15

Avoiding Trivial Solutions in Unfolding

The occurrence of trivial solutions in unfolding was recognized soon after the introduction of MDS. It was one of the reasons for introducing Stress-2. However, as indicated in the previous chapter, Stress-2 does not solve the degeneracy problem totally. In this section, we discuss several methods that have been proposed in the literature to avoid trivial unfolding representations. They all adapt the unfolding procedure in such a way that the ideal point interpretation is retained. The solutions can be categorized into three classes: (a) adapting the unfolding data, (b) adjusting the transformation, and (c) modifying the loss function.

15.1 Adjusting the Unfolding Data

One way to avoid a trivial solution in unfolding is to make sure that the transformation cannot contain a nonzero intercept and a slope of zero by adapting the data. Here, we discuss two of these options. The first one, is an *ordinal-ratio* approach to unfolding.

Ordinal-Ratio Approach

The idea behind this approach is to use an ordinal and a ratio transformation simultaneously on the same data. Thus, the data are duplicated, one data set is transformed by a ratio transformation, the other one by an

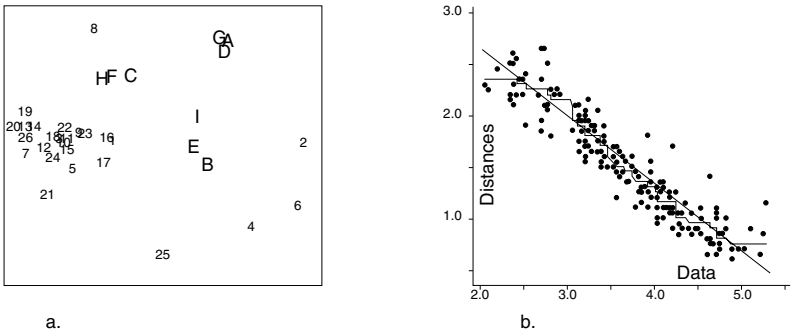


FIGURE 15.1. Unfolding representation (a) of data in Table 14.2, using mixed ordinal-linear σ_2 loss function, and its Shepard diagram (b).

ordinal transformation, and both sets of disparities are approximated by a single matrix of distances.

Let $L(o)$ be the loss function that defines an ordinal approach and $L(r)$ the corresponding loss function for ratio unfolding. For example, $L(o)$ may be σ_2 with disparities as target distances under, say, the primary approach to ties, and $L(r)$ is σ_2 with target distances computed by a ratio transformation. Then, we simply define the total loss as

$$L = a \cdot L(o) + b \cdot L(r), \tag{15.1}$$

where a and b are weights such that $a, b > 0$ and $a + b = 1$. L is equal to 0 only if both $L(o)$ and $L(r)$ are equal to 0. Note that if $L(r)$ is zero, $L(o)$ will also be zero because the ratio transformation is an admissible ordinal transformation. L will be small if both $L(o)$ and $L(r)$ are small, or if one is very small and the other is not very large. The ordinal transformation tries to model the data as usual in an ordinal manner. However, as a ratio transformation does not allow for an intercept, the trivial transformation with a nonzero intercept and zero slope cannot occur.

One drawback of this approach is that one needs to have dissimilarities in order to do a ratio transformation. For similarities, a ratio transformation with a negative slope is required to map larger similarities into smaller distances. Yet, such a transformation leads to negative disparities, which can never be properly modeled by nonnegative distances. Therefore, if we have similarities, we either have to convert similarities into dissimilarities before the unfolding analysis or revert to an *ordinal-interval* approach with $L = a \cdot L(o) + b \cdot L(i)$ and $L(i)$ the loss for unfolding of data that are interval-scaled. This ordinal-interval approach is not guaranteed to always avoid the trivial solution but the example discussed below shows a successful application.

Let us apply this approach to the brewery data, using KYST with weights $a = b = 0.5$ and loss function $L = a \cdot L(o) + b \cdot L(i)$. This yields a solution

with the Shepard diagram in Figure 15.1b. There are two regression curves now: a monotonic one, related to $L(o)$, and a linear one, related to $L(i)$. The (vertical) scatter of the points about the monotonic curve makes up one component of L, and the scatter of these same points about the linear regression line makes up the other. Hence, minimizing L tends to avoid a solution with a crude step function in the Shepard diagram, because this would make $L(i)$ large. On the other hand, the regression slope must have the desired sense to make $L(o)$ small.

The configuration resulting from this mixed ordinal-linear unfolding is presented in Figure 15.1a. It allows the usual ideal-point interpretation, but differs radically from the previous interval representation in Figure 14.16a. We now observe, for example, that brewery A is very far from the attribute point 21, which, as can be seen from studying the proximities, has the usual meaning that A possesses relatively little of this property. On the other hand, we again find that the breweries form three groups, because this closeness relation remains unaffected by the slope of the regression line.

It should be noted that, even though the loss criteria $L = a \cdot L(o) + b \cdot L(i)$ and $L = b \cdot L(i) + a \cdot L(o)$ are algebraically equivalent, they may lead to different results in an iterative optimization procedure. If the KYST program is used, for example, we find that if $L(o)$ appears as the first criterion in the weighted sum, then a solution like the one reported above is obtained; if $L(i)$ is the first criterion, then the approach does not work as desired. In other words, a solution with a Shepard diagram like Figure 14.14 results, where the monotone regression curve is a horizontal straight line. In general, such differences can result from various features of the optimization method.

