# Advances in multidimensional unfolding
Busing, F.M.T.A.

# ADVANCES IN MULTIDIMENSIONAL UNFOLDING

# ADVANCES IN MULTIDIMENSIONAL UNFOLDING

Proefschrift

ter verkrijging van de graad van Doctor
aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 21 april 2010 klokke 15:00 uur

door Franciscus Martinus Theodorus Antonius Busing
geboren te Amsterdam in 1963

weet je
wat nu zo zonde is
dat ik dit op schrijf
zonder jou

— Bert Schierbeek

Dedicated to the loving memory of Ton Busing
1931 – 1990

## PUBLICATIONS

Some ideas have appeared previously in the following publications from the same author:

Busing, F.M.T.A., Commandeur, J.J.F., & Heiser, W.J. (1997). PROXSCAL: A multidimensional scaling program for individual differences scaling with constraints. In W. Bandilla & F. Faulbaum (Eds.), *Softstat '97, advances in statistical software* (pp. 237–258). Stuttgart, Germany: Lucius.

Heiser, W.J., & Busing, F.M.T.A. (2004). Multidimensional scaling and unfolding of symmetric and asymmetric proximity relations. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 25–48). Thousand Oaks, CA: Sage Publications, Inc.

Busing, F.M.T.A., & Van Deun, K. (2005). Unfolding Degeneracies' History. In K. Van Deun, *Degeneracies in multidimensional unfolding* (pp. 29–75). Unpublished doctoral dissertation, Catholic University Leuven.

Van Deun, K., Groenen, P.J.F., Heiser, W.J., Busing, F.M.T.A., & Delbeke, L. (2005). Interpreting degenerate solutions in unfolding by use of the vector model and the compensatory distance model. *Psychometrika*, 70(1), 23–47.

Busing, F.M.T.A., Groenen, P.J.F., & Heiser, W.J. (2005). Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika*, 70(1), 71–98.

Busing, F.M.T.A. (2006). Avoiding degeneracy in metric unfolding by penalizing the intercept. *British Journal of Mathematical and Statistical Psychology*, 59, 419–427.

Busing, F.M.T.A., & de Rooij, M. (2009). Unfolding incomplete data: Guidelines for unfolding row-conditional rank order data with random missings. *Journal of Classification*, 26, 329–360.

Busing, F.M.T.A., Heiser, W.J., & Cleaver, G. (2010). Restricted unfolding: Preference analysis with optimal transformations of preferences and attributes. *Food Quality and Preference*, 21(1), 82–92.

# CONTENTS

# INTRODUCTION

Multidimensional unfolding is an analysis technique that creates configurations for two sets of objects based on the pairwise preferences between elements of these two sets. The distances between the objects correspond as closely as possible with the given preferences between them, such that high preferences correspond to small distances and low preferences to large distances. For example, in 1972, 42 respondents (21 MBA students and their spouses) rank ordered 15 breakfast items according to their preference. Unfolding now portrays both respondents and items as points in a configuration, as illustrated in Figure 1.1, such that respondents are closest to their first ranked item and furthest from their last ranked item. Moving away in any direction from a respondent's point thus decreases his/her preference for an item. The respondent's point itself, the so-called *ideal point*, thus makes up the highest point on the respondent's preference surface, which has the shape of a single-peaked function.

The rank numbers for the 15 breakfast items, albeit 1 to 15 for all respondents, are thus retrieved by the distances in the configuration, whether the respondents actually like the breakfast items or not. Furthermore, it is not said that when a blueberry muffin is most preferred by two respondents the muffin is also equally liked by these respondents. To cope with the differences in the preference scale within and between respondents, the actual rank numbers are allowed to be changed in magnitude as long as the order per respondent is maintained. The subsequent respondent-conditional monotone transformation of the rank numbers is optimally determined by the least squares unfolding technique, and consequently provides a metric solution from nonmetric data, creating distances from rank orders, respectively.

Notwithstanding the conceptual appeal, unfolding has not been used much in applications in the last few decades. As Heiser and Busing (2004) put it: "Applications of multidimensional unfolding lag seriously behind, undoubtedly due to the many technical problems that formed a serious obstacle to successful data analysis …" (Heiser & Busing, 2004, p. 27). The serious obstacle concerns *degenerate solutions*: Solutions that are perfect in terms of optimization of the least squares loss function, but useless in terms of interpretation of the unfolding solution. It is a problem that has become a trademark for unfolding. A mature analysis technique should operate faultlessly and this could hardly be claimed of unfolding. The freedom of the monotone transformation, the transformation that changes the nonmetric rank orders
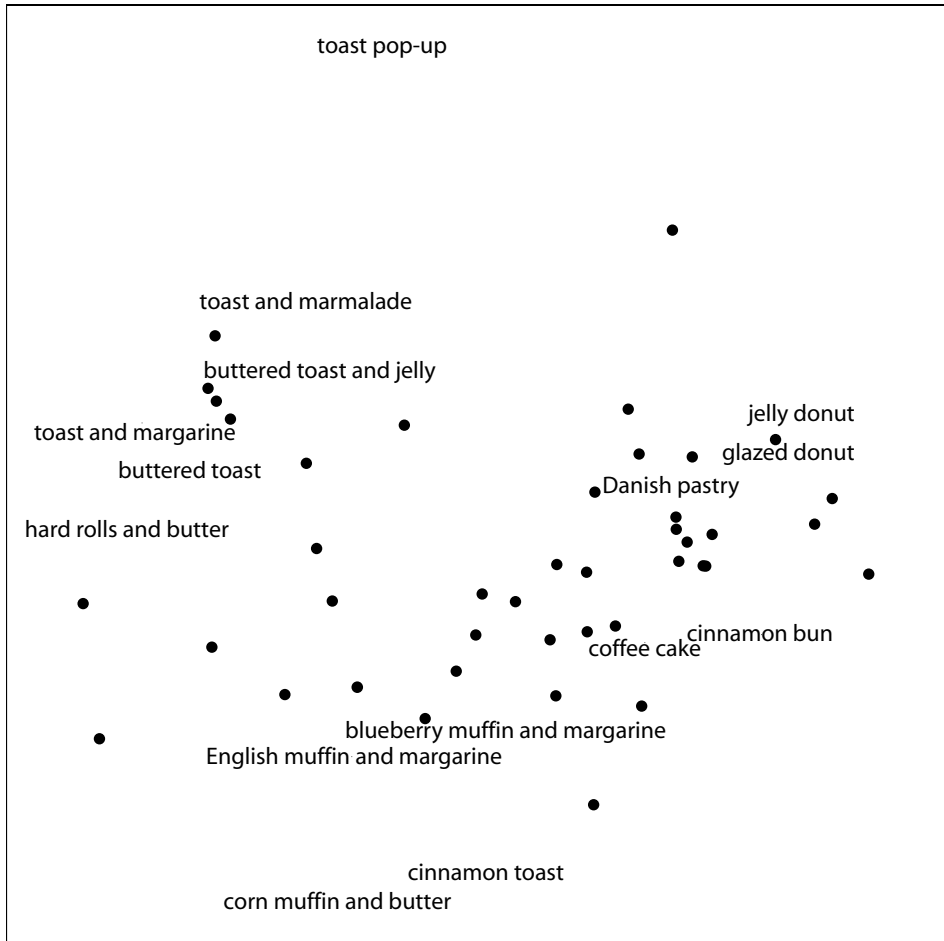
Figure 1.1    PREFSCAL unfolding solution for the breakfast data (Green and Rao, 1972) with 42 respondents (represented by dots) and 15 breakfast items.

into metric pseudo-distances, allows the different rank numbers to become (almost) identical. With the distances between the respondents and the items also (almost) identical and in addition equal to the transformed rank numbers, the solution can be achieved which is perfect in terms of fit, but completely worthless in terms of interpretative use. The (nonmetric) unfolding model is no longer identified as the freedom of transformation is such that any arbitrary data set results in a degenerate solution.

This monograph discusses the type of unfolding analysis that suffers from the degeneracy problem. It is characterized by an alternating least squares

minimization procedure for a multidimensional unfolding model that allows for optimal transformations of the data, irrespective the conditionality of the data. As such, it deviates from other types of unfolding analysis on model specification and minimization procedure. Unfolding IRT models (Andrich, 1988, 1989; Roberts & Laughlin, 1996; Roberts, Donoghue, & Laughlin, 2000), for example, exhibit single-peaked, nonmonotonic functions for unidimensional polytomous responses, intended for items that discriminate respondents, whereas probabilistic unfolding (Sixtl, 1973; Zinnes & Griggs, 1974; de Soete, Carroll, & DeSarbo, 1986; DeSarbo, de Soete, & Eliashberg, 1987; Ennis, 1993; MacKay & Zinnes, 1995; Hojo, 1997; MacKay, 2001; Hinich, 2005; MacKay & Zinnes, 2008) uses a different modeling strategy, using maximum likelihood estimation to obtain the model parameters.

The remainder of this monograph is composed as follows. Chapter 2 discusses the history of the degeneracy problem at length, specifically the scientific contributions that uncover, discuss, and resolve the obstinate problem. Multidimensional unfolding can not be considered a fully fledged analysis technique with this inconvenient problem on the side. The next two chapters, Chapters 3 and 4, offer solutions for the degeneracy problem, the former for metric unfolding only and the latter for all possible data transformations. With the degeneracy problem eliminated, multidimensional unfolding can be developed into a valuable analysis technique. Chapter 5 discusses one such a development: The addition of independent predictor variables to the unfolding model not only enhances the interpretation of the solutions, it also enables us to make predictions. Depending on the available information, this restricted unfolding model uses demographical information on respondents to predict respondent locations and item attributes to predict item locations, or vice versa, that is, the model uses additional locations to predict the variable values. Chapter 6 investigates the extent to which preferences can be missing while still maintaining a proper unfolding solution. This monograph finally discusses some topics for further research.

The technical appendix describes the implementation of the algorithm developed in Chapter 4. A strong extract of the program 'in development', PREFSCAL , belongs to the CATEGORIES module of IBM SPSS STATISTICS since version 14.0 (autumn 2005). The glossary provides insight in the (degenerate) solution types used throughout the monograph. It may be said that multidimensional unfolding is a truly amazing technique, which can handle all kinds of distance-like data, uses a simple and transparent minimization method (implementation of PREFSCAL), and produces commonly understandable graphical results. Too bad it was not working from the beginning.

2

This chapter discusses the contributions that were made to the problem of degenerate solutions in multidimensional unfolding during the twentieth century. First, the conceptual and technical foundations of multidimensional unfolding are given. Then, the work of Roskam (1968), Kruskal and Carroll (1969), Lingoes (1977), Heiser (1981), Borg and Bergermaier (1982), de Leeuw (1983), DeSarbo and Rao (1984), Heiser (1989), Kim, Rangaswamy, and DeSarbo (1999) is discussed. We conclude with a summary and some recent developments.

## 2.1 INTRODUCTION

In this chapter, a short historical overview of the developments in the domain of multidimensional unfolding in the twentieth century is given, with special attention for the problem of degenerate solutions . Multidimensional unfolding (MDU) is a technique that maps the row and column entities involved in ranking data jointly onto a low-dimensional space in such a way that the order of the distances reflects the rank orders. MDU is known to result in degenerate solutions. These are solutions that fit well and that are characterized by a clustering of the points such that an interpretation of the configuration becomes infeasible. From the overview, it will be clear that the problem of degeneracies popped up together with the first feasible algorithms, and that the problem is a very persistent one. However, almost forty years of stubborn attempts to overcome it seem to be justified, as these have led to important conceptual and technical refinements of a beautiful method.

First, we will discuss the conceptual and technical foundations of multidimensional unfolding, and at the same time we will delineate what we consider multidimensional unfolding. Then, the main part of the chapter follows, which is organized in the following way: A chronological order is maintained, organized around the important contributions that were made with respect to degenerate solutions. Each new contribution is discussed and 'illustrated' with an empirical example on the preferences of 21 MBA students and their wives for 15 breakfast items (P. E. Green & Rao, 1972) (see Table 2.1). We have chosen these data as they became some kind of norm in the domain: The success of an unfolding technique is measured by its performance for the

breakfast data. Our analyses of these data are based on strong convergence criteria, since, as mentioned in de Leeuw (1983, p. 5), Heiser conjectured that published nontrivial unfolding solutions are probably nontrivial because the iterations were stopped before the process had properly converged.

## 2.2 FOUNDATIONS OF MULTIDIMENSIONAL UNFOLDING

The unfolding method itself was at the heart of important contributions that were made to the general idea of scaling in the psychological and social sciences in the first half of the twentieth century: In that period, it was realized that measurement is possible for things that are not directly related to physical continua. As we will see, multidimensional unfolding (MDU) is the merger of two lines of development within this broad domain of scaling: Coombs and his coworkers introduced the concept of multidimensional unfolding, but a solution to the problem found its origins in multidimensional scaling (MDS). Part of what will be described here, was inspired by Delbeke (1968) and de Leeuw and Heiser (1980).

### Conceptual foundations

The history of *Unfolding* started in 1950 when Coombs, a student of Thurstone, published a paper in Psychological Review that showed how mere preference rankings contain metric information. This work built further on the ideas of indirect measurement by the method of paired comparisons, mainly inspired by Thurstone, and on the ideas of Guttman (1944, 1946): With his famous Guttman Scale, Guttman showed how both subjects as well as items can be scaled, while he only relied on qualitative data and made no distributional assumptions. Coombs (1950) developed a new type of scale which introduced a joint continuum, called *J scale*, on which both individuals and stimuli have fixed positions, and which "falls logically between an interval scale and an ordinal scale"(Coombs, 1950, p. 145). The position of the subject represents his *ideal* such that when asked which of two stimuli he prefers, this will be the one which is nearer to his own position on the continuum. The term *Unfolding* stems from the following metaphor: "Imagining a hinge located on

Table 2.1   Breakfast items and plotting codes.

| Code | Breakfast Item | Code | Breakfast Item | Code | Breakfast Item |
| --- | --- | --- | --- | --- | --- |
| TP | toast pop-up | CT | cinnamon toast | CB | cinnamon bun |
| BTJ | buttered toast and jelly | HRB | hard rolls and butter | DP | Danish pastry |
| EMM | English muffin and margarine | TMd | toast and marmalade | GD | glazed donut |
| CMB | corn muffin and butter | BT | buttered toast | CC | coffee cake |
| BMM | blueberry muffin and margarine | TMn | toast and margarine | JD | jelly donut |

the J scale at the $C_i$ value of the individual and folding the left side of the J scale over and merging it with the right side. The stimuli on the two sides of the individual will mesh in such a way that the quantity $|C_i - Q_j|$ will be in progressively ascending magnitude from left to right. The order of the stimuli on the folded J scale is the I scale for the individual whose $C_i$ value coincides with the hinge." (Coombs, 1950, p. 147). Unfolding is the reverse operation, where the preference orders of the subjects (the I scales) form the data and the objective is to find the J scale.

The unfolding idea was extended to the multidimensional case by Bennett and Hays (1960) and Hays and Bennett (1961). The first paper introduced the multidimensional unfolding model and focused on the problem of determining the minimum dimensionality required to represent the data. An example of preferences for hobbies was used to introduce the *Multidimensional Unfolding Model*: "The model states that each hobby can be characterized by its own position on each of several underlying attributes …. The model states further that every subject can be characterized by his own maximum preferences on each of these attributes, and that he will rank the hobbies according to their increasing distances from the ideal hobby defined by his own maximum preference on each attribute …let the attributes be the axes of a multidimensional space, and interpret 'distance' literally as the distance from the point representing the subject's ideal …to another point representing one of the hobbies" (Bennett & Hays, 1960, pp. 27–28). The remainder of the paper discussed how to find the minimum dimensionality needed to represent the preference rankings, while the 1961 paper discussed how to derive the configuration. Note that these papers formed the basis of the chapter on multidimensional unfolding of Coombs' influential 1964 book.

Coombs' work, and that of his coworkers, had an enormous impact on the conceptual level. However, the solution methods proposed are not tractable: As noted in Shepard (1962a), these methods yield nonmetric solutions (that is, subjects are not represented by fixed positions but by isotonic regions) for ordinal data and rely on certain rules of thumb, so that it is very difficult to set up algorithms that can be implemented in computer programs. To overcome these problems, *metric unfolding* was developed, initiated by Coombs and Kao (1960) who factor-analyzed the matrix of correlations between subject rankings, supposing that in this way the coordinates of the preference space can be found after eliminating an extra dimension labeled as a 'social utility dimension'. Ross and Cliff (1964) refined this idea by showing that a principal components analysis of the double centered matrix of squared distances allows to recover the rank of the space, and finally Schönemann (1970) proposed an algebraic solution for the metric unfolding model. A high price had to be paid for this solution, namely the beautiful idea that metric (numerical) information, i.e., distances, can be derived from qualitative (ordinal) data had

to be given up: the ordinal data are simply treated as numerical data. However, as will be discussed here below, it is possible to solve the problem in the true spirit of Coombs and Bennett and Hays, that is, a joint mapping of ranking data into a multidimensional space such that the order of the distances reflect the rank orders.

*Technical foundations*

In the same period that Coombs and Hays worked on the nonmetric multidimensional unfolding model, a big leap was made in the domain of multidimensional scaling, "an approach that has become feasible, only recently, with the advent of digital computers of sufficient speed and capacity" (Shepard, 1962a, p. 128). Important contributions of Shepard's paper were: The explicit formulation of the objective of the algorithm under construction, namely that a configuration is sought such that the distances are monotonically related to the data or *proximity measures* (a collective noun for observed similarities or dissimilarities), the demonstration that the ranked data "are generally sufficient to lead to a unique and quantitative solution" (Shepard, 1962a, p. 128), and the development of a computer algorithm that meets the objective. Shepard (1962a, 1962b) succeeded in achieving the objective put forward by Coombs, namely obtaining a metric solution from nonmetric data. However, his work still missed a rigorous numerical foundation and his computer algorithm contained several ad hoc elements (see Shepard, 1974).

Kruskal (1964a, 1964b) gave multidimensional scaling a firm theoretical foundation by introducing a "natural quantitative measure of nonmonotonicity" (Kruskal, 1964a, p. 26). This is the well known *Stress*, possible acronym for standardized residual sum-of-squares, with RAW STRESS defined as the root sum-of-squares

$$\text{R-STRESS} = \sqrt{\sum_{i<j} \left( \gamma_{ij} - d_{ij} \right)^2}, \tag{2.1}$$

where the $\gamma_{ij}$ are the optimally transformed data and the $d_{ij}$ are the distances between a stimulus $i$ and stimulus $j$. Further on in this chapter, $i$ is used as an index for subjects and $j$ gets its own summation sign for the stimuli. Formula (2.1), however, refers to (one-mode) multidimensional scaling. With the introduction of STRESS, the sound idea of finding a solution by optimizing a measurable criterion entered the domain of scaling. Kruskal not only introduced a loss function but he also showed how it could be minimized. The ability of analyzing incomplete data was also an important feature, especially for unfolding. Monotone regression was introduced as a technique to find optimally transformed data that minimizes STRESS for fixed distances. Nineteen hundred sixty four was the year that heralded in an era of research into nonmetric models for proximity data: A *nonmetric breakthrough* was realized.

An integration of the conceptual and technical insights, is found in the work of Gleason (1967) and Roskam (1968). Gleason developed a general model for multidimensional scaling that includes the analysis of conditional off-diagonal proximity data as a special case. An application of his program to empirical data can be found in the work of Delbeke (1968). Roskam is discussed in the following section.

In Table 2.2, an overview is given of the key contributions to multidimensional unfolding: For each contribution, the year of apparition, the author(s), the important findings, and the related computer program are provided.

## 2.3    ROSKAM, 1968

The dissertation of Roskam (1968) introduced a loss function currently known as *Stress-2* and represented a first systematic study of the nonmetric unfolding model as a tool for the analysis of preference data. This was the first time that the need of a proper adaptation of the loss function in order to avoid trivial solutions was pointed out. Nevertheless, even when using STRESS-2, Roskam reported unsatisfying results. Shortly after receiving his PHD, he developed together with Lingoes the MINISSA program, an acronym that stands for Michigan-Israel-Netherlands-Integrated-Smallest-Space-Analysis. Both the dissertation and the software are discussed hereafter. More biographical information and some references to the work of Roskam can be found in Bezembinder (1997).

Table 2.2    Overview of key papers and computer programs.

| Year | Author(s) | Contribution | Program |
|------|-----------|--------------|---------|
| 1968 | Roskam | Systematic study of nonmetric unfolding; Development of Stress-2 to avoid trivial solutions; Notification of importance of conditionality. | MINIRSA |
| 1969 | Kruskal Carroll | Development of Stress-2 to avoid trivial solutions; First mention of the problem of degenerate solutions. | KYST |
| 1977 | Lingoes | Imputation of the diagonal blocks and ordinary MDS analysis to avoid degeneracies. | SSAP |
| 1981 | Heiser | Restriction with bounds for the unrestrained ordinal transformations to avoid degeneracies. | SMACOF-3 |
| 1982 | Borg Bergermaier | Combining interval and ordinal transformations to avoid degeneracies. | (KYST) |
| 1983 | de Leeuw | Theoretical proof of the failure of Stress-2; Classification of degeneracy types. | |
| 1984 | DeSarbo Rao | Fixed cell weights emphasizing certain cells to avoid degeneracies; Fast algorithm minimizing Stress-2. | GENFOLD-2 |
| 1989 | Heiser | Improvement of bounded monotone regression, avoiding user specification of extra parameters. | SMACOF-3 |
| 1999 | Kim Rangaswamy DeSarbo | A priori nonmetric transformation followed by a metric analysis to avoid degeneracies. | NEWFOLD |

*"Metric Analysis of Ordinal Data in Psychology"*

In essence, Roskam's dissertation is a systematic application of the principles laid down by Kruskal to several existing formal models for conjoint data: The distance model, the compensatory distance model, the linear model, and the additive model. It mainly treats the analysis of rectangular data matrices, which is typical for unfolding data. Roskam also knew the work of Guttman and Lingoes, and taking hints from them, he expanded the work of Kruskal by accounting for the conditionality of the data, at the same time McGee (1968) permitted matrix-conditional transformations for individual differences models. This led to the development of a sound unfolding algorithm, and Roskam was the first one to thoroughly investigate the unfolding model. An important insight of Roskam was to use the *variance of the distances as a normalizing factor, in order to avoid the occurrence of degenerate solutions of the equal distance type*. The unconditional form of the loss function was introduced by Kruskal (1965) in the context of factorial experiments. Unconditional functions are characterized by the fact that they do not rely on a partition of the data whereas, for example, row-conditional functions rely on calculations (mainly transformations) performed row-wise. Kruskal (1965) used the variance as a scale factor for reasons of computational efficiency. Roskam's conditional STRESS formula is given by,

$$\text{STRESS-2} = \sqrt{\frac{1}{n} \sum_i \frac{\sum_j (\gamma_{ij} - d_{ij})^2}{\sum_j (d_{ij} - \bar{d}_i)^2}}, \qquad (2.2)$$

with $i = 1, \ldots, n$ the row entries (judges) and $j = 1, \ldots, m$ the column entries (items). Note that normalizing is done for each judge, so that the type of trivial solution where all items are equidistant from the judges but at a different distance for different judges, the so-called object-point degeneracy (see de Leeuw, 1983), cannot occur. To our knowledge, Roskam was not aware of this phenomenon. His only motivation to normalize per judge was the row-conditional nature of preference data.

Roskam (1968) gave some thoughts on trivial and degenerate solutions. He pointed out problems related to the weak order introduced by the monotone regression procedure: On the one hand, he noted that trivial solutions should be prevented by a proper normalization of the STRESS function (such that it is not possible that all items coincide or are equidistant), on the other hand he also noted that this does not necessarily exclude that some points will coincide. What Roskam meant precisely with degenerate and trivial solutions is not very clear: It seems that he used the word *trivial* for solutions that have zero STRESS due to some collapsed points and *degenerate* for solutions that are completely trivial.

In the chapter on unfolding, Roskam presented results that show an objects-circle degeneracy (items on the circumference of a circle and judges in the middle) which, however, was not recognized as a shortcoming of the unfolding algorithm (2.2) used. On the contrary, these disappointing results led Roskam to consider the distance model as probably inappropriate for preference and other types of two-mode data: "It will be noted that the points are more or less on the perimeter of the ellipse. Arrangements like these are encountered often …. The space appears to have an empty region. This may contradict the assumptions of the distance model …. If indeed the space cannot be filled, one must reject the distance model as an adequate theory in such cases" (Roskam, 1968, p. 75).

*MINIRSA*

Roskam knew the work of Kruskal very well and when working together with Lingoes in Michigan, he extensively compared the algorithms developed by Kruskal (M-D-SCAL, see Kruskal & Carmone, 1969) and by Guttman and Lingoes (SSA). This collaboration resulted in a monograph supplement of Psychometrika (Lingoes & Roskam, 1973) and in the MINISSA (Roskam & Lingoes, 1970) program. MINISSA is structurally equivalent to the program developed by Kruskal, but uses a hybrid computational approach to the minimization problem, involving techniques originated by both Kruskal and Guttman: On the one hand, the optimally transformed data are found using the monotone regression procedure introduced by Kruskal, and on the other hand, coordinates are found using the (adapted) C-matrix method of Guttman (1968), which assures convergence (as proven by de Leeuw & Heiser, 1977). So the strengths of both algorithms are combined in the MINI series. Other MINI algorithms were constructed, including MINIRSA for the analysis of off-diagonal matrices (published under Roskam's name, as mentioned by Lingoes & Roskam, 1973).

An aspect of MINIRSA that is worth mentioning, is the importance attached to the choice of the initial configuration. One reason for this importance is to avoid degenerate solutions (Lingoes & Roskam, 1973, p. 8). We analyzed the breakfast data with default values for the MINIRSA program that can be downloaded from the MDS(X) site at `http://www.newmdsx.com/mini-rsa/minirsa.htm`. The resulting configuration is depicted in Figure 2.1. The breakfast items are represented by the letter codes. The descriptions of the labels are presented in Table 2.1. The respondents are represented by dots. This configuration is a near-degenerate solution of the objects-sphere type as most of the breakfast items lie on a circle centered around a lot of judges.
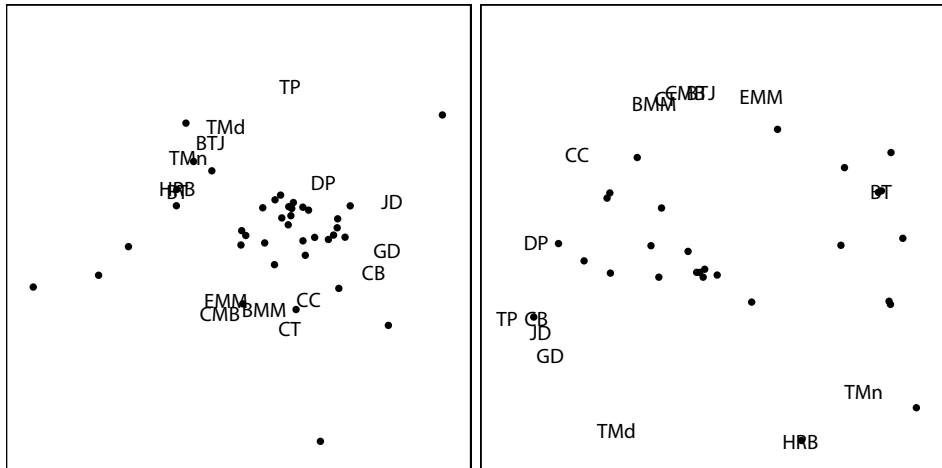
Figure 2.1    MINIRSA solution for the breakfast data (left-hand panel) and KYST unfolding solution of the breakfast data using Stress-2 and a rational start (right-hand panel).

