

23

Modeling Asymmetric Data

Distances are always symmetric, but proximities may be asymmetric. Proximities, therefore, cannot always be fully represented by the distances among points in an MDS space. If one feels that the proximities deviate from being symmetric due to error only, this is not a problem. In that case, one may somehow symmetrize the proximities (e.g., by first averaging the corresponding p_{ij} s and p_{ji} s and then running the MDS on these averages). If one hypothesizes, however, that the nonsymmetries are meaningful, one needs special models for analyzing such data. In this chapter, we consider a number of such models. First, it is shown that an asymmetric proximity matrix can always be decomposed into a symmetric and a skew-symmetric component. The symmetric component can be then be subjected to ordinary MDS. For the skew-symmetric part, we discuss special visualization techniques for either the nonsymmetric component by itself or for embedding the nonsymmetric component into an MDS representation of the symmetric component. In the rest of the chapter, we treat a variety of models that analyze asymmetric proximities directly or indirectly. Many other models for asymmetric data exist. For a good overview of models for asymmetric data, we refer to Zielman and Heiser (1996).

23.1 Symmetry and Skew-Symmetry

We compare the values in row A of Table 4.2 with those in column A. Row A shows the confusion rate for code A, presented first, with codes B, C,

and so on, respectively, presented afterwards. In column A, the code A is always the second stimulus in the comparison. Comparing corresponding elements in column A and row A, we note, for example, that $p(A, R) = .35$, and $p(R, A) = .13$. Hence, A is definitely confused more often with R if it is presented before A than if it follows A in time.

Thus, the Morse code data are definitely not symmetric. However, in the analysis of these data so far, asymmetries played no further role. They were simply discarded as error, and only the symmetric part was analyzed. But is that good science? We know that asymmetries are not uncommon in cognition. A child, for example, is typically seen as similar to a parent, but one would not say that the parent resembles the child. This asymmetry is explained as a prototype-specimen relation: the specimen resembles the prototype, but the prototype does not resemble the specimen. Other examples and more theorizing are reported by Tversky (1977), for example. So, it is at least conceivable that the asymmetries in the Morse code data are not purely random but systematic.

To arrive at an answer to that question, we first note that every square matrix \mathbf{P} can be uniquely decomposed into a symmetric matrix and a *skew-symmetric* matrix: That is, every asymmetric proximity matrix \mathbf{P} can be uniquely decomposed into

$$\mathbf{P} = \mathbf{M} + \mathbf{N}, \tag{23.1}$$

where \mathbf{M} is symmetric and \mathbf{N} is skew-symmetric. This means that $\mathbf{M} = \mathbf{M}'$ and $\mathbf{N} = -\mathbf{N}'$. The two components of \mathbf{P} are

$$\mathbf{M} = (\mathbf{P} + \mathbf{P}')/2, \text{ and} \tag{23.2}$$

$$\mathbf{N} = (\mathbf{P} - \mathbf{P}')/2. \tag{23.3}$$

Note that the diagonal elements of \mathbf{N} are always zero, because for those elements it holds that $n_{ii} = (p_{ii} - p_{ii})/2 = 0$.

To demonstrate this decomposition numerically, consider the following example, where \mathbf{P} is equal to the first four rows and columns of Table 4.2. Then, \mathbf{P} has the decomposition

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} 92 & 4 & 6 & 13 \\ 5 & 84 & 37 & 31 \\ 4 & 38 & 87 & 17 \\ 8 & 62 & 17 & 88 \end{bmatrix} = \mathbf{M} + \mathbf{N} \\ &= \begin{bmatrix} 92.0 & 4.5 & 5.0 & 10.5 \\ 4.5 & 84.0 & 37.5 & 46.5 \\ 5.0 & 37.5 & 87.0 & 17.0 \\ 10.5 & 46.5 & 17.0 & 88.0 \end{bmatrix} + \begin{bmatrix} 0.0 & -0.5 & 1.0 & 2.5 \\ 0.5 & 0.0 & -0.5 & -15.5 \\ -1.0 & 0.5 & 0.0 & 0.0 \\ -2.5 & 15.5 & 0.0 & 0.0 \end{bmatrix}. \end{aligned} \tag{23.4}$$

To show that the decomposition is unique, assume that $\mathbf{P} = \mathbf{M}_1 + \mathbf{N}_1$ is another such decomposition with $\mathbf{M}_1 = \mathbf{M}'_1$ and $\mathbf{N}_1 = -\mathbf{N}'_1$. Then, we have $\mathbf{P}' = \mathbf{M}'_1 + \mathbf{N}'_1 = \mathbf{M}_1 - \mathbf{N}_1$, and it follows that $\mathbf{P} + \mathbf{P}' = 2\mathbf{M}_1$ and

$\mathbf{P} - \mathbf{P}' = 2\mathbf{N}_1$. Inserting this into (23.2) and (23.3), respectively, we find that $\mathbf{M} = \mathbf{M}_1$ and $\mathbf{N} = \mathbf{N}_1$, which proves uniqueness of the decomposition (23.1).

Furthermore, the decomposition in \mathbf{M} and \mathbf{N} allows one to partition the sum-of-squares into a part due to symmetry and a part due to skew-symmetry. That is,

$$\begin{aligned} \sum_{i,j} p_{ij}^2 &= \sum_{i,j} \left[\frac{1}{2}(p_{ij} + p_{ji}) + \frac{1}{2}(p_{ij} - p_{ji}) \right]^2 \\ &= \sum_{i,j} \frac{1}{4} \left[(p_{ij} + p_{ji})^2 + (p_{ij} - p_{ji})^2 + 2(p_{ij} + p_{ji})(p_{ij} - p_{ji}) \right] \\ &= \sum_{i,j} m_{ij}^2 + \sum_{i,j} n_{ij}^2 + 2 \sum_{i,j} m_{ij}n_{ij} \\ &= \sum_{i,j} m_{ij}^2 + \sum_{i,j} n_{ij}^2. \end{aligned}$$

The cross-product term $\sum_{i,j} m_{ij}n_{ij}$ vanishes because

$$\begin{aligned} \sum_{i,j} m_{ij}n_{ij} &= \frac{1}{4} \sum_{i,j} (p_{ij} + p_{ji})(p_{ij} - p_{ji}) \\ &= \frac{1}{4} \left[\sum_{i,j} p_{ij}^2 - \sum_{i,j} p_{ji}^2 + \sum_{i,j} p_{ij}p_{ji} - \sum_{i,j} p_{ij}p_{ji} \right] = 0. \end{aligned}$$

Thus, \mathbf{M} and \mathbf{N} are orthogonal because $\text{tr } \mathbf{MN} = 0$. The decomposition of the sum-of-squares suggests analyzing asymmetric data in two separate steps: the analysis of the symmetric part and the analysis of the skew-symmetric part. For the Morse code data, the sum of the squared proximities without the diagonal equals 698,309.0, from which 671,489.5 (96%) is due to symmetry and 26,819.5 (4%) is due to asymmetry. This implies that the symmetric part of the data is dominant, and asymmetry plays a minor role, but may still reveal interesting relations.

23.2 A Simple Model for Skew-Symmetric Data

The simplest model for representing skew-symmetric data \mathbf{N} locates every object on a line such that the *signed* distance for every pair of coordinates represents the corresponding elements of \mathbf{N} . Expressed algebraically, this model postulates that $n_{ij} = x_i - x_j$, or, in matrix form,

$$\mathbf{N} = \mathbf{x}\mathbf{1}' - \mathbf{1}\mathbf{x}', \quad (23.5)$$

where \mathbf{x} has sum zero. Choosing $\mathbf{x} = n^{-1}\mathbf{N}\mathbf{1}$ or, in other words, the averages of the n_{ij} values within each row i over all columns, is the least-squares

solution for (23.5). Obviously, this model is so restricted that it does not fit many data. Hence, we now turn to more general models.