Augmenting the Within-Objects Blocks

A second way to avoid a trivial solution in unfolding by "changing the data," builds on the idea that unfolding is equivalent to MDS with missing data as visualized in Figure 14.1. The main idea here is to augment the data matrix with one or both of the missing "within"-sets data. Steverink, Van der Kloot, and Heiser (2002) proposed to insert Kemeny distances for the within-individuals data. In addition, they allow for different transformations within the blocks of the data matrix. The choice of transformation is critical: it must exclude the possibility of zero within-sets disparities and constant between-sets disparities to avoid the trivial equal-distances solution. For example, ordinal transformations for the between-sets proximities should be combined with the absolute transformation for the within-persons proximities to guarantee avoiding the trivial solution.

Kemeny distances for the within-persons data appear particularly suitable for unfolding preferential choice data. They are derived from preference rankings as follows. First, each person i gets a score for each pair of items

TABLE 15.1. Illustration of computing the Kemeny distance of two persons.

Pair k,l	$z_1(k,l)$	$z_2(k,l)$	$\lvert z_i(k,l) - z_j(k,l)\rvert$
AB	1	1	0
AC	1	-1	2
BC	1	-1	2
		Sum	4

TABLE 15.2. Illustrative example of an unfolding data matrix where the within-persons data are Kemeny distances. Note that the between-sets data values are the preference orders of the persons for the objects.

	A	B	C	1	2
A	–	–	–	3	2
B	–	–	–	2	3
C	–	–	–	1	1
1	3	2	1	0	4
2	2	1	3	4	0

k and l on the function $z_i(k,l)$:

$$\begin{array}{rll} 1 & \text{if } A > B & (\text{person } i \text{ prefers } A \text{ over } B), \\ 0 & \text{if } A = B & (\text{person } i \text{ is indifferent to } A \text{ and } B), \\ -1 & \text{if } A < B & (\text{person } i \text{ prefers } B \text{ over } A). \end{array}$$

Second, these z_i-scores are aggregated over all pairs to yield the Kemeny distance between persons i and j:

$$d_{\text{Kem}}(i,j) = \sum_{k<l} |z_i(k,l) - z_j(k,l)|. \tag{15.2}$$

As an illustration, consider a situation where three objects A, B, and C are judged by two persons: the preference rank-order of person 1 is $A > B > C$, and $C > A > B$ of person 2. Then, there are only three possible pairs of objects, that is, AB, AC, and BC. Table 15.1 shows the steps taken to compute their Kemeny distance, which equals 4 in this case. For this mini example, the data matrix augmented by within-persons distances is presented in Table 15.2.

The augmentation approach described above was applied to the brewery data, where the between-sets similarities were transformed ordinally (and unconditionally) and Kemeny distances were computed among the 26 attributes. The results are presented in Figure 15.2. As predicted, the trivial solution with equal distances does not occur. The breweries are located in three clusters in the center, a solution that is similar to that of the ordinal-interval approach in Figure 15.1. The right panel of Figure 15.2 shows the

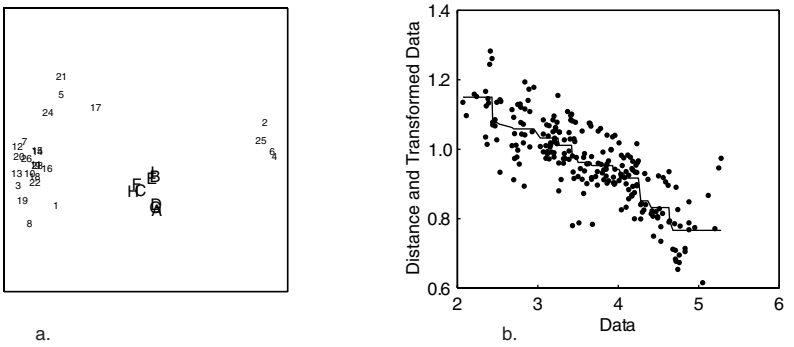


FIGURE 15.2. Unfolding representation (a) of data in Table 14.2, using the augmentation approach, and its Shepard diagram (b) for the between-sets data.

Shepard diagram of the between-sets data. The ordinal transformation is quite reasonable and does not have big jumps. However, the scatter of the distances about the regression curve is somewhat high, indicating that not all points are fitted perfectly. We also see that the disparities range from about .75 to 1.15. This means that even a brewery with the highest score on an attribute will be located at a moderate distance from the attribute. This aspect is shown in the left panel of Figure 15.2 by the fact that all attributes are distant from the center where the breweries are located.

A problem arises when the between-blocks data are transformed row-conditionally, which is a natural option for preference rank-order data. Applying the augmentation approach will yield a proper scatter of the attributes and a cluster of brewery points on top of each other. The within-block data for the attributes are properly represented, but the between-sets data (the original preference rank-orders) are trivially represented in the same way as the degeneracy in Figure 14.7b. Steverink et al. (2002) proposed to solve this problem by augmenting the data matrix with a within-columns data block as well.

Here, we propose a different type of augmentation by a within-columns data block. As the preference rank-orders are known for each subject, one can compute city-block distances between the columns using the rank-orders as coordinates. Thus, for each row, a unidimensional distance matrix is computed between the columns. Then, taking the sum of all those distance matrices over the rows gives a city-block distance matrix between the column objects. We take two additional steps. First, Steverink et al. (2002) indicate that the Kemeny distance can also be seen as a city-block distance matrix. Because we are fitting these data by Euclidean distances, we transform both within-blocks to Euclidean distances matrices by simply taking the square root of all elements (for a rationale, see Gower & Legendre, 1986). The second step involves making the range of the values in the two within-blocks equal. This adaptation is important because the

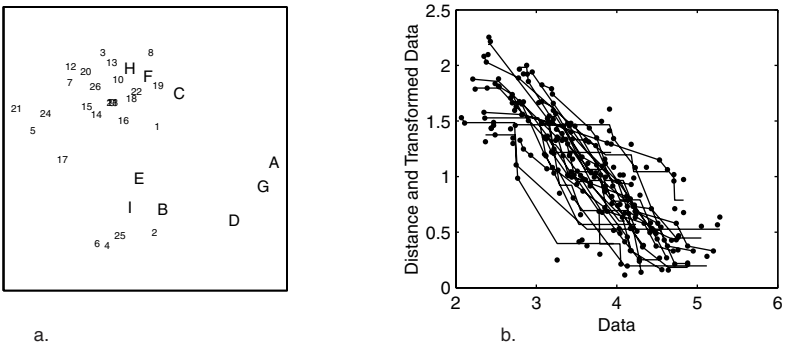


FIGURE 15.3. Ordinal row-conditional unfolding representation (a) of data in Table 14.2, augmentation both within blocks, and its Shepard diagram (b) for the between-sets data.

within-blocks are not transformed and it makes sure that the ranges of distances for the two sets of points are equivalent. Therefore, we divide each within-block by its maximum value.