## 2.4    KRUSKAL AND CARROLL, 1969

Kruskal made a major contribution to the domain of multidimensional scaling in general by formalizing the work of Shepard, and more particularly by introducing the STRESS function. His 1964 papers concerned multidimensional scaling but later, together with Carroll, he also considered the unfolding case, for which Carroll, in more than one occasion, laid down the taxonomy. Carroll defined a degenerate configuration as a nominal perfect solution, one that is guaranteed to yield zero STRESS independent of the data. Carroll states: "Thus it follows that the only way in which a nonmetric analysis of any off-diagonal matrix should be done is to split by rows (i.e., treat the matrix as conditional, even if it is not) and use STRESSFORM2", which summarized the 1969 publication which we will discuss here below.

*"Geometrical models and badness-of-fit functions"*

R-STRESS, as defined in (2.1), is dependent on the size of the configuration; shrinking the configuration will decrease STRESS. Initially, Kruskal and Carroll proposed two normalizing factors: The sum of squared distances and the variance of the distances (Kruskal, 1964a). In his 1964 papers, Kruskal chose the first factor and STRESS was there defined by

$$\text{STRESS-1} = \sqrt{\frac{\sum_{i<j}(\gamma_{ij} - d_{ij})^2}{\sum_{i<j} d_{ij}^2}}.$$

Later on (Kruskal & Carroll, 1969), the two STRESS functions based on different normalization functions were compared. It is at this point that the names STRESS FORMULA ONE (STRESS-1) and STRESS FORMULA TWO (STRESS-2) were introduced. Kruskal's STRESS-2 is exactly the same formula as the one proposed by Roskam expressed in (2.2). The use of STRESS FORMULA TWO was recommended to avoid trivial solutions in the unfolding case. Unfolding was not yet really seen as a special case of multidimensional scaling: "In a situation which *closely resembles* [emphasis added] unfolding, namely where the only dissimilarities which have been observed are between objects of two different types and no dissimilarities have been observed between the objects of each type." (Kruskal & Carroll, 1969, pp. 661–662). Unfolding was clearly presented as a special case of multidimensional scaling by Kruskal and Shepard in their 1974 paper, where it was named the so-called 'off-diagonal rectangular sub-matrix generalization'.

The preference for STRESS-2 was motivated by the following observation: A two-point solution where all subjects fall together in one point and all objects fall together in another point (see Figure 4.1) would have a STRESS-1 equal to zero giving a trivial solution with a perfect fit. With STRESS-2 this configuration cannot occur. In the same paper, Kruskal and Carroll stressed the importance of calculating STRESS-2 for each judge separately with separate monotone regressions for each judge (row-conditional) and taking the mean of these values as an overall badness-of-fit measure. In case that the denominator in (2.2) would be replaced by a summation of the individual variances, another trivial solution is possible: The two-plus-two-point solution where all judges except one fall together and all objects except one fall together (the so-called two-plus-two-point configuration, see Figure 4.1). Note that this situation differs from the one where the denominator is set equal to the variance calculated over all subjects: In this case, an object-point trivial solution will occur, as mentioned in Section 2.3 on Roskam.

In spite of all these precautions (and others, like taking the square root over the mean squared STRESS-2 instead of taking the mean of STRESS-2), degenerate solutions could not be avoided: "Our personal belief is that our badness-of-fit function is still not the right one to use in this situation. We are looking for some mathematically satisfying way of changing it which would appear to provide a way out. So far we have not been able to find it." (Kruskal & Carroll, 1969, p. 670).

*KYST*

KYST is a program for multidimensional scaling and unfolding analysis. It represents a merger of M-D-SCAL, the first program(s) written by Kruskal to perform multidimensional scaling, and TORSCA (F. W. Young & Torgerson,

1967). The program and an accompanying manual can be downloaded from the NETLIB site at `http://www.netlib.no/netlib/mds/`. Here we used it to perform an unfolding analysis of the breakfast data. The initial configuration was obtained with a classical Torgerson Scaling. The resulting configuration is depicted in Figure 2.1: The breakfast items approximately lie on a circle with a lot of subjects situated in the center. This is a near-degenerate solution of the objects-sphere type.

## 2.5 LINGOES, 1977

The major contribution of Lingoes to the domain of multidimensional unfolding is formed by the computer programs he developed together with Guttman and with Roskam. In collaboration with Guttman, he developed the Guttman-Lingoes, or G-L, series of programs, which include programs for multidimensional scaling or *smallest space analysis* (SSA), but also for multidimensional scalogram analysis (MSA) and for conjoint measurement (CM). Among these programs, we find an early unfolding program, SSAR-II , a program for the smallest space analysis of "off-diagonal rectangular sub-matrices involving the much weaker constraints of maintaining order information within rows (columns) only" (Lingoes, 1966, p. 322). Later, Lingoes and Roskam developed MINISSA and MINIRSA (see Section 2.3 on Roskam). Lingoes explicitly contributed to the problem of degenerate solutions by developing an approach based on the idea of completing the MDS matrix (Lingoes, 1977) This publication will be discussed in the next subsection, although it should be noted that it is based on a reprint of material published by the Centre National de la Recherche Scientifique that must have appeared in 1971 or 1972, as informed to us by the CNRS. We did not find the original publication, however.

*"A general nonparametric model for representing objects and attributes in a joint metric space"*

Within the nonmetric G-L program series our special interest goes to the programs that handle extended data matrices (Lingoes, 1977), the G-L SSAP series. The initial data matrix is either a score matrix (ordinal) or an attribute matrix (binary). A square symmetric matrix, suitable for MDS (multidimensional scaling), is obtained by measuring the association between the row elements and also between the column elements. For example, the similarity between two judges can be measured by calculating the Spearman rank correlation between their rankings. In this way, the matrix of *between-subject* dissimilarities can be derived from the preference data. Lingoes proposed to use the same measure to derive the matrix of *between-object* dissimilarities. Joining the two derived matrices to the preference scores matrix yields a *super-matrix*
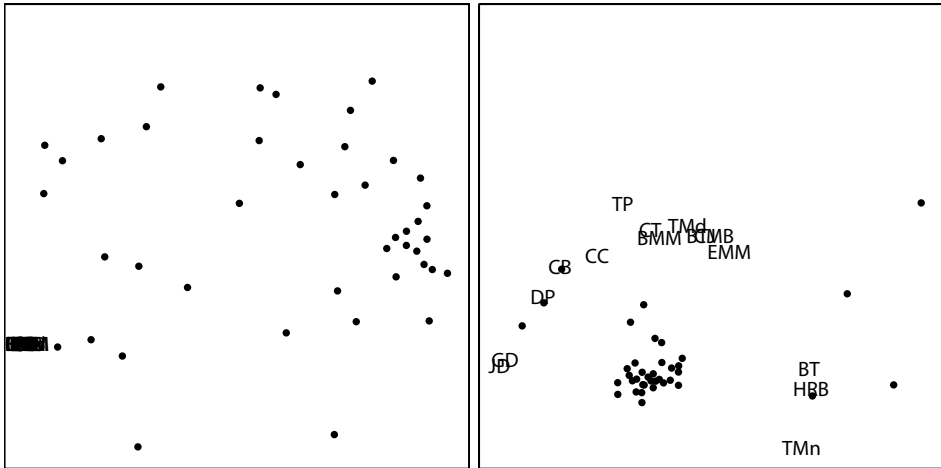
Figure 2.2   SSAP-II unfolding solution of the breakfast data (left-hand panel) with all breakfast items in a clutter of black ink on the left side and a mixed ordinal-interval unfolding solution of the breakfast data (right-hand panel).

of *conjoined matrices* which "retain all of their separate properties in respect to order-ability and comparability" (Lingoes, 1977, p. 481). This means that the two diagonal blocks are treated as matrix conditional while the off-diagonal block is treated as row-conditional. No comparisons are made between blocks. Note that Lingoes proposed this approach as a means to solve the problem of degenerate solutions. He conjectured that for techniques that only use "inter-set information, the solutions may at times be so weakly constrained that patterning is either lost or obscured or even degeneracy may result in some cases" (Lingoes, 1977, p. 480).

*SSAP-II*

We illustrate the SSAP-II program with the breakfast data. As we did not find the original program, we wrote one following the guidelines in Lingoes (1977): The loss function is R-STRESS, normalized by the sum of squared distances, which, following Lingoes is minimized in an iterative and alternating way where the transformed data are computed by using the rank-image approach and where the coordinates where computed by using the Guttman transform. As a measure of association, we used Spearman's rank correlation. With a rational start, we obtained after 100 iterations the configuration depicted in Figure 2.2. This is clearly an object-point degeneracy, where the items are clustered at the bottom-left of the plot.

## 2.6 HEISER, 1981

Heiser started working on algorithms and (restrictions in) multidimensional scaling and unfolding in the late seventies, collaborating with de Leeuw (de Leeuw & Heiser, 1977; Heiser & de Leeuw, 1979b; de Leeuw & Heiser, 1980, 1982). A convergent multidimensional scaling algorithm was developed, based on work of Guttman (see de Leeuw & Heiser, 1977), using an iterative majorization approach: "This algorithm is an improvement over ALSCAL in two major ways (a) It is simpler, faster, and more elegant; and (b) the algorithm fits distances instead of squared distances, which is more desirable. ..., [it] will become the least squares program of choice, particulary if made available in a major statistical system." (F. W. Young, 1987, p. 33).

A convergent *unfolding* algorithm was laid down in Heiser (1987a), based on earlier work in de Leeuw and Heiser (1977); Heiser and de Leeuw (1979b) and de Leeuw and Heiser (1980), but his first attempt to overcome the degeneracy problem appeared in his dissertation (Heiser, 1981). This comprehensive work on unfolding discusses many topics, of which 'restrictions on the transformations' discusses a procedure to overcome the degeneracy problem.

*"Unfolding analysis of proximity data"*

In his dissertation, Heiser showed that a nonmetric algorithm is biased towards transformations that render equal transformed proximities and concluded that the solution space for ordinal transformations in nonmetric unfolding is too big: "We shouldn't have made these cones that big in the first place" (Heiser, 1981, p. 221), a similar conclusion as Lingoes (1977) when he mentioned 'weakly constraint unfolding solutions'. A flat transformation, that is, a degenerate solution, should be avoided by tightening up the cones, i.e., by restricting the solution space. Heiser decided to explore *bounded monotone regression*, which defines a smaller class of 'smooth' functions. For this purpose, Heiser defines lower bounds ($\alpha$) and upper bounds ($\beta$) for the transformed data ($\Gamma$), based on the raw data ($\Delta$) with the smallest dissimilarity set to zero (Heiser, 1981, p. 223), such that $\beta(\delta_l - \delta_{l-1}) \geqslant \gamma_l - \gamma_{l-1} \geqslant \alpha(\delta_l - \delta_{l-1})$. With $\alpha = 0$ and $\beta = \infty$, this reduces to an ordinary monotone regression problem with non-negativity restrictions, and with $\alpha = \beta = 1$, it reduces to metric unfolding. For reasons of symmetry, Heiser chose $\beta = 1/\alpha$ with $0 \leqslant \alpha \leqslant 1$, making $\alpha$ smaller means a bigger cone, and introducing degeneracies for $\alpha \to 0$. To determine an optimal $\alpha$, Heiser realized that, although minimizing a variant of STRESS-2 led to certain degenerate solutions, *not* minimizing this function, but computing it as a separate statistic along with the minimization of R-STRESS, may provide a sensitive measure of degeneracy, a measure that can at least be employed to define an optimal value for $\alpha$, if there are no other grounds to choose. Heiser

showed that bounded monotone regression can be successfully employed, and he did so on multiple data sets. "Thus it seems that the bounded regression approach enables us to avoid the non-informative circles and spheres which pop up all the time with ordinary unfolding programs. Maybe it should be emphasized that we did not really 'solve' the problem, in the sense of improving technical aspects of the algorithm. We simply *defined another problem*, which we solve, but which lacks the elegance of uniqueness" (Heiser, 1981, pp. 230–331).

*SMACOF-3b*

The algorithms originating from de Leeuw and Heiser (1977) are implemented in a series of programs called SMACOF , acronym for scaling by majorizing a complex function (see also de Leeuw & Heiser, 1980). The metric unfolding variant is called SMACOF-3 (Heiser & de Leeuw, 1979a, 1979b), whereas the *nonmetric* multidimensional unfolding spin-off is called SMACOF-3b (Heiser, 1987b). Unfortunately, the code doesn't exist anymore. An example of the successor of bounded monotone regression will be shown in Section 2.10.

One interesting option in SMACOF-3 is the centroid start, where the column objects are restricted to be in the centroids of those rows objects that have the highest preferences for those particular column objects. These restrictions are only used to provide better initial configurations (Heiser & de Leeuw, 1979a), following Lingoes and Roskam (1973, p. 8), or to provide better interpretation (Heiser, personal communication, May 18, 2005). The centroid restrictions, however, are an extreme case of an approach further developed in quite a different way by DeSarbo and Rao (1984, p. 155) and as such also applicable to avoid degeneracies.

## 2.7   BORG AND BERGERMAIER, 1982

Borg is the author and editor of several books on multidimensional scaling and the author of a number of journal papers on the same topic. Within this domain, his focus is on facet theory, applied problems, and the scaling of individual differences. One of his papers, co-authored by Bergermaier (see Borg & Bergermaier, 1982, but also, Borg & Groenen, 2005, Chapter 14), deals with the problem of degenerate solutions in unfolding: A solution is proposed that is based on a mixed ordinal-interval approach.

*"Degenerationsprobleme im Unfolding und Ihre Lösung"*

Borg and Bergermaier (1982), who applied KYST to minimize STRESS-2, observed that ordinal unfolding may yield degenerate solutions and that interval

unfolding may yield the wrong slope, that is, more preferred items are more distant in the configuration, an artefact that can be avoided by using non-negative least squares. They proposed, however, to use a hybrid ordinal-linear approach: "Ordinal unfolding guarantees that the regression line has the right slope, while interval unfolding succeeds in avoiding degeneracies. Thus, it appears natural to combine both models into a hybrid model." (Borg & Groenen, 2005, p. 249). Such a hybrid model can be realized by minimizing

$$\textsc{stress-2}^{\text{hybrid}} = \alpha \times \textsc{stress-2}^{\text{ordinal}} + (1 - \alpha) \times \textsc{stress-2}^{\text{interval}} \quad (2.3)$$

with $0 \leqslant \alpha \leqslant 1$. This type of loss function can be minimized with KYST: "Sometimes it is desirable to do a scaling or an unfolding using linear (or polynomial) regression, but it is necessary to assure that the regression function is essentially monotone over the region containing the data values. While KYST cannot manage quite this, it can approximate it." (Kruskal, Young, & Seery, 1978, p. 28).

*Mixed ordinal-interval approach (KYST)*

We used the hybrid model proposed by Borg and Bergermaier (1982) to unfold the breakfast data. To attain this goal, we used KYST for a mixed ordinal-interval row-conditional unfolding, with $\alpha = 0.5$, minimizing (2.3). The resulting configuration is plotted in Figure 2.2: Although the solution is not completely degenerate, it still is difficult to interpret and tends to a degeneracy of the objects-sphere type. This partially degenerate solution comes as no big surprise as, in the mean time, it is known that even unfolding with an interval transformation may lead to degenerate solutions (see Chapter 3). Borg and Groenen (2005) changed the ordinal-interval approach into a working ordinal-ratio approach, due to the fixed (zero) intercept of the ratio transformation.

## 2.8   DE LEEUW, 1983

Early contributions of de Leeuw to MDS were to the development of algorithms, with special attention for convergence properties (de Leeuw, 1977a; de Leeuw & Heiser, 1977, 1980; Takane, Young, & de Leeuw, 1977). This has led to the ALSCAL (Takane et al., 1977; F. W. Young & Lewyckyj, 1979) and SMACOF algorithms (see Section 2.6 on Heiser) that both guarantee monotone convergence of the loss function values. In de Leeuw, 1977a, a convergence proof was given for an MDS algorithm that defines loss on the untransformed distances, and not, as is the case with ALSCAL, on the squared distances. The metric version of this algorithm turns out to be identical to Guttman's C-matrix method (see Guttman, 1968). A convergent nonmetric algorithm was then obtained by

combining the metric step with monotone regression. De Leeuw (1983) made an important contribution to unfolding by proving that even the use of a smart loss function, such as the conditional version of STRESS-2, cannot prevent the occurrence of degenerate solutions. In fact, this paper was the first one to formally prove how problematic the approach to unfolding as a special form of multidimensional scaling is: "The conclusion is that nonmetric unfolding, as currently formalized, is an inherently ill-posed problem and that a different approach is called for." (de Leeuw, 1983, p. ii).

*"On degenerate nonmetric unfolding solutions"*

In de Leeuw, 1983, first an overview is given of the different STRESS functions that have been used in unfolding analysis. The construction of these functions was led by one principle, namely avoiding trivial solutions that occur by making loss undefined (that is, equal to $0/0$) at these trivial solutions. With this objective in mind, the conditional version of STRESS-2 was introduced both by Roskam (1968) and Kruskal and Carroll (1969). No trivial solution was found for STRESS-2, but degenerate solutions appeared often (and, as communicated by Heiser, in de Leeuw, 1983, p. 5, one may wonder if the non-degenerate solutions that were reported are suboptimal solutions, in the sense that they were obtained with too few iterations). De Leeuw made a clear distinction between *trivial* and *degenerate* solutions: Trivial solutions have zero STRESS, are not interpretable, and can be avoided by a proper normalization, while degenerate solutions have often non-zero STRESS, are not interpretable, and *cannot* be avoided by a proper normalization. De Leeuw (1983, p. 5) showed that "the whole idea of hoping that a clever choice of the denominator solves all problems is basically unsound. There is no reason at all why the iterative process should keep away from $0/0$."

   The formal proof of the problem can be described briefly as follows. De Leeuw started from trivial solutions like the objects-circle to which he added small perturbations. He then proved two theorems by using l'Hospital rule to study the behavior of STRESS-2 along differentiable paths in the neighborhood of trivial solutions. The first theorem makes clear that, when the perturbations decrease to zero (this is, the solution converges to the trivial solution), stress converges to a finite value and not to $0/0$, such that the solution is not steered away from the trivial solution. The second theorem shows that a configuration can be found arbitrarily close to a trivial solution with arbitrarily small derivatives, or, the function can converge to a minimum in the very near neighborhood of a trivial solution.

   By studying the behavior of STRESS-2 in the neighborhood of frequently occurring degeneracies of the objects-circle, object-point and two-point type, de Leeuw showed that respectively a vector model, a signed compensatory
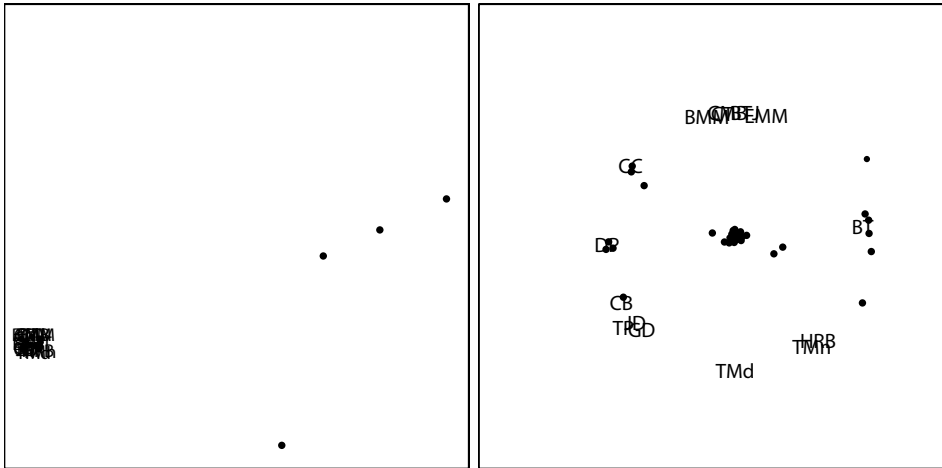
Figure 2.3    Configuration for the breakfast data when minimizing Stress-2 with a convergent algorithm. The right-hand panel depicts a detail of the left-hand panel.

distance model, or a row-conditional version of the additive model is fitted. This paper, originally a technical report of the Department of Data Theory, University of Leiden, May 1983, can be obtained at `http://repositories .cdlib.org/uclastat/papers/2006010109/`, UCLA, Department of Statistics Papers.

*Application: Breakfast data*

We illustrate the statements made by de Leeuw for the unfolding analysis of the breakfast data. Here, we used an algorithm that minimizes STRESS-2 by an alternation between monotone regression and an update of the coordinates based on iterative majorization (van Deun, Groenen, Heiser, Busing, & Delbeke, 2005): In practice, STRESS-2 is decreased in each step. The resulting configuration is depicted in Figure 2.3: For most of the subjects, it shows the same type of objects-circle degeneracy found previously when minimizing STRESS-2 with MINIRSA and KYST. Conform to de Leeuw (1983), we found a configuration that is partially degenerate, with a few subjects that are distant in the configuration which corresponds to a vector model representation (the left-hand panel of Figure 2.3), and with the breakfast items on a circle where the center is formed by most of the subjects which corresponds to a signed compensatory model representation (the right-hand panel of Figure 2.3).

DeSarbo wrote his doctoral dissertation (DeSarbo, 1978), an unpublished memorandum (DeSarbo & Carroll, 1983), and several articles on (weighted) least squares unfolding (DeSarbo & Carroll, 1980; DeSarbo & Rao, 1984; DeSarbo & Carroll, 1985). In these publications, DeSarbo describes two related models: Two-way unfolding (DeSarbo & Rao, 1984) and three-way unfolding (DeSarbo & Carroll, 1985) models. In these papers, weighting is suggested as a mean to avoid degenerate solutions.

From 1986 on, as far as unfolding is concerned, DeSarbo specializes in probabilistic multidimensional unfolding models, threshold models (DeSarbo & Hoffman, 1987), and maximum likelihood estimation for paired comparison data, (asymmetric) binary choice data, and pick any/j data (DeSarbo & Cho, 1989). Degeneracy in unfolding is not an issue for some time, until his cooperation with Kim and Rangaswamy (Kim et al., 1999).

*"GENFOLD2: A set of models and algorithms for the GENeral unFOLDing analysis of preference/dominance data"*

DeSarbo and Rao (1984) is the first published version of DeSarbo (1978), although it was already in an AMA proceedings article in 1979 (personal communication, 2005), and in DeSarbo and Carroll (1983), and fully published in DeSarbo and Carroll (1985). DeSarbo and Rao (1984) describe a general set of unfolding models for analyzing two-way preference or dominance data. The set contains many models or options, such as, internal and external unfolding (Carroll, 1972), constrained and unconstrained analysis (see also de Leeuw & Heiser, 1980), conditional and unconditional as well as metric and nonmetric transformations, and simple, weighted, and generalized unfolding models (Carroll & Chang, 1967). The objective function for all models is weighted R-STRESS with squared distances and the function is minimized by alternating weighted least squares. The three-way variant only estimates the metric unfolding model (DeSarbo & Carroll, 1985).

In order to avoid degeneracy, DeSarbo and his co-authors propose to use weights for the data. Since the possible cause of degeneracy is considered to be the error in the data, dissimilarities are allowed to be weighted depending on their reliability. For ratio data, the weights may be defined as $w_{ij} = \delta_{ij}^{-p}$, whereas for interval and ordinal data the weighting function might be more meaningful using the (row) ranks of the data $r(\delta_{ij})$ instead of $\delta_{ij}$, or bimodal or step weighting functions might be specified (DeSarbo & Rao, 1984, p. 156). Both the weighting function and the value of $p$ can be made by trial and error, i.e., the choice of $p$ and the accompanying function is a empirical issue, depending upon the data.
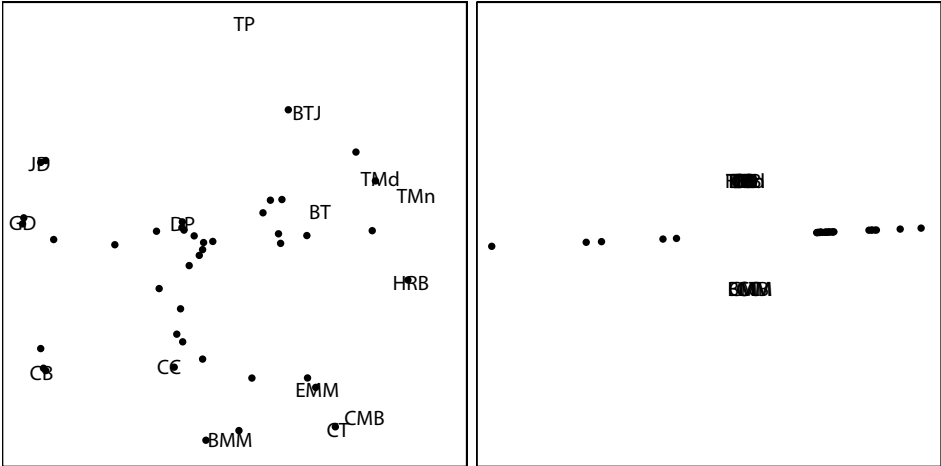
Figure 2.4    GENFOLD solution with p=2 for the breakfast data minimizing Stress-2 (left-hand panel) and Stress-1 (right-hand panel).

In a Monte Carlo study, evidence was provided for the robustness of the methodology, although a proper, more extensive MC study has yet to be done. Nevertheless, applications with Pain Reliever Preference Data, Residential Communication Devices Data, Reading Profile Data, and the Miller-Nicely Data show that the GENFOLD procedure is able to provide interpretable configurations, without a general form of the weights and a trial-and-error choice of p.

*GENFOLD-2*

GENFOLD-2 (DeSarbo & Rao, 1984) (Kim, Rangaswamy, & DeSarbo, 1999, even mention a GENFOLD) was never made publicly available, but the loss function is simple enough to be minimized with another unfolding program, of which we have chosen KYST, as long as data weighting is available. In KYST, the weights can be specified as a function of the data, such that it conforms to $w_{ij} = \delta_{ij}^{-p} = r(\delta_{ij})^{-p}$, as the breakfast data contain complete rank order information for each row, and with $p = 2$. In Figure 2.4, left-hand panel, the unfolding solution for the breakfast data is obtained for STRESS-2. Although DeSarbo and Rao (1984, p. 168) mention that the appropriate loss function to be used in the case of non-metric analyses is STRESS-2, Figure 2.4, right-hand panel, shows the solution obtained with STRESS-1. This allows us to differentiate between non-degeneracy due to the specific weighting function (as proposed by DeSarbo & Rao, 1984) and non-degeneracy due to normalization on the variance (as

proposed by Roskam, 1968). Clearly, without the latter, weighting the data is not enough to prevent degeneracy.

## 2.10 HEISER, 1989

Heiser (1981) showed that bounded monotone regression offered a way out of non-informative circles and spheres, but introduced unwanted additional parameters. In the years following his dissertation, Heiser continued working on this problem, which finally leaded to a smooth monotone regression procedure (Heiser & Meulman, 1983b; Heiser, 1985, 1986, 1987b, 1989): Bounded monotone regression with *internal bounds*.

