Chapter 5
Variants of MDS Models

Abstract Various forms of MDS are discussed: ordinal MDS, metric MDS, MDS with different distance functions, MDS for asymmetric proximities, individual difference MDS models, MDS for more than one proximity value per distance, and weighting proximities in MDS.

Keywords Ordinal MDS · Interval MDS · Ratio MDS · Drift vector model INDSCAL · IDIOSCAL · Unfolding

5.1 The Type of Regression in MDS

A main difference of various MDS models is the type of regression that these models use. The most popular MDS model in research publications has been *ordinal* MDS, sometimes also—less precisely—called *nonmetric* MDS. Ordinal MDS computes an m-dimensional configuration $\mathbf{X}$ so that the order of the distances over $\mathbf{X}$ deviates as little as possible from the order of the proximities. Hence, in ordinal MDS and assuming that we have dissimilarities δ_{ij} as data (or that the proximities have been converted into dissimilarities), the representation function

$$f : \delta_{ij} \rightarrow d_{ij}(\mathbf{X}) \tag{5.1}$$

is *monotone* so that

$$f : \delta_{ij} < \delta_{kl} \rightarrow d_{ij}(\mathbf{X}) \leq d_{kl}(\mathbf{X}) \tag{5.2}$$

for all pairs i and j, and k and l, respectively, for which data (dissimilarities) are given. Proximities that are not defined ("missing data") are skipped. That is, if p_{ij} is missing, it imposes no restriction onto the MDS solution so that the distance $d_{ij}(\mathbf{X})$ can be chosen arbitrarily.

An important distinction between two forms of ordinal MDS is how *ties* (i.e., equal data values) are treated. The default in most programs is that ties can be *broken* in the MDS solution. That means that equal proximities need *not* be mapped into

I. Borg et al., *Applied Multidimensional Scaling and Unfolding*,
SpringerBriefs in Statistics, https://doi.org/10.1007/978-3-319-73471-2_5

equal distances. This is called the *primary approach to ties*. The *secondary* approach to ties ("keep ties tied") leads to an additional requirement for ordinal MDS, namely

$$f : \delta_{ij} = \delta_{kl} \rightarrow d_{ij}(\mathbf{X}) = d_{kl}(\mathbf{X}). \tag{5.3}$$

The primary approach to ties is usually more meaningful. Consider, for example, the data discussed in Sect. 2.2 where subjects had to judge the similarity of different countries on 9-point rating scales. Here, ties are unavoidable, because the rating scale had only nine different levels: With 66 different pairs of countries, this will automatically lead to the same proximity values for some pairs, even if the respondent feels that the respective countries are not really equally similar. Using a 66-point rating scale would not help either, because no respondent can make reliable distinctions on such a scale. Moreover, each single judgment is somewhat fuzzy and not absolutely reliable. Hence, equal ratings often mean something like "about equal" or "practically equal" but not simply "equal."

A second class of MDS models, called *metric* MDS, goes back to the beginnings of MDS in the 1950's (Torgerson 1952). Such models specify an analytic (usually monotone) function for f rather than requiring that f must be only "some" monotone function. Specifying analytic mapping functions for f has the advantage that it becomes easier to trace the mathematical properties of such models. Moreover, metric MDS avoids some technical problems of ordinal MDS such as, in particular, degenerate solutions (see Sect. 7.5). On the other hand, they typically lead to solutions with poorer fit to the data, because it is generally more difficult to represent data in more restrictive models. Yet, this may not be a drawback, because an excellent fit can also mean that more error is represented in the MDS solution. Ordinal MDS tends to over-fit the data, while metric MDS may iron out error in the data so that the solution becomes more robust and replicable.

The standard model of metric MDS is *interval MDS*, where

$$f : \delta_{ij} \rightarrow a + b \cdot \delta_{ij} = d_{ij}(\mathbf{X}), \tag{5.4}$$

with a and b ($\neq 0$) as free parameters. Interval MDS attempts to preserve the data in the distances such that the relations of *differences* ("intervals") among the data are preserved.[1] This makes sense, for example, if the data are interval-scaled. In that case, no meaningful information is lost if the data are scaled by multiplying them by some nonzero constant b and by adding an arbitrary constant a to each data value. All statements about the data that remain invariant under such linear transformations are considered meaningful; all other statements (e.g., statements about the ratio of data values) are not meaningful.