The proposed procedure of augmenting both within-blocks is applied to the brewery data in Figure 15.3. In the analysis, the within-blocks data were not transformed but the between-sets data were obtained by an ordinal row-conditional transformation. The Shepard diagram in the right panel of Figure 15.3 shows that the transformations are far from constant. A similar pattern as before emerges for the configuration (Figure 15.3a), with three clusters of breweries. Note that in this analysis we may explicitly interpret the distances between all points and not only the between-sets distances because we have (generated) data for all dissimilarities.

A disadvantage of the augmentation approach proposed above is that it may be seen as doing two separate metric MDS analyses on the within-blocks data. The between-sets data are of minor importance and merely determine the translation of one of the sets with respect to the other. On the other hand, all three blocks of the data use the same rank-order information of the between-sets data. More experience with this approach is needed to see how well it performs in practice.

Other suggestions to fill the within-blocks data have been proposed by Rabinowitz (1976), Heiser and De Leeuw (1979), and Van Deun, Heiser, and Delbeke (2004).

15.2 Adjusting the Transformation

A different way to avoid the trivial equal-distances unfolding solution is to restrict the transformation so that the nonzero intercept and zero slope transformation is excluded. This goal could be either achieved by a bound

on the intercept or some restriction on the slope that excludes a slope of zero. One obvious transformation satisfying this restriction is the ratio transformation. Clearly, no intercept is estimated and the slope is equal to one, so that the zero slope and nonzero intercept cannot occur. It is a simple manner to avoid the trivial solution in unfolding, but it may not recognize the ordinal nature of data that are often used in unfolding, such as preference rank-orders.

A variant of this idea was proposed by Kim, Rangaswamy, and DeSarbo (1999), who use a two-step procedure. In their first step, they preprocess the original dissimilarities by a transformation (λ_{ij}) of the original data. These λ_{ij}s satisfy several properties. First, λ_{ij} should be strictly monotone with the dissimilarities such as, for example, a linear or a strictly ordinal transformation. Second, λ_{ij} for the most preferred item in each row is set to zero. Third, the transformations are the same for all rows. The form of the transformation is left to the user, as long as it satisfies the three conditions stated above. In the second step, after this preprocessing of the data, the λ_{ij}s are used as input data in Stress, allowing for row-conditional ratio transformations.

The reason why this approach avoids the trivial solution is that the d-hats of each row cannot become the same constant, as the d-hat corresponding to the most preferred stimulus per row is equal to zero and the remaining d-hats per row necessarily are nonzero because of the ratio transformation. It should be said, though, that this method has some arbitrariness in the way the user specifies the λ_{ij}s. Different specifications of the λ_{ij}s for the same data will lead to different solutions.

For preference rank-orders, it is more preferable to apply a transformation that has more freedom than the ratio transformation but is still able to avoid the nonzero constant and zero slope transformation. The smoothed monotone regression approach of Heiser (1985, 1989a) can do this (see also Section 9.2). The basic idea is that the absolute difference of the differences $\hat{d}_k - \hat{d}_{k-1}$ and $\hat{d}_{k-1} - \hat{d}_{k-2}$ in the transformation should be smaller than the average d-hat. Note that for $k = 1$ and $k = 2$, there will be references to nonexisting elements $\hat{d}_0$ and $\hat{d}_{-1}$ that are substituted by zero. The important consequence of this substitution is that the smallest d-hat cannot be larger than the average d-hat. Thus, this approach has an internal upper bound on the smallest d-hat, while restricting the size of the steps that can be made in the transformation. These restrictions combined with the requirement that the sum of squared d-hats are equal to some nonzero constant assure that the constant d-hat solution is excluded so that the trivial solution cannot occur.

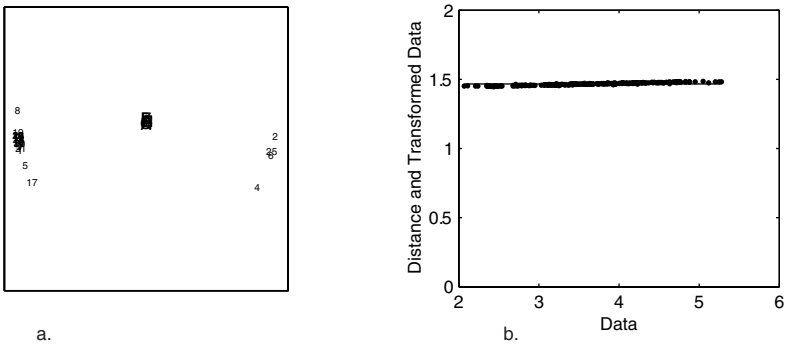


FIGURE 15.4. A degenerated linear unfolding solution obtained by KYST (Panel a) of the brewery data in Table 14.2. Panel b displays the Shepard diagram).