*"Order invariant unfolding analysis under smoothness restrictions"*

Already in his dissertation, Heiser realized that the two additional parameters for the bounded monotone regression were a nuisance. Although flexible, the detailed manipulation of the bounds was not attractive for a general procedure or strategy. Instead of external or user-specified bounds, Heiser searched for more natural or internally determined bounds, and found them in the form of a *mean step*. Details on computation, treatment of ties, and application of this approach to square symmetric nonmetric multidimensional scaling can be found in Heiser (1985). In later publications, the procedure is applied to the unfolding case (cf. Heiser, 1986, 1987b, 1989). The general idea of smooth monotone regression, as bounded monotone regression with the mean step is called, is the following. Assume there is only one vector with dissimilarities to be transformed and the dissimilarities are in increasing order. While monotonicity is a condition on the first order differences, i.e., $\gamma_l - \gamma_{l-1} \geqslant 0$, smoothness is defined as a condition on the second order differences, as $|\theta_l - \theta_{l-1}| \leqslant \overline{\theta}$, where $\theta_l = \gamma_l - \gamma_{l-1}$ and $\overline{\theta}$ is the *mean step*. In words: Each step may not deviate more from the previous step than the mean step. Even with this smoothness restriction, considerable amounts of nonlinearity, such as quadratically and logarithmically increasing values, are still possible. The technical report further describes the treatment of ties and discusses algorithmic considerations, such as the use of explicit normalization on the transformed proximities and the switch to a faster minimization strategy. This last improvement can not diminish the huge computational burden of the smooth monotone regression procedure, which in those days already became overwhelming for 25 objects, i.e., for $(25 \times 24)/2 = 300$ dissimilarities.

   Heiser (1989) described different forms of degenerate solutions and why these solutions occur so often in unfolding. Normalization on the variance seems to be the best choice, but not for an unconditional transformation of the data. With row-conditional transformations, even when using the
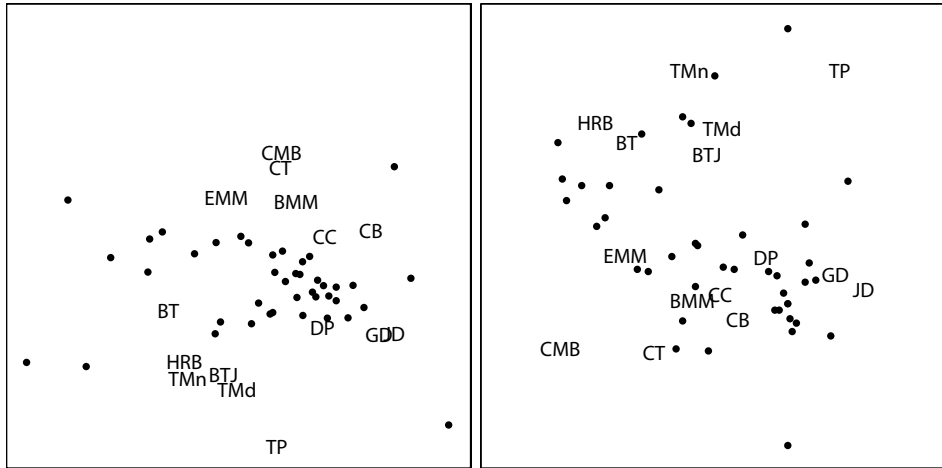
Figure 2.5    SMACOF-3b solution for the breakfast data (left-hand panel) and NEWFOLD solution for the breakfast data (right-hand panel).

variance normalization (per row), degenerate solutions occur and can take all kinds of forms. Smooth monotone regression can be used to avoid "a distance distribution in which all mass is concentrated at one or two values" (Heiser, 1989, p. 15) and doesn't even 'need' the variance normalization. For the applications to the study of the 1960 presidential campaign (Sherif, Sherif, & Nebergall, 1965) and power in the classroom (Gold, 1958), "use was made of the FORTRAN program SMACOF-3b, which has been designed to minimize the NORMALIZED RAW STRESS under the smoothness restrictions, with the sum of squared transformed proximities as the NORM" (Heiser, 1989, p. 19).

### SMACOF-3b

Although SMACOF-3b does not exist anymore, PREFSCAL (Busing, Heiser, Neufeglise, & Meulman, 2005) is used here to perform the unfolding analysis with smooth monotone regression. PREFSCAL, with the penalty function incorporated in the algorithm disabled, uses an identical minimization function as SMACOF-3b: PREFSCAL uses implicit normalization instead of explicit normalization and a slightly different update algorithm (see Technical Appendix B). In Figure 2.5, the solution for the breakfast data is obtained for NORMALIZED RAW STRESS with smooth monotone regression. The solution does not appear to be degenerate, but it took more than 2 minutes with the default convergence criteria and more than 30 minutes with the strictest convergence criteria (with the ordinary monotone regression procedure, it took about 0.4 seconds).

The main idea of Kim's dissertation (Kim, 1990) was published as Kim, Rangaswamy, and DeSarbo (1999) and presented earlier at two marketing science conferences in 1989 and 1990. Concerning multidimensional scaling and unfolding, there were no further publications by these authors, although there is work in progress in the field of external unfolding (FIXFOLD): A MICROSOFT WINDOWS version of FIXFOLD is released in the next release of the Marketing Engineering software at `http://www.mktgeng.com`.

*"A quasi-metric approach to multidimensional unfolding for reducing the occurrence of degenerate solutions"*

Kim, Rangaswamy, and DeSarbo (1999) describe an approach that reduces the occurrence of degenerate solutions. A non-degenerate solution is characterized by intermixedness of both sets of objects in the configuration. Such a configuration is pursued by "maximally differentiating the point in the joint space while, at the same time, maintaining correspondence as closely as possible to the rank order of the preference" (Kim et al., 1999, p. 150), i.e., by maximizing the preference differentials, the differences between consecutive preferences. In the model, the preference differentials are bounded between lower and upper limits, which are implicitly incorporated in the objective function and in the scaling algorithm. "To prevent degeneracy and assure intermixedness, our algorithm uses the raw data to set up an *a priori* matrix of target distances between the ideal point and the stimulus points, to be satisfied by a resulting configuration. This matrix, denoted as $\Delta$, incorporates the lower bound on the preference differentials, …" (Kim et al., 1999, p. 152). This part of the approach is similar to the bounded regression approach used by Heiser (see Section 2.6).

The target set of distances for nonmetric unfolding can be equally spaced or unequally spaced. The equal option conforms to a linear function between differentials and distances, while the unequal option specifies the differentials randomly (drawn from a normal distribution), with increased spacing depending on the preference ranks. The target set of distances for metric unfolding simply sets the lowest value of the data equal to zero, by lowering the data with its minimum value. After this *a priori* transformation of the data, Kim et al. continue with a row-conditional metric analysis without estimating an additive constant.

A Monte Carlo simulation study shows that, "on average, *the proposed model dominates the competing models on all measures*" and "is generally robust across a number of experimental factors" (Kim et al., 1999, pp. 160–163). In two applications, consumer studies of preference for MBA programs

and analgesic preference, the proposed procedure performs better than KYST, ALSCAL, SMACOF-3, and GENFOLD-3, and "does not appear to be trading off as much preference recovery with non-degenerate solutions" (Kim et al., 1999, pp. 163–172).

*NEWFOLD*

NEWFOLD is described in Kim et al. (1999) and specific details of the optimization procedure can be found in Kim (1990). NEWFOLD is a DOS program and will only run in console mode under MICROSOFT WINDOWS. It handles both metric and non-metric data for both internal and external unfolding. The starting configuration for the program can either be rational for stimuli, random, or user-provided. After the a priori computation of the target set of distances, a conjugate-gradient method with restarts is used to find the optimal locations for the respondent and product points. In Figure 2.5, the solution for the breakfast data is obtained for NEWFOLD with the unequal option for nonmetric data and a rational starting stimuli configuration. The points of both sets are well intermixed, but the trade-off with correspondence (between data and differentials) is bound to provide solutions with worse fit statistics than solutions with optimally transformed data.

## 2.12   SUMMARY

We can summarize this history on the degeneracy problem in multidimensional unfolding as follows. Roskam, who was the first one to systematically investigate multidimensional unfolding with a sound algorithm, discovered the need to adapt the STRESS function used in multidimensional scaling, in order to avoid trivial solutions like the objects-sphere. This has led to the development of what is currently known as STRESS-2, a function characterized by a normalization on the variance of the distances per row of the data matrix. Both Roskam (1968) and Kruskal and Carroll (1969) reported disappointing results as solutions kept popping up that are degenerate: Although the loss function effectively avoids trivial solutions, very often configurations were found that highly resemble the trivial solutions found with inappropriate normalizations of the STRESS function (STRESS-1). In 1983, de Leeuw proved that minimizing STRESS-2 is no guarantee against degeneracies as it does not stir solutions away from trivial solutions. He also showed that the unfolding algorithms are even attracted to the neighborhood of trivial solutions, and that in these neighborhoods other models are fitted.

Early on, the realization that adaptations of the loss function did not yield the desired results, drew attention to other aspects of the multidimensional unfolding procedure. Lingoes, probably in 1971 (see Lingoes, 1977), turned

his attention to the fact that the unfolding problem is weakly constrained by only supplying inter-set information: To avoid degeneracies, he constrained the data more by completing the diagonal blocks of the matrix that forms the input of the multidimensional scaling analysis. The weakness stems from the monotone regression procedure: It is based on averaging over distances that are not in the proper order, with the result that a lot of transformed data become equal. These equal data, in turn, result in a lot of equal distances, a situation that is typical of degenerate solutions. In this framework, the work of Heiser (1981, 1989), Borg and Bergermaier (1982), and Kim et al. (1999) is situated, who all restrict the transformation in some way, making the weakly constrained unfolding problem less weak.

A last aspect of the unfolding situation that was considered, is found in the work of DeSarbo and Rao (1984) who used weights to reduce the error in the data, which they thought to be responsible for degeneracies.

## 2.13   RECENT DEVELOPMENTS

The search for non-degenerate solutions did not end with the contribution of Kim et al. (1999). Nevertheless, recent developments are tributary to the ideas developed in the contributions we discussed here.

The work of Lingoes (1977) is in line with Steverink, Heiser, and van der Kloot (2002), Borg and Groenen (2005) and van Deun, Heiser, and Delbeke (2007), who developed unfolding as a multidimensional scaling analysis of a completed super-matrix. In line with Lingoes (1977), van Deun et al. (2007) proposed a MDU technique that relies on a MDS analysis of a completed super-matrix. Because a block conditional approach yields degenerate solutions, they stressed the need of comparable dissimilarities such that an unconditional MDS analysis is warranted. Their proposal is used as one of the possible initial configurations for PREFSCAL. A MATLAB procedure can be obtained from katrijn.vandeun@psy.kuleuven.be. Except for extending the work of Lingoes (1977), Borg and Groenen (2005) also correct the approach taken in Borg and Bergermaier (1982). Since it is now known that unfolding with interval transformations can also lead to degenerate solution (see Chapter 3), the ordinal-interval approach is replaced with an ordinal-ratio approach.

The work of de Leeuw (1983) is extended by van Deun, Groenen, Heiser, et al. (2005). They illustrate how degenerate solutions are informative and fit the data well, and how these solutions can be made interpretable by resorting to another type of representation than a distance type. The insight of de Leeuw (1983) that a vector model is fitted in the neighborhood of an object point degeneracy and a research suggestion from DeSarbo and Rao (1984), inspired van Deun, Groenen, and Delbeke (2006) to solve the occurrence of (some) degeneracies by using a hybrid vector ideal point model for the representation

of preference data. To solve their model, a least squares loss function was introduced and they developed an accompanying algorithm called VIPSCAL , acronym for vector ideal point scaling. VIPSCAL is available in MATLAB from `katrijn.vandeun@psy.kuleuven.be`.

In the same tradition as Borg and Bergermaier (1982), Busing (2006) proposes to adjust the transformation function, by adding a penalty on an unwanted high intercept to force an uphill slope for the transformation(s) to avoid degenerate solutions. This idea is further developed in Chapter 3. Finally, there is the work of Roskam (1968), Kruskal and Carroll (1969), Heiser (1981, 1989), and Groenen (1993) that has inspired Busing, Groenen, and Heiser (2005) to develop a PENALIZED STRESS function to overcome the degeneracy problem. The details of this approach are discussed in Chapter 4.

3

It has long been thought that degeneracy in unfolding only concerned nonmetric unfolding. In the next chapter we will establish that degeneracy occurs for all transformations which include estimation of an intercept and a slope. Consequently, degeneracy also plagues metric unfolding, since one member of the metric transformation family, the interval transformation, includes estimation of both an intercept and a slope. In this chapter, a simple solution is proposed to the degeneracy problem for metric unfolding by penalizing for an undesirable intercept. An application of this approach will illustrate its potential.

## 3.1   INTRODUCTION

Unfolding is a technique that analyzes proximity data between two sets of objects (Bennett & Hays, 1960; Coombs, 1964). Well-known examples of such proximity data are rank orders of breakfast items scored by MBA students and their wives (P. E. Green & Rao, 1972), paired comparison preferences from students for tea, with different tea temperatures and different amounts of sugar (Carroll, 1972), and citation frequencies between scientific journals (Weeks & Bentler, 1982). In all these cases, the proximities can be interpreted in terms of distances between two objects from different sets. As such, small distances should correspond to students and their highest ranked breakfast item, to students and their most preferred cup of tea, or between journals with many co-citations. On the other hand, large distances should be reserved for students and their lowest ranked breakfast items, for students and their least favorable cups of tea, or for journals with no or very few co-citations. Unfolding finds distances that correspond closest to the proximities.

Programs for unfolding are scarce. Programs for unfolding which are able to produce non-degenerate solutions are even more difficult to find (characteristics of (non)degenerate solutions will be described later). The commercially available programs, ALSCAL (Takane et al., 1977) and SYSTAT (Wilkinson, 1999), tend to produce degenerate solutions, although premature termination of the algorithm often provides a solution on its way towards a degenerate solution. GENFOLD (DeSarbo & Rao, 1984) is not available at all, besides, it uses a weighting scheme that is unable to avoid degenerate solutions (see Busing, Groenen,

& Heiser, 2005). The quasi-metric approach employed in NEWFOLD (Kim et al., 1999) avoids trivial solutions, but closer inspection of the algorithm shows that the procedure restricts itself to only one kind of metric transformation, a limited transformation that also avoids degenerate solutions in other programs. Currently, two loss functions are able to avoid degenerate solutions, STRESS-2 (Kruskal & Carroll, 1969) and P-STRESS (Busing, Groenen, & Heiser, 2005), implemented in KYST (Kruskal et al., 1978) and IBM SPSS PREFSCAL, respectively, but these functions need a penalty on the variance or variation to do so, and even then, STRESS-2 doesn't always succeed in avoiding a degenerate solution (see Borg & Groenen, 2005).

In this chapter, a simple but effective solution is proposed to the degeneracy problem in metric unfolding: Penalize the undesirable intercept to avoid identical transformed proximities, a characteristic feature of degenerate solutions. The proposed procedure is easily implemented in mathematical general purpose programs, like MATLAB or S-PLUS, or can be setup in a general MDS program.

In the following, first, an illustrative example is shown of a degenerate solution, followed by a formal description of metric unfolding, which, among others, includes unfolding with an interval transformation. Then, cases will be discussed in which metric unfolding leads to degenerate solutions. Finally, the penalty proposal is formulated in more formal terms and applied to the earlier example to show the benefit of the proposal.

## 3.2    EXAMPLE

For the following unfolding example, Price and Bouffard (1974) asked 52 students from introductory psychology classes at Indiana University to rate the 225 combinations of 15 behaviors and 15 situations on a scale from 0, the behavior is *extremely inappropriate* in this situation, to 9, the behavior is *extremely appropriate* in this situation. The average rates over persons ranged from 0.62, for fighting in a church, to 8.85, for sleeping in your own room. The data values were subtracted from 9.0, the highest possible value, to obtain dissimilarity data so that, a lower score (or small distance) now corresponds to more appropriate behavior in a situation. The results of an unfolding analysis *with an interval transformation of the proximities* is shown in Figure 3.1.

The configuration (left-hand panel) consists of an imaginary circle with the various types of behavior in the center (only indicated by ###) and the situations on the (left side) edge of the circle. The distance from the situations to the behaviors is identical for all behavior in all situations. This fact is also reflected in the transformation plot (right-hand panel), where the horizontal transformation line indicates identical transformed proximities for any of the original proximities. This solution is perfect in terms of fit (after 223
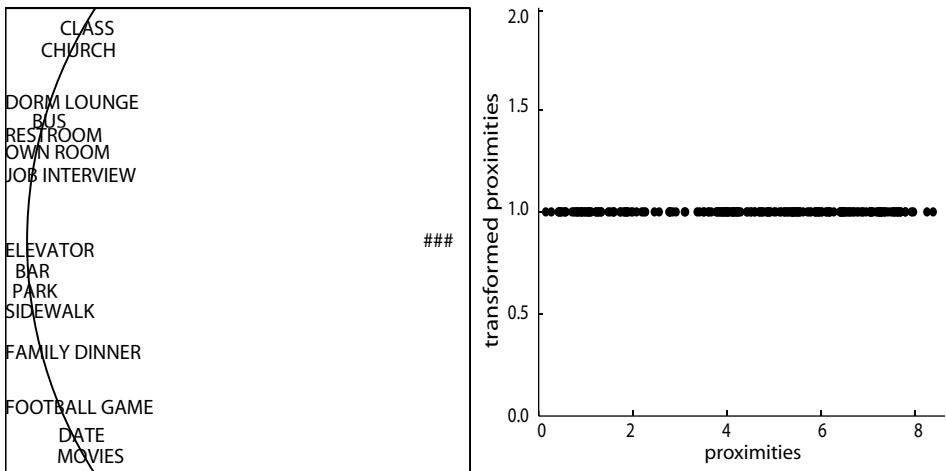
Figure 3.1    Interval unfolding solution for the behavior-situation appropriateness data. The situations are (uppercase): CLASS, DATE, BUS, FAMILY DINNER, PARK, CHURCH, JOB INTERVIEW, SIDEWALK, MOVIES, BAR, ELEVATOR, RESTROOM, OWN ROOM, DORM LOUNGE, and FOOTBALL GAME. The behaviors (in the configuration only indicated by ###): Run, Talk, Kiss, Write, Eat, Sleep, Mumble, Read, Fight, Belch, Argue, Jump, Cry, Laugh, And Shout.

iterations, NMSE = 0.000000, as will become clear later on), but completely useless from an interpretational point of view. This phenomenon is called a *degenerate* solution and can occur with any kind of data, while using an interval transformation.

With identical loss functions, commercially available software programs, like ALSCAL and SYSTAT, produce similar solutions. Although the modest default convergence criteria of these programs terminate the iterative process early, the final configurations are clearly degenerate.

### 3.3    METRIC UNFOLDING

Multidimensional unfolding is a technique that finds low-dimensional configurations for two sets of objects. Here, the proximities $\delta$ between the two sets, given in vector form, are expected to be dissimilarities. The objective of unfolding is to find coordinates for the two sets of objects such that the distances $d$ between the coordinates correspond as closely as possible to the proximities $\delta$, or a transformation $\gamma = f(\delta)$ thereof. Correspondence is usually measured as the mean squared error between $\gamma$ and $d$, that is,

$$\text{MSE}(\gamma, d) = \|\gamma - d\|^2, \tag{3.1}$$

35

where $\| \cdot \|$ represents the Euclidean norm. Since minimization is concerned with both $\gamma$ and $\mathbf{d}$, these quantities are iteratively updated (Shepard, 1962a; Kruskal, 1964a). This update strategy requires some form of normalization to avoid the so-called one point solution (Kruskal & Carroll, 1969). Here, after each iteration, (3.1) is *explicitly* normalized by letting $\|\mathbf{d}\|^2 = nm$, where $n$ and $m$ are the number of objects in each set, although an implicit normalization would also suffice. The consequence of the explicit normalization, which fixes the sum-of-squares of the distances to $nm$, is the dilation of the configuration to a fixed size, which avoids the situation where everything becomes equal to zero, i.e., the one point solution. The normalized version of (3.1), i.e., $\|\mathbf{d}\|^{-2}\|\boldsymbol{\delta} - \mathbf{d}\|^2$, is also known as Kruskal's *Stress-1* and is used as a minimization function in SYSTAT and KYST, whereas a squared version, i.e., $\|\boldsymbol{\delta}^2\|^{-2}\|\boldsymbol{\delta}^2 - \mathbf{d}^2\|^2$, is also known as Young's *S-Stress-1*, one of ALSCAL's loss functions.

The metric transformation function is defined as

$$\gamma = b_1 + b_2 \boldsymbol{\delta}, \tag{3.2}$$

subject to $b_2 \geqslant 0$ to assure a positive linear relation between $\boldsymbol{\delta}$ and $\gamma$ and subject to $b_1 \geqslant -b_2 \delta_{\min}$, with $\delta_{\min}$ as the smallest element of $\boldsymbol{\delta}$, to prevent negative transformed proximities. Using $\widetilde{\boldsymbol{\delta}} = \boldsymbol{\delta} - \delta_{\min}$ subject to $b_1 \geqslant 0$ instead of $\boldsymbol{\delta}$ subject to $b_1 \geqslant -b_2 \delta_{\min}$ amounts to the same problem, but simplifies the estimation of $b_1$ and $b_2$, as described in Appendix 3.A. Despite the non-negativity restriction, (3.2) satisfies the axioms for an interval scale measurement (see Stevens, 1946) and is subsequently referred to as an interval transformation. The term interval unfolding is used to indicate a metric unfolding using an interval transformation of the type specified in (3.2). Fixing either or both $b_1$ and $b_2$ to a constant provides another member of the metric transformation family (see Table 3.1).

Combining (3.1) and (3.2) and adding the explicit normalization factor $\|\mathbf{d}\|^2$ defines the complete objective function for metric unfolding as

$$\text{NMSE}(b_1, b_2, \mathbf{d}) = \|\mathbf{d}\|^{-2}\|b_1 + b_2\widetilde{\boldsymbol{\delta}} - \mathbf{d}\|^2, \tag{3.3}$$

subject to $\|\mathbf{d}\|^2 = nm$, $b_1 \geqslant 0$, and $b_2 \geqslant 0$.

Table 3.1    Metric transformation family.

| Intercept $b_1$ | Slope $b_2$ | |
| --- | --- | --- |
| | Fixed | Free |
| Fixed | no transformation | ratio transformation |
| Free | intercept transformation | interval transformation |

## 3.4   DEGENERACY

Degenerate unfolding solutions occur when transformations are allowed free estimation of both intercept and slope, in which case the unfolding model is not identified. Busing, Groenen, and Heiser (2005) state that an *absolute* degenerate solution possesses two characteristic features: The mean squared error between the (transformed) proximities and the distances is zero and all distances are equal to a constant. If we exclude the single point solution, which is obviously a degenerate solution, by the sum-of-squares normalization, the degenerate configuration usually shows two or four points at equal distance, containing objects of just one set per point. The transformation plot of an absolute degenerate solution is recognized by a horizontal line, corresponding to a zero slope and an intercept equal to some positive constant (in our case $b_1 = 1$). The value of the constant depends, however, on the chosen normalization.

Nevertheless, a degenerate unfolding solution, as described in the introduction, only occurs if both $b_1$ and $b_2$ are estimated freely. In that case, NMSE reduces to zero, with $b_1 = 1$ and $b_2 = 0$, setting $\gamma = 1$, and $\mathbf{d} = \mathbf{1}$, where $\mathbf{1}$ is a unit vector. Note that for $b_1 = 1$, the normalization restriction is satisfied since $\|\mathbf{1}\|^2 = nm$. Fixing either or both parameters to a constant will identify the model and subsequently avoid degeneracy. If $b_1$ is fixed to zero, a degenerate solution can only occur for $b_2 = 0$, setting all proximities and distances equal to zero, but this situation is avoided by the sum-of-squares normalization. This transformation, with only a free slope, coincides with a ratio transformation (see Table 3.1) and will not lead to a degenerate solution. Alternatively, $b_2$ can be fixed to a non-zero constant and only $b_1$ needs to be estimated (an intercept transformation). A similar approach was followed by de Soete and Heiser (1993) with $b_2 = 1$. Since the slope of the transformation remains fixed in this case, equal transformed proximities cannot occur and a degenerate solution will be avoided. Consequently, fixing both parameters will not lead to a degenerate solution.

## 3.5   PENALIZING THE INTERCEPT

For interval unfolding, the intercept $b_1$ can be penalized for deviating from zero, which subsequently prevents a zero slope $b_2$. In other words, the intercept is 'pulled down', but since $\|\mathbf{d}\|^2$ must remain equal to $nm$, the transformation line cannot remain horizontal, and a non-zero slope results. With a small penalty, $b_1$ will still approach one, but with a large penalty, $b_1$ will tend to zero, setting the smallest transformed proximity equal to zero. Choosing a moderate penalty will place $b_1$ between zero and one, at one point corresponding to a ratio transformation.

The introduction of a penalty forces us to make a small adjustment to (3.3), adding a quadratic term as,

$$\textsc{pmse}(b_1, b_2, \mathbf{d}, \kappa) = \|\mathbf{d}\|^{-2} \left( \|b_1 + b_2 \widetilde{\boldsymbol{\delta}} - \mathbf{d}\|^2 + \kappa \|b_1\|^2 \right), \qquad (3.4)$$

with penalty parameter $\kappa \geqslant 0$. For $\kappa = 0$, (3.4) is identical to (3.3). The current penalty corresponds with the penalty employed in ridge regression. Ridge regression (Hoerl & Kennard, 1970; Tikhonov & Arsenin, 1977), however, shrinks all estimators by penalizing their size, while the current procedure only shrinks the intercept parameter. Rearranging terms shows that $b_1$, $b_2$, and $\mathbf{d}$ can be estimated with the same routines and under the same restrictions ($\|\mathbf{d}\|^2 = nm$, $b_1 \geqslant 0$, and $b_2 \geqslant 0$) as before, since

$$\begin{aligned} \textsc{pmse}(b_1, b_2, \mathbf{d}, \kappa) &= \|\mathbf{d}\|^{-2} \left( \|b_1 + b_2 \widetilde{\boldsymbol{\delta}} - \mathbf{d}\|^2 + \kappa \|b_1\|^2 \right) \\ &= \|\mathbf{d}\|^{-2} \left( \|\mathbf{H}\mathbf{b} - \mathbf{d}\|^2 + \kappa \|b_1\|^2 \right) \\ &= \|\mathbf{d}\|^{-2} \|\mathbf{H}^*\mathbf{b} - \mathbf{d}^*\|^2, \end{aligned}$$

where $\mathbf{H} = [\mathbf{1} \ \widetilde{\boldsymbol{\delta}}]$, $\mathbf{b} = [b_1 \ b_2]'$, and $\mathbf{H}$ and $\mathbf{d}$ are augmented with $[\sqrt{nm\kappa} \ 0]$ and $[0]$ to obtain $\mathbf{H}^*$ and $\mathbf{d}^*$, respectively. Note that the normalization of the distances remains the same, since $\|\mathbf{d}^*\|^2 = \|\mathbf{d}\|^2$. Minimization of (3.4) is easily implemented in high-level languages like MATLAB. Appendix 3.C contains the code of the m-file used for the analyses in this chapter.

Groenen (personal communication, 2002) noted that the above procedure can also be performed by an ordinary unfolding or MDS program, such as KYST (Kruskal et al., 1978), SPSS PROXSCAL (Meulman, Heiser, & SPSS, 1999), or SPSS PREFSCAL (Busing, Heiser, et al., 2005), as long as proximity weights ($w$) and fixed coordinates are allowed. To do this, add one dummy object to each set, say $r$ and $c$, for which it holds that $\delta_{rc} = 0$, $d_{rc} = 0$ using fixed coordinates, $w_{ij} = 1$ for all $i, j$, $w_{rj} = 0$ for all $j$, and $w_{ic} = 0$ for all $i$. Finally, the weight between $r$ and $c$ is used as the penalty by setting $w_{rc} = nm\kappa$. Since there is no relation between the dummy objects $r$ and $c$ and the original objects, since these weights are zero, the influence of the dummy objects in (3.3) is equal to $\kappa b_1^2$. An example of the above suggestion is given in Appendix 3.B, using IBM SPSS PREFSCAL.

