Appendix D

Multidimensional Scaling

D.1 Introduction

Multi-Dimensional Scaling (MDS) is a set of data analysis methods which allow the data interpreter to infer the dimensions of the perceptual space of subjects. Wickelmaier (2003). Or in other words the goal of MDS is to map data points in $\mathbb{R}^P$ to a lower dimensional manifold (Hastie *et al.,* 2008).

Appendix D is a description of multidimensional scaling based on a good introduction to this technique presented by Wickelmaier (2003) who makes a very good overview of MDS for non-statisticians. MDS was used in this thesis to find dissimilarities between facies models generated using multiple-point statistics, from different training images.

The raw data entering into a MDS analysis are typically a measure of the similarity or dissimilarity. The primary outcome of an MDS analysis is a spatial configuration, in which the objects are represented as points. The points in this spatial representation are arranged in such a way, that their distances correspond to the similarities of the objects: similar object are represented by points that are close to each other, dissimilar objects by points that are far apart.

The input data in MDS analysis is called proximities (Wickelmaier, 2003) or dissimilarities (Hastie *et al.,* 2008); these are the estimation of the similarities or dissimilarities of the investigated objects.

D.2 Proximities (dissimilarities)

The input data for MDS analyses are called proximities. Proximities indicate the overall similarity or dissimilarity of the objects under investigation. An MDS analysis looks for a spatial configuration of the objects, so that the distances between the objects match their proximities as closely as possible (Wickelmaier, 2003). Data are commonly arranged in a square matrix, the so-called *proximity matrix*. There are two major groups of methods for deriving proximities: direct and indirect methods.

D.2.1 Direct methods

Subjects might either assign a numerical (dis)similarity value to each pair of objects, or provide a ranking of the pairs with respect to their (dis)similarity.

(Dis)Similarity ratings

Typically, in a rating task subjects are asked to express the dissimilarity of two objects by a number. In the case of a dissimilarity scale, a low number indicates a strong similarity between two objects, and a high number a strong dissimilarity. In order to obtain the ratings, all possible pairs of objects have to be presented to the participant (a total number of $n(n\text{-}1)/2$, where n is the number of objects). This assumes, however, that the dissimilarity relation is symmetrical, and thus the order within each pair is of no relevance.

It is also possible to handle asymmetric proximity data, where for example object a is more similar to object b than b to a. Asymmetric proximities might for example arise when collecting confusion data (see below).

Investigating asymmetric proximity relations, however, doubles the number of pairs to be presented. The advantage of direct ratings is that the data are

immediately ready for an MDS analysis. Therefore, both an individual investigation of each participant, and an aggregate analysis based on averages across the proximity matrices are possible. A disadvantage of dissimilarity ratings is the rapidly growing number of paired comparisons, as the number of objects increases.

D.2.2 Indirect methods

Indirect methods for proximity data do not require that an interpreter assigns a numerical value to the elements of the proximity matrix directly. Rather, the proximity matrix is derived from other measures, e. g. from confusion data or from correlation matrices.

Confusion Data

Confusion data arise when the researcher records how often subjects mistake one stimulus for another. Consider an experiment where the letters of the alphabet are briefly presented via loudspeaker. The task of the subject is to recognize the letter.

From the data a proximity matrix can be derived: letters that are rarely confused get a high dissimilarity value, letters that are often confused a low one. An advantage of using confusion data is that the similarity of stimuli is judged on a perceptual level without involving much cognitive processing. Thus, very basic perceptual dimensions might be revealed using this technique. On the other hand, confusion data are often asymmetric and do not allow for an individual analysis. Most notably, there must be a good chance of confusing one object with the other, which excludes perfectly discriminable stimuli from being investigated using this method.

Correlation matrices

Yet another application of MDS is to use it for visualizing correlational data. When objects are measured on different scales and the measurements are correlated with each other, a matrix of correlation coefficients evolves. Even with just a few objects, such a matrix becomes complex, and it is hard to detect

patterns of correlation. An MDS solution plots the objects on a map, so that their correlational structure is accessible by visual inspection.

Table D.1 shows an example of a correlation matrix: it lists the correlation coefficients between crime rates collected in the 50 U. S. states (Borg and Groenen, 1997). From the data alone it is not easily seen which crime rates are related.