15.3 Adjustments to the Loss Function

Several authors have tried to avoid the trivial unfolding solution by adjusting the loss function. The first proposal was to use Stress-2; that is,

$$\sigma_2(\mathbf{X}) = \left(\frac{\sum_{i<j}[\delta_{ij} - d_{ij}(\mathbf{X})]^2}{\sum_{i<j}[d_{ij}(\mathbf{X}) - \bar{d}]^2} \right)^{1/2}.$$

The denominator of $\sigma_2(\mathbf{X})$ measures the variance of the distances about the mean distance. Therefore, the denominator will be close to zero if all distances are almost the same. This implies that if the distances become similar during the iterations, $\sigma_2(\mathbf{X})$ becomes larger and larger. Hence, equal-distances solutions should be avoided.

Ordinal unfolding by the KYST program using Stress-2 resulted in a configuration with three clusters of breweries and attributes located at two different distances (see Figure 14.13). Although the ordinal solution may not be totally satisfactory, it certainly does not display the equal-distances solution. However, a linear transformation with KYST (with strict convergence settings) does yield a constant distance solution (see Figure 15.4). To understand why this happens, we need to consider both the numerator and denominator of Stress-2 as the distances become almost equal. In that case, both the denominator and the numerator approach zero, so that no immediate conclusions can be drawn about the behavior of Stress-2. Mathematical analysis should bring more insight into this situation. We get back to this issue in the next chapter. For now it suffices to remark that apparently linear unfolding using Stress-2 does not avoid equal distances. Stress-2 may stay away from the trivial equal distance solution but it is not guaranteed to do so.

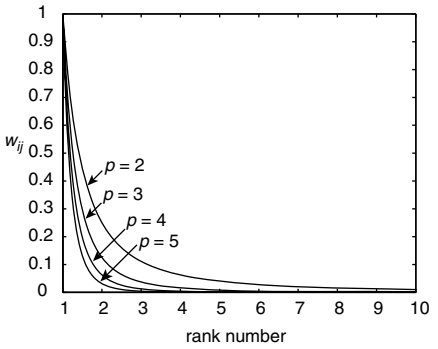


FIGURE 15.5. Influence of a weight w_{ij} as a function of the rank-order as proposed by DeSarbo and Rao (1984).

Weighting Strategies

DeSarbo and Rao (1984) proposed to use specialized weighting schemes in raw Stress to avoid the trivial solution: $w_{ij} = \delta_{ij}^{-p}$, where δ_{ij} is a ranking number (1 = most preferred) and $p > 0$ determines the influence of the object to the Stress function. Figure 15.5 shows how much an object contributes to Stress as a function of its ranking number for $p = 2, \ldots, 5$. In Figure 15.5, we see that even for $p = 2$, the residuals of the second most preferred object are weighted by only 25% compared to the most preferred object. An object with ranking number 3 is weighted by about 11%, and so on. Thus, even with a small p of 2, only the three most preferred objects of each individual (row) determine the solution. In the case of $p = 5$, the second most preferred stimulus contributes only 3%. This means that for each row there is essentially only a single stimulus that contributes to Stress. As a consequence, there will be only very few effective constraints between the points of both sets so that the points can be quite freely located in space. For this approach, the quality of the solution mainly depends on the quality of the starting configuration. A similar condition is true for the transformation. Only the transformations of the first few most preferred stimuli can be interpreted; the others hardly contribute to Stress.

Contrary to its claim, the weighting method does not exclude the trivial solution. The reason is that for any transformation that allows for a zero slope and constant intercept, distances can be obtained that are equal to the intercept. For a formal proof, we refer to Busing, Groenen, and Heiser (2005).

Penalizing the Intercept

For linear unfolding, Busing (2005) proposed a simple idea to avoid a transformation with a nonzero intercept and zero slope. His idea is to add a penalty to the Stress function to avoid a large value of the intercept. This

idea can be formalized by the following loss function.

$$\sigma_i(a, b, \mathbf{X}) = \sum_{(i,j)} [a + b\delta_{ij} - d_{ij}(\mathbf{X})]^2 + \omega a^2, \tag{15.3}$$

where ω is a nonnegative value indicating the strength of the penalty. Clearly, for $\omega = 0$ the old Stress function is retained. In the limiting case of $\omega = \infty$, the intercept a will become zero and minimizing $\sigma_i(a, b, \mathbf{X})$ reduces to ratio unfolding.

Penalizing the intercept only makes sense for dissimilarity data. If the data are similarities, we expect a transformation with a large intercept and a negative slope, so that large similarities correspond to small nonnegative d-hats and small distances, whereas small similarities correspond to large nonnegative d-hats and large distances. This means that the intercept is expected to be large, which contradicts the idea of penalizing the intercept. To overcome this problem, the similarities have to be transformed into dissimilarities before applying the current approach. As a consequence, the Shepard diagram will be increasing because dissimilarities are used in (15.3).

Penalizing the intercept is not applicable to just any transformation. For example, the approach is not effective for ordinal unfolding, because the transformation is free to find a constant transformation for all but the smallest dissimilarity. However, penalizing the intercept can be effective for spline transformations (see Section 9.6), provided that the spline is quite restricted.

We applied this approach to the brewery data of Table 14.2. Because the data (p_{ij}) in this case are similarity ratings from 1 = not true to 6 = very true, they had to be transformed into dissimilarities first. This was done by setting $\delta_{ij} = 7 - p_{ij}$ so that the dissimilarities were again in the range from 1 to 6, where 1 now indicates "very true" and 6 "not true". In this application, ω was set to 5 after some experimentation. The results are presented in Figure 15.6. Again we see the split of the breweries into the three different clusters that turn up in the other solutions as well. The Shepard diagram in panel b of Figure 15.6 is increasing indeed and has an intercept that is reasonably small compared to the slope. We may conclude that for linear unfolding, the simple approach of penalizing the intercept is effective to avoid the trivial solution.