[1]Consider Table 1.1. Auto Theft and Murder are correlated with .11; Rape and Larceny with .60; the difference between these correlations is .49. This is about the same as the correlation between Assault and Burglary (.52). So, in the interval MDS solution in Fig. 1.4, the difference of the distances between the points for Auto Theft and for Murder should be about equal to the distance between Assault and Burglary.

More MDS models follow easily by choosing other mapping functions f (e.g., an exponential or a power function). However, if f is not at least weakly monotone, then such functions do not lead to easily interpretable results.

An MDS model that is stronger than interval MDS is *ratio MDS*, often considered the most restrictive model in MDS. It drops the additive constant a of interval MDS as an admissible transformation and searches for a solution that preserves the proximities up to an overall scaling factor b ($b \neq 0$). In case of the rectangle data analyzed in Chap. 2, this model could be seriously considered, because here the response scale started at "0=equal, identical" and hence scores of zero are meaningful if the subjects used the scale correctly.

The user chooses a particular MDS model f for a variety of reasons:

- Scale level. If theoretical or empirical reasons speak for a certain scale level of the data, then it usually makes sense to pick a corresponding MDS model. The choice of an appropriate scale level also depends on the zeitgeist to some extent. In the 1970s, for example, ordinal MDS was heavily pushed, whereas today metric (even: ratio) MDS has become the default (at least for statisticians) since it forces the user to justify any substantively blind optimizing transformations of the data.
- Minimize assumptions. The researcher wants to assume as little as possible about the relation of the data to MDS distances. Rather, he/she wants to let the data speak for themselves. Or he/she wants to get something for as little as possible. A typical case is a small inter-correlation matrix that is scaled with ordinal MDS and then interpreted in terms of dimensions or in terms of regions with wildly curving boundaries and many misplaced points. This can be useful as a first step in a field of research where little is known, but, of course, it should be replaced with more restrictive models in the long run.
- Robustness versus over-fitting. One often scales given proximities with both ordinal MDS and interval MDS. Ordinal MDS leads to smaller Stress values than interval MDS, but it may simply over-fit the data and, occasionally, it can also lead to meaningless degenerate solutions (see Sect. 7.5). Hence, when running both ordinal MDS and interval MDS, one can cross-validate the solutions and test for artifacts.
- Nonlinear mappings. The proximities are sometimes predicted to have a nonlinear relation to the distances in an MDS space. One example is Thurstone scaling, discussed on p. 47. In that case, one may not know how to specify the regression function analytically, or no program exists that would fit such a model, or one may want to test a certain prediction about the regression trend but not enforce it. So, ordinal MDS is used and then the Shepard diagram is studied closely for the shape of the regression trend. One may also first replace the data with their ranks and then use interval MDS: Weeks and Bentler (1979) have shown in simulations that this *rank-linear MDS* successfully recovers configurations whose distances have highly nonlinear (but still monotone) relations to the data.
- Dimensionality. Using weak MDS models leads to relatively small Stress values. This is often taken as evidence that one does not need higher-dimensional representation spaces. After all, if the Stress is small, then there is little left to explain.

This reasoning is somewhat too formal, though, because what one is really interested in are meaningful and replicable solutions. The Stress of the solution is but a technical index.

- Marketing. The user wants to get a small Stress value, because he/she fears that otherwise his/her results will not be publishable. This is, of course, a poor reason, because the Stress must always be evaluated relative to the particular MDS model, the dimensionality of the solution space, the number of points, and many other criteria such a robustness, stability, and replicability (Mair et al. 2016).

MDS models with other regression functions than those based on Steven's four classical scale levels exist too. One example is MDS with *spline transformations* on the data. Splines are piecewise (connected in k knots) polynomial functions of the n-th degree that lead to smooth (but not necessarily linear or monotone) regression lines in Shepard diagrams. The knots and the degree of the polynomials control the spline.

To illustrate what happens when running different types of MDS with real data, we use the inter-correlations in Table 2.2. Figure 5.1 exhibits the Shepard diagrams of 2d MDS solutions for these correlations. They show that ratio MDS fails completely, not because the regression trend is not nearly linear, but because the regression line must run through the origin: Only then are the data mapped into distances that preserve their ratios. The spline regression here is forced to run through the origin too. It is almost completely linear and has almost the same Stress as interval MDS.