Crime	No.	1	2	3	4	5	6	7
Murder	1	1	0.52	0.34	0.81	0.28	0.06	0.11
Rape	2	0.52	1	0.55	0.7	0.68	0.6	0.44
Robbery	3	0.34	0.55	1	0.56	0.62	0.44	0.62
Assault	4	0.81	0.7	0.56	1	0.52	0.32	0.33
Burglary	5	0.28	0.68	0.62	0.52	1	0.8	0.7
Larceny	6	0.06	0.6	0.44	0.32	0.8	1	0.55
Auto theft	7	0.11	0.44	0.62	0.33	0.7	0.55	1

Table D.1: Correlations of crime rates over 50 U. S. states.

The MDS representation in Figure D.1 simplifies the task. The distances in the Figure correspond to the correlation coefficients, so that a high correlation is represented by a small distance, and vice versa. In addition to the graphical representation, the MDS analysis provides an explanation of the correlations by interpreting the axes of the MDS space: the x-axis might be interpreted as "person versus property", the y-axis as "hidden versus street".

Applying MDS to correlational data might reveal the relations between the objects more vividly than merely reporting correlation coefficients. A drawback of this method is that the proximities need to be constructed from additional measurements. The other methods of deriving proximities do not require such

measurements. Thus, an MDS analysis is possible, even if scales, dimensions or attributes of the stimuli under concern are unknown beforehand. In fact, it is the goal of the analysis to derive such dimensions.
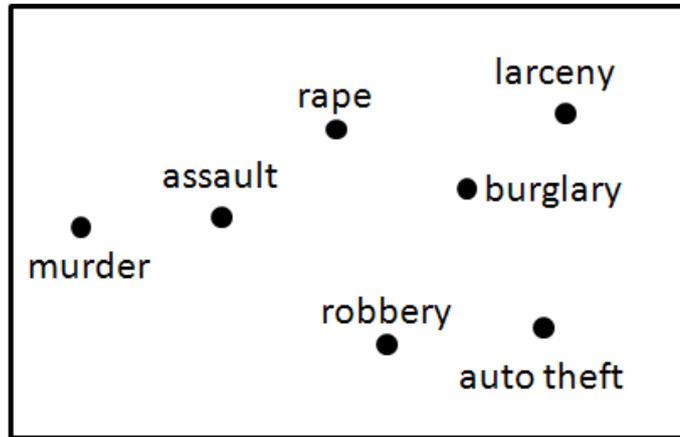

Figure D.1: 2-D MDS representation of the correlations in Table D.1 (Borg and Groenen, 1997).

Both direct and indirect methods of deriving proximity data yield the proximity matrix, which serves as an input for MDS programs. In many practical applications it will be straight forward to ask the participants directly for their judgments of the (dis)similarity of objects. Indirect methods, on the other hand, might be well suited to the investigation of basic perceptual dimensions, or in the case when additional measures of the objects under study already exist.

D.3 How does MDS work?

The goal of an MDS analysis is to find a spatial configuration of objects when all that is known is some measure of their general (dis)similarity. The spatial configuration should provide some insight into how the subject(s) evaluate the objects in terms of a (small) number of potentially unknown dimensions. Once the proximities are derived the data collection is concluded, and the MDS solution has to be determined using a computer program.

Many MDS programs make a distinction between classical and nonmetric MDS. Classical MDS assumes that the data, the proximity matrix, say, display metric properties, like distances as measured from a map. Thus, the distances in a classical MDS space preserve the intervals and ratios between the proximities as good as possible. For a data matrix consisting of human dissimilarity ratings such a metric assumption will often be too strong. Nonmetric MDS therefore only assumes that the order of the proximities is meaningful. The order of the distances in a nonmetric MDS configuration reflects the order of the proximities as good as possible while interval and ratio information is of no relevance.