PREFSCAL: *Penalizing Equal d-hats*

Another penalty approach was taken by Busing et al. (2005). As trivial unfolding solutions are characterized by constant d-hats, one obvious way to avoid a trivial solution is penalizing the Stress function for equal d-hats. An advantage of this approach is that all standard transformations (see Chapter 9) can be applied. Also, the resulting unfolding configuration can be interpreted in terms of the ideal point model.

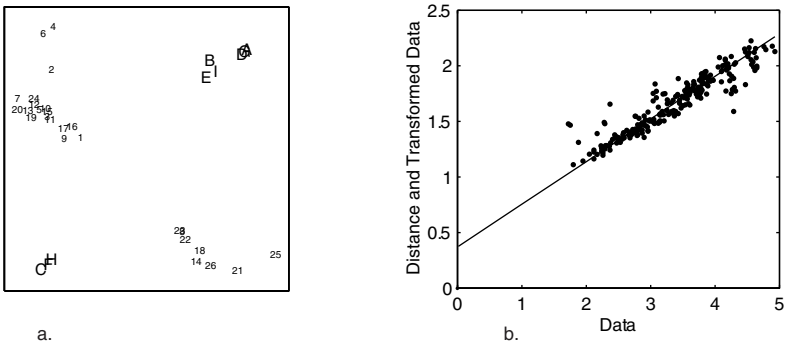


FIGURE 15.6. Linear unfolding representation (a) of data in Table 14.2, by penalizing the size of the intercept, and its Shepard diagram (b).

To identify constant d-hats, Busing et al. (2005) suggests using the variation coefficient of Pearson (1896), which is defined by

$$\nu(\hat{\mathbf{d}}) = \frac{\text{standard deviation}(\hat{\mathbf{d}})}{\text{mean}(\hat{\mathbf{d}})} = \frac{\left(K^{-1}\sum_k (\hat{d}_k - \bar{\hat{d}})^2\right)^{1/2}}{K^{-1}\sum_k \hat{d}_k}, \qquad (15.4)$$

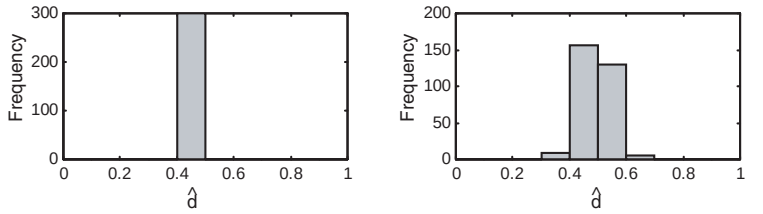
where $\bar{\hat{d}} = K^{-1}\sum_k \hat{d}_k$ and k is an index that runs over all d-hats. The coefficient of variation is a measure that indicates the spread with respect to the mean. It can be derived that $\nu(\hat{\mathbf{d}})$ is independent of the scale of $\hat{\mathbf{d}}$, so that $\nu(\hat{\mathbf{d}}) = \nu(a\hat{\mathbf{d}})$ for any $a > 0$.

To see what the variation coefficient does, we simulated four different distributions of 300 d-hats, varying the mean, the standard deviation, and the modality. Both from Figure 15.7a and from (15.4) it can be seen that a zero standard deviation yields a zero variation coefficient. If the spread around the mean is small relative to the mean, then $\nu(\hat{\mathbf{d}})$ is also small (panel b. of Figure 15.7). As the spread around the mean gets larger relative to the mean, then $\nu(\hat{\mathbf{d}})$ also increases (Figures 15.7c and 15.7d). A maximum value of $\nu(\hat{\mathbf{d}}) = (K-1)^{1/2}$ is attained if all but one of the d-hats are zero.

The variation coefficient can be used as a diagnostic for identifying solutions with constant d-hats. The PREFSCAL model proposed by Busing et al. (2005) exploits this diagnostic by using it as a penalty. To be more precise, their PREFSCAL model minimizes penalized Stress that is defined as

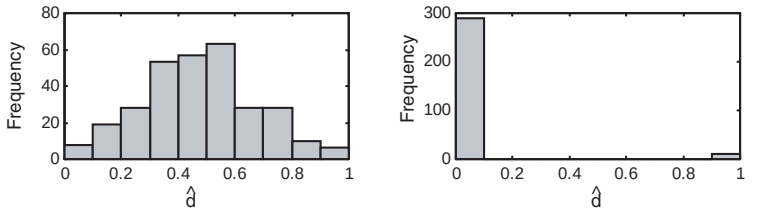
$$\sigma_p(\hat{\mathbf{d}}, \mathbf{X}) = \sigma_n^{\lambda}(\hat{\mathbf{d}}, \mathbf{X}) \left(1 + \frac{\omega}{\nu^2(\hat{\mathbf{d}})}\right), \qquad (15.5)$$

where $\sigma_n(\hat{\mathbf{d}}, \mathbf{X})$ is normalized Stress defined by (11.1) and λ and ω are two penalty parameters to be specified under the restrictions $0 < \lambda < 1$ and $\omega > 0$. The parameter λ is called a lack-of-penalty parameter that



a. std $\ll$ mean
std = 0, mean = .50, $\nu = 0$

b. std $<$ mean
std = .05, mean = .49, $\nu = .10$



c. std $\approx$ mean
std = .20, mean = .48, $\nu = .41$

d. std $\gg$ mean
std = .18, mean = .03, $\nu = 5.39$

FIGURE 15.7. Value of the variation coefficient ν as a function of the mean and the standard deviation of four hypothetical distributions of 300 $\hat{d}$s.

influences the balance between the penalty $1 + \omega\nu^{-2}(\hat{\mathbf{d}})$ and $\sigma_n(\hat{\mathbf{d}}, \mathbf{X})$: the closer λ gets to zero, the stronger the penalty. The parameter ω determines when the penalty gets active: for small ω, say, $\omega = .1$, the $\sigma_p(\hat{\mathbf{d}}, \mathbf{X})$ will hardly be influenced by the penalty for d-hats as in Figure 15.7b, whereas a large ω, say, $\omega = 5$, ensures strong influence of the penalty for the same d-hats. Based on extensive simulations, Busing et al. (2005) recommend choosing $\lambda = .5$ and a value of $\omega = .5$, although ω may need some fine tuning depending on the data.

