97
SS	14.86	7.79	1.50	24.14	12.43	10.22	22.65

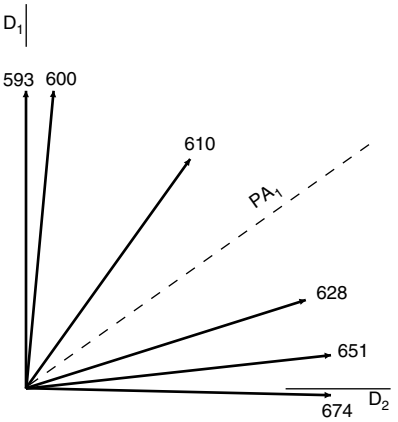

FIGURE 18.7. Vector representation of point coordinates in Table 18.7.

entation (Thurstone, 1947). Simple structure requests that the point coordinates are either very large or very small, and intermediate values are avoided. Table 18.7 shows such simple-structure coordinates resulting from rotating the first two principal axes. For the rotated dimensions, D_1 and D_2, the vector for color 593 has a coordinate value of 1.95 on the second axis, while its projection onto the first axis has zero length. Thus, this vector is collinear with the second axis. For 674, the converse is almost true.

Because the third dimension accounts for so little, we may simply ignore it, and concentrate on the plane spanned by the first two PAs. This plane is presented in Figure 18.7, together with the principal axes and the simple-structure coordinate system that give rise to the values in Tables 18.7. The endpoints of our six color vectors fall almost onto a circle about the origin in the order of their wavelengths. Thus, the perceived dissimilarities in the colors are represented by the different orientations of the vectors. The fact that all vectors have roughly the same length is, according to Ekman (1963), a consequence of the fact that the colors were all matched in brightness and saturation.

How well does Figure 18.7 represent the data? A global answer is provided by comparing the data with the scalar products implied by the given vector configuration. The latter are computed from the first two principal axes in Table 18.7 or, equivalently, from D_1 and D_2 in Table 18.7. One finds, for example, that the reconstructed scalar product is $\widehat{b}(593, 600) = (1.14)(1.29) + (-1.58)(-1.47) = 3.793$, and this is almost the same as the data value $b(593, 600) = (3.68 + 3.89)/2 = 3.785$. Given all $\widehat{b}$ and b values, we can combine them into a global fit measure. One possibility is to use the correlation coefficient of the $\widehat{b}$ and the b values. It yields $r = 0.9805$. Another measure is obtained by adding the squared differences of all $\widehat{b}$ and b values and dividing this sum by the sum-of-squares of the b values. There are no standards for evaluating such a loss function, but it suggests other representation criteria; for example, the $\widehat{b}$ values could be replaced by the rank-image values of the data, defining a loss function for a procedure that maps the data *ordinally* into scalar products. This criterion was used in SSA-III, a program for *nonmetric factor analysis* (Lingoes & Guttman, 1967).

Successive Extractions of Dimensions from Scalar Products

Rather than computing an MDS solution for scalar products in one fixed dimensionality, one can extract this solution dimension by dimension. This allows further tests.

Consider the preliminary **B**-matrix in Table 18.4. One may interpret its asymmetries as essentially due to random noise and thus generate a "better" **B**-matrix by averaging the corresponding b_{ij}- and b_{ji}-values. For the

TABLE 18.8. Upper half with diagonal shows error values $b_{ij} - \widehat{b}_{ij}$ for 3D vector configuration; lower half shows asymmetries, $(b_{ij} - b_{ji})/2$ of values in Table 18.4.

nm	593	600	610	628	651	674
593	.08	-.08	-.01	.05	-.02	-.01
600	.11	.09	-.01	-.02	.01	.01
610	-.01	-.09	.06	-.07	.03	.01
628	-.03	.08	-.10	.16	-.17	.07
651	.12	.11	-.03	.00	.30	-.16
674	.05	.12	-.09	.02	.00	.10

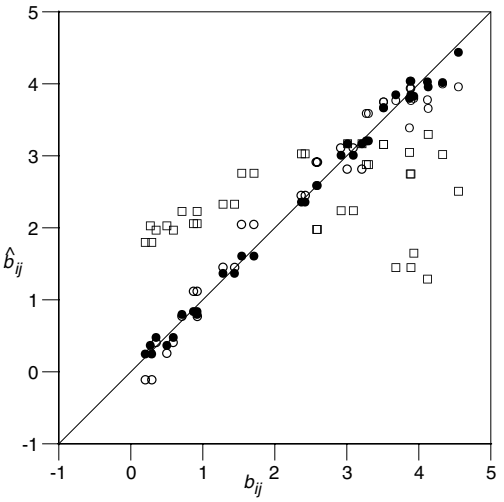

FIGURE 18.8. Plot of residuals of empirical scalar products of Table 18.4 estimated by vector configurations in 1D (squares), 2D (open circles), and 3D (filled circles).