D.3.1 Classical MDS

Consider the following problem: looking at a map showing a number of cities, one is interested in the distances between them. These distances are easily obtained by measuring them using a ruler. Apart from that, a mathematical solution is available: knowing the coordinates x and y, the Euclidean distance between two cities a and b is defined by:

$$d_{ab} = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \qquad\qquad eq.\,D.1$$

Now consider the inverse problem: having only the distances is it possible to obtain the map? Classical MDS (Torgerson, 1952), addresses this problem. It assumes the distances to be Euclidean. Euclidean distances are usually the first choice for an MDS space. There exist, however, a number of non-Euclidean distance measures, which are limited to very specific research questions (Borg and Groenen, 1997). In many applications of MDS the data are not distances as measured from a map, but rather proximity data. When applying classical MDS to proximities it is assumed that the proximities behave like real measured distances. This might hold e. g. for data that are derived from correlation matrices, but rarely for direct dissimilarity ratings. The advantage of

classical MDS is that it provides an analytical solution, requiring no iterative procedures.

In order to illustrate Classical MDS, Wickelmaier (2003) presents and example based on the distances between Danish cities. Assume we have measured the distances Between Copenhagen (cph), Arhus (aar), Odense (ods) and Aalborg (aal) on a map. Therefore the proximity matrix measuring the distance in millimetres would look like:

	cph	aar	ods	aal
cph	0	93	82	133
aar	93	0	52	60
ods	82	52	0	111
aal	133	60	111	0

The matrix of squared proximities is:

$$P^{(2)} = \begin{vmatrix} 0 & 8649 & 6724 & 17689 \\ 8649 & 0 & 2704 & 3600 \\ 6724 & 2704 & 0 & 12321 \\ 17689 & 3600 & 12321 & 0 \end{vmatrix}$$

Since there are 4 objects ($n=4$), the matrix J is calculated by:

$$J = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} - 0.25 \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 0.75 & -0.25 & -0.25 & -0.25 \\ -0.25 & 0.75 & -0.25 & -0.25 \\ -0.25 & -0.25 & 0.75 & -0.25 \\ -0.25 & -0.25 & -0.25 & 0.75 \end{vmatrix}$$

Applying J to $P^{(2)}$ yields a double centred matrix B:

$$B = -\frac{1}{2}JP^{(2)}J = \begin{vmatrix} 5035.625 & -1533.0625 & 258.9375 & 3740.938 \\ -1533.0625 & 507.8125 & 5.3125 & 1039.9375 \\ 258.9375 & 5.3125 & 2206.8125 & -2471.062 \\ 3740.938 & 1039.9375 & -2471.062 & 5172.062 \end{vmatrix}$$

For a 2-D representation of the four cities, the first two largest eigen-values and their corresponding eigen-vectors of B have to be extracted.

$$\lambda_1 = 9724.168, \quad \lambda_2 = 3160.986, \quad e_1 = \begin{pmatrix} -0.637 \\ 0.187 \\ -0.253 \\ 0.704 \end{pmatrix}, \quad e_2 = \begin{pmatrix} -0.586 \\ 0.214 \\ 0.706 \\ -0.334 \end{pmatrix}$$

Finally the coordinates of the cities are obtained by multiplying eigen-values and eigen-vectors:

$$X = \begin{vmatrix} -0.637 & -0.586 \\ 0.187 & 0.214 \\ -0.253 & 0.706 \\ 0.704 & -0.334 \end{vmatrix} \begin{vmatrix} \sqrt{724.168} & 0 \\ 0 & \sqrt{3160.986} \end{vmatrix} \begin{vmatrix} -62.831 & -32.974 \\ 18.403 & 12.026 \\ -24.960 & 39.71 \\ 69.388 & -18.763 \end{vmatrix}$$
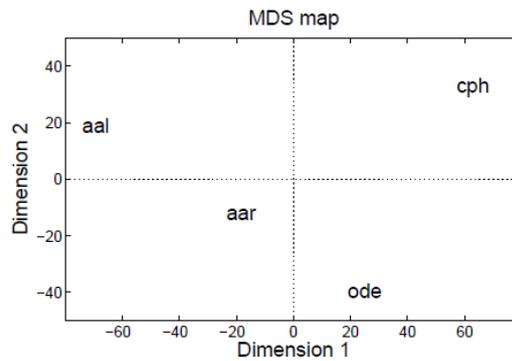