The penalty term in (15.5) obtains high values whenever almost equal d-hats occur (thus when $\nu^2(\hat{\mathbf{d}})$ is close to zero), because the inverse of the squared variation coefficient, $\nu^{-2}(\hat{\mathbf{d}})$, will become large. Thus, when minimizing $\sigma_p(\hat{\mathbf{d}}, \mathbf{X})$, the algorithm will stay away from constant d-hats, because $\sigma_p(\hat{\mathbf{d}}, \mathbf{X})$ has high values for those d-hats. Penalized Stress has the additional advantage that as we move away from the trivial solution, the penalty term becomes less influential and $\sigma_n(\hat{\mathbf{d}}, \mathbf{X})$ will dominate the minimization. The cause of this property lies in the sum of one plus $\omega\nu^{-2}(\hat{\mathbf{d}})$. Thus, whenever $\nu(\hat{\mathbf{d}})$ is large, $\nu^{-2}(\hat{\mathbf{d}})$ gets close to zero, so that the entire penalty term is close to one. Then, the minimization of $\sigma_n(\hat{\mathbf{d}}, \mathbf{X})$ is the most important part and the penalty term will hardly influence the minimization. An additional advantage of the definition of penalized Stress is that $\sigma_p(\hat{\mathbf{d}}, \mathbf{X}) = 0$ for perfect nontrivial solutions (i.e., solutions with zero

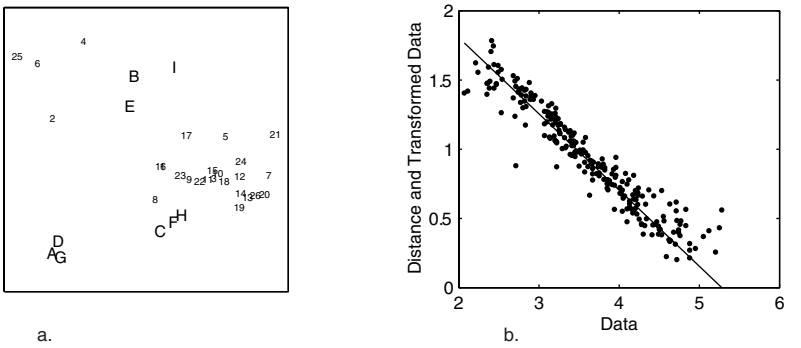


FIGURE 15.8. Linear unfolding representation (a) of data in Table 14.2 obtained by PREFSCAL, and its Shepard diagram (b).

normalized Stress). Thus, if a perfect nontrivial solution exists, penalized Stress should be able to find it. Another property of $\sigma_p(\hat{\mathbf{d}}, \mathbf{X})$ is that the minimum of $\sigma_p(\hat{\mathbf{d}}, \mathbf{X})$ is independent of the scale of $\mathbf{X}$ or $\hat{\mathbf{d}}$, by which we mean that multiplying both $\mathbf{X}$ and $\hat{\mathbf{d}}$ by a positive constant a does not change the value of σ_p. Without the property of scale independence, penalized Stress would be sensitive to the size of the unfolding problem. Thus, the PREFSCAL penalty parameters λ and ω are independent of the number of row and column objects and of the normalization of the d-hats.

Figure 15.8 displays the results of a PREFSCAL analysis on the brewery data. The PREFSCAL solution is quite similar to Figure 15.1 obtained by the ordinal-interval approach. Again, the three clusters with three breweries each emerge. However, there are some differences in the positioning of the attributes. For example, in Figure 15.1 attribute 8 is located outside the triangle spanned by the three clusters, whereas PREFSCAL locates it inside the triangle.

For row conditional transformations, constant d-hats should be avoided for each row. Therefore, the penalty should be large whenever the d-hats of a single row become constant. PREFSCAL achieves this objective by defining row conditional penalized Stress as

$$\sigma_{p.rc}(\hat{\mathbf{d}}, \mathbf{X}) = \sigma_n^{\lambda}(\hat{\mathbf{d}}, \mathbf{X}) n_2^{-1} \sum_{i=1}^{n_2} \left(1 + \frac{\omega}{\nu^2(\hat{\mathbf{d}}_i)} \right), \tag{15.6}$$

where n_2 is the number of rows in the unfolding problem and $\hat{\mathbf{d}}_i$ contains the d-hats for row i (Busing et al., 2005). Here, too, $\nu^{-2}(\hat{\mathbf{d}}_i)$ becomes large as the d-hats of row i become constant. Therefore, if any row tends to a constant, then the penalty term $n_2^{-1} \sum_{i=1}^{n_2} [1 + \omega \nu^{-2}(\hat{\mathbf{d}}_i)]$ becomes large.