## 3.6   EXAMPLE (CONTINUED)

The penalized interval unfolding solution in Figure 3.2 is not a degenerate solution, since the transformed proximities, as well as the distances, have sufficient variation. The penalty 'pulled' the intercept 'down', resulting in a significant non-zero slope (see right-hand panel Figure 3.2), but also a non-zero NMSE (after 471 iterations, NMSE = 0.035848 and PMSE = 0.040504).
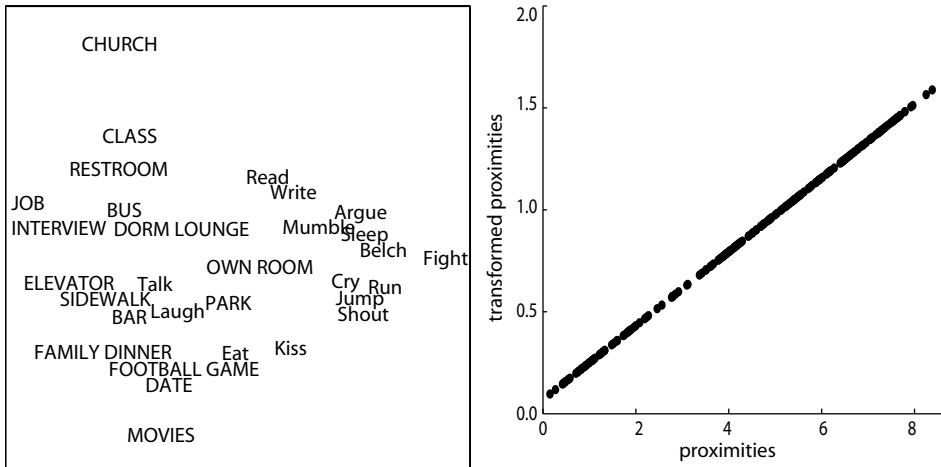
Figure 3.2    Penalized interval unfolding solution for the behavior-situation appropriateness data using pmse with κ=0.5.

The current unfolding solution shows a scatter of objects throughout the configuration space with different distances between situations and behaviors (see left-hand panel).

The interrelationship between behaviors and situations was investigated in 1974 by Price and Bouffard with a three-way analysis of variance (persons × situations × behaviors). There was a relatively large effect obtained from differences among situations, behaviors, and their interaction, which in turn accounted for fairly large proportions of variance in judgements of behavioral appropriateness. Price and Bouffard then separately classified behaviors and situations. The analysis resulted in two separate dimensions: (1) behavioral appropriateness, with fight and belch on the inappropriate end and laugh and talk on the appropriate end, and (2) situational constraint, with church and job interview on the restricted end, and park and own room on the unrestricted end.

In Figure 3.2 (left-hand panel), the unconstrained situations, like own room and park, are in the center of the configuration, indicating the many optional behaviors within reach. church, class, restroom, and job interview in the upper left-hand part are restricted situations according to Price and Bouffard and positioned here, further away from any behavior. More social events are gathered in the lower left-hand part of the configuration, closer to laugh, eat, and kiss. The behavioral appropriateness dimension from Price and Bouffard runs from left (appropriate) to right (inappropriate) through the configuration. The more quiet behaviors (read, write, mumble) are positioned
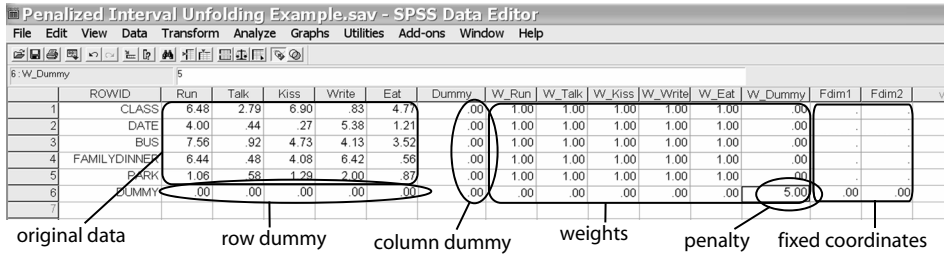
39

Figure 3.3   SPSS data example.

above the center and the louder and more physical behavior (shout, run, jump) lie below the center.

### 3.7   CONCLUSION

In this chapter, a simple but effective solution is proposed to the degeneracy problem in metric unfolding. With the aid of a simple penalty on the intercept of the transformation function, it is possible to prevent the transformation line from attaining a horizontal position, which in turn leads to variation in the transformed proximities, and, consequently, in the distances.

Instead of penalizing the intercept, one might choose to penalize the slope for attaining the unwanted zero value. This, however, is computationally more complex but will undoubtedly lead to the same result.

The current approach is not applicable for ordinal transformations, since restricting the intercept does not necessarily identify the ordinal unfolding model. An ordinal transformation is a step function and as such, every proximity can act as an unwilling intercept, in a sense that every proximity may take a step large enough to level the transformation function. Smooth monotone regression (Heiser, 1989) does impose comparable restrictions on the proximities, but also bounds the differences between one transformed proximity and the next, such that one big step is impossible.

Although this chapter only discusses unconditional and unweighted unfolding, the procedure could be generalized to weighted unfolding or row-conditional unfolding. For weighted unfolding, a weighted metric is added to the Euclidean norms in (3.4) and the procedure in Appendix 3.A is adapted accordingly. The metric transformation function (3.2) is extended for row-conditional unfolding to $\gamma_i = b_{1i} + b_{2i}\widetilde{\delta}_i$, where $i$ is a row indicator and $b_{1i}$ and $b_{2i}$ can be fixed in numerous ways (see, for example, DeSarbo & Rao, 1984, pages 165–168) to obtain different forms of row-conditional metric transformation functions.

APPENDIX 3.A   PENALIZED INTERVAL TRANSFORMATION

A non-negative least squares procedure is used to find estimates for $b_1$ and $b_2$ under the restriction that $b_1 \geqslant 0$ and $b_2 \geqslant 0$, given $\widetilde{\delta}$ and $\mathbf{d}$. Since only two parameters need to be estimated, a limited number of steps suffice. Expanding (3.4), and setting the derivative with respect to $\mathbf{b}$ equal to zero, gives two equations with two unknowns as

$$b_1 \omega + b_2 \mathbf{1}'\widetilde{\delta} = \mathbf{1}'\mathbf{d} \tag{3.5}$$

$$b_1 \widetilde{\delta}'\mathbf{1} + b_2 \widetilde{\delta}'\widetilde{\delta} = \widetilde{\delta}'\mathbf{d}, \tag{3.6}$$

where $\omega = \mathbf{1}'\mathbf{1} + nm\kappa = (1+\kappa)nm$ with $\kappa \geqslant 0$. For $\kappa = 0$, the penalty is not in effect and (3.3) is optimized instead. Substitution of $b_1$ from (3.5) in (3.6) provides an estimate for $b_2$ as

$$\widehat{b}_2 = \frac{\omega\widetilde{\delta}'\mathbf{d} - \widetilde{\delta}'\mathbf{1}\mathbf{1}'\mathbf{d}}{\omega\widetilde{\delta}'\widetilde{\delta} - \widetilde{\delta}'\mathbf{1}\mathbf{1}'\widetilde{\delta}}.$$

If $\widehat{b}_2$ is smaller than zero, then $\widehat{b}_2$ is set to zero. Using $\widehat{b}_2$ in either (3.5) or (3.6) provides $\widehat{b}_1$. If, however, $\widehat{b}_1$ is smaller than zero, then $\widehat{b}_1$ is set to zero and $b_2$ is re-estimated, using either (3.5) or (3.6), with $b_1 = 0$.

APPENDIX 3.B   EXAMPLE: IBM SPSS PREFSCAL SPECIFICATION FOR PMSE

The syntax for an interval unfolding with PREFSCAL is given below.

```
─────────────────────── Code Start ───────────────────────
1 PREFSCAL
2    VARIABLES       = Run Talk Kiss Write Eat Dummy
3    /WEIGHTS        = W_Run W_Talk W_Kiss W_Write W_Eat W_Dummy
4    /INPUT          = ROWS(rowid)
5    /PROXIMITIES    = DISSIMILARITIES
6    /INITIAL        = CLASSICAL
7    /CONDITIONALITY = UNCONDITIONAL
8    /TRANSFORMATION = LINEAR(INTERCEPT)
9    /PENALTY        = LAMBDA(1.0) OMEGA(0.0)
10   /CRITERIA       = MINSTRESS(0.0) DIFFSTRESS(0.0)
11   /RESTRICTIONS   =
12   ROW(COOR('Penalized Interval Unfolding Example.sav') Fdim1 Fdim2)
13   COLUMN(COOR('Penalized Interval Unfolding Example.sav') Fdim1 Fdim2)
14   /PRINT          = HISTORY MEASURES DECOMPOSITION
15   /PLOT           = INITIAL COMMON TRANSFORMATIONS RESIDUALS.
──────────────────────── Code End ────────────────────────
```

The data consists of the first five rows and columns (line 2) of the behavior-situation appropriateness data. Figure 3.3 shows the data setup in SPSS for the penalized interval unfolding. In this case, the SPSS data file also contains the fixed coordinates (lines 12–13) for the dummy variables (Fdim1 and Fdim2).

## APPENDIX 3.C    EXAMPLE: MATLAB CODE FOR PMSE

```
──────────────────────── Code Start ────────────────────────
1   function [X,Y,Gamma,new] = pmse (Delta,X,Y,kappa)
2
3   % pre-processing
4   [n,m] = size(Delta); n1 = ones(n,1); m1 = ones(m,1);
5   G = pinv(n*eye(m)-n*m);
6   ave = sum([X;Y])./(n+m); X = X-n1*ave; Y = Y-m1*ave;
7   D = sqrt(sum(X.*X,2)*m1'+n1*sum(Y.*Y,2)'-2*X*Y');
8   r = sqrt(n*m/sum(sum(D.^2))); D = r*D; X = r*X; Y = r*Y;
9   mse = sum(sum((Delta-D).^2));
10  DeltaTilde = Delta-min(min(Delta));
11  sumf = sum(sum(DeltaTilde));
12  ssqf = sum(sum(DeltaTilde.^2));
13
14  % main iterations
15  for iter = 1:10000
16
17    % transformation update
18    sumd = sum(sum(D));
19    cros = sum(sum(DeltaTilde.*D));
20    sumw = (1+kappa)*n*m;
21    work = ssqf*sumw-sumf^2;
22    if (work == 0) b2 = 0; else b2 = (cros*sumw-sumf*sumd)/work; end;
23    if (b2 < 0), b2 = 0; end;
24    b1 = (sumd-b*sumf)/sumw;
25    if (b1 < 0) b1 = 0; b2 = cros/ssqf; if (b2 < 0), b2 = 0; end; end;
26    Gamma = b1+b2*DeltaTilde;
27
28    % configuration update
29    E = D <= eps; B = Gamma./(D+E); B = B.*(E==0);
30    Xtilde = diag(sum(B'))*X-B*Y;
31    Ytilde = diag(sum(B))*Y-B'*X;
32    Y = G*(Ytilde+m1*sum(Xtilde)./m);
33    X = (Xtilde+n1*sum(Y))./m;
34    ave = sum([X;Y])./(n+m); X = X-n1*ave; Y = Y-m1*ave;
35    D = sqrt(sum(X.*X,2)*m1'+n1*sum(Y.*Y,2)'-2*X*Y');
36    r = sqrt(n*m/sum(sum(D.^2))); D = r*D; X = r*X; Y = r*Y;
37
38    % post-processing
39    mse = sum(sum((Gamma-D).^2));
40    penalty = kappa*n*m*b1*b1;
41    new = (mse+penalty)/(n*m);
42    if (iter > 1), if (old-new < eps), break; end; end;
43    old = new;
44  end
───────────────────────── Code End ─────────────────────────
```

4

Multidimensional unfolding methods suffer from the degeneracy problem in almost all circumstances. Most degeneracies are easily recognized: The solutions are perfect but trivial, characterized by approximately equal distances between points from different sets. A definition of an *absolutely* degenerate solution is proposed which makes clear that these solutions only occur when an intercept is present in the transformation function. Many solutions for the degeneracy problem have been proposed and tested, but with little success so far. In this chapter, we offer a substantial modification of an approach initiated by Kruskal and Carroll (1969) that introduced a normalization factor based on the variance in the usual least squares loss function. Heiser (1981) and de Leeuw (1983) showed that the normalization factor proposed by Kruskal and Carroll was not strong enough to avoid degeneracies. The factor proposed in the present chapter, based on the coefficient of variation, discourages or penalizes (nonmetric) transformations of the proximities with small variation, so that the procedure steers away from solutions with small variation in the inter-point distances. An algorithm is described for minimizing the re-adjusted loss function, based on iterative majorization. The results of a simulation study are discussed, in which the optimal range of the penalty parameters is determined. Two empirical data sets are analyzed by our method, clearly showing the benefits of the proposed loss function.

## 4.1  INTRODUCTION

Nonmetric multidimensional unfolding has been an idea that defeated most – if not all – attempts so far to develop it into a regular analysis method for preference rankings or two-mode proximity data. In Coombs' original formulation of unidimensional unfolding (Coombs, 1950, 1964), we have rankings of $n$ individuals over $m$ stimulus objects, and the objective is to find a common dimension called the quantitative J scale (joint scale), which contains ideal points representing the individuals and stimulus points representing the stimulus objects. On the J scale, each individual's preference ordering corresponds to the rank order of the distances of the stimulus points from the ideal point, the nearest being the most preferred.
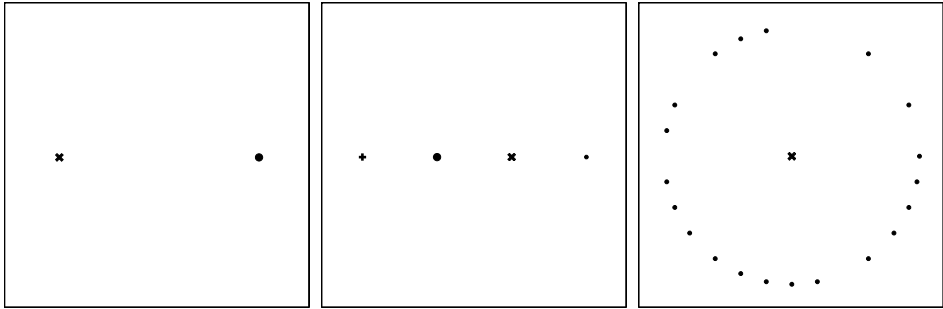
Figure 4.1  Typical degenerate solutions. The big dot represents multiple ideal points, the little dot represents a single ideal point, the cross represents multiple stimulus points, and the plus sign represents a single stimulus point.

Analytical procedures for fitting a quantitative J scale to a given set of rankings have been developed by, amongst others, McClelland and Coombs (1975); Roskam (1968), and van Blokland-Vogelesang (1989, 1993). These procedures usually consist of two steps: (1) a combinatorial search over the set of all possible partial orders on the distances between midpoints (defined as points that are equally distant to a pair of ideal points), and (2) a solution to a set of linear equations to find numerical scale values. Although in principle both steps could be generalized to the multidimensional case (Bennett & Hays, 1960; Hays & Bennett, 1961), no successful applications on real data have been reported in the literature, probably because, as Marden (1995, p. 256) has put it, "the computational aspects are a bit daunting".

Using a reformulation of Coombsean unfolding with J scales into a multidimensional scaling (MDS) method, characterized by the optimization of some (usually least squares) badness-of-fit function of the model distances, including possible transformations of the data, several authors have proposed iterative fitting methods for nonmetric unfolding (Roskam, 1968; Kruskal & Carroll, 1969; F. W. Young, 1972; Takane et al., 1977; Heiser, 1981). As already noted by Carroll (1972), many of these procedures have theoretical problems due to the badness-of-fit function being optimized that can lead to degeneracies yielding a nominally perfect badness-of-fit value, but convey no information. Degenerate solutions are recognized by the fact that all or almost all distances from ideal points to stimulus points are the same (see Figure 4.1).

It should be stressed that degeneracies do not only occur with ordinal data (the classic nonmetric case), but also with interval data, where we allow for an additive constant (Borg & Lingoes, 1987, p. 181). The fundamental cause of degeneracy is that the assumption of ordinal or interval measurement level allows transformations of the data that equalize them to a single intercept term,

which can be perfectly reproduced by an equal distance solution consisting, for example, of all individual points collapsed into one single point, and all stimulus points collapsed into another, distinct point (left-hand panel of Figure 4.1).

To avoid such degeneracies, Kruskal and Carroll (1969) proposed alternative badness-of-fit functions that incorporate normalization factors based on the variance of the distances, thus posing a penalty on equal distance solutions with zero variance. However, Kruskal and Carroll themselves reported that, although their method works nicely with artificial data, it had so far performed less than fully satisfactorily with real data. In fact, they say that "the method may now be good enough to be successful with some data in some cases, even though we do not have such examples" (Kruskal & Carroll, 1969, p. 669). Similar observations of continuing degeneracies, even when normalizing on the variance, have been reported by Heiser (1981, ch. 7), DeSarbo and Rao (1984), and Borg and Groenen (1997, p. 247).

In this chapter, we propose a new method for multidimensional unfolding of ordinal and interval data that solves the degeneracy problem for most practical purposes. We will first discuss the MDS formulation of unfolding and the fundamental cause of degeneracy in more detail. Current solutions for the degeneracy problem are reviewed in that light. The approach in this chapter is similar to Kruskal and Carroll's (1969) in using a badness-of-fit function that steers away from undesired solutions. The major difference with Kruskal and Carroll's approach is that instead of using a penalty based on the variance we use a penalty based on the coefficient of variation. Our motivation for this choice is first explained for the unconditional case, and then the penalty function is extended to the more common conditional case, in which the data are transformed for each individual separately. The two cases result in a similar badness-of-fit function that is the simple product of two factors, called *penalized Stress*. Details of our computational method to minimize PENALIZED STRESS are treated in the technical appendix. PENALIZED STRESS contains two parameters that control the focus of its behavior, and we report the results of a simulation study offering guidelines for the choice of these parameters. Finally, we illustrate our unfolding method with two applications and conclude the chapter with a summary.

## 4.2   BADNESS-OF-FIT FUNCTIONS IN UNFOLDING

Any MDS approach to unfolding is characterized by the badness-of-fit function it tries to minimize. Let us denote the coordinates of the ideal points that represent the individuals by $x_{ik}$, with $i = 1, \ldots, n$ and $k = 1, \ldots, p$, where $p$ equals the pre-chosen dimensionality of the solution. For each individual, the $x_{ik}$ are collected in the $p$-vector $x_i$, and the $x_i$'s are in turn collected in

the $n \times p$ matrix $\mathbf{X}$. Furthermore, the coordinates of the stimulus points that represent stimulus objects are given in the $m \times p$ matrix $\mathbf{Y}$, with elements $y_{jk}$ ($j = 1, \ldots, m$ and $k = 1, \ldots, p$) and rows $\mathbf{y}_j$. The data of the problem are either a set of dissimilarities $\delta_{ij}$, or a set of similarities $\rho_{ij}$, where the indices have the same range as before. In both cases, we need to find a new set of quantities $\gamma_{ij}$ of transformed data. The latter quantities are called pseudo-distances (Kruskal, 1977), and they are defined either as $\gamma_{ij} = f(\delta_{ij})$, with $f(\cdot)$ some monotonically *increasing* function, or as $\gamma_{ij} = g(\rho_{ij})$, with $g(\cdot)$ some monotonically *decreasing* function. Various special cases of $f(\cdot)$ and $g(\cdot)$ can be distinguished, the most common of which are a linear transformation without intercept (for ratio-scale data), a linear transformation with intercept (for interval-scale data), an arbitrary monotone function (for ordinal-scale data), and a monotone spline transformation (for quasi-nonmetric data, using the terminology of Winsberg & Carroll, 1989).

The pseudo-distances are to be found by the method so that they match as closely as possible with the Euclidean distances $d(\mathbf{x}_i, \mathbf{y}_j)$ between ideal points and stimulus points, defined as

$$d(\mathbf{x}_i, \mathbf{y}_j) = \sqrt{\sum_{k=1}^{p} (x_{ik} - y_{jk})^2}.$$

The mismatch or badness-of-fit between the pseudo-distances and the distances is usually measured in MDS by the ratio of a (weighted) *least squares function* and some *normalization function* (Kruskal & Carroll, 1969). If we denote the weights by $w_{ij}$, then their row sums are $w_{i+} = \sum_{j=1}^{m} w_{ij}$ and their total sum is $w_{++} = \sum_{i=1}^{n} w_{i+}$. For individual $i$, the weighted mean squared error is given by

$$\sigma_{\text{row}}^2(\boldsymbol{\gamma}_i, \mathbf{x}_i, \mathbf{Y}) = \frac{1}{w_{i+}} \sum_{j=1}^{m} w_{ij} \big(\gamma_{ij} - d(\mathbf{x}_i, \mathbf{y}_j)\big)^2, \tag{4.1}$$

and for the total group of individuals we obtain

$$\sigma_r^2(\boldsymbol{\Gamma}, \mathbf{X}, \mathbf{Y}) = \frac{1}{w_{++}} \sum_{i=1}^{n} w_{i+} \sigma_{\text{row}}^2(\boldsymbol{\gamma}_i, \mathbf{x}_i, \mathbf{Y}). \tag{4.2}$$

The notation $\sigma_{\text{row}}^2(\boldsymbol{\gamma}_i, \mathbf{x}_i, \mathbf{Y})$ and $\sigma_r^2(\boldsymbol{\Gamma}, \mathbf{X}, \mathbf{Y})$ is used to make explicit that the mean squared error is considered to be a function not only of the configurations $\mathbf{X}$ and $\mathbf{Y}$, but also of the pseudo-distances, collected in the $n \times m$ matrix $\boldsymbol{\Gamma}$. If we minimize (4.2) over the set of functions $f(\cdot)$ or $g(\cdot)$, the resulting pseudo-distances are often denoted by $\widehat{d}_{ij}$, called d-*hats*, and the resulting function $\sigma_r^2(\widehat{\mathbf{D}}, \mathbf{X}, \mathbf{Y})$ is now only a function of $\mathbf{X}$ and $\mathbf{Y}$, since $\widehat{\mathbf{D}}$ is explicitly

available given $\mathbf{X}$ and $\mathbf{Y}$. The square root of $\sigma_r^2(\widehat{\mathbf{D}}, \mathbf{X}, \mathbf{Y})$ is in the MDS literature commonly referred to as *raw Stress*, a tribute to Kruskal (1964a, 1964b).

Several normalization functions can be used in the badness-of-fit function to control the range of values of either $\boldsymbol{\Gamma}$ or $\mathbf{X}$ and $\mathbf{Y}$, or both. For example, the square of Kruskal's STRESS-1 is given by

$$\sigma_1^2(\mathbf{X}, \mathbf{Y}) = \frac{\sigma_r^2(\widehat{\mathbf{D}}, \mathbf{X}, \mathbf{Y})}{\eta_1^2(\mathbf{X}, \mathbf{Y})}, \tag{4.3}$$

where $\eta_1^2(\mathbf{X}, \mathbf{Y}) = w_{++}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} d^2(\boldsymbol{x}_i, \boldsymbol{y}_j)$. The normalization function $\eta_1^2(\mathbf{X}, \mathbf{Y})$ is effective in avoiding the situation in which all objects tend to collapse into one single point, since in that case $\eta_1^2(\mathbf{X}, \mathbf{Y})$ would become very small, and hence (4.3) would become very large. In unfolding, usually another normalization function is recommended, since STRESS-1 still allows degenerate solutions to occur in the ordinal-scale and interval-scale case. The cause of this fundamental problem is the fact that $\widehat{d}_{ij}$ may become equal for all $i, j$ if the set of transformations includes an intercept. The next subsection gives the exact conditions under which this phenomenon arises.

*Necessary and Sufficient Conditions for the Incidence of Degenerate Solutions*

In the following, we want to prove formally that the presence of an intercept in the transformation is a necessary and sufficient condition for a degeneracy in unfolding. We have to be precise about the definition of degenerate solutions and about the context in which they arise. First, we treat the unconditional case, in which there is only one transformation, and then the conditional case, in which there are $n$ separate transformations, one for each row of the data matrix.

A solution $(\mathbf{X}, \mathbf{Y})$ is called *absolutely degenerate* in the unconditional case if and only if R-STRESS (4.2) is zero and $d(\boldsymbol{x}_i, \boldsymbol{y}_j) = d$ for all $i, j$, with $d$ some positive constant. Although the data may contain ties ($\delta_{ij} = \delta_{i'j'}$ for some $i, j$ and $i', j'$), it is assumed that they are not completely tied; that is, we assume that we don't have $\delta_{ij} = \delta$ for all $i, j$. Such data would defeat the purpose of unfolding, which is to account for variation in the dissimilarities. It is also initially assumed that all weights are strictly positive, or $w_{ij} > 0$ for all $i, j$. The set of transformations $\Omega$ is always a cone, that is, it always satisfies the condition that $\boldsymbol{\Gamma} \in \Omega$ implies $\beta \boldsymbol{\Gamma} \in \Omega$, for any non-negative scalar $\beta$. Finally, presence of an intercept in the transformation means that $\Omega$ is closed under nondecreasing linear transformations, that is, it satisfies the condition that $\boldsymbol{\Gamma} \in \Omega$ implies $(\beta \boldsymbol{\Gamma} + \alpha \mathbf{E}) \in \Omega$, for non-negative $\alpha$ and $\beta$, where $\mathbf{E}$ is an $n \times m$ matrix of ones. It is easy to verify that the latter condition indeed applies to both interval-scale data and ordinal-scale data. We are now ready for the first proposition.

**Proposition 1** *In unconditional weighted least squares unfolding an absolutely degenerate solution exists if and only if the set of transformations is closed under nondecreasing linear transformations.*

*Proof.* Suppose $\Omega$ is closed under nondecreasing linear transformations. Then it is valid to select from the cone the pseudo-distances $\alpha E \in \Omega$ for some positive $\alpha$, so that R-STRESS becomes

$$\sigma_r^2(\alpha E, X, Y) = \frac{1}{w_{++}} \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} \left( \alpha - d(x_i, y_j) \right)^2.$$

Taking $y_j = 0$ and $x_i = \alpha \|a_i\|^{-1} a_i$, where $a_i$ is some arbitrary p-vector, yields $d(x_i, y_j) = \alpha$, so that all residuals become zero and hence R-STRESS becomes zero for any choice of $w_{ij}$. Conversely, suppose that we have some configuration satisfying $d(x_i, y_j) = \alpha$. To obtain a R-STRESS of zero we must have, for all $i, j$,

$$w_{ij}(\gamma_{ij} - \alpha)^2 = 0.$$

Since by assumption $w_{ij} > 0$, these equations are satisfied only if there exist pseudo-distances $\gamma_{ij} = \alpha$ for all $i, j$, which implies that we must have $\alpha E \in \Omega$. $\square$

Proposition 1 establishes existence, but not uniqueness. Thus, there may be other solutions that are absolutely degenerate (for instance, translations and rotations of the particular one specified in the proof of the proposition), and there may be other solutions with zero STRESS but without constant distance. In case that $w_{ij} = 0$ for some $i, j$, there similarly may be zero STRESS solutions that are not absolutely degenerate, even though they do trivialize all dissimilarity information that is taken into account in the loss function. Although Proposition 1 is formulated in terms of R-STRESS, it also holds for Kruskal's STRESS-1 defined in (3), since an absolutely degenerate solution has $\eta_1^2(X, Y) = \alpha^2 > 0$.