Figure D.2: Classical MDS 2-D representation of the four Danish cities

D.3.2 Nonmetric MDS

The assumption that proximities behave like distances might be too restrictive, when it comes to employing MDS for exploring the perceptual space of human subjects. In order to overcome this problem, Shepard (1962) and Kruskal (1964) developed a method known as *nonmetric multidimensional scaling*. In nonmetric MDS, only the ordinal information in the proximities is used for constructing the spatial configuration. A monotonic transformation of the proximities is calculated, which yields scaled proximities. Optimally scaled proximities are sometimes referred to as *disparities $d = f(p)$*. The problem of nonmetric MDS is how to find a configuration of points that minimizes the squared differences

between the optimally scaled proximities and the distances between the points. More formally, let p denote the vector of proximities (i. e. the upper or lower triangle of the proximity matrix), $f(p)$ a monotonic transformation of p, and d the point distances; then coordinates have to be found, that minimize the so-called *stress*

$$STRESS = \sqrt{\frac{\sum(f(p)-d)^2}{\sum d^2}}$$

$$eq.\ D.2$$

MDS programs automatically minimize stress in order to obtain the MDS solution; there exist, however, many (slightly) different versions of stress.

D.4 Judging the goodness of fit

The amount of stress may also be used for judging the goodness of fit of an MDS solution: a small stress value indicates a good fitting solution, whereas a high value indicates a bad fit. Kruskal (1964) provided some guidelines for the interpretation of the stress value with respect to the goodness of fit of the solution (Table D.2).

Stress	Goodness of fit
>0.20	poor
0.10	fair
0.05	good
0.025	excellent
0.00	perfect

Table D.2: Stress and goodness of fit.

Caution: These simple guidelines are easily misused. In order to avoid misinterpretation, the following should be kept in mind:

- There are many different stress formulae in the MDS literature. The guidelines, however, apply only to the stress measure computed by equation (D.2).

- Stress decreases as the number of dimensions increases. Thus, a two-dimensional solution always has more stress than a three-dimensional one.

Since the absolute amount of stress gives only a vague indication of the goodness of fit, there are two additional techniques commonly used for judging the adequacy of an MDS solution: the scree plot and the Shepard diagram (Borg and Groenen, 1997; Hair *et al.,* 1998).

In a scree plot, the amount of stress is plotted against the number of dimensions. Since stress decreases monotonically with increasing dimensionality, one is looking for the lowest number of dimensions with acceptable stress. An "elbow" in the scree plot indicates that more dimensions would yield only a minor improvement in terms of stress. Thus, the best fitting MDS model has as many dimensions as the number of dimensions at the "elbow". Figure 3 shows a paradigmatic scree plot and a Shepard diagram. The elbow in the scree plot suggests a three-dimensional MDS space.
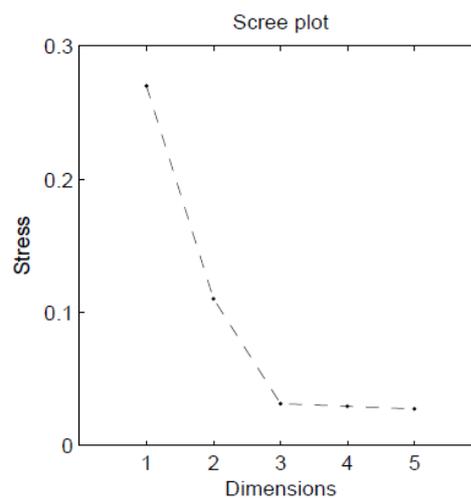

Figure D.3: Screen plot displaying an elbow at three dimensions.

In Section 7.2 a Screen plot was used to select the number of dimension that best fit the MDS model, but instead of Stress it was plot energy which can be estimated as: Energy=100* (1- STRESS)

Basics of a nonmetric MDS algorithm

The core of a nonmetric MDS algorithm is a twofold optimization process. First the optimal monotonic transformation of the proximities has to be found. Secondly, the points of a configuration have to be optimally arranged, so that their distances match the scaled proximities as closely as possible. The basic steps in a nonmetric MDS algorithm are:

1. Find a random configuration of points, e. g. by sampling from a normal distribution.

2. Calculate the distances d between the points.

3. Find the optimal monotonic transformation of the proximities, in order to obtain optimally scaled data $f(p)$.

4. Minimize the stress between the optimally scaled data and the distances by finding a new configuration of points.

5. Compare the stress to some criterion. If the stress is small enough then exit the algorithm else return to 2.