23.3 The Gower Model for Skew-Symmetries

An interesting decomposition of a skew-symmetric matrix has been given by Gower (1977) and Constantine and Gower (1978). The singular value decomposition of any skew-symmetric matrix \mathbf{N} has the special form

$$\mathbf{N} = \mathbf{PK}\Phi\mathbf{P}', \tag{23.6}$$

where \mathbf{P} is orthonormal, Φ has singular values ordered in pairs $(\phi_1, \phi_1, \phi_2, \phi_2, \dots)$, and \mathbf{K} is a permutation-reflection matrix that has along its diagonal 2×2 blocks with off-diagonal values -1 and 1 . The decomposition of \mathbf{N} in (23.4) is

$$\left[\begin{array}{cc|cc} .16 & .00 & .00 & -.99 \\ -.97 & .19 & .00 & -.16 \\ -.01 & -.04 & 1.00 & -.00 \\ -.19 & -.98 & -.04 & -.03 \end{array} \right] \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{array} \right] \left[\begin{array}{c|c} 15.72 & \\ \hline & 15.72 \\ \hline & .91 \\ & .91 \end{array} \right] \left[\begin{array}{cccc} .00 & .19 & -.04 & -.98 \\ -.16 & .97 & .01 & .19 \\ \hline .99 & .16 & .00 & .03 \\ .00 & .00 & 1.00 & -.04 \end{array} \right].$$

The dimensions here come in pairs with equal singular values. Such a pair of dimensions is called a *bimension*. Each bimension spans a plane. Its points can be taken from the pairs of columns (1,2), (3,4), and so on of \mathbf{PK} or from the respective columns of \mathbf{P} . The configurations are the same in both cases. Their interpretation hinges on how any two points are related to each other in terms of (a) the angle subtended by the vectors that they define, and (b) the area of the triangle spanned by these vectors. The area of the triangle represents the size of the asymmetry, and sense of the angle represents the sign of the asymmetry. Note that because the singular values in each bimension are equal, the bimension may be freely rotated or reflected without changing the fit. Thus, in the Gower model, the axes cannot be interpreted.

To illustrate these notions, let us apply the Gower decomposition to the skew-symmetric part of the Morse code confusion data (Rothkopf, 1957). Zielman and Heiser (1996) did the same analysis on the Morse codes that represent the 10 digits only. The first bimension (with singular value 67.37, showing 34% of the total skew-symmetry) of the full table is presented in Figure 23.1, a display we call a *Gower diagram*. In this display, all the rows of \mathbf{P} are plotted as vectors and we have to use a clockwise rotation for positive estimates of n_{ij} . To understand how to interpret this plot, consider the triangle between the origin, point H, and point V. The area of the triangle is an estimate of the size of the asymmetry. Because clockwise rotations indicate positive estimates, going from H to V indicates that if H is the first stimulus in the pair, it is more often confused with V than vice versa.

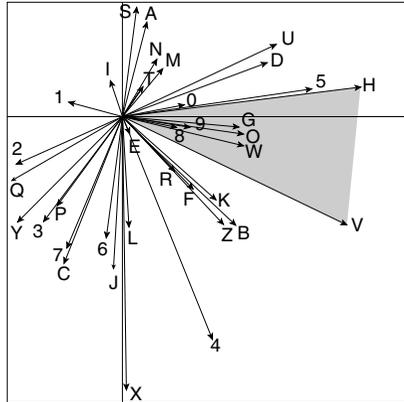


FIGURE 23.1. Gower diagram of the skew-symmetric part of the Morse code data (first dimension). The area of the triangle between the origin and the two points H and V is an estimate of the value n_{HV} . The clockwise rotation indicates that $n_{HV} > 0$ and $n_{VH} < 0$.

Once identified, an asymmetry like this one can be interpreted substantively. In this case, it is easy to understand, because the sequence HV is \dots followed by $\dots-$, whereas VH is $\dots-$ followed by \dots . Clearly, the middle $-$ makes it easier for the subjects to distinguish the two signals.

We also note from Figure 23.1 that the big contributors to asymmetry are the signals X, 4, V, and H, where, for example, H4, VX, and V4 generate higher confusion rates than, respectively, 4H, XV, and 4V. In contrast, for E, I, and T it does not really matter whether they occur as the first or second signal, because they lie close to the origin and do not have much asymmetry with other signals.

Some of the properties of Gower’s model are as follows.

- If n is even, then there are $n/2$ dimensions; if n is odd, then there are $n/2 - 1$ dimensions; that is, $\phi_n = 0$.
- Points that lie on the same line through the origin do not have asymmetry, thus spanning a triangle with zero area.
- A point close to the origin has little asymmetry with all other points, and, hence, triangles where these points together with the origin and any other point form the corners tend to be small, in general.
- If there is a line through the origin in a dimension such that all of the vectors project positively on this line, then reordering \mathbf{N} by the order of the vectors of the dimension yields a matrix with all negative elements in the lower (or upper) triangular matrix and the positive elements in the upper (or lower) triangular matrix. No circular triads are present in the data in this case.

- If points lie on a line not through the origin of the vectors, then the points form an additive scale. Let the order of three points along such a line be A, B, C . Then, the area spanned by the triangle OAC equals $OAB + OBC$, which is clearly additive.
- When computing the solution for the Gower model, we do not know a priori what the direction of interpretation will be. If the submatrix of \mathbf{K} equals $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, then an anticlockwise rotation from vector i to j indicates a positive estimate of n_{ij} and a negative estimate for n_{ij} . If the computational procedure gives $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ as a submatrix, then we have to apply a clockwise rotation to identify a positive estimate of n_{ij} .

23.4 Modeling Skew-Symmetry by Distances

The power of the Gower decomposition is that a graph of only n objects is obtained in a plane where the angles subtended by the vectors have a fixed meaning. A disadvantage is that areas represent the size of asymmetries, because judgments on areas are cognitively demanding. Interpreting distances is easier. Therefore, we propose a new model for visualizing a skew-symmetric matrix that expresses asymmetries by Euclidean distances.

The distance model for skew-symmetry uses Euclidean distances between points i and j to estimate the size of the skew-symmetric effect $|n_{ij}|$. In addition, similar to the Gower model, the direction of rotation is important. If the angle measured clockwise between the vector to point i and the vector to point j is less than 180° , then n_{ij} is estimated by $d_{ij}(\mathbf{X})$. Conversely, if this angle is between 180° and 360° measured clockwise (or, equivalently, between 0 and -180° , in the counterclockwise sense), then n_{ij} is estimated by $-d_{ij}(\mathbf{X})$. Thus, the model predicts that starting from a point i all points j that are in the half plane with positive angles between 0° and 180° (measured clockwise) have a positive estimate for n_{ij} . All points j that have negative angles between 0° and -180° (measured clockwise) produce a negative estimate for n_{ij} .