For the conditional case, we call a solution $(X, Y)$ absolutely degenerate if and only if R-STRESS (4.2) is zero and $d(x_i, y_j) = d_i$ for all $i, j$, with $d_i$ some row-specific positive constant. With respect to ties, we exclude cases in which $\delta_{ij} = \delta_i$ for any $i$. We now work with row-specific cones $\Omega_i$ that are closed under nondecreasing linear transformations. Proposition 2 is given without proof, since the argument is completely analogous to that of Proposition 1.

**Proposition 2** *In conditional weighted least squares unfolding an absolutely degenerate solution exists if and only if all sets of transformations are closed under nondecreasing linear transformations.*

In this case, the class of absolutely degenerate solutions is even larger than in the unconditional case, because we may now take $x_i = \alpha_i \|a_i\|^{-1} a_i$ for any positive $\alpha_i$, together with $y_i = 0$. Thus, $x_i$ is in fact completely arbitrary, and solutions of this type have the property that the matrix of pseudo-distances satisfies $\Gamma = \alpha e_m$, with $e_m$ an $m$-vector of ones and $\alpha$ an $n$-vector of intercepts.

For the conditional case, an even larger class of degenerate solutions exists, which is often met in practice, when only a few rows are absolutely degenerate. In this case, a solution $(X, Y)$ is called *partially degenerate* if and only if R-STRESS (4.1) is zero and $d(x_i, y_j) = d_i$ for some $i$ ($1 \leqslant i < n$), with $d_i$ some row-specific positive constant. Examples of this type of degeneracy will be shown in the applications.

*Overview of Solutions for the Degeneracy Problem*

Solutions for the degeneracy problem, proposed in the literature so far, can be classified into three approaches, each modifying one of the conditions in Propositions 1 and 2. First, there are the bounded regression approach of Heiser (1981, sec. 7.3), the quasi-metric approach of Kim et al. (1999), the smoothed monotone regression approach of Heiser (1989), and the mixed ordinal-linear approach of Borg and Lingoes (1987, sec. 11.6). These approaches modify the definition of the cones $\Omega$ or $\Omega_i$, so that the intercept is no longer free to vary and attached to zero. For example, Kim et al. use lower bound values of zero to assure that the most preferred stimulus point will fall close to the corresponding ideal point.

Except for Heiser's (1989) approach, these approaches are not truly non-metric or ordinal anymore in the sense of Kruskal (1964b) and Kruskal and Carroll (1969). Specifically, if the cones are a function of the actual data values (instead of only their rank order), the implication is that if one changes the original data with an admissible transformation, the unfolding solution does not remain the same, because the definition of the cones has been changed. Although these methods successfully avoid absolutely degenerate solutions, they tend to introduce bias in situations where the transformations coincide with the borders of the cones. The approach of Kim et al. avoids this situation altogether by using an a-priori transformation of the data and continuing with a row-conditional metric analysis without estimating an additive constant.

Second, there is the approach of DeSarbo and Rao (1984), which works with data dependent weights in an attempt to de-emphasize large dissimilarities that supposedly are especially prone to error. Propositions 1 and 2 show that this approach is bound to fail, since absolutely degenerate solutions are not excluded by any weighting structure.

Finally, there are the approaches of Kruskal and Carroll (1969) and the present approach, which modify R-STRESS in such a way that absolutely degenerate solutions no longer correspond to low values of the modified badness-of-fit function and hence are not candidate solutions anymore.

*Normalization on the Variance*

The usual recommendation to prevent absolutely degenerate solutions is to use the 'split-by-rows' option and normalization on the variance of the distances. Writing $\sigma_{row}^2(\widehat{d}_i, x_i, Y)$ for the result of minimizing (4.1) over the pseudo-distances of the $i$-th individual, and defining the variance of the distances with respect to the ideal point of individual $i$ as $\eta_2^2(x_i, Y) = w_{i+}^{-1} \sum_{j=1}^m w_{ij}(d(x_i, y_j) - \overline{d}_i)^2$, where $\overline{d}_i = w_{i+}^{-1} \sum_{j=1}^m w_{ij} d(x_i, y_j)$, the square of Kruskal's STRESS-2 generalized to the weighted row-conditional or split-by-rows case is defined as

$$\sigma_2^2(X, Y) = \frac{1}{w_{++}} \sum_{i=1}^n \frac{w_{i+} \sigma_{row}^2(\widehat{d}_i, x_i, Y)}{\eta_2^2(x_i, Y)}. \tag{4.4}$$

Splitting the STRESS by rows and using the variance as a normalization function in (4.4) avoids implosion into a single point, and supposedly avoids degeneracy at the same time, since an absolutely degenerate solution has $\eta_2^2(x_i, Y) = 0$, so that (4.4) would grow to infinity when approaching a degeneracy. However, numerous testruns with KYST (Kruskal et al., 1978) using STRESS-2, of which we will show two examples in the application section, still appear to give approximately degenerate solutions in many cases. To clarify this unexpected finding, de Leeuw (1983) has shown, by what essentially is an application of l'Hopital's rule, that in arbitrarily small neighborhoods of degenerate solutions $\sigma_2^2(X, Y)$ can have a low value and that the algorithm optimizes a bilinear model rather than a distance model.

Heiser (1981) has experimented with normalization on the variance of the pseudo-distances. He used the equally weighted version of the badness-of-fit function

$$\sigma_3^2(\Gamma, X, Y) = \frac{1}{w_{++}} \sum_{i=1}^n \frac{w_{i+} \sigma_{row}^2(\gamma_i, x_i, Y)}{\eta_3^2(\gamma_i)}. \tag{4.5}$$

with $\eta_3^2(\gamma_i) = w_{i+}^{-1} \sum_{j=1}^m w_{ij}(\gamma_{ij} - \overline{\gamma}_i)^2$, where $\overline{\gamma}_i = w_{i+}^{-1} \sum_{j=1}^m w_{ij}\gamma_{ij}$. Note that (4.5) does not include Kruskal's d-hats, since optimizing over the pseudo-distances cannot be done on the R-STRESS anymore, but should involve the normalization function as well, requiring a specialized procedure. This approach has the advantage that finding $X$ and $Y$ for any given $\Gamma$ can be done by a single and stable algorithm. It tries to avoid degeneracy indirectly, by staying

away from undesired transformations of the data. Unfortunately, Heiser (1981) reported solutions that were very similar to solutions obtained with (4.4), and hence (4.5) did not solve the degeneracy problem either.

## 4.3    UNFOLDING BY PENALIZING ON THE COEFFICIENT OF VARIATION

The current research started from two ideas. The first was to try to change the balance between numerator and denominator of the badness-of-fit function by introducing a power function for only one of them (a device successfully used in Groenen and Heiser (1996)). The second was to look at another measure of variability, since the variance is invariant under addition of a constant, and therefore it does not discriminate well between solutions with small average distance between ideal points and stimulus points and solutions with uniformly large average distance, which are typically obtained in degenerate solutions. The traditional normalization factors are insufficient to avoid degeneracy and other measures of variability might provide more adequate penalties for entering undesirable regions of the parameter space (see Trosset (1998) for interpreting normalization factors as penalty functions in the context of nonmetric multidimensional scaling). In the next two sections, we will first introduce our penalty function for the unconditional case, in which the badness-of-fit (and the transformation) is not split by individuals, and after that treat the more common conditional case, in which we do split the transformation by individuals, but aggregate the R-STRESS and the penalty separately, as in (4.3).

*Unconditional Case*

For distributions of positive quantities that have a scale whose origin is not arbitrary, but meaningful – like mass, frequency, but also similarity and distance – variability is best measured with respect to the origin, rather than with respect to the mean. Karl Pearson's *coefficient of variation* (Pearson, 1896) is a scale-free, relative measure of variability that takes the standard deviation as a fraction of the mean. For any $T$-vector $\mathbf{a}$ with non-negative values $a_t$, the variation coefficient $\upsilon(\mathbf{a})$ can be written as

$$\upsilon(\mathbf{a}) = \frac{s(\mathbf{a})}{\overline{a}} = \sqrt{\frac{T^{-1} \sum_{t=1}^{T} a_t^2}{\left(T^{-1} \sum_{t=1}^{T} a_t\right)^2} - 1}. \qquad (4.6)$$

It is clear from (4.6) that $\upsilon(\mathbf{a})$ achieves its minimal value when all $a_t$ are equal, that is, $\upsilon(\mathbf{a}) = 0$ iff $a_t = \overline{a}$ for all $t$. It is equal to one if $\mathbf{a}$ has an evenly divided bimodal distribution with one mode at zero, and attains an upper bound of $\sqrt{T-1}$ if all but one of the $a_t$ are zero (Katsnelson & Kotz, 1957).
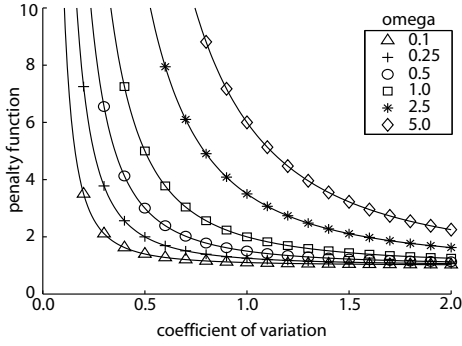
Figure 4.2   Penalty function $\mu(\Gamma)$ versus variation coefficient $\upsilon(\Gamma)$ for different values of $\omega$ (see legend).

We will define our badness-of-fit function $\sigma_p^2(\Gamma, X, Y)$, called *penalized Stress*, to have the following form:

$$\sigma_p^2(\Gamma, X, Y) = \sigma_r^{2\lambda}(\Gamma, X, Y)\mu(\Gamma), \tag{4.7}$$

where $\lambda$ is a lack-of-penalty parameter ($0 < \lambda \leqslant 1$) that controls the balance between the R-STRESS $\sigma_r^2(\Gamma, X, Y)$ and the penalty function $\mu(\Gamma)$. When $\lambda$ increases, the influence of the penalty term decreases, and vice versa. For the penalty term, we follow the same approach as Heiser (1981) in penalizing on some function of $\Gamma$, so that for any given value of $\Gamma$ minimizing $\sigma_p^2(\Gamma, X, Y)$ amounts to minimizing $\sigma_r^2(\Gamma, X, Y)$. The penalty function is chosen to be a multiplicative factor, and such that it approaches 1 when $\upsilon(\Gamma)$ becomes large, and grows to infinity when $\upsilon(\Gamma)$ goes to zero. A simple set of functions that has these properties is defined as

$$\mu(\Gamma) = 1 + \frac{\omega}{\upsilon^2(\Gamma)}, \tag{4.8}$$

where $\omega$ is a range parameterrange: When $\omega$ is small, $\mu(\Gamma)$ is especially effective when $\upsilon(\Gamma)$ is small, while for large values of $\omega$, larger values of the variation coefficient will cause relatively effective penalties as well.

The shape of (4.8) as a function of $\upsilon$ is illustrated in Figure 4.2 for different values of $\omega$. In all cases, the penalty will be highest if the mean of the transformed data is relatively large and the standard deviation relatively small – a condition that typically arises under degeneracy.

*Conditional Case*

For ratings or rankings of $n$ individuals over $m$ stimulus objects it is natural to require $n$ separate transformations that are conditional on the rows of the

data matrix. Since the R-STRESS (4.2) is additive over individuals, it can be minimized over $\gamma_i$ row after row, and there is no need to adjust that part of PENALIZED STRESS. However, the penalty function (4.8) does need adjustment, because we want to penalize lack of variation in each and every transformation. Denoting the variation of the pseudo-distances in row $i$ by $\upsilon(\gamma_i)$, we use the harmonic mean to aggregate them, yielding the penalty function

$$\mu_c(\Gamma) = 1 + \frac{\omega}{\left(\frac{1}{n}\sum_{i=1}^{n}\upsilon^{-2}(\gamma_i)\right)^{-1}}, \tag{4.9}$$

for the conditional case. We use the harmonic mean of the variation coefficients, because of its property that it is never greater than the geometric mean, which in turn is never greater than the arithmetic mean (cf. Hardy, Littlewood, & Polya, 1952, p. 26), because it is relatively strongly affected when one of the individual variation coefficients becomes very small, and because the resulting penalty term can be conveniently minimized. By simple algebra, we have the result

$$1 + \frac{\omega}{\left(\frac{1}{n}\sum_{i=1}^{n}\upsilon^{-2}(\gamma_i)\right)^{-1}} = 1 + \omega\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\upsilon^2(\gamma_i)} = \frac{1}{n}\sum_{i=1}^{n}\left(1 + \frac{\omega}{\upsilon^2(\gamma_i)}\right), \tag{4.10}$$

showing that $\mu_c(\Gamma)$ can be written as an additive function over individuals. So the penalty function based on the harmonic mean of the individual variation coefficients is equal to the arithmetic mean of the individual penalty terms.

We think it is essential to aggregate the badness-of-fit and the penalty terms separately, instead of aggregating their row-wise products as in (4.4) and (4.5), because in the latter case low variability for some individual can be compensated by low R-STRESS of the same individual. This compensation makes a method based on (4.4) or (4.5) less reactive to degeneracies at the individual level than our method based on (4.9). Our method also ensures a correct weighting of individuals when fitting $X$ and $Y$, because it does not contaminate $w_{i+}$ with the individual penalty terms, which give relatively little weight to individuals with highly variable pseudo-distances.

*An Alternating Update Strategy using Iterative Majorization*

PENALIZED STRESS can be minimized by an alternating procedure, following the general strategy first used in MDS by Takane et al. (1977), in which we alternate between finding an update for the configuration given the current estimate of the transformation(s) and finding an update for the transformation(s), given the current estimate of the configuration. In the present case, each of these two steps is carried out by iterative majorization. For the theory of iterative majorization in MDS, we refer to de Leeuw (1977a) and Heiser (1995)

as basic papers, and to Groenen (1993) and Groenen and Heiser (1996) for some extensions used to minimize PENALIZED STRESS. The minimization of PENALIZED STRESS is extensively described in the technical appendix.

### 4.4    SIMULATION STUDY

The parameters $\lambda$ and $\omega$ in PENALIZED STRESS both influence the strength of the penalty. However, for a practical implementation and reasons of parsimony, it is preferable to restrict at least one parameter to a constant, or suggest limits for their values. To determine reasonable values for $\lambda$ and $\omega$, we conducted the following simulation study.

The data in this simulation study were generated as follows. First, 30 row coordinates and 15 column coordinates in two dimensions were drawn from a standard normal distribution (the Polar Box-Muller method was used to obtain standard normal deviates (see, for example, Dagpunar, 1988), with uniform pseudo-random numbers from maximally equidistributed combined LFSR generators (L'Ecuyer, 1999)) and the Euclidean distances between these coordinates were computed (a similar design was used by Coombs and Kao (1960) and Kruskal and Carroll (1969)). Error was added by multiplying the simulated distances with the exponential of a percentage of a standard normal distribution, corresponding to imposing log-normal error on the distances (see Wagenaar & Padmos, 1971). Two error percentages were chosen, 0% and 25%, corresponding to an error-free and an error-perturbed condition. Transformations were conducted for two types of data: Interval data was transformed using linear regression with an intercept (using non-negative least squares to ensure non-negative pseudo-distances), and monotone regression (Kruskal, 1964b) was used to transform ordinal data. Transformations in the unconditional case were based on all data, whereas in the row-conditional case, the data were transformed for each row separately. A wide range of values was chosen for $\lambda$ and $\omega$, based on the results of several pilot studies. Values for $\lambda$ ranged from 0.1 to 1 with increments of 0.1 and values for $\omega$ were chosen as 0.1, 0.25, 0.5, 1, 2.5, and 5. A strong penalty is marked by small values for $\lambda$ in combination with large values for $\omega$, whereas a weak penalty has the opposite values for both parameters.

We have also considered other factors in pilot studies, but since the influence of the number of rows, columns, and dimensions and the level of error on

Table 4.1    Summary of independent factors and levels for the simulation study.

| Factor | # | Levels | Factor | # | Levels |
|---|---|---|---|---|---|
| Error Percentage | 2 | 0%, 25% | $\lambda$ | 10 | 0.1, 0.2, 0.3, …, 0.9, 1.0 |
| Data Level | 2 | Interval, Ordinal | $\omega$ | 6 | 0.1, 0.25, 0.5, 1.0, 2.5, 5.0 |
| Conditionality | 2 | Unconditional, Row-Conditional | | | |

the behavior of λ and ω was negligible, single neutral values for these factors suffice. The independent factors in the simulation study are summarized in Table 4.1. We used a full factorial design for the five factors. For each of the 480 cells, 62 replications were drawn. The initial configuration used in the analysis is described in Step 1 of the algorithm.

To assess the level of degeneracy, we cannot rely on a single measure. A good solution has low STRESS and reasonable variation in both pseudo-distances and distances. For the variation of the pseudo-distances, it is essential to discriminate between unconditional and row-conditional variation, as the variation of row-conditional transformations depends on single rows. This difference does not apply for the distances, because the variation in a configuration is mostly determined unconditionally, that is, by looking at the complete configuration. The measures that need simultaneous inspection are: R-STRESS $\sigma_r^2(\Gamma, X, Y)$, the (conditional) variation coefficient of the pseudo-distances, that is, $\upsilon(\Gamma)$ for the unconditional case and $\upsilon_c(\Gamma) = [n^{-1} \sum_{i=1}^{n} \upsilon^{-1}(\gamma_i)]^{-1}$ for the row-conditional case, and the variation coefficient of the distances $\upsilon(D)$.

For the error-free condition, it is expected that R-STRESS becomes zero, while both variation coefficients remain distinctively non-zero. Although the error-perturbed condition lacks the pre-defined outcome of zero R-STRESS, R-STRESS is expected to improve beyond the R-STRESS imposed initially during the simulation process, but remains distinctively non-zero. For the error-perturbed condition, we also require that solutions must have variation coefficients that are distinctively non-zero.

*Results*

The results for the error-free condition are as expected. In all cases, R-STRESS was zero or very close to zero, while the variation coefficients of both distances and pseudo-distances ranged between 0.4 and 0.6. This outcome applied to all values of λ and ω. An alternative explanation for the non-occurrence of degeneracies in the error-free condition might be the good initial configuration. Kruskal and Carroll (1969) mention that even with an older loss function (STRESS-1, a function that later proved to be unable to prevent degeneracies), F. W. Young and Torgerson (1967) succeeded in avoiding degeneracies, due to a very good, rationally generated starting configuration, and then moving into a nearby local minimum. This alternative explanation can be tested by using random starts.

In Table 4.2, the results are shown of a small additional simulation study using random starts, with the same independent factors as before, but with fewer levels for λ and ω. With certain choices of λ and ω, the algorithm typically finds the same non-degenerate solution from randomly generated initial

configurations that it finds from a very good rationally generated initial config-
uration. This suggests that non-degeneracy is built into the PENALIZED STRESS
objective function, rather than fortuitously appearing in local solutions near
good initial configurations. Specifically, it turns out that strong penalty pa-
rameters have a tendency towards transformations with equal increments,
especially in the row-conditional case. This artifact of strong penalties (e. g.,
$\lambda = 0.1$ and $\omega = 5.0$) sets the smallest pseudo-distance to zero, and puts the re-
maining pseudo-distances at equal increments, which ultimately prevents the
algorithm from finding the perfect solution. There is a considerable difference
between unconditional and row-conditional transformations. One of the rea-
sons for this difference could be the amount of freedom in the transformations.
Generally, the unconditional case is stricter than the row-conditional case.
With row-conditional ordinal data, for example, there are at most $nm(m-1)/2$
inequalities to be satisfied, compared to a maximum $nm(nm-1)/2$ in the
unconditional case. Consequently, the restrictions on the solution are harder
to satisfy in the latter case, and therefore local minima may occur more of-
ten with unconditional transformations. In conclusion, we can say that the
procedure is capable of finding the perfect configuration, given error-free
data, without the aid of a good (rational) start, provided we choose $\lambda$ in the
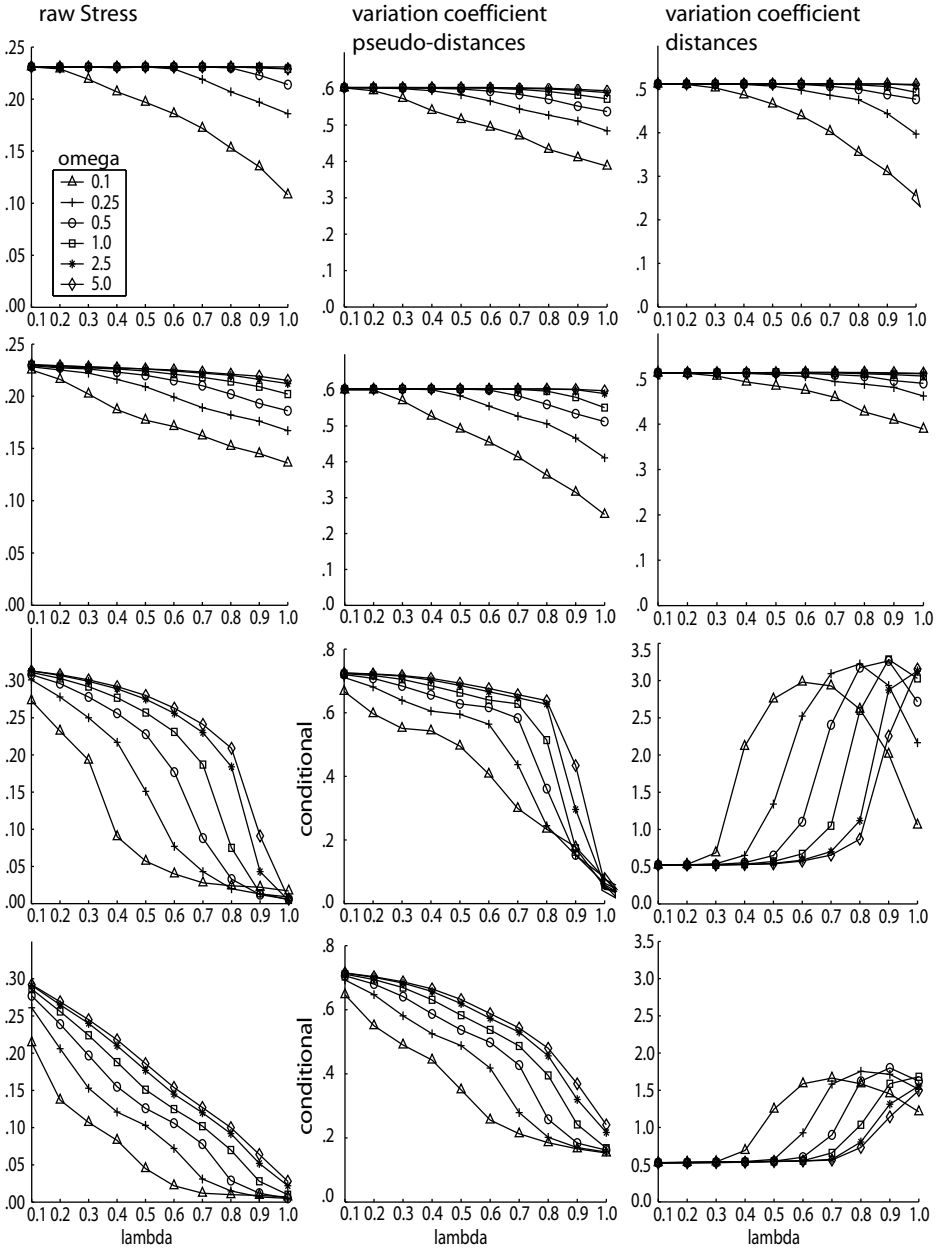neighborhood of 0.5.

For the error-perturbed condition, the results are presented in Figure 4.3.
Note that the average R-STRESS for the error-perturbed condition imposed
during the simulation process is 0.245.

For the unconditional case (top six panels of Figure 4.3), only the weakest
penalty tends to some form of degeneracy, indicated by diminishing R-STRESS

Table 4.2    Recovery percentage for error-free data for 2500 random configurations.

| Data Level | $\lambda$ | Unconditional | | | Row-Conditional | | |
|---|---|---|---|---|---|---|---|
| | | $\omega =0.1$ | $\omega =0.5$ | $\omega =5.0$ | $\omega =0.1$ | $\omega =0.5$ | $\omega =5.0$ |
| Interval | 0.1 | 10 | 10 | 11 | 61 | 0 | 0 |
| | 0.5 | 26 | 14 | 14 | 95 | 99 | 97 |
| | 1.0 | 8 | 15 | 14 | 42 | 94 | 99 |
| Ordinal | 0.1 | 8 | 10 | 10 | 89 | 0 | 0 |
| | 0.5 | 28 | 10 | 12 | 97 | 98 | 97 |
| | 1.0 | 7 | 24 | 12 | 75 | 99 | 99 |

Figure 4.3    (facing page)    Multiple line plots for the following conditions: Unconditional interval (top three
panels), unconditional ordinal (next three panels), row-conditional interval (next three panels), and row-
conditional ordinal (bottom three panels). The panels represent (from left to right): $\sigma_r(\Gamma, \mathbf{X}, \mathbf{Y})$, $\upsilon_c(\Gamma)$, and
$\upsilon(\mathbf{D})$, with lines showing the averages over 50 replications for different values of $\omega$ (the legend for all panels
is displayed in the top-left-hand panel).

raw Stress

variation coefficient
pseudo-distances

variation coefficient
distances



omega
△ 0.1
+ 0.25
⊕ 0.5
□ 1.0
∗ 2.5
◇ 5.0

conditional

conditional

lambda

lambda

lambda

and diminishing variation coefficients for both distances and pseudo-distances. The choice $\lambda = 0.5$ and $\omega > 0.1$ performs well for all unconditional analyses.

For the row-conditional case, things are more complicated, although both interval and ordinal transformations show the same pattern, as was the case for the unconditional transformations. Strong penalties fail to improve R-STRESS beyond the R-STRESS imposed during the simulation process, probably due to a strong tendency towards transformations with equal increments. As the penalty becomes weaker, R-STRESS decreases, until it is really close to zero for the weakest penalties, indicating the presence of degeneracies.

The conditional variation coefficients of the pseudo-distances confirm the presence of degeneracies for the weaker penalties: Average values rapidly decrease as $\lambda$ becomes larger than 0.6. The conditional variation coefficient of the pseudo-distances and the variation coefficient of the distances (an unconditional coefficient) show opposite results: With a decreasing penalty, the values of the former decrease while the latter increase. This result is not caused by the fact that one coefficient uses pseudo-distances and the other uses distances, but is due to the use of unconditional and conditional variation coefficients. The ultimate cause is the specific type of degeneracy that occurs here. For the weak penalties, degeneracies appear in every single row. When $n - 1$ rows have low variation with a mean close to zero and one row also has low variation but with a substantial mean, the conditional variation coefficient will be close to zero, while the unconditional variation coefficient will be large, closer to its upper bound. The row-configuration of such a solution shows a cluster of $n - 1$ points and, at a relative large distance, the other, remaining point. This type of degeneracy is already known from ordinal MDS, where it occurs when objects are clustered, with smaller within-cluster distances than between-cluster distances. With a relatively high dimensionality with respect to the number of clusters, all objects within a cluster will end up in a single point, causing the type of degeneracy described above, which incidentally implies that the penalty function developed here should also be able to prevent degeneracies in ordinal MDS.