Figure 5.1 shows, moreover, that ordinal MDS and interval MDS arrive at similar conclusions. The regression trend in ordinal MDS is almost linear, except for local steps and dents that are not interpretable and most likely not replicable. Yet, the Stress for ordinal MDS is clearly smaller than the Stress for interval MDS, simply because the regression line is closer to the points in the Shepard diagram in case of ordinal

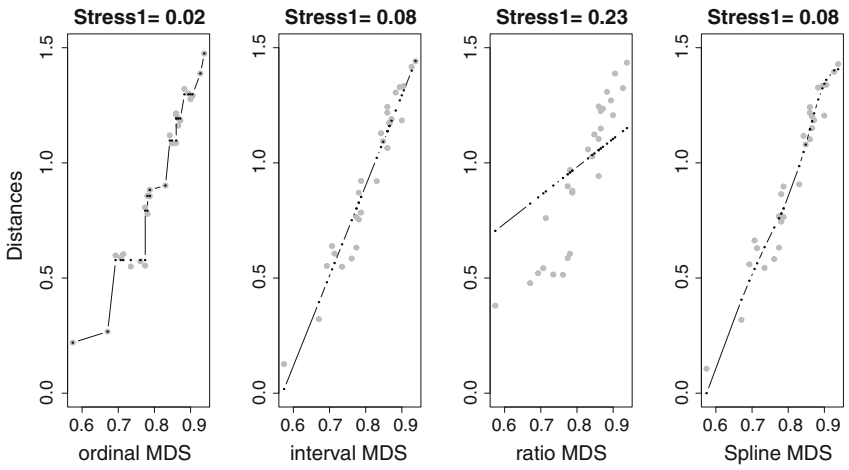

Fig. 5.1 Shepard diagrams for four types of regression; data from Table 2.4

MDS. Experience with empirical data shows that ordinal MDS and interval MDS often lead to highly similar results. Simulation studies come to the same conclusion (see Fig. 3.2). Ratio MDS is typically found to be much more demanding than either ordinal or interval MDS.

5.2 Euclidean and Other Distances

A second criterion for classifying MDS models is choosing a particular distance function. In psychology, the family of Minkowski metrics (specified in formula 2.3) used to be popular for modeling subjective similarity judgments of different types of stimuli (analyzable vs. integral stimuli) under different conditions (such as time, pressure). However, applications of MDS in the current literature almost always use Euclidean distances, because only they guarantee that the geometry is not misleading. Few MDS programs are even able to compute solutions with distances other than Euclidean distances. If they offer other distance functions too, then typically only the city-block metric.

Euclidean distances, as all other Minkowski distances, imply a *flat* geometry. In special cases, it can be useful to construct MDS representations in *curved* spaces. As an example, one can think of distances on a sphere. Here, the distance between two points is the shortest path ("geodesic path") in the two-dimensional curved space (i.e., on the sphere), which is the length of a cord spanned between two points over the surface of the sphere or, expressed more sloppily, airline distances. Curved geometries can sometimes be useful (e.g., in psychophysics), but they are never used in general data analysis situations. Circular scales do, however, play an important role in psychology, but they do not require true curved-space analysis. Rather, they use circles or balls embedded in Euclidean spaces. See, for example, "spherical MDS" on p. 72 and "circular unfolding" on p. 103.

5.3 MDS of Asymmetric Proximities

Distances are always symmetric, i.e., $d_{ij} = d_{ji}$, for all i, j. Proximities that are not symmetric can, therefore, not be represented by distances in MDS models. Yet, as long as the asymmetries are just error-based, no real problem arises. MDS simply irons out these errors. Or the user eliminates or reduces them by averaging corresponding data values.

Asymmetries may, however, be reliable and meaningful. Examples are the asymmetries in an import–export matrix, where country i imports more from county j than vice versa. Then, social networks can be studied in terms of how much each person i likes the other person j: Liking is rarely fully symmetric, and asymmetries can be very meaningful. A third example is simply the order of presentation when

making pairwise comparisons of i and j: How similar is the Morse code di-di-da to the subsequent da-di, and how similar is da-di to the following di-di-da (see p. 70)?

A simple approach of dealing with asymmetric proximities in the MDS context is the *drift vector model*. The model requires decomposing the proximity matrix $\mathbf{P}$ into a symmetric part $\mathbf{S}$ and an asymmetric part $\mathbf{A}$. The symmetric part is computed by averaging corresponding cells, $\mathbf{S} = (\mathbf{P} + \mathbf{P}')/2$, with elements $s_{ij} = (p_{ij} + p_{ji})/2$. This matrix is then scaled as usual with MDS. The rest of the proximities is $\mathbf{A} = \mathbf{P} - \mathbf{S}$, with elements $a_{ij} = p_{ij} - s_{ij}$. This matrix is *skew-symmetric*. It can be represented on top of the MDS solution for $\mathbf{S}$ by attaching an arrow on each point i that is directed toward point j, or away from point j, depending on the sign of the asymmetry. The length of this arrow is chosen as $k \cdot |a_{ij}|$, with k some convenient overall scaling factor (e.g., $k = 1/\text{mean}(p_{ij})$).