To fit the distance model to skew-symmetric data, we need to determine for point i whether point j lies in the positive or in the negative rotational half of the plane. Let \mathbf{x}_i be the 2×1 vector with the row coordinates of point i . Now, a rotation of \mathbf{x}_i by -90° is obtained by reflecting the first coordinate vector and then swapping the dimensions, that is, by $\mathbf{T}'\mathbf{x}_i$ with

$$\mathbf{T}' = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \text{ and } \mathbf{T} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

If another vector \mathbf{x}_j is projected onto $\mathbf{T}'\mathbf{x}_i$, and the result is positive, then point j is on the positive side of the plane so that n_{ij} is estimated by

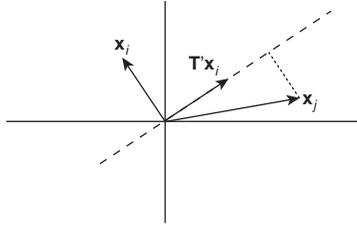


FIGURE 23.2. Illustration of determining whether point j gives a positive or negative contribution for a combination ij . Vector \mathbf{x}_j is projected on $\mathbf{T}'\mathbf{x}_i$, the -90° rotation of \mathbf{x}_i . In this example, the projection $\mathbf{x}'_j\mathbf{T}'\mathbf{x}_i$ is positive indicating a positive estimate for n_{ij} .

$d_{ij}(\mathbf{X})$. Figure 23.2 gives an example of this case. On the other hand, if the projection is negative, then point j is on the negative side of the plane so that n_{ij} is estimated by $-d_{ij}(\mathbf{X})$. The projection of \mathbf{x}_j onto $\mathbf{T}'\mathbf{x}_i$ is given by $\mathbf{x}'_j\mathbf{T}'\mathbf{x}_i$. Let the sign function be defined by

$$\text{sign}(z) = \begin{cases} 1 & \text{if } z > 0, \\ 0 & \text{if } z = 0, \\ -1 & \text{if } z < 0. \end{cases}$$

Now, the estimate of n_{ij} can be obtained by $\text{sign}(\mathbf{x}'_j\mathbf{T}'\mathbf{x}_i)d_{ij}(\mathbf{X})$. A formal model for the distance model for skew-symmetry is obtained by minimizing the sum of squared differences between n_{ij} and its estimate, that is, by minimizing

$$L(\mathbf{X}) = \sum_{i=1}^n \sum_{j=1}^n [n_{ij} - \text{sign}(\mathbf{x}_i\mathbf{T}\mathbf{x}_j)d_{ij}(\mathbf{X})]^2 \quad (23.7)$$

over \mathbf{X} .

We applied this model to the skew-symmetric part of the Morse code data and the results are given in Figure 23.3. Small distances in the plot generally indicate small skew-symmetries. The plot is dominated by large asymmetries that are shown by large distances. For example, we see that H and X have a large distance. Because we are using a clockwise rotation, the sequence HX leads to higher confusion rates than presenting X first and H afterwards. Comparing this solution to the Gower diagram in Figure 23.1 shows that there is not so much difference. However, the advantage of the distance model for skew symmetry is that the distance of two points shows to what extent the proximities of two objects are asymmetric.

Some caution is required here. The loss function (23.7) was fitted with a general-purpose optimization routine in MatLab. Such a routine may not be optimal for this loss function. In particular, we expect that this loss function may be quite sensitive to local optima. Further study is needed to see how severe this problem is.

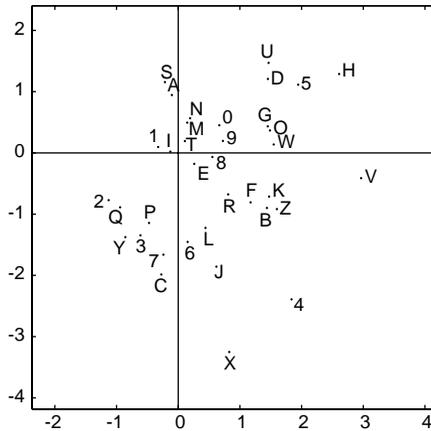


FIGURE 23.3. Plot of the distance model for skew-symmetry on the Morse code data. The distance in a clockwise direction estimates positive values for n_{ij} and those in an anticlockwise direction estimate minus the distance for n_{ij} .

As is the case for the Gower decomposition, the dimensions in the current model come in pairs, the so-called bimension. The distance model for skew-symmetry could be extended to two or more pairs of such dimensions. It seems natural to compute the distance for each bimension separately, inasmuch as the interpretation is done bimensionwise.

23.5 Embedding Skew-Symmetries as Drift Vectors into MDS Plots

Another simple method for asymmetries is simultaneously displaying the symmetric part and the skew-symmetric part of the data. This makes it possible to see how these two data components are related to each other. The skew-symmetric values are embedded into the MDS representation of the symmetrized data by drawing arrows (*drift vectors*) from each point i to any other point j in the configuration so that these vectors correspond in length and direction to the values in row i of the skew-symmetric matrix (Borg, 1979; Borg & Groenen, 1995). Thus, on point R in Figure 4.6 we would attach a vector of length $.11 = (.35 - .13)/2$ pointing towards A. The units for the arrows are chosen so that they can be represented most conveniently in the given configuration. The arrow's direction towards A is chosen to express that A is more often confused with R when presented first than vice versa.

To avoid a cluttered picture, we can draw only the resultant of the vector bundle thus attached to each point. The resultant averages out random nonsymmetries and shows the general drift (see Figure 23.4). Length and

direction angle of drift vectors are computed as follows. We use vector notation and show the 2D case.

1. Do for all points i .
2. Do for all points $j \neq i$.
3. Given vectors \mathbf{x}_i and \mathbf{x}_j in terms of their MDS coordinates, $\mathbf{a}_{ij} = \mathbf{x}_j - \mathbf{x}_i$ is the vector from point i to point j in the MDS configuration.
4. Norm \mathbf{a}_{ij} to unit length to get \mathbf{b}_{ij} ; that is, $\mathbf{b}_{ij} = \mathbf{a}_{ij}/(\mathbf{a}'_{ij}\mathbf{a}_{ij})^{1/2}$.
5. Multiply \mathbf{b}_{ij} by element n_{ij} of the skew-symmetric component of the proximity matrix to obtain \mathbf{c}_{ij} ; that is, $\mathbf{c}_{ij} = n_{ij}\mathbf{b}_{ij}$.
6. End do.
7. Average the n vectors \mathbf{c}_{ij} to obtain the (average) drift vector for point i , \mathbf{d}_i ; that is, $\mathbf{d}_i = n^{-1} \sum_j \mathbf{c}_{ij}$.
8. For plotting \mathbf{d}_i , compute \mathbf{d}_i 's length as the root mean square of its elements and the direction angle relative to the Y -axis, $\alpha_i = \arccos(\mathbf{d}'_i\mathbf{u}/\sqrt{\mathbf{d}'_i\mathbf{d}_i})$, where $\mathbf{u}' = (0, 1)$.
9. End do.

Figure 23.4 shows the 2D MDS solution with the embedded drift vectors. It is obvious that the nonsymmetries in the confusion data are not random (see also Möbus, 1979). One notes that the arrows exhibit a definite vector field with a trend that, in substantive terms, indicates that shorter Morse code signals are more often confused with longer ones than vice versa. The vertical axis reflects the temporal length of the signals. The signal E, for example, is just one ·, while the signal for O is — — — —. Moreover, because the trend is towards the North-West, the asymmetries also reflect the composition pattern of the signals: signals in the direction of the drift vectors tend to have more short components (see Figure 4.7).

23.6 Analyzing Asymmetry by Unfolding

We now change to models that directly analyze the entire asymmetric proximity matrix. One property that all these models share is that they somehow model both the symmetric part of the data and the skew-symmetric part. One of the simplest distance models that can be used is unfolding (see Chapters 14 to 16) as, for example, has been suggested by Gower (1977).