In conclusion, the simulation study made clear that combinations of weak penalty parameters are unable to prevent degeneracies. On the other hand, strong penalties tend to induce transformations with equal increments, thus showing little improvement in STRESS and reducing the ability of the method to distinguish clearly between nonmetric (ordinal) and metric (interval) solutions. Unconditional transformations are clearly less sensitive to degeneracies and thus allow for weaker penalties. Finally, setting $\lambda$ to 0.5 specifies the function as a usual square root, whereas $\omega$ can be used as a conventional penalty parameter, with a default value of 0.5, a value that can be decreased or increased as circumstances require.

Two different applications of unfolding are presented to illustrate the behavior of PENALIZED STRESS (or P-STRESS), implemented in PREFSCAL. For this purpose, PREFSCAL was compared with publicly available programs for unfolding, or programs that claim to be able to avoid degenerate solutions, namely with ALSCAL (Takane et al., 1977), KYST (Kruskal et al., 1978), GENFOLD (DeSarbo & Rao, 1984), and NEWFOLD (Kim et al., 1999). All programs but GENFOLD are publicly available. GENFOLD was mimicked with PREFSCAL using $\lambda = 1$, $\omega = 0$, and a user-provided weighting structure $w_{ij} = r(\gamma_{ij})^{-p}$, where $r(\gamma_{ij})$ represents the (row-)rank of $\gamma_{ij}$ and $p$ is provided by the user (for more details, see DeSarbo & Rao, 1984, pp. 154–157). We have used GENFOLD with $p = 2$, which, according to DeSarbo and Rao, appears to work well.

In the first application, we analyze the brewery data of Borg and Berger-maier (1982) using an unconditional interval transformation for the similarities. In the second application, the breakfast data of P. E. Green and Rao (1972) are used to illustrate unfolding with row-conditional ordinal transformations of the data. For these analyses, the program's default initial configuration and the strictest convergence criteria were used, without limiting the number of iterations used, to obtain results that are as accurate as possible.

For a numerical comparison, it is advisable to consider several measures simultaneously. First, there are several fit measures to consider, indicating how well the distances fit the pseudo-distances (see Technical Appendix G, page 213). Instead of R-STRESS, which was used in the simulation study, we now use STRESS-2 (4.4) as a badness-of-fit measure, since STRESS-2 is more commonly known and scale-independent. Several goodness-of-fit measures are provided to give additional information about the recovery of information. Variance accounted for (VAF), which is equal to the square of Pearson's product-moment correlation coefficient, is computed over all values, irrespective of the conditionality of the analysis. For the remaining goodness-of-fit measures, the conditionality is taken into account as the arithmetic mean is computed over rows in case of the row-conditional analyses of the breakfast data. This is true for Pearson's product-moment correlation coefficient ($r$), Spearman's rank correlation coefficient ($\rho$; RHO), and Kendall's tau-$b$ coefficient ($\tau_b$; TAU).

Then, variability measures are provided to indicate the spread of either pseudo-distances or distances (see Technical Appendix G, page 220). For this purpose, we provide both the variance (var) and the coefficient of variation ($\upsilon$). The distances are considered unconditional, since a configuration is looked upon as a whole, while the interest in the pseudo-distances depends on the conditionality chosen for the analysis. These conditional measures are taken to be the harmonic mean over rows, providing a conservative measure of

variability, since the harmonic mean tends to put more emphasis on rows with little variation.

Finally, we provide a non-degeneracy index (D-INDEX), suggested by Shepard (1974), and an intermixedness index (I-INDEX), suggested by DeSarbo, Young, and Rangaswamy (1997) (for formulas, see Technical Appendix G, pages 220–221). Shepard described "a rough index of the *non*degeneracy of a solution" as the ratio of the number of distinct distances to the total number of distances. Here, a distance is distinct from another distance, when the difference between the two distances is more than one tenth of the sum of the two distances. DeSarbo et al. introduced three indices, $I_1 = \ln(\overline{d}_y/\overline{d}_{xy})$, $I_2 = \ln(\overline{d}_x/\overline{d}_{xy})$, and $I_3 = \ln(\overline{d}_y/\overline{d}_x)$, where $\overline{d}_x$ and $\overline{d}_y$ are the average within-set-distances and $\overline{d}_{xy}$ is the average between-set-distance. These indices aim to indicate a well-interspersed configuration when all three indices are near zero, or, equivalently, when the sum-of-squares, denoted by I-INDEX, is close to zero.

*Brewery Data*

In Borg and Bergermaier (1982), the brewery data set is used to demonstrate some of the degeneracy problems in unfolding. These data were obtained by asking beer drinkers to rate 9 breweries on 26 attributes on a 6-point scale ranging from 1='not true at all' to 6='very true'. Averaged over individuals, the values are taken to be similarities on an interval level. In a few consecutive steps, using KYST (Kruskal et al., 1978), Borg and Bergermaier showed that STRESS-2 with a metric transformation without additive constant comes closest to an acceptable solution.

Figures 4.4 and 4.5 show the configurations (left-side panels) and transformation plots (right-side panels) for the brewery data with an unconditional interval transformation (which *includes* an additive constant). GENFOLD (with p = 2) provides an absolutely degenerate solution (see Figure 4.4, bottom panels): STRESS-2 is zero and all distances are equal to some positive constant, resulting in zero variability for the distances (see Table 4.3). The transformation plot (right-side panel) shows an horizontal (zero) slope, again indicating no variability in the pseudo-distances. The configuration shows that the breweries (indicated by numbers) are situated in the middle of an imaginary circle with the attributes (indicated by black dots) on the circumference of the circle. The radius of the circle is equal to the intercept of the interval transformation, which is a clear case of an absolutely degenerate solution. The weighting

Figure 4.4   (facing page)   Configurations (left-side panels) and transformation plots (right-side panels) for the brewery data set with an unconditional interval transformation using ALSCAL (top panel), KYST (middle panels), and GENFOLD with p=2 (bottom panels).
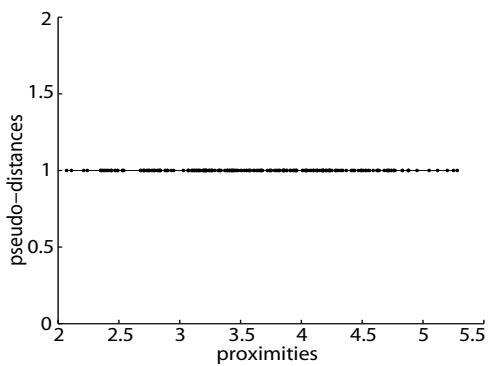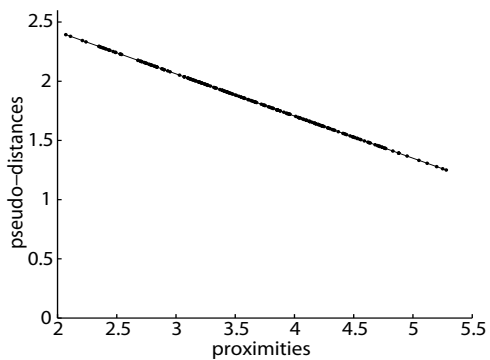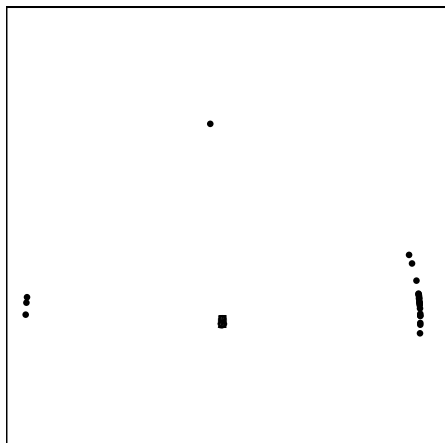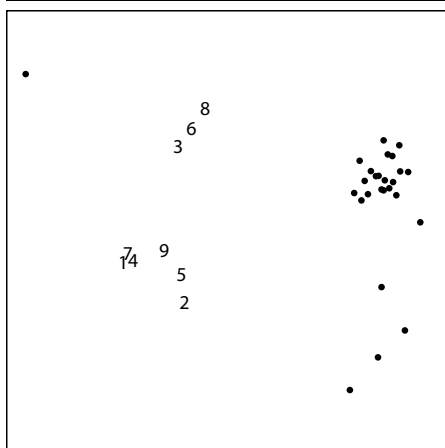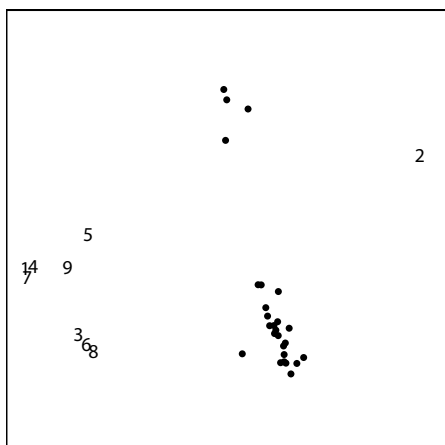
structure was not able to avoid a degenerate solution, as was already predicted in the overview of solutions for the degeneracy problem in Section 4.2. Irrespective of the minimization method used (PREFSCAL uses majorization and alternating least squares), the current GENFOLD solution is a global minimum as the weighted sum-of-squares function reached its lowest value possible. With various values of p, absolutely degenerate solutions were found invariably, even though sometimes other zero STRESS solutions occurred as well, especially when p was large ($p > 5$). In the latter case, the effective number of data points dropped below the number of estimated parameters, because an increasing majority of the weights tends to zero as $p \to \infty$.

The solutions of ALSCAL and KYST are closely related, which might be a consequence of the correspondence in loss functions. Both solutions show some variability in the distances (see Table 4.3), but the two sets (breweries and attributes) are clearly separated in the configuration, as can be seen in Figure 4.4 (left-side panels) and in Table 4.3, indicated by the intermixedness index I-INDEX. The fit measures for both analyses are quite good; KYST has even the lowest STRESS-2 value of all interpretable analyses. This is not a coincidence, since KYST minimizes STRESS-2, and is supposed to find a solution with a lower STRESS-2 value, as ALSCAL minimizes S-STRESS-2, which uses *squared* distances and *squared* pseudo-distances instead of the unsquared versions. Note that the transformation plot is not available for ALSCAL, since the program erroneously provides distances instead of pseudo-distances in its output.

Table 4.3    Various measures for ALSCAL, KYST, GENFOLD, NEWFOLD, and PREFSCAL for the brewery data set.

| | ALSCAL | KYST | GENFOLD | NEWFOLD | PREFSCAL |
|---|---|---|---|---|---|
| **Badness-of-Fit** | | | | | |
| Stress-2 | 0.540 | 0.310 | 0.000 | 0.907 | 0.326 |
| **Goodness-of-Fit** | | | | | |
| VAF | 0.885 | 0.904 | 0.000 | 0.341 | 0.911 |
| R | 0.941 | 0.951 | 0.000 | 0.584 | 0.955 |
| RHO | 0.952 | 0.955 | 0.011 | 0.681 | 0.967 |
| $\text{TAU}_b$ | 0.824 | 0.835 | 0.000 | 0.495 | 0.844 |
| **Variability** | | | | | |
| $\text{var}(\mathbf{D})$ | 0.020 | 0.022 | 0.000 | 0.158 | 0.143 |
| $\upsilon(\mathbf{D})$ | 0.144 | 0.150 | 0.000 | 0.433 | 0.404 |
| $\text{var}(\mathbf{\Gamma})$ | N/A | 0.020 | 0.000 | 0.165 | 0.165 |
| $\upsilon(\mathbf{\Gamma})$ | N/A | 0.142 | 0.000 | 0.444 | 0.445 |
| **Non-Degeneracy and Intermixedness** | | | | | |
| D-INDEX | 0.658 | 0.701 | 0.004 | 0.838 | 0.855 |
| I-INDEX | 1.344 | 1.676 | 582.651 | 0.276 | 0.195 |

N/A = not available.
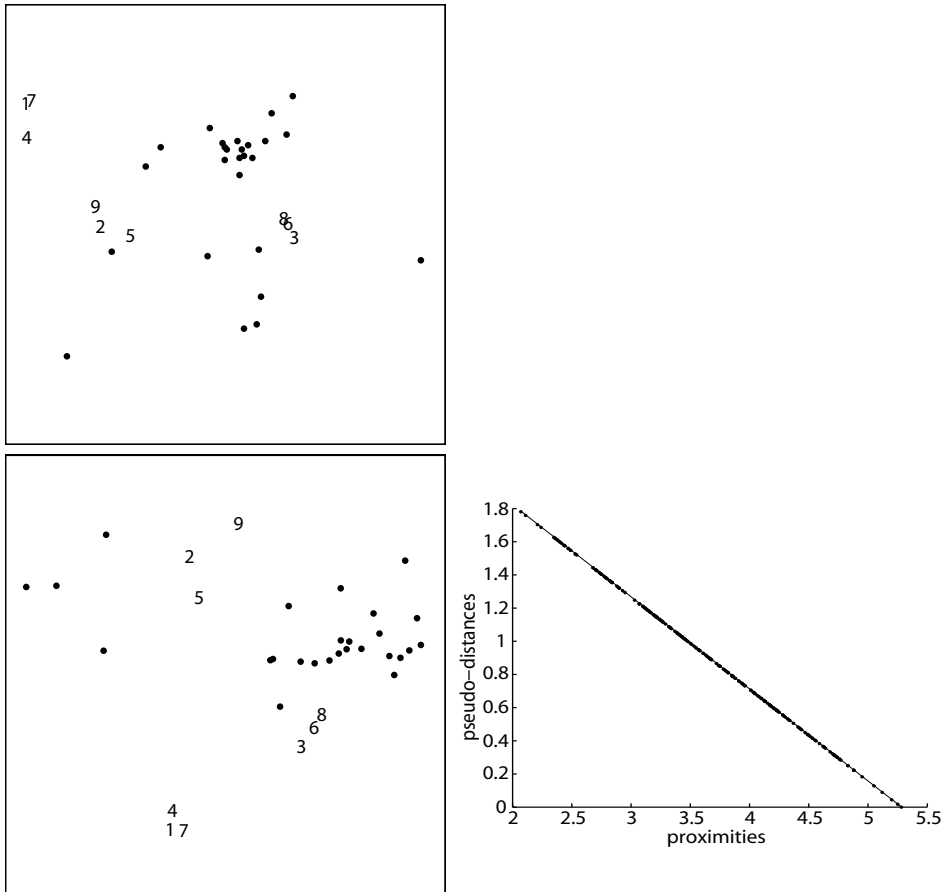
Figure 4.5    Configurations (left-side panels) and transformation plot (right-side panel) for the brewery data set with an unconditional interval transformation using NEWFOLD (top panel) and PREFSCAL with $\lambda$=0.5 and $\omega$=0.5 (bottom panels).

The solutions of NEWFOLD and PREFSCAL (Figure 4.5) show more variability in the distances. The breweries and the attributes are less clearly separated than with ALSCAL and KYST. The improved spread of points is confirmed with the higher variability measures in Table 4.3. Also, Shepard's D-INDEX and DeSarbo's I-INDEX have improved over ALSCAL and KYST. However, the fit measures for NEWFOLD are not good; these measures are even worse than those for ALSCAL and KYST. This could be expected, since NEWFOLD, after its a-priori transformation, does not pursue an optimal interval transformation. In the interval case, the a-priori transformation consists of subtracting the smallest dissimilarity from the data and NEWFOLD continues with a row-conditional

metric unfolding, *without* estimating an intercept. This quasi-metric analysis assures a non-degenerate solution, but without estimating an optimal intercept, it is bound to produce worse fit statistics.

The PREFSCAL solution in Figure 4.5 (bottom panels) provides the best fit statistics, non-degeneracy index D-INDEX, and intermixedness index I-INDEX, while maintaining sufficient variability in both pseudo-distances and distances. Yet the PREFSCAL configuration is in close agreement with the metric solution without additive constant obtained by Borg and Bergermaier.

*Breakfast Data*

P. E. Green and Rao (1972) obtained rankings from 21 Wharton School MBA students and their wives concerning their overall preference for 15 breakfast items. Every individual placed the stimulus number of an item in the rank positions from highest (1 ='most preferred') to lowest (15 ='least preferred'), indicating their closeness or proximity towards a breakfast item, which resulted in row-conditional ordinal (preference) data. Figure 4.6 shows the configurations (left-side panels) and transformation plots (right-side panels) obtained with ALSCAL, KYST, and GENFOLD, respectively. The breakfast items are indicated by their plotting code and the individuals are shown as black dots.

In Figure 4.6, both ALSCAL and GENFOLD show degenerate solutions. For ALSCAL, the breakfast items are situated on the circumference of an imaginary circle, with the majority of individuals in the center of the circle. This configuration shows very little variability in the distances, which is confirmed by the variability measures and Shepard's D-INDEX (see Table 4.4). The GENFOLD solution is an absolute degenerate solution in the conditional sense (cf. Proposition 2): The distances to the breakfast items are identical within individuals as $d(\mathbf{x}_i, \mathbf{y}_j) = d_i$ for all $i, j$, but not necessarily across individuals (see Figure 4.6, bottom right-hand panel). All individual transformations have horizontal lines, but the lines differ in intercept, which is acknowledged in Table 4.4, as the conditional variability measures and Shepard's D-INDEX are zero or close to zero. The value of STRESS-2 is remarkable. Closer inspection of the results shows that the STRESS-2 function value becomes 0/0. Actually, the value becomes $\epsilon/\epsilon^2 = \epsilon^{-1}$, where $\epsilon$ is a very small number: The numerator indicates that the weighted sum-of-squares loss function is close to zero ($\epsilon$),

Figure 4.6    (facing page)    Configurations (left-hand panels) and transformation plots (right-hand panels) for the breakfast data set with row-conditional ordinal transformations using ALSCAL (top panel), KYST (middle panels), and GENFOLD with p=2 (bottom panels). The breakfast items (and plotting codes) are given in Table 2.1.
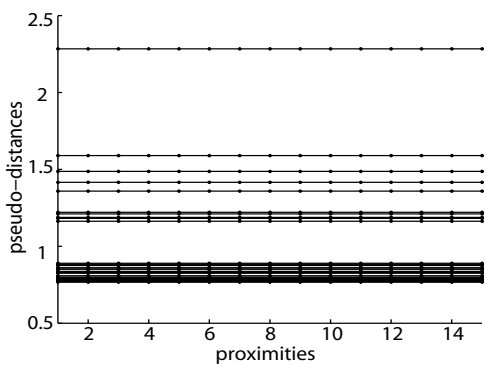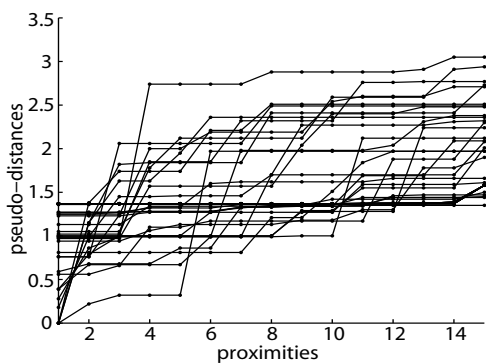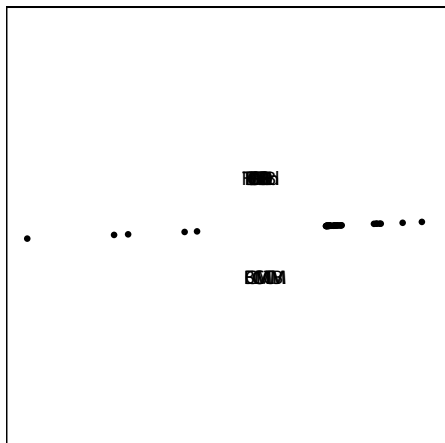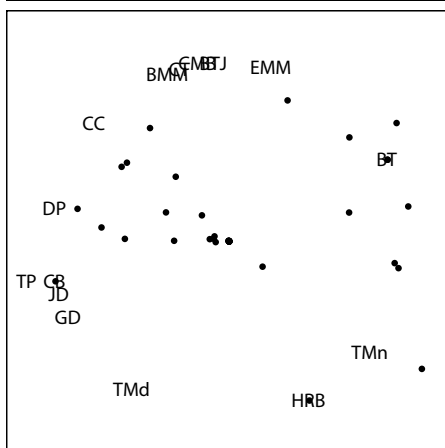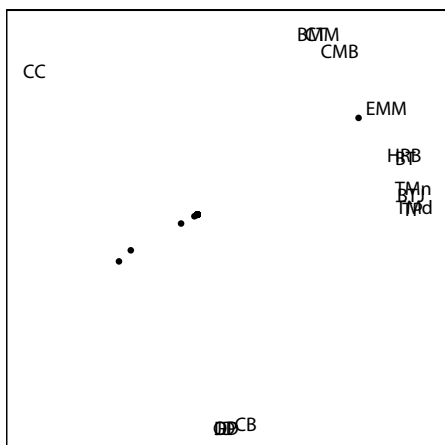
but with a denominator even closer to zero ($\epsilon^2$), the actual STRESS-2 function value becomes huge. This condition also explains the perfect value for VAF.

Although, upon first inspection, the KYST configuration (Figure 4.6) shows no apparent degeneracy, the transformation plots for 15 out of 42 individuals (right-side panel) indicate at least a partially degenerate solution: Irrespective of the ranking of the items, all transformed proximities obtain one and the same value, except for rank 15, which slightly deviates under influence of the normalization factor. In the transformation plot, these transformations are indicated by an horizontal line with an intercept of about 1.37. In the configuration, the ideal point of such an individual, situated in the center of the configuration, has nearly the same distance towards all breakfast items, except for TP, which is slightly larger. KYST shows a substantial improvement over ALSCAL and GENFOLD, though. Considering the variability measures for KYST in Table 4.4, both variance and variation of the distances are satisfactory: The inter-set distances in the configuration show enough variability. The conditional variability measures, however, indicate that within individuals there is very little variability. We deduced the same fact earlier from both the configuration and the transformation plot.

Both NEWFOLD and PREFSCAL (Figure 4.7), on the other hand, show no apparent form of degeneracy. Both individuals and items are well spread throughout the configuration, which is also indicated by the variability measures in Table 4.4, and the individual transformations show distinct non-zero slopes. NEWFOLD does not provide final pseudo-distances (for the numerical

Table 4.4    Various measures for ALSCAL, KYST, GENFOLD, NEWFOLD, and PREFSCAL for the breakfast data set.

| | ALSCAL | KYST | GENFOLD | NEWFOLD | PREFSCAL |
|---|---|---|---|---|---|
| Badness-of-Fit | | | | | |
| Stress-2 | 792.8 | 0.399 | 176212 | 0.974 | 0.560 |
| Goodness-of-Fit | | | | | |
| VAF | 0.034 | 0.833 | 1.000 | 0.481 | 0.807 |
| $\bar{R}$ | 0.463 | 0.911 | NAN | 0.747 | 0.874 |
| $\overline{RHO}$ | 0.478 | 0.615 | 0.523 | 0.730 | 0.798 |
| $\overline{TAU}_b$ | 0.360 | 0.608 | NAN | 0.580 | 0.709 |
| Variability | | | | | |
| $\text{var}(\mathbf{D})$ | 0.013 | 0.122 | 0.091 | 0.197 | 0.192 |
| $\upsilon(\mathbf{D})$ | 0.114 | 0.374 | 0.317 | 0.495 | 0.483 |
| $\text{var}_c(\mathbf{\Gamma})$ | N/A | 0.003 | 0.000 | 0.369 | 0.233 |
| $\upsilon_c(\mathbf{\Gamma})$ | N/A | 0.078 | 0.000 | 0.750 | 0.575 |
| Non-Degeneracy and Intermixedness | | | | | |
| D-INDEX | 0.094 | 0.492 | 0.054 | 0.741 | 0.749 |
| I-INDEX | 12.542 | 0.544 | 0.872 | 0.055 | 0.184 |

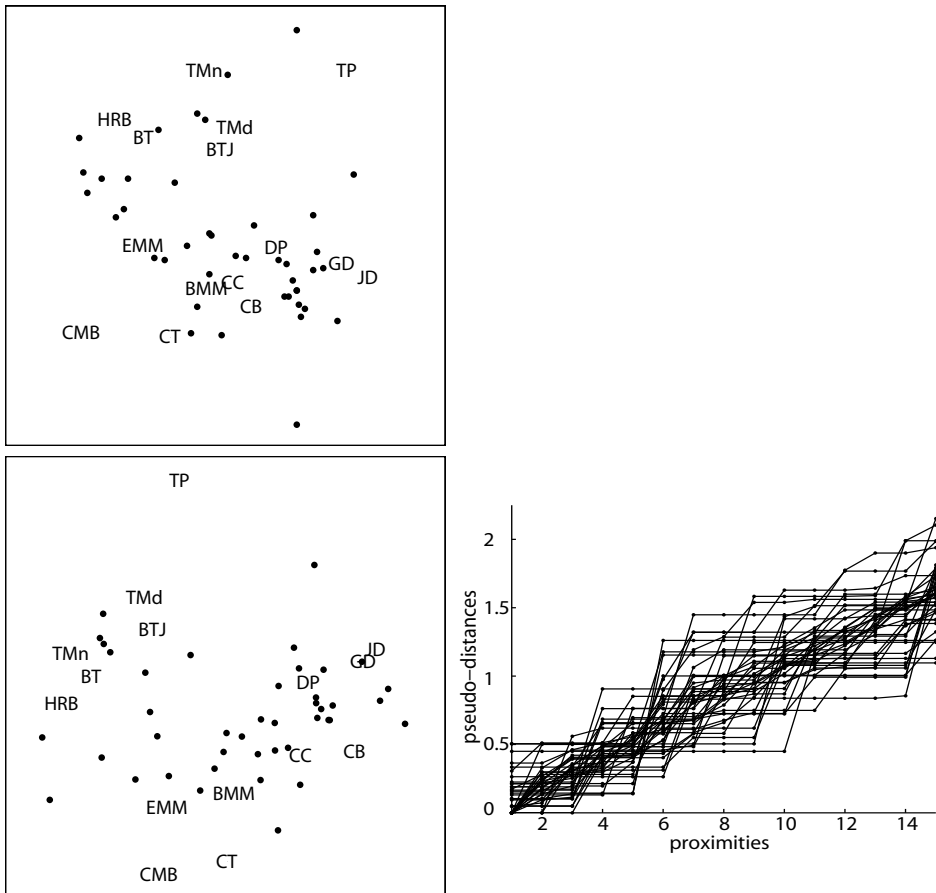N/A = not available; NAN = not a number (due to divisions by zero).

Figure 4.7    Configurations (left-side panels) and transformation plot (right-side panel) for the breakfast data set with row-conditional ordinal transformations using NEWFOLD (top panel) and PREFSCAL with $\lambda$=0.5 and $\omega$=0.3 (bottom panels). For the breakfast plotting codes, see Table 2.1.

results, the a-priori pseudo-distances were used instead), so a transformation plot could not be created, but a similar plot as for PREFSCAL might be expected. Again, PREFSCAL outperforms NEWFOLD on the fit measures. Since NEWFOLD does not perform a nonmetric analysis, but a row-conditional metric analysis *without* an intercept after an a-priori transformation, this could be expected. In the ordinal case, the a-priori transformation consists of replacing rank order one with a zero and drawing the differences between subsequent rank orders from a normal distribution (Kim et al., 1999, and A. Rangaswamy, personal communication, August 20, 2003). Introducing random error and using a metric transformation leads to relatively poor fit statistics. On the whole,

PREFSCAL performs well on all measures, providing a solution with variability in both distances and conditional pseudo-distances as well as acceptable fit statistics.