We demonstrate this model with a small example. Let $\mathbf{P}$ be a matrix of similarity values (e.g., the number of references in row journal i to column journal j):

$$\mathbf{P} = \begin{bmatrix} 0 & 4 & 6 & 13 \\ 5 & 0 & 37 & 21 \\ 4 & 38 & 0 & 16 \\ 8 & 31 & 18 & 0 \end{bmatrix} = \mathbf{S} + \mathbf{A} = \begin{bmatrix} 0 & 4.5 & 5.0 & 10.5 \\ 4.5 & 0 & 37.5 & 26.0 \\ 5.0 & 37.5 & 0 & 17.0 \\ 10.5 & 26.0 & 17.0 & 0 \end{bmatrix} + \begin{bmatrix} 0.0 & -0.5 & 1.0 & 2.5 \\ 0.5 & 0.0 & -0.5 & -5.0 \\ -1.0 & 0.5 & 0.0 & -1.0 \\ -2.5 & 5.0 & 1.0 & 0.0 \end{bmatrix}.$$

For $\mathbf{S}$, interval MDS yields the point configuration in Fig. 5.2. The values of $\mathbf{A}$ are represented as arrows in this plot. They are inserted one by one into the MDS solution. For example, on point #2, an arrow with a length of 5 units is attached pointing toward point #4. The resultant of all arrows attached to a point is the *drift vector* of that point, represented by the thick arrows with larger arrow heads in Fig. 5.2.

One notices in this vector field that, for example, journals #2 and #3 refer to each other relatively often and also quite symmetrically (because these points are close, and because the drift vectors are short). For journals #2 and #4, the mutual referencing is clearly smaller and, moreover, it is also rather asymmetric: Journal #2 looks more toward #4 than vice versa.

One can experiment somewhat with how one wants to represent the symmetries and the asymmetries (e.g., show all arrows or only resultant vectors; only vectors with the same meaning; use different scale factors for lengths of arrows). This can easily be done by using and modifying the R script for plotting Fig. 5.2 in the supplementary code file.

Let us look at a real data example. Rothkopf (1957) studied to what extent 598 test persons confused different acoustic Morse signals. He used 36 different signals, the 26 letters of the alphabet, and the natural numbers from 0 to 9. In the experiment, each person had to judge whether two signals, i and j, presented acoustically one after the other, were the same or not the same. Both (i, j) and (j, i) had to be judged in the experiment. The percentage of "Same!" judgments (i.e., the confusion probability) for each pair is taken as a measure of the psychological similarity of each pair.

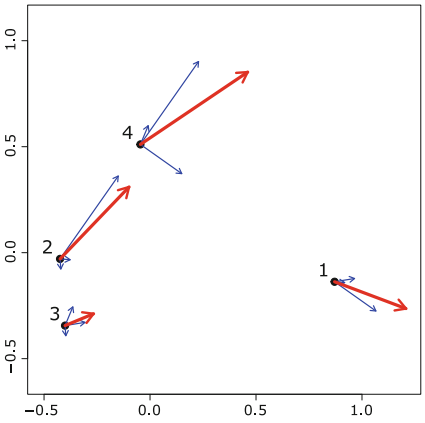

Fig. 5.2 Asymmetry vectors over an MDS solution; thick arrows with large heads are resultants (drift vectors)

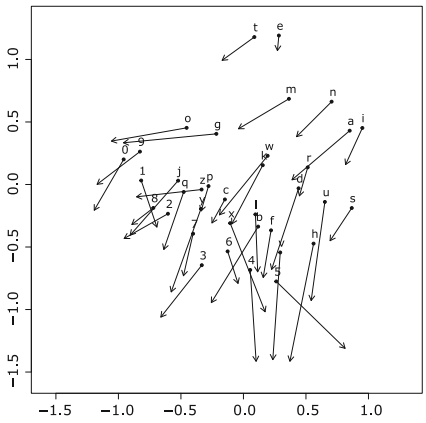

Fig. 5.3 MDS with drift vectors for Morse code confusion data

The confusion probabilities are not symmetric. For example, the signal for i (di-di) is more frequently confused with a subsequent signal for s (di-di-di) than s is confused with a subsequent i (35% vs. 16%). But do these asymmetries exhibit a systematic pattern or are they just random? We answer this question using the `driftVectors()` function of the SMACOF package[2]:

```
data(morse2)   ## morse2 = 1 - confusion probabilities
fit.drift <- driftVectors(morse2, type="ordinal")
fit.drift
plot(fit.drift)
```