To see how well PREFSCAL performs in conditional unfolding, we allowed separate transformations for each attribute of the brewery data (the rows in Table 14.2). We specified a monotone spline transformation of the sec-

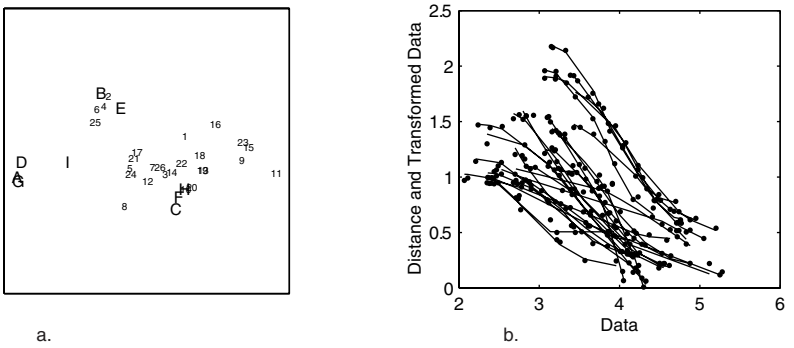


FIGURE 15.9. Representation of row conditional unfolding of data in Table 14.2 obtained by PREFSCAL(panel a), and its Shepard diagram (panel b).

ond degree with one interior knot, which is less restrictive than a linear or quadratic transformation, but more restrictive than an ordinal transformation. A striking feature of the solution (left panel of Figure 15.9) is that the breweries are located among the attributes, whereas in the other solutions discussed so far there is a clear separation of the breweries and the attributes. The right panel of Figure 15.9 contains the combined Shepard diagram of all attributes. It can be seen that the fit is high (Kruskal's Stress-1 is .0001) because most of the points are closely located to the a curve indicating that the difference between distance and d-hat will be small for these brewery and attribute pairs. The transformation curves generally are smooth and have variation coefficients markedly different from zero. Therefore, they are obviously not horizontal and not degenerated.

At the time of writing, PREFSCAL is scheduled to appear in SPSS in 2005. However, in the PREFSCAL program in SPSS, the row-conditional penalized Stress is defined slightly different from (15.6). In the program, $\sigma_n^{\lambda}(\hat{\mathbf{d}}, \mathbf{X})$ in (15.6) is replaced by an implicitly normalized form of Stress for each of the rows (Busing, 2004); that is, the PREFSCAL program in SPSS minimizes

$$\left(n^{-1} \sum_i \frac{\|\hat{\mathbf{d}}_i - \mathbf{d}_i\|^2}{\|\hat{\mathbf{d}}_i\|^2} \right)^{\lambda} n_2^{-1} \sum_{i=1}^{n_2} \left(1 + \frac{\omega}{\nu^2(\hat{\mathbf{d}}_i)} \right). \tag{15.7}$$

The reason for this difference is that it is computationally more convenient and can handle additional constraints on the configuration more easily. Both (15.6) and (15.7) are otherwise the same.

15.4 Summary

To give an overview of the quality and main properties of the methods discussed in this chapter, we have constructed Table 15.3. Most of the

TABLE 15.3. Comparison of approaches aimed at avoiding trivial solutions.

	Un-conditional	Row-conditional	Transformation	Trivial Solution Excluded	Quality
Adjusting Data					
Ratio-ordinal	+	+	Ordinal	Yes	+
Interval-ordinal	+	+	Ordinal	No	+/–
Augmenting within-persons block	+	–	All for between-sets ratio for within-sets	No	+/–
Augmenting both within-sets	+	+	All for between-sets ratio for within-sets	Yes	+
Adjusting the Transformation					
Ratio transformation	+	+	Ratio	Yes	+
Approach of Kim et al. (1999)	+	+	Ratio	Yes	+
Smoothed monotone regression	+	+	Restricted ordinal	Yes	+
Adjusting the Loss Function					
Stress-2	+	+	All	No	–
Weighting approach by DeSarbo	+	+	All	No	–
Penalizing the intercept	+	+	Interval	Yes	+
Penalized Stress by PREFSCAL	+	+	All	Yes	+

methods either have limited applicability and may depend highly on the software that is available. For example, the ratio-ordinal or the augmentation method both require that different types of transformations can be specified for the data. KYST can do that, but other programs cannot. Some of the methods discussed use forms of ratio transformations that may not be suited for preference rank-order data. The most promising approach to fit the ideal point model for unfolding seems to be the PREFSCAL model that gives good quality solutions for all standard transformations used in MDS.

When applying one of the methods for unfolding described in this chapter, one caution is needed. It is our experience that convergence criteria of the unfolding algorithms have to be set much more strictly than for ordinary MDS programs. Failing to do so may lead to a premature halt of the algorithm. The obtained solution may look nontrivial at a first glance, but continuing the algorithm with stricter convergence criteria may well lead to the trivial solution. Therefore, it is wise to make the algorithm run for many iterations so that one is sure to have avoided the trivial solution.

15.5 Exercises

Exercise 15.1 Table 15.4 on p. 333 shows the average ratings of 90 students from 15 different countries for 21 nations (columns) on 18 attributes (rows). The data were collected and reported by Wish, Deutsch, and Biener (1972). The rating scales are bipolar 9-point scales such as "collectivistic vs. individualistic" (scale 2). For most scales, only one label is shown: the

other end of the scale is obvious (as in "rich", where the other scale end is "poor").

(a) To analyze these data, first use ordinal unconditional unfolding. Represent these data in a plane. The solution is most likely degenerated into points-on-circles and/or into clusters of attributes and countries, respectively.

(b) Check whether linear unfolding helps to avoid the degeneracies. Discuss how the linear unfolding solution differs from the one for ordinal unfolding.

(c) Try out some of the methods discussed in this chapter to avoid the degeneracies. For example, compute within-country proximities and within-attribute proximities. Augment the above data matrix with these coefficients, and then run unfolding on this matrix.

Exercise 15.2 Use the data from Table 14.2 to compute coefficients for the similarity of breweries and of attributes, respectively. Then run an unfolding analysis of the data matrix in Table 14.2 after "completing" it with within-breweries and with within-attributes similarities (as suggested in Figure 14.1). Do you succeed in avoiding the degeneracies observed in Figures 14.13 and 14.15, respectively?