MDS representation of **B**, we now proceed stepwise, extracting one principal component after the other,[4] until the representation seems sufficiently precise. Precision may be defined by requiring that the b_{ij}s do not differ from the scalar products computed on the MDS coordinates, $\widehat{b}_{ij}$, by a magnitude that lies in the range of the asymmetries. Because these asymmetries were assumed to be due to error, further principal components would only represent structure that cannot be distinguished from error. Table 18.8 shows that the criterion is satisfied by a 3D solution. Figure 18.8 shows the residuals of the estimation of the scalar products in 1D, 2D, and 3D. The sum of squared errors is in 1D 63.17, in 2D 2.53, and in 3D .29. According to these criteria, a representation in at most 3D seems adequate, because the error sum-of-squares in 3D is less than the sum-of-squares of the asymmetries in Table 18.8 (= .336).

MDS Representations of v- and s-Data

We have seen that scalar products determine not only n stimulus points but also a unique origin. However, scalar products are often employed in a purely ancillary fashion, because they allow direct computation of a vector configuration by algebraic means. If we begin with distance estimates, we can convert them into scalar products by picking some point to serve as an origin. In that case, the origin has no direct empirical meaning. If the data are scalar products that are not just indices such as correlations computed over persons, say, but measurements constructed from contained-in judgments, then the origin has a meaning, as we have seen for the color data.

But will there be other differences between the MDS solutions derived from scalar-product and distance data? Yes, because of restrictions built into v-judgments. Figure 18.3b shows that the contained-in judgments have a lower bound when the two respective stimulus vectors **i** and **j** are perpendicular, so that $c_{ij} = 0$. In the color circle in Figure 4.1, this is the case, for example, for the colors with wavelengths 674 nm and 584 nm. Indeed, Table 18.2 shows for the very similar stimulus pair (674 nm and 593 nm) that the contained-in rating is almost equal to 0. But what can the subject say when asked to evaluate to what extent the color with wavelength 555 nm is contained in the color 674 nm? For these colors, the respective vectors in the color circle form an obtuse angle. Even more extreme are the complementary colors red and green, which are opposite each other in the color circle: what portion of red is contained in green? Because the subject cannot respond with v-values of less than 0 (unless the procedure is gen-

[4]This extraction process amounts to a spectral decomposition of **B**, $\mathbf{B} = \lambda_1 \mathbf{q}_1 \mathbf{q}_1' + \ldots + \lambda_m \mathbf{q}_m \mathbf{q}_m'$, where λ_i is the ith eigenvalue and $\mathbf{q}_i$ the corresponding eigenvector. See formula (7.12).

eralized in some way), it seems plausible that we end up with $b(674,584) = 0$, $b(674,490) = 0$, and also $b(584, 490) = 0$. But this means that the scaling problem corresponds to a situation like that in Figure 18.2, where all three angles are equal to 90°. To fit the three quarter-circles together necessitates a 3D space. So, for the complete color circle, we should expect a 4D MDS representation if it is based on v-data.

18.6 Exercises

Exercise 18.1 Given a matrix $\mathbf{B}$, how can one check, by matrix computation, whether $\mathbf{B}$ is a scalar-product matrix?

Exercise 18.2 Consider the following geometric problems.

(a) What is the angle between $\mathbf{x} = (2, -2, 1)$ and $\mathbf{y} = (1, 2, 2)$?

(b) What is the projection of $\mathbf{x}$ onto $\mathbf{y}$?

(c) What is the projection of $\mathbf{x}$ onto the plane spanned by (1, 0, 0) and (1, 1, 0)?

Exercise 18.3 What multiple of $\mathbf{a} = (1, 1)$ should be subtracted from $\mathbf{b} = (4, 0)$ to make the result orthogonal to $\mathbf{a}$? Sketch a figure.

Exercise 18.4 Draw two vectors $\mathbf{a}$ and $\mathbf{b}$ in the plane, both emanating from the same origin, such that $\mathbf{a}+\mathbf{b}$ is perpendicular to $\mathbf{a}-\mathbf{b}$. What properties have to hold for $\mathbf{a}$ and $\mathbf{b}$ to make this possible?

Exercise 18.5 Consider the experiment by Sixtl (1967) reported on p. 397. He asked subjects to assess how much of emotion x is contained in emotion y. The exact question posed to the subjects was: "How much does x have of y?" Whether this instruction was further explained is not reported. The emotions were shyness, compassion, desire, love, humbleness, tenderness, anxiety, aggressiveness, wrath, and disgust.

(a) Discuss the task to which these subjects had to respond. Devise additional or alternative instructions that would make it very clear to them what they were expected to deliver.

(b) What type of questions concerning this task do you expect the subjects to raise in this context?

(c) Compare the above experimental method to one where proximities are collected. What type of data collection would you prefer? Which one is more likely to yield better data?

(d) Assume that the contained-in judgments generate the type of data that proponents of this method are hoping to get. What are the additional insights that these data would then allow over and beyond direct similarity ratings, say?

Exercise 18.6 Data collection by way of contained-in judgments has been restricted to a range of 0% to 100%. This implies that the respondents cannot distinguish stimuli that are "orthogonal" to each other from those that are opposite to each other, for example. They would both be rated as 0%. Devise a method that does away with this restriction. Work out the instructions that you would use to instruct the respondents about their task, and discuss your approach in the context of both the color similarities and the similarity of emotional experiences discussed in this chapter.