The solution is shown in Fig. 5.3. The configuration of points represents the symmetric part of the data quite well (Stress = 0.192). The resultant drift vectors form a vector field that indeed exhibits a systematic trend: All arrows point more or less into the same direction. Substantively, this means that long signals are more often confused with short ones than vice versa (see p. 70f. for more information on these data).

[2]Note that for converting a complete $n \times n$ matrix of similarities, **P**, into dissimilarities, you cannot use `sim2diss()`, because it does not work on the whole matrix. Use your own conversion. For example, run `diss <- max(P) - P`, and then use `diss` in `driftVectors`.

5.4 Modeling Individual Differences in MDS

A popular variant of MDS is the *dimensional weighting model*, often called the INDSCAL model by the name of its original computer program (Carroll and Chang 1970). We explain the basic idea of this model using an experiment on color perception. Helm (1964) asked a sample of test persons to assess the similarity of ten chips with different colors (same brightness and same saturation). For each individual, similarity scores were obtained for each pair of colors. Some test persons were deuteranopic to some extent; i.e., they were not able to clearly distinguish green and purplish-red ("red-green blind").

Rather than first averaging the 16 data sets and then scaling the averaged data with standard MDS, we can use them directly in the dimensional weighting model. In this approach, we have to assume that there exists a *group space*, $\mathbf{X}$, that represents what all persons have in common. Moreover, each individual i has his/her individual space, $\mathbf{X}_i$. The model claims that the various individuals in the sample are not really that different. Rather, each *individual space* is a simple transform of the group space. That is, $\mathbf{X}_i = \mathbf{X}\mathbf{W}_i$, where $\mathbf{W}_i$ is a diagonal matrix with positive elements. Geometrically, this means that person i's individual space is generated by stretching or compressing the group space along its dimensions. INDSCAL fits this model by searching for both an $\mathbf{X}$ and for a set of weight matrices $\mathbf{W}_i$ such that the distances among the rows of the $\mathbf{X}_i$'s optimally correspond to the (admissibly transformed) dissimilarities of the persons (the dhat's). Expressed more formally, the distances are

$$d_{ijk}(\mathbf{X}) = \left(\sum_{a=1}^{m} w_{ak}(x_{ia} - x_{ja})^2 \right)^{1/2}, \qquad w_{ak} \geq 0, \tag{5.5}$$

where the parameter $k = 1, ..., N$ stands for different individuals or cases.

For the Helm data, the group space (using SMACOF; see commands below) is computed as presented in the left panel of Fig. 5.4: It is the expected color circle, slightly squeezed in the vertical direction as the fitted circle (dashed line) shows. The middle panel contains the model's solution for the most deuteranopic person CD4: Here, the color circle is stretched along Dimension 1 or, which has the same effect, it is compressed along Dimension 2. This reflects that this person cannot discriminate well between green and the purplish-red colors. The right panel exhibits the solution for person N6a. This person, a color-normal person, stretches the group space somewhat along Dimension 2.

```
res.helm <- indscal(helm, type="interval")
res.helm  ## gives Stress-1 etc.
plot(res.helm)  ## plots the group space
names(res.helm) ## shows elements of object res.helm
res.helm$cweights[[16]]  ## prints weight matrix for person 16
```

The weights for the 16 data sets are visualized in Fig. 5.5, often called the *subject space* of an INDSCAL solution. We here see that person CD4 stretches the group

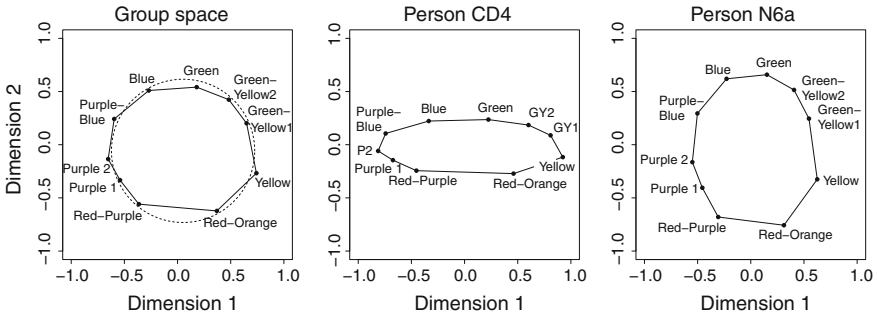

Fig. 5.4 Dimensional weighting MDS for Helm color data: group space (left panel); weighted group spaces of person CD4 who is red-green-deficient (middle); person N6a (right)