Exercise 15.3 Consider the contingency table below that is reported by Garmize and Rychlak (1964). Its entries show the frequencies with which different persons gave particular interpretations (rows) to Rorschach inkblot pictures when induced (by role play) into one of the moods shown in the columns.

Interpretation	Fear	Anger	Depression	Love	Ambition	Security
Bat	33	10	18	1	2	6
Bear	0	0	2	0	0	0
Blood	10	5	2	1	0	0
Boot(s)	0	1	2	0	0	0
Bridge	1	0	0	0	0	0
Butterfly	0	2	1	26	5	18
Cave	7	0	13	1	4	2
Cloud(s)	2	9	30	4	1	6
Fire	5	9	1	2	1	1
Fur	0	3	4	5	5	21
Hair	0	1	1	2	0	0
Island	0	0	0	1	0	0
Mask	3	2	6	2	2	3
Mountains	2	1	4	1	18	2
Rock(s)	0	4	2	1	2	2
Smoke	1	6	1	0	1	0

(a) Unfold these data with ordinal and metric models and test out different ways to avoid degeneracies.

TABLE 15.4. Average ratings of 90 students from 15 different countries for 21 nations on 18 attributes (Wish et al., 1972).

Country	Aligned with U.S.A.	Collect.-Individualistic	Peaceful	Individual Rights	I Like	Good	Similar to Ideal	Full of Opportunity	Stable	People Satisfied	Internally United	Influential Culture	Educated People	Rich	Industrialized	Powerful	On Way Up	Large
U.S.A.	9.0	7.4	4.5	7.5	7.5	6.6	5.5	7.5	7.2	6.1	4.8	7.4	4.4	8.8	8.8	8.9	6.6	8.8
U.K.	8.5	5.8	6.7	7.9	7.7	7.2	5.8	5.8	7.7	6.6	6.9	6.8	8.1	7.2	8.3	6.5	5.3	4.0
W.Germany	8.1	5.9	5.7	6.4	6.5	6.3	4.7	6.2	7.0	6.7	6.3	5.4	7.9	7.8	8.3	7.1	7.6	5.6
France	7.2	5.6	5.8	6.1	6.4	6.1	5.4	5.7	5.4	5.6	5.1	6.4	7.9	6.4	6.6	6.1	6.0	5.6
Israel	7.4	3.1	3.3	6.4	6.1	5.7	4.4	6.2	6.5	6.6	7.6	5.4	6.8	6.2	6.1	5.9	7.6	1.9
Japan	7.2	4.9	6.4	6.6	6.8	6.9	5.2	6.7	7.4	6.5	7.2	5.9	6.6	7.4	8.3	7.0	8.0	5.0
South Africa	6.0	6.6	5.1	2.9	3.6	3.4	2.2	2.8	4.6	3.4	2.8	3.2	6.9	6.3	5.6	4.7	4.7	5.7
Greece	6.9	5.9	6.2	3.2	6.4	4.8	3.6	3.7	3.5	3.8	4.2	6.2	3.4	3.6	3.7	3.0	4.5	3.0
Spain	6.6	6.1	6.4	3.4	5.5	4.2	2.9	3.5	5.5	4.4	4.5	5.0	4.6	3.5	3.9	3.1	4.3	4.6
Brazil	6.4	5.6	6.8	4.4	6.5	5.2	3.1	3.9	3.6	3.6	4.3	3.6	3.0	4.2	3.8	3.8	5.9	7.2
Mexico	6.8	5.2	7.0	4.6	6.4	5.7	3.5	4.1	5.2	4.6	5.7	4.7	3.3	3.8	4.0	3.4	5.7	5.4
Ethiopia	5.6	5.2	6.8	4.1	6.2	5.4	2.7	3.6	5.4	4.8	5.9	3.1	2.8	3.0	2.5	2.8	5.4	4.1
India	6.0	4.6	6.8	4.6	6.2	5.5	2.9	3.4	5.0	3.5	3.4	5.6	2.5	2.1	3.0	3.6	5.6	8.1
Indonesia	5.1	5.5	4.8	3.3	5.2	4.3	2.5	3.6	3.4	3.9	3.7	3.2	2.8	3.4	3.0	3.4	5.2	5.3
Congo	5.0	5.7	4.8	3.0	4.8	3.8	1.7	3.0	2.4	3.5	2.6	3.0	2.2	3.1	2.2	2.6	4.5	5.4
Egypt	3.6	4.1	3.1	3.8	5.1	4.2	2.5	3.5	4.1	4.5	5.6	5.0	3.0	3.4	3.5	3.7	4.7	5.1
China	1.1	2.0	2.4	2.1	4.2	3.9	2.7	3.2	4.5	3.8	4.2	5.2	3.4	3.1	4.7	7.0	6.4	8.7
Cuba	2.1	2.6	3.7	2.9	5.0	4.4	2.9	3.7	4.5	4.3	6.1	4.0	3.7	3.6	3.8	3.5	5.7	2.0
Yugoslavia	3.9	2.6	6.6	4.1	6.5	5.5	3.9	4.4	6.2	5.6	6.0	3.7	5.4	4.0	5.0	3.8	6.4	4.1
Poland	3.2	2.4	5.7	3.1	5.6	4.9	3.3	4.0	6.6	4.9	6.3	3.6	6.1	4.6	5.9	3.9	5.8	4.5
USSR	2.7	1.5	3.7	2.6	5.3	5.0	4.0	4.3	7.3	5.6	6.8	6.6	7.1	7.1	7.9	8.5	7.8	8.8

(b) Discuss in what ways the data could be preprocessed or weighted, noting, for example, that there are many zeros and also many very low frequencies.

(c) Check out what applying weights on the data does to your unfolding solutions.