We look at an example of brand switching data. These data are derived from supermarket scanner data as described by Bell and Lattin (1998). In this example, we want to investigate how households change in buying 15 different cola soft drinks. The daily purchases of cola soft drinks were recorded for 488 U.S. households over a period of 104 weeks from June 1991

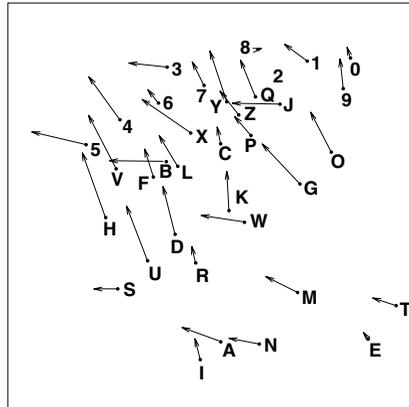


FIGURE 23.4. 2D Morse code MDS configuration with drift vectors to model the asymmetry in the Morse codes.

to June 1993. A household is considered to make a change whenever the products were of a different type or brand for two subsequent purchases. If a household has more than one purchase on a day, then we divide this switch evenly over all the products that have been bought that day. Table 23.1 shows the rounded brand switching data for colas. The rows of the table indicate the type of cola bought before and the column indicates the type of cola that is currently bought. Thus, the changes are made from the row product to the column product.

Brand switching data can be interpreted as similarities, because large values indicate that households easily switch between the two products and, hence, consider them similar. For unfolding, we need to transform these similarities into dissimilarities. One such transformation can be obtained by the gravity model discussed in Section 6.4. This model stems from astronomy and relates the gravitational force p_{ij} to a squared Euclidean distance d_{ij}^2 by the relation $p_{ij} = km_i m_j / d_{ij}^2$, where k is a constant and m_i and m_j are the masses of the two bodies. We equate m_i and m_j with the row and column sums of the brand switching matrix. Then,

$$\delta_{ij} = \left(\frac{m_i \cdot m_j}{p_{ij}} \right)^{1/2} \tag{23.8}$$

gives an asymmetric dissimilarity matrix on which unfolding can be performed. If p_{ij} equals zero, then δ_{ij} is declared missing. The brand switching data converted in the sense of the gravity model are given in Table 23.2. Note that Zielman and Heiser (1993) and Groenen and Heiser (1996) have used the gravity model in a similar context.

We have applied unfolding to the cola brand switching data. The joint representation is given in Figure 23.5, where the rows (the colas from which the change is made) are plotted as solid points and the columns (the colas

TABLE 23.1. Brand switching data among 15 different colas. The row indicates from which product the change is made, the column contains the product to which is changed.

From	To														
	a.	b.	c.	d.	e.	f.	g.	h.	i.	j.	k.	l.	m.	n.	o.
a. Coke decaf	41	11	2	8	0	2	15	8	14	0	9	11	0	6	2
b. Coke diet decaf	9	341	32	3	4	8	55	78	31	1	63	16	17	14	4
c. Pepsi diet decaf	3	27	160	15	8	2	18	15	32	2	31	13	2	12	7
d. Pepsi decaf	7	3	17	89	2	3	16	8	4	0	3	27	1	6	3
e. Canfield	1	7	6	2	119	6	20	8	19	0	16	15	2	21	7
f. Coke	4	4	2	1	4	73	37	8	12	3	8	33	3	36	6
g. Coke classic	14	53	16	16	22	38	675	98	56	10	48	187	33	172	20
h. Coke diet	5	74	14	12	7	5	108	716	123	26	92	31	11	27	18
i. Pepsi diet	14	35	36	3	15	11	56	120	422	20	86	82	29	38	10
j. RC diet	0	5	0	1	3	3	6	30	5	12	17	6	4	14	1
k. Rite diet	13	70	29	6	12	5	49	87	92	19	471	40	11	34	8
l. Pepsi	8	18	9	26	19	29	204	26	91	5	29	663	24	217	51
m. Private label	2	14	4	3	1	2	35	13	22	1	20	19	364	23	1
n. RC	7	10	13	7	19	34	171	30	31	10	36	230	22	440	41
o. Wildwood	3	3	7	3	10	9	26	22	11	2	4	48	2	35	215

TABLE 23.2. Brand switching data of Table 23.1 converted in the sense of the gravity model (23.8).

From	To														
	a.	b.	c.	d.	e.	f.	g.	h.	i.	j.	k.	l.	m.	n.	o.
a. Coke decaf	20	89	150	56	0	122	113	143	94	0	116	129	0	153	159
b. Coke diet decaf	99	37	86	210	203	139	135	105	145	274	100	245	144	230	258
c. Pepsi diet decaf	123	93	27	67	103	200	170	171	102	139	102	195	302	178	140
d. Pepsi decaf	59	206	62	20	152	120	133	173	214	0	242	100	315	186	158
e. Canfield	181	155	120	156	23	98	136	199	112	0	120	154	256	114	118
f. Coke	88	199	201	214	120	27	97	193	137	93	165	100	202	84	124
g. Coke classic	117	136	178	133	127	94	57	137	159	127	168	105	152	96	169
h. Coke diet	182	108	177	144	211	242	132	47	100	74	113	241	246	227	167
i. Pepsi diet	96	137	97	252	126	143	161	102	47	74	103	130	133	168	196
j. RC diet	0	120	0	144	93	91	163	67	144	31	77	159	119	91	205
k. Rite diet	98	96	106	175	139	209	170	117	100	74	43	183	212	175	216
l. Pepsi	152	231	234	103	135	106	102	263	123	177	214	55	176	85	105
m. Private label	185	159	213	185	358	245	149	226	152	241	156	198	27	158	454
n. RC	144	273	171	175	119	86	98	216	185	111	169	82	162	52	103
o. Wildwood	132	300	141	161	99	101	151	152	187	149	305	109	324	112	27

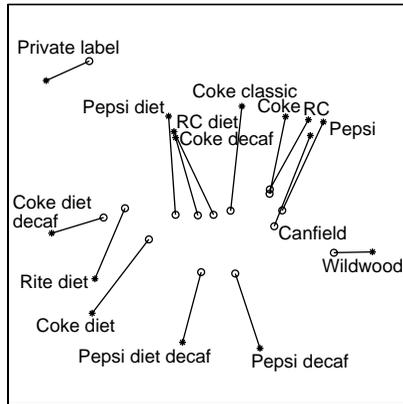


FIGURE 23.5. Unfolding on the brand switching data of Table 23.1 after being converted by the gravity model. The solid points denote the the colas chosen at time t , the open circles represent colas chosen at time $t + 1$.

to which the change is made) as open circles. To interpret this diagram, one studies, for example, how Coke diet buyers change: note that they tend to move to Riet diet, Coke diet decaf, Pepsi diet, RC diet, and Pepsi diet decaf, because they are the nearest in the plot. They are not likely to change to Private label, Wildwood, and Canfield because these brands are far away. A striking feature of the solution is that there is not much changing going on for customers buying the Private label because it is located far away from all other colas. In the same manner, one can focus on any other cola and see how customers change to the competing colas.