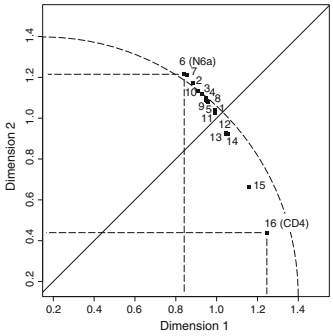

Fig. 5.5 Dimension weights of 16 individuals for group space shown in Fig. 5.4, left panel

space along Dimension 1 relative to Dimension 2 by a factor of roughly 3 (1.25 : 0.44). The weight ratio of person N6a is about 2:3. One must be careful, though, in interpreting these weights: They are only meaningful *relative to the group space*, and the group space is, unfortunately, *not unique*. If the group space is stretched or compressed along its dimensions, different weights are found for each person, while the overall fit of the MDS solution remains the same. What one can interpret as data-driven, therefore, is merely that person CD4 weights the vertical dimension less than person N6a, since this relation remains invariant under horizontal or vertical stretchings of the group space. That N6a weights Dimension 2 three times as much as CD4, or that he/she weights the dimensions in a certain weight ratio, is only true relative to the group space in Fig. 5.4. This also explains why the color-normal person N6a seems to put more weight on Dimension 2: The reason is that the group space also contains the data of the color-deficient persons which makes the configuration somewhat elliptical rather than circular as it would be true for color-normal persons only.

The overall Stress of the INDSCAL solution with its 2d group space for the ten colors and its sixteen 2×2 diagonal weight matrices is .123. The output object `res.helm` contains additional fit indexes such as Stress-per-Point measures. Most important is the fit of each person in this model. Here we find for person CD4 a Stress value of

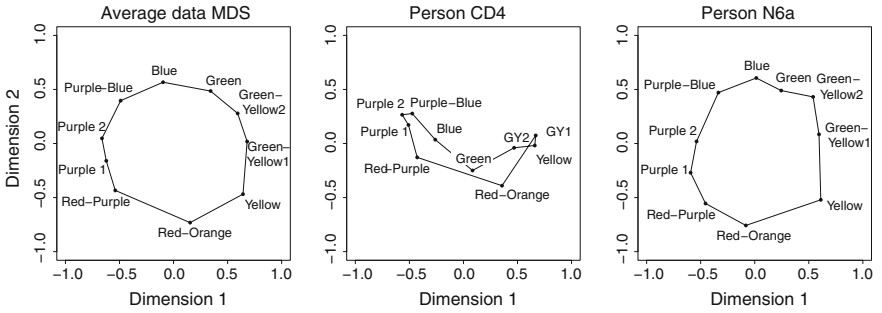

Fig. 5.6 Three separate standard interval MDS solutions for average color data, person's CD3 data, and person's N6a data, respectively; configurations of CD3 and N6a Procrustean-fitted to average data solution

.106 and for person N6a a value of .108. These values can be compared, for example, with the Stress values of standard MDS that scales each person's data separately. For the averaged data, interval MDS finds a solution (Fig. 5.6, left panel) with a Stress of .069. For person CD4, MDS finds a solution (Fig. 5.6, center panel; configuration Procrustean-fitted to average data solution) with a Stress of .074, and for N6a with .069. Thus, the INDSCAL model seems to be doing quite well and one could hope that the dimensional weighting model with its simple individual differences theory irons out error that is over-fitted in the standard MDS solutions. On the other hand, the results of scaling the average data and each person's data separately with standard MDS lead to essentially the same substantive insights.