For an interpretation of the PREFSCAL configuration in Figure 4.7, two dimensions can be identified. The horizontal dimension discriminates between soft and hard bread or toast. The vertical dimension has nonfat items on one end and the more fattening items on the other, lower end. Some clusters can be identified too. On the upper left-hand side there are toast and hard breads with margarine or butter, in the middle on the right there are sweet donuts and coffee cakes, and on the lower side of the configuration there are muffins.

## 4.6 SUMMARY

Degenerate solutions are a common problem in nonmetric unfolding. The origin of this problem kept many researchers occupied, but without a clear conclusion. DeSarbo and Carroll (1985) believed "that a possible cause for degeneracy is the error in the data" and although Carroll (1972) mentioned that "the problem is theoretically tractable", according to Kim et al. (1999) "the exact causes of degeneracy are currently unknown". This chapter has formally identified what the exact necessary and sufficient conditions for degeneracy are, insofar as degeneracy is defined as a solution with zero STRESS and constant distances. Shepard (1974) proposed the percentage of distinct distances among all distances as a "rough index of the *non*degeneracy of a solution". Although the measure is able to indicate an absolutely degenerate solution, the magnitude of a degeneracy is more difficult to grasp. In order to determine the extent of a degeneracy, DeSarbo et al. (1997) have defined degeneracy indices that measure how well the two sets of points are interspersed. The sum-of-squares of these indices come close to a single measure, but still indicates intermixedness, not degeneracy. In conclusion, we have to rely on multiple measures to determine whether a solution is near-, quasi-, strongly, or extremely degenerate.

An approach initiated by Kruskal and Carroll (1969) used a normalization or penalty function to avoid solutions in which the interset distances become equal. Although solving these problems with a penalty function was qualified as "basically unsound" by de Leeuw (1983), Kruskal and Carroll (1969) believed that their method might be successful, even though for real data the method performed less than fully satisfactorily. About the use of STRESS-2 they remarked: "However, the second scale factor is zero for each judge except the distinguished judge. Consequently, with one exception, the individual badness-of-fit values are 0/0. While this is not a positively reassuring sign, it does suggest that a computation which seeks to minimize the overall badness-of-fit will avoid this kind of configuration." We believe that De Leeuw's results

do not indicate that Kruskal and Carroll's approach was conceptually flawed, but that the remedy was not strong enough, and their considerable optimism for the future now turns out to be warranted.

In this chapter, it is argued that the cause for degeneracies resides in the set of admissible transformations. When this set includes an additive constant, all data values are allowed to become equal. Admissibility of an additive constant is not limited to nonmetric (ordinal) transformations, but also applies to interval or spline transformations. However, if we resort to methods that work by restricting the set of admissible transformations we do loose generality and have to make quite a number of arbitrary specifications. So it is better to find some other way to make solutions with a large additive constant unattractive, without *ad hoc* interventions. In this chapter, we have shown that our adaptation of Kruskal and Carroll's approach into PENALIZED STRESS does work. Using the coefficient of variation, a scale-free and relative measure of variability, in the adapted penalty function, and using the harmonic mean for determining the penalty in the row-conditional case, we obtain a general badness-of-fit function that successfully avoids a degenerate solution in a wide range of circumstances. Additionally, two strength parameters were introduced in PENALIZED STRESS to be able to adjust the balance between R-STRESS and the penalty. The simulation study made clear that one of these parameters could be restricted to a constant value, whereas the other parameter was best chosen in a specific interval. Two applications showed that our method compares favorably with other methods, not only with artificial data, but also with real data.

# 5

RESTRICTED UNFOLDING

The fundamentals of preference mapping are revisited in the context of a new restricted unfolding method that has potential for wide application to product optimization for consumers. Since more of an attribute is not necessarily preferred, the unfolding distance model provides estimates of ideal points for consumers and therefore provides a more adequate representation of the preference relationships, compared to conventional internal preference mapping. Compared to other ideal point methods, the new unfolding technique offers advantages in terms of allowing for the ordinal nature of the ratings, rather than implicitly assuming that ratings are linear. The proposed restricted unfolding model incorporates property fitting, both passive, as a separate, second step, and active, as a restriction on the product locations. This is also available as a restriction on the respondents' locations and as such establishing a link between internal and external preference mapping.

## 5.1 INTRODUCTION

There are many different methods of preference mapping analysis and it has been a subject of continuous development through from the 1960's to the present day. Often alternative types of analysis need to be carried out and compared before choosing the most appropriate model for any given set of data. Traditionally there are two basic classes of preference mapping techniques. A recent comparison of both types is discussed in van Kleef, van Trijp, and Luning (2006): " Internal preference analysis gives precedence to consumer preferences and uses perceptual information as a complementary source of information. External analysis, on the other hand, gives priority to perceptual information by building the product map based on attribute ratings and only fits consumer preferences at a later stage." (van Kleef et al., 2006, page 388). From a conceptual point of view, internal preference analysis achieves a multidimensional space representing differences among the products based on the preference data, whereas external preference analysis finds a multidimensional representation based on attribute ratings. From a technical point of view, both analyses are quite similar. Internal preference

analysis estimates consumer respondent vectors and product points by internal preference mapping (Carroll & Chang, 1970; Carroll, 1980) or (categorical) principal component analysis, and then estimates vectors for the attributes with (multiple) regression analysis, given the product points, an additional analysis also known as property fitting (Carroll & Chang, 1964a). The term *internal* is due to Carroll (1972), referring to the simultaneous (internal) estimation of both respondent and product coordinates. External preference analysis, on the other hand, first estimates attribute vectors and product points, essentially by the same techniques as internal preference analysis, and then estimates consumer respondent points or vectors, given the product points, by some form of (multiple) regression analysis, also known as external unfolding (see Carroll, 1972; Schiffman, Reynolds, & Young, 1981; Meulman, Heiser, & Carroll, 1986). The term *external*, also due to Carroll (1972), entails the second step, in which the respondent coordinates are estimated, given (externally) the products points. Examples of both types of preference analysis can be found in McEwan and Thomson (1989), Daillant-Spinnler, MacFie, Beyts, and Hedderley (1996), Arditti (1997), Murray and Delahunty (2000), and Thompson, Drake, Lopetcharat, and Yates (2004).

The model proposed in this chapter, the restricted unfolding model, finds points for respondents and products, possibly restricted by additional variables. The model is similar to internal preference analysis as it gives precedence to the preference data and uses additional attribute data to enhance interpretation. However, the proposed model deviates from existing models on several important aspects, namely concerning the representation of individuals (ideal point or vectors), the type of model (restricted or not), the number of (analysis) steps (one, two, or both), the type of transformations (especially monotone transformations), and concerning measures against degenerate unfolding solutions. The model elaborates on and combines the work of Carroll and Chang (1964a), Carroll (1972), de Leeuw and Heiser (1980), DeSarbo and Rao (1984), and Busing, Groenen, and Heiser (2005).

The remainder of this chapter discusses the restricted unfolding model in detail and elaborates on an extensive example on preferences for different types of tomato soup. Possible extensions for product development are reintroduced and comparisons are made with existing models, namely with Principal Component Analysis (PCA), Landscape Segmentation Analysis (LSA), and Euclidean Distance Ideal Point Mapping (EDIPM). The chapter concludes with a discussion.

## 5.2 THE RESTRICTED UNFOLDING MODEL

The *unrestricted* unfolding model finds a lower-dimensional representation of respondents and products, where the distances between both sets correspond

as closely as possible with the preferences of the respondents for the products, that is, $\Delta \cong d(\mathbf{X}, \mathbf{Y})$, where $\Delta$ represents the preferences, $d(\mathbf{X}, \mathbf{Y})$ represents the Euclidean distances between the respondent coordinates $\mathbf{X}$ and the product coordinates $\mathbf{Y}$, and $\cong$ represents a least squares relation. Large distances correspond to the products least preferred and, consequently, small distances correspond to the most preferred products. Products that appear close together are analogously preferred just as respondents with similar preference profiles share the same positions in space.

The preferences, often measured at the ordinal measurement level, are allowed to be optimally transformed to fit the distances as closely as possible, i.e., $\Gamma = f(\Delta)$, where $f(\cdot)$ is a class of monotone transformation functions. This class of functions includes linear and ordinal transformation functions, but also, for example, monotone spline transformation functions. The unrestricted unfolding model with optimal transformations of the preferences is then given as $\Gamma \cong d(\mathbf{X}, \mathbf{Y})$. Transformations of the preferences $\Delta$ are usually performed per respondent (i.e. row-conditional) due to the inter-respondent incomparability of the preferences, that is, $\gamma_i = f_i(\delta_i)$ for each respondent $i$. Unconditional transformations, allowing comparisons over respondents, are also available.

For the *restricted* unfolding model, the coordinates for their part can be restricted to form a linear combination of variables $\mathbf{E}$, such that $\mathbf{X} = \mathbf{E}_x \mathbf{B}_x$ or $\mathbf{Y} = \mathbf{E}_y \mathbf{B}_y$, giving the model as $\Gamma \cong d(\mathbf{E}_x \mathbf{B}_x, \mathbf{E}_y \mathbf{B}_y)$. The variables, as columns in $\mathbf{E}$, often referred to as prediction variables, explanatory variables, external variables, or attribute variables, are also allowed to be optimally transformed to fit (the projection of) the coordinates, such that $\mathbf{Q} = g(\mathbf{E})$, where $g(\cdot)$ is a class of transformations, not necessarily monotone. The full restricted unfolding model with optimal transformation of both preferences and variables is then given as $\Gamma \cong d(\mathbf{Q}_x \mathbf{B}_x, \mathbf{Q}_y \mathbf{B}_y)$.

The restricted unfolding model corresponding with the example in the next section is given as

$$\Gamma \cong d(\mathbf{X}, \mathbf{Q}\mathbf{B}). \tag{5.1}$$

The coordinates of the products $\mathbf{Y}$ are thus a linear combination of the transformed variables $\mathbf{Q}$, with $\mathbf{B}$ as the matrix with regression coefficients. This relation corresponds with property fitting (Carroll & Chang, 1964a) or, equivalently, the second step in internal preference analysis (van Kleef et al., 2006). Both procedures determine directions in the configuration based on the variables $\mathbf{E}$. However, the unified restricted unfolding model actually restricts the coordinates, whereas property fitting is a separate analysis and poses no restrictions on the configuration.

The optimal direction in the configuration for variable $\mathbf{q}$ is determined by

$$\mathbf{p} \cong \mathbf{q}, \tag{5.2}$$

where $\mathbf{p}$ is the projection of the product coordinates $\mathbf{Y}$ onto direction vector $\mathbf{a}$, i.e., $\mathbf{p} = \mathbf{Ya}$. The direction vector $\mathbf{a}$ for variable $\mathbf{q}$ is given by $\mathbf{a} = (\mathbf{Y'Y})^{-1}\mathbf{Y'q}$ (see Chang & Carroll, 1969). Although (5.2) is not minimized directly by minimizing (5.1), Meulman and Heiser (1984) showed that minimizing (5.1) results in product coordinates $\mathbf{Y} = \mathbf{QB}$ offering projections $\mathbf{P}$ with improved approximations of the variables $\mathbf{Q}$ as compared to property fitting. Moreover, the direction vectors, collected as columns in the matrix $\mathbf{A}$, are used to plot the directions of the variables in the configuration, as with property fitting, since these provide the optimal variable values after projection of a location (coordinate) onto the direction. The directions provided by $\mathbf{B}$ would be used for optimal interpolation, that is, finding an optimal location based on variable values.

For the restricted unfolding model, an alternating least squares and iterative majorization framework is used to estimate all parameters. In order to avoid degenerate solutions, a persistent problem that pursued unfolding for decades (see van Deun, 2005, for a recent review), the coefficient of variation is used in a penalty function as described in Busing, Groenen, and Heiser (2005) and used in IBM SPSS PREFSCAL (Busing, Heiser, et al., 2005). The technical appendix (page 147 and further) discusses the algorithm, as well as the differences and extensions to Busing, Groenen, and Heiser (2005), in detail. A more general reference for multidimensional unfolding and iterative majorization is Borg and Groenen (2005).

5.3  CASE STUDY

A consumer study was carried out to guide the optimal product characteristics of tomato soup. Nine different formulations of soup were developed and each was tasted and assessed by 298 consumers recruited to take part in the study. A small portion of each soup was consumed, with palate cleansing between each product, to minimize the effects of taster fatigue. Each consumer was classified according to gender (male, female), age group (22-31, 32-40) and soup consumption frequency (low, medium, high). Each product was rated on a 9-point liking scale, ranging from 1=dislike extremely to 9=like extremely. The particular feature of this case study was that a formulation-based design was used to create $3 \times 3$ factorial set of products with the two formulation factors chosen to vary systematically over two key sensory dimensions and which for simplicity are referred to here as (1) Flavor Intensity and (2) Sourness, both with levels set as 'Low', 'Medium' and 'High' (see Table 5.1). In subsequent analyses and graphics, individual products are identified by the particular combination of Flavor Intensity (Low=I$_1$, Medium=I$_2$, and High=I$_3$) and Sourness (Low=S$_1$, Medium=S$_2$, and High=S$_3$). Table 5.1 (upper part) shows the average liking over all consumers. The most liked product overall was the

medium Flavor Intensity and low Sourness (I2-S1) tomato soup. An mixed model analysis of variance (Table 5.1, lower part) shows that all fixed effects are significant, with smaller F-values for Flavor Intensity and the interaction between Flavor Intensity and Sourness, indicating smaller differences between the liking of all Flavor Intensity levels, contrasted by the strong significant differences between the levels of Sourness. The random effect parameters are all significant: respondents have different preferences for soups, with respect to both Flavor Intensity and Sourness levels. These effects will be explored in more detail by the restricted unfolding analysis to follow.

From a trained sensory panel additional data was collected on the nine different types of soup. From an initial set of 29 sensory attributes, a reduced

Table 5.1    Descriptive statistics (upper part with median, mean, and standard deviations of overall likings and share of choices) and a mixed model analysis of variance (lower part) comparing 9 different tomato soups on flavor intensity and sourness.

| Descriptive Statistics | | | | | |
|---|---|---|---|---|---|
| Flavor Intensity | Sourness | Median | Mean | Standard Deviation | Share of Choices |
| Low | Low | 6 | 6.01 | 2.08 | 14% |
| | Medium | 6 | 5.93 | 2.00 | 13% |
| | High | 5 | 4.97 | 2.14 | 5% |
| Medium | Low | 7 | 6.32 | 2.06 | 17% |
| | Medium | 6 | 5.95 | 1.92 | 11% |
| | High | 6 | 5.52 | 1.97 | 8% |
| High | Low | 6 | 6.01 | 2.14 | 15% |
| | Medium | 6 | 5.91 | 1.98 | 11% |
| | High | 5 | 5.25 | 2.02 | 6% |

Share of choices = percentage of first choices.

| Mixed Model Analysis of Variance | | | | |
|---|---|---|---|---|
| Fixed Effects | Numerator DF | Denominator DF | F | Sig. |
| Flavor Intensity | 2 | 594 | 5.25 | .005 |
| Sourness | 2 | 594 | 57.40 | .000 |
| Flavor Intensity x Sourness | 4 | 1188 | 2.41 | .048 |

| Random Effects | Estimate | Standard Error | Wald Z | Sig. |
|---|---|---|---|---|
| Respondents | 0.94 | 0.12 | 7.50 | .000 |
| Respondents x Flavor Intensity | 0.50 | 0.08 | 5.99 | .000 |
| Respondents x Sourness | 0.27 | 0.07 | 3.76 | .000 |
| Residuals | 2.44 | 0.10 | 24.37 | .000 |

set of 10 was included in the analysis with the selection based on removing attributes which were less discriminating between the products and creating a reduced set that captured the underlying sensory space effectively. The actual identities of the selected attributes are not included in the analysis output because of potential issues of commercial sensitivity. From the (significant) correlations between the attributes, two groups of related attributes can be deduced: the Flavor Intensity group consists of A01, A02, A03, A08, and A10, while the Sourness group consists of A06, A07, and A09, where A06 has a negative correlation with Sourness.

*Restricted Unfolding*

A restricted unfolding model, the model provided in (5.1), for the tomato soup data results in a two-dimensional preference map for 298 respondents and 9 soups. The two dimensions are optimal, which was assessed through a scree plot. In Figure 5.1, the respondents are represented by dots and the soups by the levels of the two variables, Flavor Intensity and Sourness. The preference scale was reversed, without loss of generality, to get the preferences in line with the distances: small distances now correspond to high preference and vice versa. More specifically, the distances between respondents and soups correspond with the monotonically transformed preferences of the respondents. The applied transformation relaxes the *equally spaced* preference scale, while maintaining the order restriction on the preferences. Figure 5.1 shows that there is a concentration of respondents near the low sour soups (S1). This is consistent with the mean overall likings from Table 5.1 as well as with the mixed model analysis of variance result. The respondents are distributed about evenly over the levels of Flavor Intensity, matching the smaller F-value for Flavor Intensity and Flavor Intensity-Sourness interaction in Table 5.1. The other sources of variation are reflected by the fact that respondents differ in position with respect to the nine soups, Flavor Intensity, and Sourness.

The two variables used to define the types of soup are represented as directions (or dimensions or axes) in the configuration. The original variables are uncorrelated ($r = 0.000$), which does not preclude correlated directions in the solution (actual $r = -0.023$) due to changing variable values by optimal variable transformations. The projections of the soup coordinates onto these directions correspond with the levels (values) of the variables. The variables are equally spaced in formulation terms (actual concoction of substance), but not necessarily in perceptual terms. The variables are optimally monotonically transformed to meet this feature. The variable Sourness, initially with categories 1 (Low), 2 (Medium), and 3 (High), is transformed monotonically, *keeping ties tied*. In the plot, this phenomenon results in identically projected values for soups of the same Sourness (for example, the transformed value for
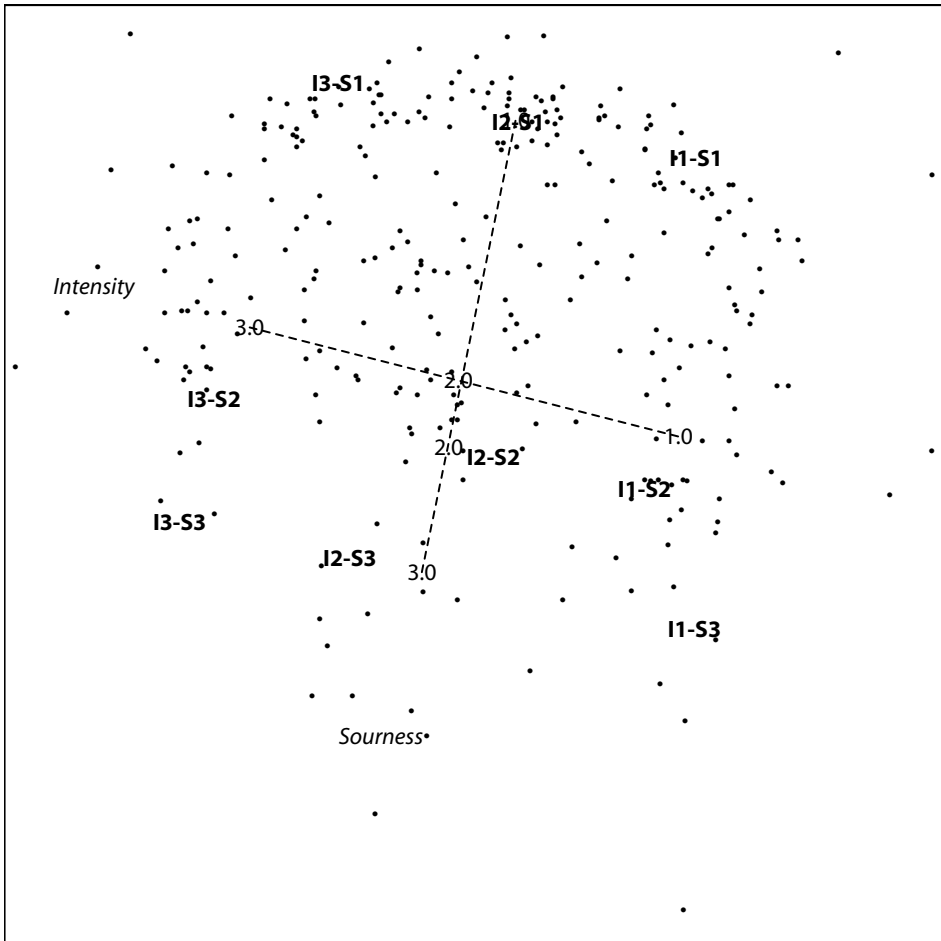
Figure 5.1   Restricted unfolding solutions for the tomato soup data with two active variables (flavor intensity and sourness) restricting the product configuration.

low sour soups becomes -2.76, see Table 5.2). The monotone transformation, however, accommodates differences in intervals, as the distance between projection of the low and medium sour soups differs from the distance between projections of the medium and high sour soups. Since the transformation of the factor Flavor Intensity allows *ties to be untied*, this fact is not observed for Flavor Intensity: medium Flavor Intensity and low Flavor Intensity appear closer together for medium sour soups than for high sour soups. However, as with the keep ties tied option for the monotone transformation of Sourness, the order of soups after projection remains *restricted* to the initial order of

the variables: a low Flavor Intensity soup never overtakes a medium Flavor Intensity soup on the Flavor Intensity direction. Respecting unequal spacing in perceptual terms should also allow the transformation for Sourness to untie ties, but this is omitted for illustrative purposes. The actual choice for either handling of ties depends on both substantive (data properties, interpretation, frugality) and statistical (fit, variation) considerations.

Although the current solution has a variance accounted for (VAF; Average squared correlation between the transformed preferences and the distances) of 0.82, it is more appropriate to use the sum-of-squares accounted for (SSAF; Average of the sum of squared differences between the transformed preferences and the distances divided by the product of the sum-of-squares of the transformed preferences and the distances) or even a rank order coefficient, considering the monotonically transformed data. Kruskal's STRESS-1 (Kruskal, 1964a) equals 0.17, which corresponds to a SSAF of 0.97 (see Busing & de Rooij, 2009), and Spearman's rho, providing the average rank order correlation between the transformed preferences and the distances, equals 0.85. In 71% of the cases, the respondent is closest to its highest preferred type of soup (FIRST; Proportion indicating the correspondence of the highest preference(s) with the smallest distance(s) relative to the total number of respondents; see Technical Appendix G). For the unrestricted unfolding solution, these values are 0.80, 0.98, 0.84, and 65%, respectively, for VAF, SSAF, RHO, and FIRST. The variation of the distances and the transformed preferences, an important ingredient of the loss function PENALIZED STRESS (see Busing, Groenen, & Heiser, 2005), equal 0.47 and 0.46, respectively. In terms of fit and variation, this solution is a quite good solution, especially when compared to the unrestricted solution, as restricting the product configuration even improves some statistics while maintaining others on a comparable level. An optimal model follows the data characteristics concerning transformation function(s) and conditionality, has the best or comparable fit and variation statistics, allows for easy interpretation

Table 5.2    Initial and transformed variables for the tomato soup data.

| Soup Type | Initial Flavor Intensity | Transformed Flavor Intensity | Initial Sourness | Transformed Sourness |
| --- | --- | --- | --- | --- |
| I1-S1 | 1 | -1.81 | 1 | -2.76 |
| I1-S2 | 1 | -2.15 | 2 | 0.72 |
| I1-S3 | 1 | -3.01 | 3 | 2.04 |
| I2-S1 | 2 | 0.08 | 1 | -2.76 |
| I2-S2 | 2 | -0.57 | 2 | 0.72 |
| I2-S3 | 2 | 0.64 | 3 | 2.04 |
| I3-S1 | 3 | 1.98 | 1 | -2.76 |
| I3-S2 | 3 | 2.41 | 2 | 0.72 |
| I3-S3 | 3 | 2.42 | 3 | 2.04 |

and prediction, and is parsimonious. The current restricted unfolding solution corresponds adequately to these characteristics.

*Active and Passive Variables*

The restricted unfolding model presented in Figure 5.1 uses two variables describing the products to restrict the configuration. These variables are incorporated in the model, and one single analysis suffices to link preferences and variables. These variables are called *active* variables, which means that the variables participate actively in finding an optimal configuration, keeping a strict relation between coordinates and variables, i.e., $\mathbf{Y} = \mathbf{QB}$. During the iterative optimization process, variables are transformed and regression coefficients updated to result in optimally transformed variables, optimally fitting the (by then transformed) preferences.

Property fitting, on the other hand, entails a separate analysis, fitting variables concerning products or respondents to a fixed configuration. In this case, the external variables are called *passive* variables, variables that have no influence on the (fixed) configuration. Still, these variables can be optimally transformed (van der Kooij, 2007), which makes it feasible to enter nominal or ordinal variables in the equation. Optimal transformations also result in improved fit, but only so far as the fixed configuration permits.

Figure 5.2 shows the same restricted unfolding solution, except that for an improved interpretation or prediction of the soups, the 10 attribute variables from the trained sensory panel were fitted to the configuration. The measurement level is assumed to be numerical for these average ratings, and thus a numerical transformation is chosen. The demographic variables describing the respondents are ordinal (age groups and consumption frequency), ties allowed to be untied, and nominal (gender). In Figure 5.2, these demographical variables are passive, but it is also possible to use respondent variables to restrict the configuration, for example to estimate an ideal point discriminant model (Takane, Bozdogan, & Shibayama, 1987).

All variables, active and passive, respondent and product related, can be described in terms of direction (direction vectors) and strength (variance accounted for, VAF). The variables are graphically represented by the straight lines (numerical transformed variables) and dotted lines (monotonically transformed variables), black for active variables and gray for passive variables. The endpoints of the lines provide the minimum and maximum values of the original variables as reference, unless one of the endpoint falls outside the perimeter, in which case the midpoint value is provided.