23.7 The Slide-Vector Model

The unfolding model for asymmetry estimates many parameters. To reduce the number of parameters, a constrained form of unfolding can be used. One such model is the *slide-vector* model that constrains the row and column points to be equal up to a translation. This restriction implies that the solution consists of the points that represent the choice objects and one uniform shift of the entire space, the *slide-vector* \mathbf{z} , in a fixed direction. This model was first proposed by De Leeuw and Heiser (1982, who attributed it to a personal communication with Kruskal, 1973) and was thoroughly worked out by Zielman and Heiser (1993). Note that Carroll and Wish (1974b) refer to the same model when they speak of the drift vector model. The rationale behind the model is that the data can be thought to consist of symmetric distances augmented by a strong wind: changes against the wind direction will take more effort, whereas changes in the direction are easier. In fact, Figure 23.4 suggests just that model.

To lay out the slide-vector model formally, we define \mathbf{X} to be the row coordinates and \mathbf{Y} the column coordinates in the unfolding problem. Then, the restriction in the slide-vector model amounts to $y_{is} = x_{is} - z_s$. The definition of the distance in the slide-vector model is given by

$$d_{ij}(\mathbf{X}, \mathbf{Y}, \mathbf{z}) = \left(\sum_s (x_{is} - y_{js})^2 \right)^{1/2} = \left(\sum_s (x_{is} + z_s - x_{js})^2 \right)^{1/2}. \quad (23.9)$$

Thus, the row points \mathbf{X} are equal to the column points \mathbf{Y} translated by the slide-vector \mathbf{z} .

Clearly, if $\mathbf{z} = \mathbf{0}$, then (23.9) reduces to the ordinary symmetric Euclidean distance. For two objects that are at the same position, $d_{ij}(\mathbf{X}, \mathbf{Y}, \mathbf{z})$ reduces to $(\sum_s z_s^2)^{1/2}$, which again is symmetric. In fact, if diagonal values δ_{ii} are fitted in the unfolding problem, then these entries are all estimated by the length of the slide vector.

The slide-vector model can easily be fitted by considering unfolding as MDS with missing within-group dissimilarities (see Section 14.1) combined with external constraints (see Section 10.3). In matrix notation, the slide-vector restrictions are given by $\mathbf{Y} = \mathbf{X} - \mathbf{1z}'$. Stacking the row and column coordinates underneath each other gives

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X} - \mathbf{1z}' \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{I} & -\mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{z}' \end{bmatrix} = \mathbf{E} \begin{bmatrix} \mathbf{X} \\ \mathbf{z}' \end{bmatrix}.$$

Thus, the slide-vector model is fitted by providing the external constraints \mathbf{E} in this way.

Consider a small illustrative example of three objects for generating the matrix of external constraints \mathbf{E} . Suppose we want to apply the slide-vector model in two dimensions. Then, the full MDS matrix becomes

$$\Delta = \begin{bmatrix} 0 & 0 & 0 & \delta_{11} & \delta_{12} & \delta_{13} \\ 0 & 0 & 0 & \delta_{21} & \delta_{22} & \delta_{23} \\ 0 & 0 & 0 & \delta_{31} & \delta_{32} & \delta_{33} \\ \delta_{11} & \delta_{21} & \delta_{31} & 0 & 0 & 0 \\ \delta_{12} & \delta_{22} & \delta_{32} & 0 & 0 & 0 \\ \delta_{13} & \delta_{23} & \delta_{33} & 0 & 0 & 0 \end{bmatrix} \text{ and } \mathbf{W} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Note that the matrix \mathbf{W} indicates that the within-block dissimilarities are missing, so that we are dealing with unfolding. The between-block dissimilarities contain the asymmetric data. The matrix of coordinates is simply

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ y_{11} & y_{12} \\ y_{21} & y_{22} \\ y_{31} & y_{32} \end{bmatrix},$$

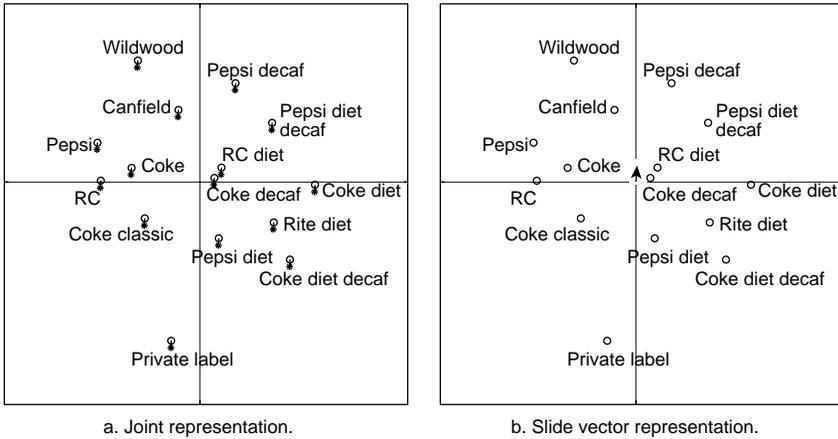


FIGURE 23.6. The slide-vector model fitted to the brand switching data of Table 23.1 after being converted by the gravity model. Panel (a) shows the joint representation of rows (*) and columns (o). Panel (b) shows the representation with one set of points and the slide vector. The arrow in the center that indicates the slide vector is rather small.

that is, restricted to be equal to

$$\mathbf{E} \begin{bmatrix} \mathbf{X} \\ \mathbf{z}' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ z_1 & z_2 \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ x_{11} - z_1 & x_{12} - z_2 \\ x_{21} - z_1 & x_{22} - z_2 \\ x_{31} - z_1 & x_{32} - z_2 \end{bmatrix}.$$

Thus, the slide-vector model can be fitted by providing the matrix of external variables \mathbf{E} above to an MDS program that allows for linear restrictions on the configuration and allows for missing values of the dissimilarities.

We now return to the colaswitching data of Table 23.1 and their transformations by the gravity model in Table 23.2. The slide-vector model was fitted by PROXSCAL in SPSS as it allows for external variables such as those defined by \mathbf{E} . The resulting configuration is given in Figure 23.6. Because the slide-vector model is a constrained version of unfolding, there are two possible representations. The joint representation in Figure 23.6a shows the row and column points together. It can be clearly seen that the column points are indeed equal to the row point up to a translation. The second representation (see Figure 23.6b) only shows a single set of coordinates together with the slide-vector \mathbf{z} .

The example shows that the slide vector is rather small for these data. The model mostly captures the symmetric part of the data and shows that a uniform trend in the asymmetries, however large they may be, is relatively small.

It seems that switching takes place mostly between colas of the same type. For example, there is switching with the group of Coke, Pepsi, RC, and Canfield, within the group Pepsi diet, Coke diet decaf, Rite diet, and Coke diet, but less switching between these groups. In this solution, too, we see that Private label is farthest away from all other colas indicating that those households do not switch easily to other colas.

A disadvantage of the slide-vector model is that it is quite restrictive compared to unconstrained unfolding. Instead of $n \times p$ coordinates for \mathbf{Y} in unrestricted unfolding, the slide-vector model only estimates p parameters for the slide vector. It seems that the slide-vector model works better if asymmetries are large relative to the symmetric part of the data. We also expect the slide-vector model to perform better for small data sets because the restrictions on the coordinates are weaker than for large n .

23.8 The Hill-Climbing Model

The formal advantage of the slide-vector model is that it easily fits into the constrained unfolding framework so that it can be fitted by a standard program such as PROXSCAL. The joint representation of \mathbf{X} and the constrained \mathbf{Y} is easy to interpret even though it doubles the number of points. The more parsimonious representation of only \mathbf{X} and \mathbf{z} seems harder to interpret. To remedy the latter problem, we propose an adaptation of the slide-vector model. We are not aware of references in the literature that have proposed this model earlier.