If one drops the idea of common dimensions for all individuals,[3] a new model arises, the IDIOSCAL model (also called subjective metric model, elliptical distances model, or subjective transformations model). It admits more general person-specific transformations of the group space. Formally, in $\mathbf{X}\mathbf{W}_i = \mathbf{X}_i$, the matrix $\mathbf{W}_i$ is *not* restricted to be diagonal, but can be any real-valued $m \times m$ matrix. Such a matrix can be interpreted in various ways. One interpretation is that person i first rotates the dimensions of the common space in his/her own way and then stretches/compresses the space along these dimensions.[4] Yet, such a transformation demolishes the major selling point of the INDSCAL model, namely its *unique* dimensions. Not only has an INDSCAL solution the *same* dimensions for *all* individuals, but *any* rotation of these dimensions generally leads to higher Stress. Users of INDSCAL—market researchers in particular—had always hoped that these dimensions were the "true" psychological dimensions underlying the observations.

[3] Computationally, this is done by requesting `constraint="idioscal"` in the `smacofIndDiff()` function or by simply using the `idioscal()` function.

[4] $\mathbf{W}_i$ can always be (uniquely) split by singular value decomposition into the product $\mathbf{U}\mathbf{D}\mathbf{V}'$, where $\mathbf{U}$ and $\mathbf{V}'$ are rotations/reflections, and $\mathbf{D}$ is a diagonal matrix of dimension weights. Hence, $\mathbf{X}\mathbf{W}_i = \mathbf{X}\mathbf{U}\mathbf{D}\mathbf{V}'$ means the group space is first rotated/reflected by $\mathbf{U}$, then weighted by $\mathbf{D}$, and then rotated once more.

In the example above, *meaningful* dimensions were indeed identified, but one should not expect that this will always be the case, simply because such dimensions may not exist. The user should also know that the fit of the INDSCAL model is not necessarily much worse if all dimension weights are set to the same value. Borg and Lingoes (1978) report an example where the dimension weights scatter substantially even though they explain very little additional variance over unit weights. In such a case, the unique orientation of the dimensions is not very strong either.

The idea of the dimension weighting model can also be realized in a more stepwise and bottom-up approach. This avoids some of the interpretation problems. To do this, one first scales each of the given N data matrices, one by one, by standard MDS. One then uses Procrustean transformations (see Sect. 7.6 on p. 84f. for details) to fit the N resulting configurations to each other by admissible transformations (rotations, reflections, size adjustments, and translations). The average of all the fitted configurations is then taken as the "common" configuration (*centroid configuration*), or, alternatively, one uses the MDS configuration based on the averaged data as the group space. Subsequently, one identifies the dimensions of the group space that, if weighted, optimally explain the individual MDS solutions. This *hierarchical* approach is used by the Procrustean-Individual-Differences-Scaling (PINDIS) model (Lingoes and Borg 1978). It also continues the bottom-up fitting process with IDIOSCAL and other more general models, all of them not very useful in practice. However, the hierarchical approach suggests checking how much better the fit of a more general model is compared to one with more restrictions. The user can ask such a question using the `smacofIndDiff()` function by first running the above script with `constraint="identity"` (see next section for more on that option), then with `constraint="indscal"`, and finally with `constraint="idioscal"`. The three resulting solutions have Stress values of .146, .123, and .121, respectively. So, for the Helm color data, individual dimension weighting leads to a noticeable Stress reduction, but admitting more general idiosyncratic transformations is hardly worth it.

5.5 Scaling Replicated Proximities

In most MDS applications, we are dealing with exactly one data value per distance. Modern MDS programs allow using not just one proximity (p_{ij}) for each distance d_{ij} but two or more ($p_{ij}^{(k)}$, $k = 1, 2, ...$). Consider Fig. 5.7, a stack of $n \times n$ complete proximity matrices. Often, the values in such a data cube are first averaged over all persons and then averaged once more over the two halves of the resulting complete matrix. The data in the rectangle experiment described on p. 19 were generated in this fashion. Alternatively, we could inform the MDS program that what we have is one complete data matrix for each of N persons and that we want the program to find a solution where each d_{ij} represents, as precisely as possible, up to $N \cdot (n^2 - n)$ proximities, where the "up to" means that missing data are possible. Only the main

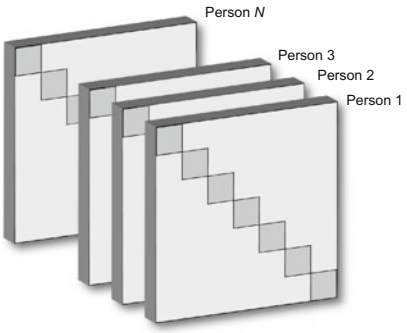

Fig. 5.7 Data cube consisting of complete proximity matrices for N persons

diagonals in each proximity matrix can *never* impact the MDS solution because $d_{ii} = 0$, for any i, is true in any distance geometry.