The new model is based on the *hill-climbing* metaphor: walking uphill is more difficult than walking downhill, whereas on a plateau walking from point A to B takes the same effort as walking from B to A. This idea can be modeled by choosing the distance measure as

$$d_{ij}(\mathbf{X}, \mathbf{z}) = \left(\sum_s (x_{is} - x_{js})^2 \right)^{1/2} + \frac{\sum_s (x_{is} - x_{js})z_s}{\left(\sum_s (x_{is} - x_{js})^2 \right)^{1/2}}, \quad (23.10)$$

or in matrix notation

$$d_{ij}(\mathbf{X}, \mathbf{z}) = \|\mathbf{x}_i - \mathbf{x}_j\| + \frac{(\mathbf{x}_i - \mathbf{x}_j)' \mathbf{z}}{\|\mathbf{x}_i - \mathbf{x}_j\|}. \quad (23.11)$$

A least-squares model estimating (23.11) is given by

$$L(\mathbf{X}, \mathbf{z}) = \sum_{i=1}^n \sum_{j=1}^n \left(\delta_{ij} - \left[\|\mathbf{x}_i - \mathbf{x}_j\| + \frac{(\mathbf{x}_i - \mathbf{x}_j)' \mathbf{z}}{\|\mathbf{x}_i - \mathbf{x}_j\|} \right] \right)^2. \quad (23.12)$$

The rationale behind this model is that the projection of the difference vector $\mathbf{x}_i - \mathbf{x}_j$ of going from point i to point j on a slope given by the

slope vector \mathbf{z} measures to what extent it is more difficult or easier to go from point i to j than the Euclidean distance only. If the difference vector is orthogonal to the slope vector \mathbf{z} , then no asymmetry is modeled. If the difference vector is parallel to the slope vector, then maximum asymmetry is achieved. Note that $d_{ii}(\mathbf{X}, \mathbf{z}) = 0$ by definition so that the diagonal values cannot be modeled. The denominator $\|\mathbf{x}_i - \mathbf{x}_j\|$ in (23.11) is chosen so that the length of the Euclidean distance between i and j does not influence the amount of asymmetry.

The orientation of the difference vector, and thus the positioning of the points, is influenced by the asymmetry in the data because the orientation of the difference vector determines the projection on the slope vector. It may be verified that (23.12) can be decomposed into a symmetric part \mathbf{M} with elements $m_{ij} = (\delta_{ij} + \delta_{ji})/2$ and a skew-symmetric part \mathbf{N} with elements $n_{ij} = (\delta_{ij} - \delta_{ji})/2$; that is,

$$L(\mathbf{X}, \mathbf{z}) = 2 \sum_{i=1}^n \sum_{j=i+1}^n (m_{ij} - \|\mathbf{x}_i - \mathbf{x}_j\|)^2 + \sum_{i=1}^n \sum_{j=1}^n \left(n_{ij} - \frac{(\mathbf{x}_i - \mathbf{x}_j)' \mathbf{z}}{\|\mathbf{x}_i - \mathbf{x}_j\|} \right)^2. \quad (23.13)$$

This reformulation shows that the distances directly model the symmetric part of $\mathbf{\Delta}$, and the projections model the skew-symmetric part. The size of the symmetric part and the skew-symmetric part can be expressed in terms of the sum-of-squares $\|\mathbf{M}\|^2$ and $\|\mathbf{N}\|^2$. The relative difference between those measures influences the solution (23.13) in how much symmetry and how much skew-symmetry of the data is fitted. Consider the following adaptation of (23.13); that is,

$$L_w(\mathbf{X}, \mathbf{z}) = \frac{2\alpha}{\|\mathbf{M}\|^2} \sum_{i=1}^n \sum_{j=i+1}^n (m_{ij} - \|\mathbf{x}_i - \mathbf{x}_j\|)^2 + \frac{1-\alpha}{\|\mathbf{N}\|^2} \sum_{i=1}^n \sum_{j=1}^n \left(n_{ij} - \frac{(\mathbf{x}_i - \mathbf{x}_j)' \mathbf{z}}{\|\mathbf{x}_i - \mathbf{x}_j\|} \right)^2, \quad (23.14)$$

where $0 \leq \alpha \leq 1$ is a fixed weight that sets the relative importance of the two parts. Choosing $\alpha = 1$ fits only the symmetric part \mathbf{M} as in regular MDS. For $\alpha = 0$, only the skew-symmetric part \mathbf{N} is fitted. If $\alpha = .5$, then both parts are equally important in the solution. Note that the length of \mathbf{z} only reflects the amount of skew-symmetry that is captured from \mathbf{N} . For the interpretation, it is the direction of \mathbf{z} that shows how the difference vectors project on the hill slope \mathbf{z} .

We have fitted the hill-climbing model to the cola data. Our model cannot be estimated by a constrained form of MDS and a specialized algorithm had to be developed. Here we have used a general-purpose minimization

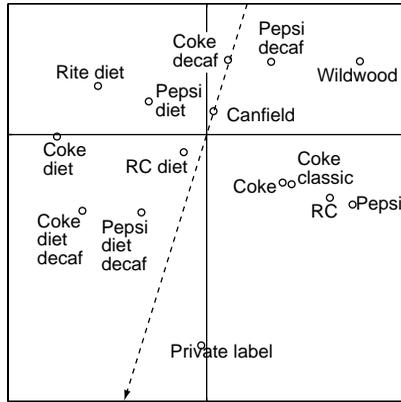


FIGURE 23.7. The hill-climbing model fitted to the brand switching data of Table 23.1 after being converted by the gravity model. The dotted line gives the direction of the slope up the hill.

function in MatLab to compute a solution. Figure 23.7 presents the results. The slope vector gives the uphill direction. The symmetric part of the data can be interpreted as usual. For example, there seems to be much switching between Coke, Coke classic, RC, and Pepsi because they are close together. To a lesser extent this also holds for the group of Coke decaf and Pepsi decaf, the group Coke diet, Pepsi diet and Rite diet, and the group of Coke diet decaf and Pepsi diet decaf.

To see how much skew-symmetry is present, the hill-climbing model predicts that changing from Private label to most other colas is easier than changing from those colas to Private label. The reason is that starting from Private label to any other cola it is downhill and the other way is uphill. We also see that there are several groups of colas whose difference vectors are almost orthogonal to the slope direction. Those groups lie at the same altitude on the hill and hardly display asymmetry in how often people change from one cola to the other or vice versa. One such group is Coke, Coke classic, RC, and Pepsi. Other groups that are mostly symmetric consist of Coke decaf and Pepsi decaf, a group with Pepsi diet and Rite diet, and a group of Coke diet decaf and Pepsi diet decaf. In a similar way, more relations could be deferred from the hill-climbing representation in Figure 23.7.

The hill-climbing model resembles to some extent the *jet-stream* model proposed by Gower (1977). This model uses the metaphor of flying times taking the jet-stream into account. Using our notation, the distances in the jet-stream model are defined by

$$d_{ij}(\mathbf{X}, \mathbf{z}) = \frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{1 + \frac{(\mathbf{x}_i - \mathbf{x}_j)' \mathbf{z}}{\|\mathbf{x}_i - \mathbf{x}_j\|}}. \tag{23.15}$$

The difference from the hill-climbing model is that the asymmetry factor appears in the jet-stream model in the denominator, whereas in the hill-climbing model it turns up as a separate term.

23.9 The Radius-Distance Model

Another model for fitting asymmetry directly is based on a representation of objects by circles with different radii (Okada & Imaizumi, 1987). The asymmetric dissimilarity δ_{ij} is modeled along the line connecting the centers i and j of two circles. The *radius-distance* from i to j is defined as the Euclidean distance d_{ij} between the centers of the circles, subtracting the starting radius of object i and adding the ending radius of object j . Thus, the radius-distance model can be fitted by minimizing

$$\begin{aligned} L(\mathbf{X}, \mathbf{r}) &= \sum_{i=1}^n \sum_{j=1}^n [\delta_{ij} - (d_{ij}(\mathbf{X}) - r_i + r_j)]^2 \\ &= \|\Delta - (\mathbf{D}(\mathbf{X}) - \mathbf{1}\mathbf{r}' + \mathbf{r}\mathbf{1}')\|^2, \end{aligned} \quad (23.16)$$

where $\mathbf{D}(\mathbf{X})$ has elements $d_{ij}(\mathbf{X})$ which refer to the usual Euclidean distance and \mathbf{r} is a vector containing nonnegative radii r_i . As the hill-climbing model, the radius distance model always fits the diagonal elements by 0 because for the symmetric part $d_{ii}(\mathbf{X}) = 0$ and for the skew-symmetric part $r_i - r_i = 0$.

An algorithm to minimize $L(\mathbf{X}, \mathbf{r})$ can be easily formulated by recognizing that the loss can be decomposed in a symmetric and skew-symmetric part (Bove & Critchley, 1993). This property means that $L(\mathbf{X}, \mathbf{r})$ may be written as

$$\begin{aligned} L(\mathbf{X}, \mathbf{r}) &= \|(\Delta + \Delta')/2 - \mathbf{D}(\mathbf{X})\|^2 \\ &\quad + \|(\Delta - \Delta')/2 - (\mathbf{r}\mathbf{1}' - \mathbf{1}\mathbf{r}')\|^2. \end{aligned} \quad (23.17)$$

Therefore, the symmetric part can be fitted by a regular MDS on $(\Delta + \Delta')/2$. The solution for the skew-symmetric part $\mathbf{N} = (\Delta - \Delta')/2$ requires a bit more care because of the nonnegativity constraints on the radii r_i . Some rewriting allows the skew-symmetric term of (23.17) to be expressed as

$$\|\mathbf{N} - (\mathbf{r}\mathbf{1}' - \mathbf{1}\mathbf{r}')\|^2 = \text{tr } \mathbf{N}'\mathbf{N} + 2n\mathbf{r}'\mathbf{J}\mathbf{r} - 4\mathbf{r}'\mathbf{J}\mathbf{N}\mathbf{1}, \quad (23.18)$$

where $\mathbf{N} = (\Delta - \Delta')/2$ and $\mathbf{J} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}'$ is the centering matrix. There is a simple analytic solution to (23.18). In Section 23.2, the unconstrained minimizer for (23.18) was given as $\mathbf{r}_u = n^{-1}\mathbf{J}\mathbf{N}\mathbf{1}$. Note that the centering matrix \mathbf{J} can be left out because $\mathbf{N}\mathbf{1}$ has column sum zero due to the skew-symmetry of \mathbf{N} . It is clear that \mathbf{r}_u does not satisfy the restriction that all

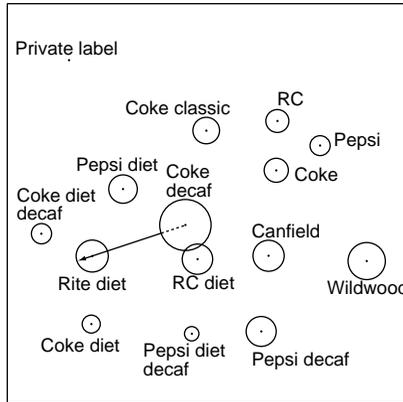


FIGURE 23.8. The radius-distance model of Okada and Imaizumi (1987) fitted to the brand switching data of Table 23.1 after being converted by the gravity model.

$r_i \geq 0$. However, any $\mathbf{r}_c = \mathbf{r}_u + c\mathbf{1}$ with a $c \geq \min_i r_i$ will be a feasible solution. The reason is that adding a constant does not change (23.18), because \mathbf{r} is premultiplied by \mathbf{J} and the nonnegativity constraints will be satisfied. Therefore, we choose $c = \min_i r_i$ so that the smallest radius equals zero.

Figure 23.8 shows the results of the distance radius model to these data. The symmetric relations can be easily interpreted by considering the centers of the circles. Large distances indicate little mutual switching whereas colas at close distance imply more mutual switching. The asymmetric part is taken care of by the differences in circle sizes. The fitted distance going from Coke decaf to Rite diet is indicated by the arrow in Figure 23.8. Going the other way around, from Rite diet to Coke decaf, the distance is computed from the border of the Rite diet circle to the far end of the Coke decaf circle. Because the circle of Coke decaf is larger than the circle of Rite diet, the distance Rite diet to Coke decaf is larger than the distance Coke decaf to Rite diet. This indicates that more households are changing from Coke decaf to Rite diet than vice versa. In a similar way, all the relations can be interpreted.

A nonmetric version of the radius-distance model was proposed by Okada and Imaizumi (1987) who included a gradient-based algorithm. However, the algorithmic approach outlined here can still be followed. First, replace the δ_{ij} by \hat{d}_{ij} as was done in Chapter 9 when going from metric to nonmetric MDS. Then alternately update one set of parameters while keeping the other fixed. Thus, given \mathbf{X} and \mathbf{r} , update $\hat{\mathbf{d}}$. Normalize $\hat{\mathbf{d}}'\hat{\mathbf{d}}$ to n^2 so that the trivial solution of $\hat{\mathbf{d}} = \mathbf{0}$, $\mathbf{X} = \mathbf{0}$, and $\mathbf{r} = \mathbf{0}$ is avoided. Next, update \mathbf{X} and \mathbf{r} given $\hat{\mathbf{d}}$. The update of \mathbf{X} is given by (8.29) where $(\hat{\mathbf{D}} + \hat{\mathbf{D}}')/2$ should be used instead of the dissimilarities. The update of \mathbf{r} can be obtained by \mathbf{r}_c discussed above. Remain iterating until convergence is obtained.

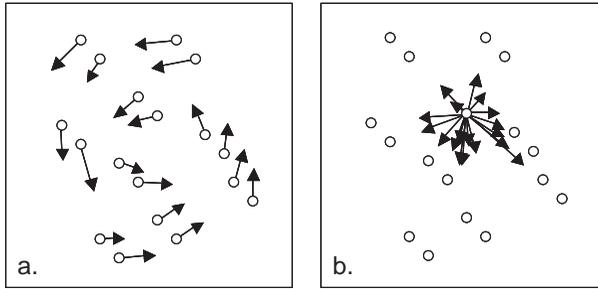


FIGURE 23.9. A circular vector field (panel a) and a configuration with a dominant point that “feeds” into all other points (panel b).

Okada (1990) discusses how the radius-distance model can be extended to ellipsoids instead of circles. This relaxation has the advantage that for each dimension and each object the radius can be different thereby allowing to estimate the skew-symmetric component in a better way. Okada (1990) also presents a gradient based algorithm.