As an example, consider two sets of inter-correlations among work values collected in West and East Germany, respectively (see p. 84 for more on these data). These data can be loaded by calling `data(EW_eng)` in SMACOF: This activates a list of two correlation matrices, one for West Germany and the other for East Germany. If we interpret the entries in these correlation matrices as replications or as indicators of common German work value constructs, we could first average these matrices and then scale the averaged correlations. However, this reduces the number of data that determine the MDS solution in half, and more data for each distance should lead to a more data-driven, more robust solution. To use both matrices to generate one MDS solution, we can run this:

```
EW.diss <- list(east = sim2diss(EW_eng$east), west = sim2diss(EW_eng$west))
res <- smacofIndDiff(EW.diss, type="ordinal", constraint="identity")
res; summary(res)  ## gives Stress, coordinates, SPP values
plot(res, main="East+West combined")
```

5.6 Weighting Proximities in MDS

We saw in formula 3.3 that introducing weights into the MDS loss function is easily possible. It is always used if there are missing data, because then these weights are set to zero in case of an NA value. This makes the MDS algorithm ignore the distances that correspond to (missing) proximities. The distances can take on any value without affecting the Stress of the solution. However, for real data, one sometimes has additional information about the data such as their reliability. In that case, one may want to weight the more reliable data more in the Stress function. Another case is the size of the data themselves. Judgments on the similarity of pairs of objects are sometimes said to be the more difficult the more similar the objects are. The error that this uncertainty introduces can be expressed by weighting observed dissimilar-

ities the heavier the greater they are. In other words, larger dissimilarities should determine the MDS solution more than smaller dissimilarities.

Computationally, weighting an MDS or an unfolding solution can be easily accomplished by adding a weight matrix of the size of the data matrix to the call of the MDS function or program. For example, assume we want to scale the inter-correlations in Table 2.2, **R**, and assume further that we have a table of reliabilities of these correlations, **W**. To run weighted MDS, we first have to convert the correlations into dissimilarities by `diss <- sim2diss(R)` and then call `result <- mds(diss, weightmat=W)`.

If you experiment with different weight matrices, you will find, though, that weights need considerable variance to affect the MDS configurations. One possibility is setting `W <- R^q` for some large or small `q`. For example, for `q=10` the large similarities (i.e., small dissimilarities) will be fitted much better than small similarities (i.e., large dissimilarities) so that only the small distances in the plot should be interpreted and the large distances be better ignored. For a very small `q`, e.g., `q=-10` the reverse is true. Hence, weighting can be useful in practice but it may require some experimentation. The Shepard diagram will show which dissimilarities are well fitted and which are not.

5.7 Summary

MDS is a family of different models. They differ in the way they map the proximities into distances, and in the distance functions they employ. The various regression functions preserve certain properties of the data such as the ranks of the data in ordinal MDS, the relative differences of any two data values in interval MDS, and the ratios of the data in ratio MDS. Typically, Euclidean distances are chosen as the targets in MDS. City-block distances or dominance metrics are also used in psychological modeling. Some MDS models allow using multiple proximities per distance. Asymmetric proximities can be handled by the drift vector model: It represents their symmetric part by the distances of an MDS configuration, and their skew-symmetric part by drift vectors attached to the points. A popular MDS model is INDSCAL which represents a set of N proximity matrices, one for each of N individuals, by one common MDS space and by N sets of individual weights for its dimensions.

References

Borg, I., & Lingoes, J. C. (1978). What weight should weights have in individual differences scaling? *Quality and Quantity*, *12*, 223–237.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of 'Eckart-Young' decomposition. *Psychometrika*, *35*, 283–320.

Helm, C. E. (1964). Multidimensional ratio scaling analysis of perceived color relations. *Journal of the Optical Society of America*, *54*, 256–262.

Lingoes, J. C., & Borg, I. (1978). A direct approach to individual differences scaling using increasingly complex transformations. *Psychometrika*, *43*, 491–519.

Mair, P., Borg, I., & Rusch, T. (2016). Goodness-of-fit assessment in multidimensional scaling and unfolding. *Multivariate Behavioral Research*, *51*, 772–789.

Rothkopf, E. Z. (1957). A measure of stimulus similarity and errors in some paired-associate learning. *Journal of Experimental Psychology*, *53*, 94–101.

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, *17*, 401–419.

Weeks, D. G., & Bentler, P. M. (1979). A comparison of linear and monotone multidimensional scaling models. *Psychological Bulletin*, *86*, 349–354.