23.10 Using Asymmetry Models

The MDS user should be aware of the simple but important fact that any asymmetry model is, as all models are, designed for a particular purpose only. Each such model represents one particular form of asymmetry only. It helps to detect only those patterns in the data on which it focuses. If the chosen model does not show this particular form of asymmetry for the given data, it does not imply that there are no other systematic asymmetries in the data. Consider, for example, the slide-vector model. It is made to show to what extent the data contain a general asymmetry in one direction of the space. As we can see in Figure 23.6, this form of asymmetry can be rather small. However, the data may contain other interesting asymmetries. Figure 23.9 shows two particular forms of asymmetries that would go unnoticed by the slide-vector model: the circular vector field in panel A and the configuration with one “dominant” element in panel B. (Neither of these cases is inconceivable for real data.) The circular field would be detected by the drift-vector model, for example, but the case in panel B would not lead to a (resultant) drift vector that adequately describes the asymmetries. Rather, in this case, it would be more revealing to show all the drift vectors attached to this one particular point, provided that the asymmetries can be considered big enough (relative to the symmetric part of the data) and, of course, reliable enough to warrant further studies.

When analyzing asymmetries, the user should experiment with different models to avoid missing systematic patterns that exist in the data. Models that identify the extent of one global and linear trend, in particular,

should be complemented with models that represent the more fine-grained asymmetries. To use such a hierarchy of models also allows one to assess whether it is worth it, relative to the quality of the data, to pursue models with many free fitting parameters.

23.11 Overview

In this section, we give a summary of the models for asymmetry and skew-symmetry discussed in this chapter. The reader should know that the discussion in this chapter is not exhaustive. Other models for asymmetry exist in the literature, some aimed at specific applications. In this chapter, we have restricted ourselves to mostly distance-based models for asymmetry or skew-symmetry.

Many of the models for asymmetry can be decomposed into a symmetric part and a skew-symmetric part. Some models only estimate the skew-symmetric part. Others fit the asymmetric data directly. Table 23.3 gives an overview of the models discussed in this chapter. Analyzing the skew-symmetric component separately from the symmetric part has the advantage that for the interpretation one only cares about the skew-symmetry. On the other hand, it may be useful to see the symmetric and skew-symmetric relations simultaneously. An important issue in deciding for an asymmetry model is the way of representing the asymmetry and how easy it is to interpret it. The latter remains a subjective matter.

Throughout this chapter, we only discussed the analysis of two-way asymmetric data. For three-way data, several models have been proposed in the literature. For example, Okada and Imaizumi (1997) extend the radius-distance model to the case of replications of two-way asymmetry data. De Rooij and Heiser (2000) extend distance measures to deal with the case of one-mode three-way asymmetric data.

23.12 Exercises

Exercise 23.1 Consider the data in the table below. They represent preceding–following contingencies for certain types of threat display behaviors shown by a common bird, the great tit (Blurton Jones, 1968). The numbers correspond to the proportion of times that the behavior in column j followed the behavior in row j . For example, feeding follows fluffing 3% of the time. Spence (1978) argues that these data are “a measure of how ‘close’ behavior j is to behavior i ” and uses MDS to “visually detect” possible groupings of behaviors. The asymmetry of the data is noticed by Spence, but not studied.

TABLE 23.3. Summary of the properties of the models for asymmetric data discussed in this chapter. A + or a - in the columns **P**, **M**, and **N** indicate whether the model fits asymmetric proximities directly (column **P**), the symmetric part separately (column **M**), and the skew-symmetric part separately (column **N**).

Section	Model	P	M	N	Graphical Representation
23.3	Signed-distance model	-	-	+	Signed distances between points on line
23.3	Gower decomposition	-	-	+	Areas between vectors plus orientation
23.4	Distance model for skew-symmetry	-	-	+	Distance between points plus orientation
23.5	Scaling the skew-symmetry	-	+	+	Symmetry by distance between points, skew-symmetry by a summary vector
23.6	Unfolding	+	-	-	Distances between row and column objects
23.7	Slide-vector model	+	-	-	As unfolding, but row and column points are equal up to a translation
23.8	Hill-climbing model	+	-	-	Symmetry by distance between points, skew-symmetry modeled by projection of difference vector onto the slope direction.
23.9	Radius-distance model	-	+	+	Distance between two points with the radius from the starting circle removed and the radius of the arriving circle added

	Type of Behavior	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Attack	4	17	16	11	10	13	11	0	6	0	0	9	4
2	Head down	26	0	5	14	4	13	2	8	5	0	0	5	18
3	Horizontal	25	3	0	12	13	11	3	2	10	8	0	4	9
4	Head up	5	9	8	8	14	15	5	4	13	0	2	5	12
5	Wings out	22	13	10	5	2	10	2	7	7	0	0	2	19
6	Feeding	2	5	18	13	11	3	3	5	13	8	1	16	1
7	Incomplete feeding	4	10	15	4	4	13	7	22	0	0	12	8	0
8	Hopping around	1	10	0	4	2	4	46	0	3	6	11	11	3
9	Hopping away	0	4	6	9	5	1	8	4	1	6	31	15	10
10	Crest raising	0	0	0	6	7	3	0	11	17	1	30	13	12
11	Fluffing	0	4	5	6	3	3	0	23	13	35	0	6	3
12	Looking around	5	0	5	0	3	6	12	12	11	30	8	0	9
13	Hopping towards	5	25	12	8	21	4	2	2	2	7	5	6	0

- (a) Assess, by matrix decomposition, just how asymmetric these data are.
- (b) Use the symmetric portion of the data for a two-dimensional MDS analysis. Then, add the skew-symmetric portion as vectors to a few behaviors that are strongly asymmetric, and to a few others that are only mildly asymmetric.
- (c) How would you interpret the symmetric portion of these data? (Hint: Blurton Jones speculates that behaviors within each “group” may have certain causal factors in common.)
- (d) Scale these data by the slide vector model, using PROXSCAL.

Exercise 23.2 Consider the data matrix below (Coombs, 1964). It shows the frequencies with which an article that appeared in the journal shown as a row entry cites an article in the column journal.

Journal	AJP	JASP	JAP	JCPP	JCP	JEdP	JexP	Pka
Am. J. Psy.	119	8	4	21	0	1	85	2
J. Abnorm. Soc. Psy.	32	510	16	11	73	9	119	4
J. Applied Psy.	2	8	84	1	7	8	16	10
J. Comp. Physiol. Psy.	35	8	0	533	0	1	126	1
J. Consulting Psy.	6	116	11	1	225	7	12	7
J. Educ. Psy.	4	9	7	0	3	52	27	5
J. Exp. Psy.	125	19	6	70	0	0	586	15
Psychometrika	2	5	5	0	13	2	13	58

To study the interaction behavior of these journals, we may follow Coombs, Dawes, and Tversky (1970) by first subtracting the column and the row means from the matrix entries. This leaves pure interaction values. Then, proceed as follows.

- (a) Split the matrix of interaction values into its symmetric and skew-symmetric component.

- (b) Scale the symmetric part via MDS. Interpret the solution.
- (c) Attach drift vectors to the points of the MDS configuration by hand or by using an appropriate graphics package (see, e.g., Borg & Groenen, 1995).
- (d) How do you interpret these drift vectors?

Exercise 23.3 Consider Table 23.2 on p. 505 with the asymmetric dissimilarities obtained from the brand switching between 15 colas.

- (a) Compute the skew-symmetric matrix of this table.
- (b) Compute the unidimensional skew-symmetry model (23.5). Plot the results on a line. How do you interpret this solution?
- (c) Apply Gower's decomposition to these data. Plot the first dimension. Interpret the solution.
- (d) Which model do you expect to recover the skew-symmetry the best? Why do you think so?
- (e) Compute for both models how much of the sum of squared skew-symmetry is recovered by the unidimensional skew-symmetry model and by Gower's decomposition using the first dimension. Does your computation coincide with your expectations?
- (f) Do the two models differ in their interpretation? If so, how?