Two inserted boxes in Figure 5.2 contain information about the association strength of the passive variables. The VAF per variable is computed as the squared correlation between the transformed variable ($\mathbf{q}$) and the projection
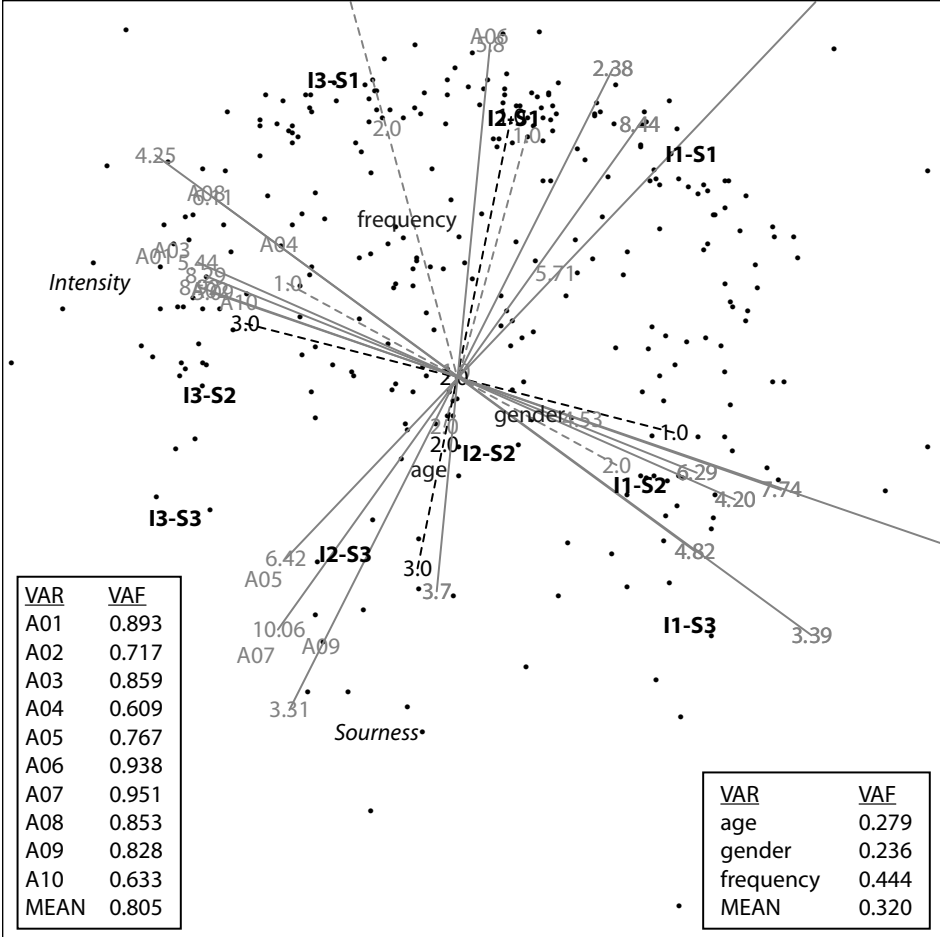
Figure 5.2   Restricted unfolding solutions for the tomato soup data with both active and passive variables, the first set restricting the product configuration, the latter two sets superimposed on respondents and product configuration, but only after convergence.

of the respective coordinates onto the direction vector of the variable (**p**). With two active variables restricting the configuration, and two dimensions, the VAF of the active variables is perfect (1.000). Concerning the passive variables, the VAF's of the demographic variables are rather small (right-hand side box), while the VAF's of the attributes are considerable (left-hand side box). The mean variance accounted for (values after MEAN) provides an overall fit measure for the variables. Predicting the respondent and soup coordinates with the aid of these variables explains 32% and 81% of the variance in respondent

and product coordinates, respectively. Selection of variables for the restricted unfolding model faces the same difficulties as variable selection for (linear) regression models (see, for example, A. Miller, 2002) and worse, since the regression model is only a subproblem in the restricted unfolding model.

## 5.4    OPTIMIZING PRODUCT DEVELOPMENT

The restricted unfolding model is well suited for the optimization of products. There exists a variety of proposals in the literature on how to determine optimal locations for new products, given the locations for existing products and respondents, both for deterministic or single choice models and for probabilistic choice models (see, for example, Shocker & Srinivasan, 1974; Albers & Brockhoff, 1977; P. E. Green & Krieger, 1989; Baier & Gaul, 1999). Providing the best procedure is beyond the scope of this chapter, so we merely demonstrate the potential use of the restricted unfolding model maximizing the share of choices using some heuristic methods, the method of "search through coarse and fine grid", which proved flexible in a variety of distance metrics, cost functions, choice models, search boundaries, and the like (Shocker & Srinivasan, 1974). Note that the restricted unfolding model optimizes distances, so the following examples are also described in terms of distances. Once an optimal location is determined, the restricted unfolding model allows for an easy description of the product in terms of product attributes or relate the product to respondent characteristics.

*Finding the optimal product from several prototypes*

The optimal position for a prototype product can be defined as the position where it will attract the most consumer respondents. Logically, in a deterministic framework, this position is located somewhere in the center of the respondents. There are, however, several possibilities for the definition of this center. Since the results from the previous section are given in two dimensions, the present centers are also defined in two dimensions, but easily generalized to more than two dimensions.

   The center that is optimally related to the distances from the unfolding solution is not the min-max center, which is in the middle of the respondents, $c_1' = (\min\{X\} + \max\{X\})/2$, nor is it the mean center $c_2' = n^{-1}\mathbf{1}'X$, which is the point that minimizes the sum of the *squared* distances to the respondents points, i.e., $\min \sum \|X_i - c_2'\|^2$, but it is the median center $c_3$, which is the point that minimizes the sum of the distances to the respondents points, i.e., $\min \sum \|X_i - c_3'\|$. This problem is known as the Fermat-Weber location problem, as it arises in the optimization of the location of sales units (Weber, 1909), or as the problem of finding the spatial median in quantitative geography
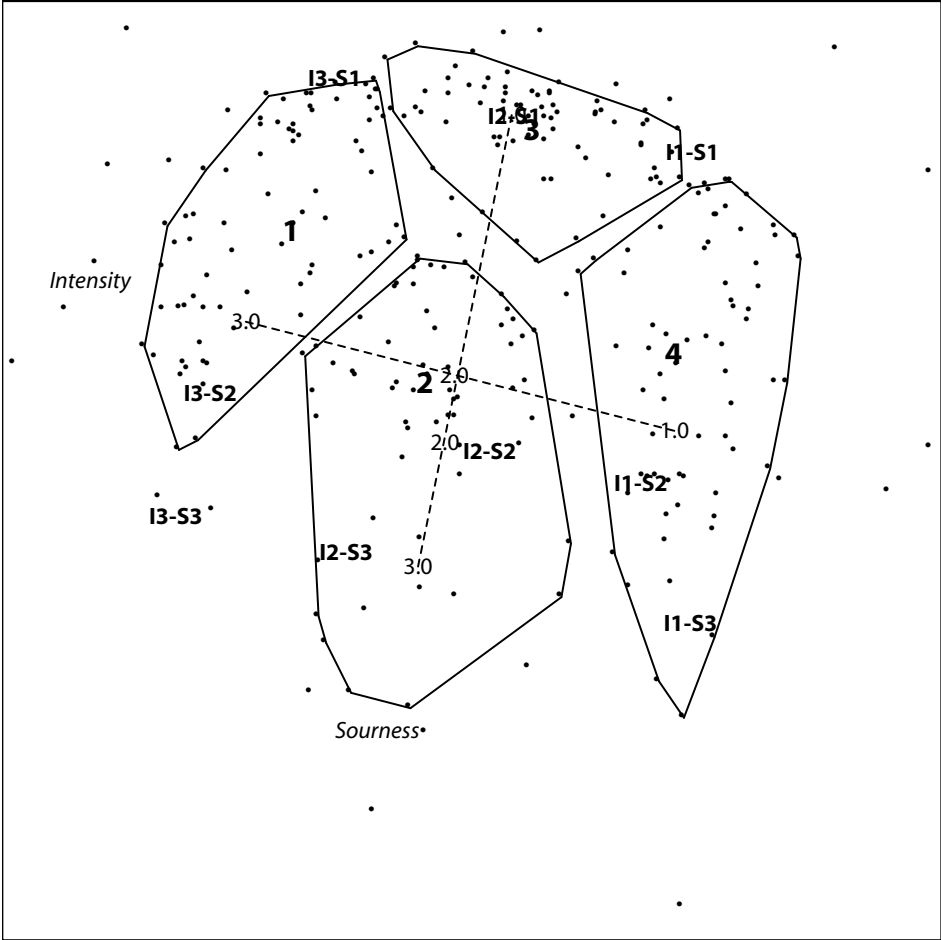
Figure 5.3    Restricted unfolding solution for the tomato soup data with four optimal clusters with centers through probabilistic d-clustering and 90% convex hulls.

or geometry (Hayford, 1902; Sviatlovsky & Eells, 1937). Descent algorithms for finding the median center $c_3$ are suggested in Gower (1974) and Bedall and Zimmermann (1979), and recommended in Brown (1985), and Chaudhuri (1996), while an iterative solution, actually one of the first examples of iterative majorization (see de Leeuw, 1977a; Groenen, 1993; Heiser, 1995, for iterative majorization in an MDS framework), is given by Weiszfeld (1937).

When the respondents fall apart into two distinct clusters, the median center probably ends up somewhere in between the two clusters, a location not preferred by either group of respondents. In that case, multiple centers are

preferred, one center for each cluster. The researcher has then to decide which cluster to pursue, aided by additional information on clusters or respondents, either by labeling, averaging, or by incorporating active or passive variables into the restricted unfolding model describing respondent characteristics. A recently proposed probabilistic d-clustering procedure determines cluster centers based on the median center (Ben-Israel & Iyigun, 2007). Additional features allow for the computation of different cluster sizes and the number of clusters (see Iyigun, 2007; Iyigun & Ben-Israel, 2008). For now, an adapted version of the Calinski and Harabasz (1974) statistic indicates that the optimal cluster solution consists of four clusters. Figure 5.3 shows the four cluster solution with 90% convex hull (nonparametric) confidence intervals. The share proportion per cluster corresponds well with the average score per cluster, as can be seen from Table 5.3. The soups with the highest mean share proportion per cluster coincides with the soups of cluster 3, which are the low sour tomato soups, specifically the medium flavor intensity and low sour tomato soup.

Now, suppose we use the median center of cluster 3 with coordinates (1.17, 3.68) as the location for the new soup and we want to determine the scale values for Flavor Intensity and Sourness related to this location. Projection of the center coordinates onto the variable vectors gives $-0.15$ for Flavor Intensity and $-2.72$ for Sourness. Back-transformation or calibration (Gower & Hand, 1996; Gower, Meulman, & Arnold, 1999), that is, $e = g^{-1}(q)$, using linear interpolation (for intermediate values) or extrapolation (for out-of-scale values), is used to reduce the transformed values to the original variable levels, such that Flavor Intensity becomes 2.00, as projection of the center ends up in between the medium Flavor Intensities, as can be deduced from Figure 5.3, and Sourness becomes 1.01, after linear interpolation. Both calibrated values can also be computed from Table 5.2, as $-0.57 \leqslant 0.15 \leqslant 0.64$ (third column)

Table 5.3  Average score and share proportion per cluster for each soup.

| Soup | Average Score per Cluster | | | | Share Proportion per Cluster* | | | | |
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | Mean |
|---|---|---|---|---|---|---|---|---|---|
| I1-S1 | 5.175 | 4.970 | 7.132 | 6.680 | | | 0.237 | 0.400 | 0.159 |
| I1-S2 | 4.888 | 6.030 | 6.013 | 6.880 | | 0.015 | | 0.480 | 0.124 |
| I1-S3 | 4.100 | 4.806 | 4.697 | 6.333 | | 0.030 | | 0.093 | 0.031 |
| I2-S1 | 6.263 | 5.552 | 7.750 | 5.627 | 0.013 | 0.104 | 0.671 | | 0.197 |
| I2-S2 | 5.638 | 6.552 | 5.842 | 5.867 | | 0.507 | | 0.027 | 0.134 |
| I2-S3 | 5.475 | 6.567 | 5.105 | 5.067 | | 0.194 | | | 0.049 |
| I3-S1 | 7.038 | 5.194 | 7.000 | 4.653 | 0.537 | | 0.092 | | 0.157 |
| I3-S2 | 6.875 | 6.104 | 5.697 | 4.907 | 0.450 | 0.119 | | | 0.142 |
| I3-S3 | 6.362 | 5.507 | 4.763 | 4.320 | | 0.030 | | | 0.008 |

*zero's are omitted.

and Flavor Intensity becomes 2.00 (second column), and as $-2.72$ is very close to $-2.76$ (fifth column), Sourness becomes 1.01 (after linear interpolation).

From a probabilistic point of view, the optimal locations maximize the probability products are chosen. The probability for product $k$ by respondent $i$ is defined in terms of inverse distances as $p_{ik} = cd_{ik}^{-1}$, where $d_{ik}$ is the distance between product $k$ and respondent $i$ and $c$ a constant that ensures that the chances add up to one. Maximizing the choice over all products is identical to minimizing the inverse of the sum over all inverse distances, that is, $\max \sum_k p_{ik} = \min(\sum_k cd_{ik}^{-1})^{-1}$. Finding the locations for all respondents is related to the harmonic means cluster function (see Zhang, Hsu, & Dayal, 1999; Zhang, 2000). A solution for the probabilistic case is not pursued here.

*Finding the optimal product among competitors*

The restricted unfolding model can also be used to identify the optimal profiles for products facing competition. For the next two examples, we will use a simple grid search with contour plots for both deterministic and probabilistic choice models (see, for example, Baier & Gaul, 1999).

For the first example, suppose the current solution (Figure 5.1) represents the current tomato soup market and we would like to launch a new product that has first choice for most respondents. Specifying a grid with potential soups on top of the solution from Figure 5.1 and interpolating equal grid values provides the contour plot given in Figure 5.4. Grid values are determined by first rank ordering the distances from all respondents to current (large dots) and potential (grid) soups, counting the number of first choices for the potential soups, and dividing the sum by the number of respondents. The core of a MATLAB function accompanying Figure 5.4 computes these grid values as Z (line 6), which can be plotted with one of the MATLAB contour functions (line 9).

The grid values in Figure 5.4 indicate the proportion of respondents with first choice for the potential soup in that area. The plot shows two major areas where the proportion is over 0.16. Also considering the areas with more than 0.12 or 0.14 and considering the direction of both variables, we considered the crossed area as the location for our new tomato soup, since this area benefits from the adjacent 0.12 and 0.14 areas on both variables.

Having identified the position for the new tomato soup $(-2.9, 1.3)$, projection of the new soup coordinates onto the two directions provides only the transformed values for the attribute variables. Again, back-transformation is used to reduce the transformed value to the original variable level (see Table 5.2). For Flavor Intensity, projection onto the direction vector gives the value 2.22, which gives 3.0 exactly for the original level of Flavor Intensity. For
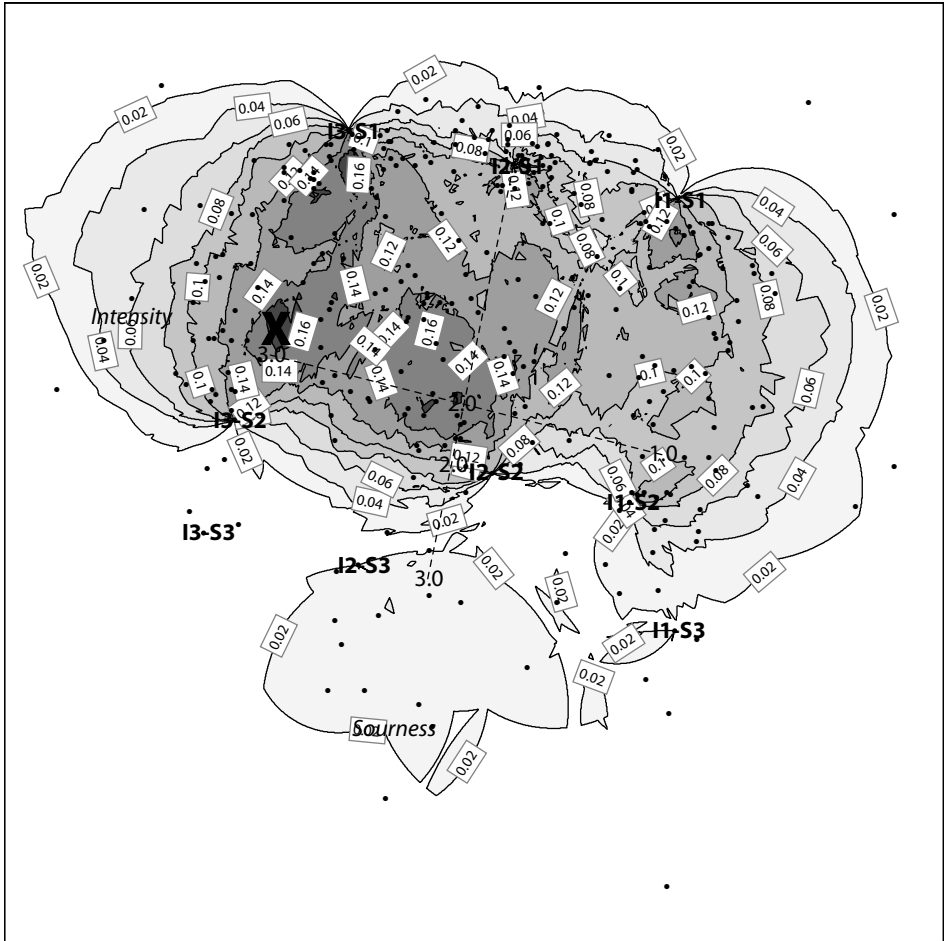
```
                              ─── Code Start ───
1  mind = min (D');                          % minimum distance per respondent
2  for i = 1:nx                              % loop over grid x-axis
3    for j = 1:ny                            % loop over grid y-axis
4      g = [x(i),y(j)];                       % grid coordinate
5      d = distance (g,X);                    % distances to all respondents
6      Z(i,j) = sum (d < mind)/n;             % proportion first choices
7    end                                      % end loop grid y-axis
8  end                                        % end loop grid x-axis
9  contourf (x,y,Z');                        % filled contour plot
                              ─── Code End ───
```

Figure 5.4  Restricted unfolding solution with discrete choice contours, corresponding MATLAB code, and an optimal product location (X). The products represent own internal prototype products in relation to potential competitor products.

Sourness, projection ends up in between two observations, with the transformed value of $-0.48$. Linear interpolation gives 1.66 as the original Sourness level for the new product. If the scale type of Sourness would be categorical, instead of numerical (imagine that the level of sourness was manipulated by varying the number of lumps of sugar), the numerical value is rounded off to the nearest available category value, in this case the value 2.

For the second example, suppose the current solution (Figure 5.1) represents the current tomato soup market and we would like to launch a new product that has the highest chance of being chosen by all respondents. In this case, we use the probability for product $k$ by respondent $i$ defined in terms of inverse distances as $p_{ik} = (1 + d_{ik} \sum_j d_{ij}^{-1})^{-1}$, where $d_{ik}$ is the distance between respondent $i$ and *new* product $k$. Specifying a grid with potential soups on top of the solution from Figure 5.1 and computing the average probability over respondents for each position (represented by z in the MATLAB code, line 6) provides the contour plot given in Figure 5.5. Part of the MATLAB function is provided below the corresponding figure.

The plot shows two areas where the average probability is over 0.13. Assuming that the position for the new soup close to the currently most preferred soup would be a too great commercial risk, the area with the cross is chosen as the best position for the new soup. Again, the new soup coordinates $(-0.62, 1.60)$ are projected onto the two variables providing the transformed values for both attribute variables. Back-transformation gives numerical values 2.05 and 1.50, respectively, for Flavor Intensity and for Sourness.

## 5.5 COMPARISON

The restricted unfolding model creates configurations with locations for respondents and products from the respondents preferences for these products, while taking into account product attributes or respondent characteristics. Commonly, these data are handled with external or internal preference analysis (see van Kleef et al., 2006). The primacy of the product locations is on the preferences, which are based on perceived benefits (Meyers & Shocker, 1981), not on the perceptions, which are similarity judgements based on characteristic attributes. Using preferences instead of perceptions is recommended by Derbaix and Sjöberg (1994) as these are more stable and certain judgements. This rules in favor of internal preference analysis (over external preference analysis) and we will therefor compare the restricted unfolding solution with three such analyses, a vector model, Principal Components Analysis (PCA), and two ideal point models, Landscape Segmentation Analysis (LSA) (Ennis, 1999, 2005; Rousseau & Ennis, 2008) from The Institute for Perception (Richmond, VA) and Euclidean Distance Ideal Point Mapping (Meullenet, Lovely, Threlfall, Morris, & Striegler, 2008).
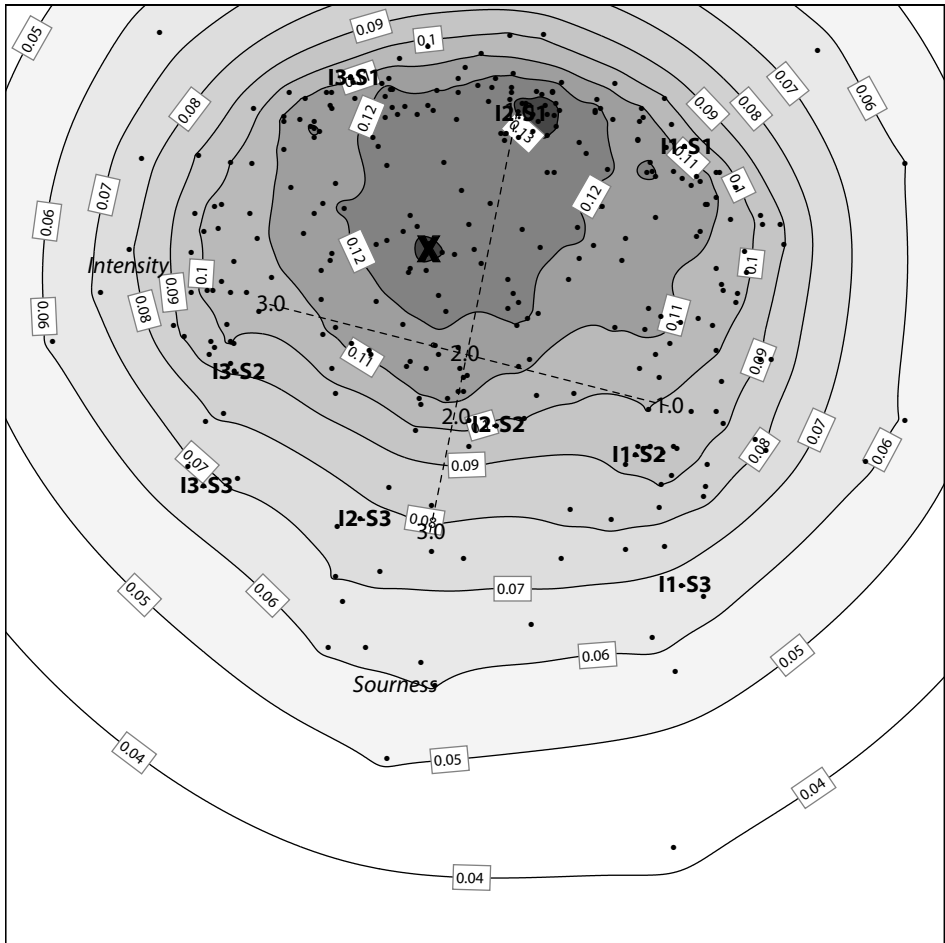
```
_____ Code Start _____
1  sumd = sum (1./D');                        % sum over inverse distances per respondent
2  for i = 1:nx                               % loop over grid x-axis
3    for j = 1:ny                             % loop over grid y-axis
4      g = [x(i),y(j)];                        % grid coordinate
5      d = distance (g,X);                     % distances to all respondents
6      Z(i,j) = mean (1./(1+d.*sumd));         % probability for soup on grid coordinate
7    end                                       % end loop grid y-axis
8  end                                         % end loop grid x-axis
9  contourf (x,y,Z');                          % filled contour plot
_____ Code End _____
```

Figure 5.5   Restricted unfolding solution with probability choice contours, corresponding MATLAB code, and an optimal product location (X). The products represent own internal prototype products in relation to potential competitor products.

*Comparison with the vector model*

The vector model represents respondents as vectors or directions instead of points in the configuration. Although the preferences are still defined in terms of distances, an intermediate projection step is needed for the vector model. This is, in contrast with a remark from van Kleef et al. (2006, p. 390), not only a practical but also a conceptual issue. The underlying preference curves for the vector model are either linear (PCA or MDPREF , see Chang and Carroll (1969); Carroll and Chang (1970)) or monotonically increasing (CATPCA , see Meulman, van der Kooij, and Heiser (2004); Meulman, Heiser, and SPSS (2005); Linting, Meulman, Groenen, and van der Kooij (2007)), meaning that a vector points in the direction of maximum preference. The further a product projects onto a vector, the more it is preferred. The projection model thus assumes that consumer respondents have extreme optimum points, since the ideal product is situated (far) outside the cloud of actual product points. The vectors indicate the ideal direction without specifying the location. As a consequence, the projection model forces the actual most preferred products to be on the edge of the product cloud, and optimal products even further (see Figure 5.6). As Ennis (2005) notes, "this type of model is well suited to account for attributes, such as luxury or off-taste, for which the consumer's ideal will fall outside any conceivable region of the sensory space into which products are placed", but not for accounting attributes "such as sweetness or flavor level, for which the consumer will reject products with too much or too little of the attribute". Consumer preference data, however, is almost only considered of the last type and projection models are thus less suited to describe the optimal locations as opposed to the distance model, where a single peaked preference curve is utilized. Here, a respondents' position coincides with its ideal product and moving away from this point in any direction decreases its preference.

Practically, the vector model configuration must always be interpreted through projection. First, products are projected onto the respondents' vector, and then, differences between the projections are interpreted in terms of distance. The intermediate step makes it more difficult to "read" a vector model configuration as compared to a distance model configuration. Sometimes, the respondent vectors are only represented by their endpoint, due to the vast amount of black ink generated by the vectors. Omitting the vectors might arouse confusion, because the impression might be raised that distances can be interpreted directly, between products points and respondent vector endpoints, which is definitely not allowed according to the model.

There are several methods to estimate the parameters of the vector model, two of the most renowned being Multidimensional Preference Scaling (MDPREF) (Chang & Carroll, 1969; Carroll & Chang, 1970), and Principal Component Analysis (PCA). The methods differ in the preliminary normalization of the

data and the scaling of the configuration and the vectors. Figure 5.6 (left-hand panel) displays the resulting PCA configuration. The product configuration shows the same pattern of soups as observed before, accounted for by the content of the soups based on Flavor Intensity and Sourness, although the clear gap between the low sour soups (S1) and the other soups has vanished. The majority of respondent vectors are directed towards the low sour soups (S1), and not towards the (low intensity – I1) sour soups (S3), which is in agreement with the data and previous results. The length of a respondent vector corresponds with the VAF per respondent, which is not the case for MDPREF, as MDPREF normalizes the vectors to equal lengths. The proportion of correct first choices (FIRST = 0.57) cannot match the restricted unfolding solution results (FIRST = 0.71). Main reasons for the difference between PCA and restricted unfolding are the handling of the data, the unfolding analysis optimally monotonically transforms the data, and the underlying preference model. Allowing optimal transformations of the preferences in PCA, that is, using Categorical Principal Component Analysis (CATPCA) and specifying an ordinal optimal scaling level for the preferences, increases the VAF from 0.42 to 0.61. Nevertheless, as discussed before, linear preference curves do not fit this type of data well and CATPCA does not overcome this drawback. The contours in Figure 5.6 (left-hand panel) indicate the proportion of first choices, similar to Figure 5.4. Illustrative for the underlying linear preference curves in vector models are the outside positions for the most preferred soups and the even further positioned optima for new soups.

*Comparison with other ideal point methods*

An internal ideal point model used quite widely is Landscape Segmentation Analysis (LSA). It employs a probabilistic similarity model (Ennis, Palen, & Mullen, 1988; Mullen & Ennis, 1991; Ennis, 1993) to position respondents and products on a map. The model uses a fixed transformation of the liking ratings, for example from a 9-point hedonic scale, as a similarity measure between the products and the consumer's ideal point. Sensory information from the same set of products may be superimposed on the resulting plot to estimate the sensory profile of positions on the map with a high density of consumer ideals. This amounts to property fitting or the passive variables approach in the restricted unfolding analysis. The right-hand panel of Figure 5.6 shows the solution of a landscape segmentation analysis (IFPRESS v7.3), where the contours were plotted using the MATLAB code shown below the figure. Here, as with the IFPRESS software (Ennis & Rousseau, 2004), the contours display the proportion of respondents per unit area in the configuration. The darker the area, the more dense the number of respondents. Note that the MATLAB code
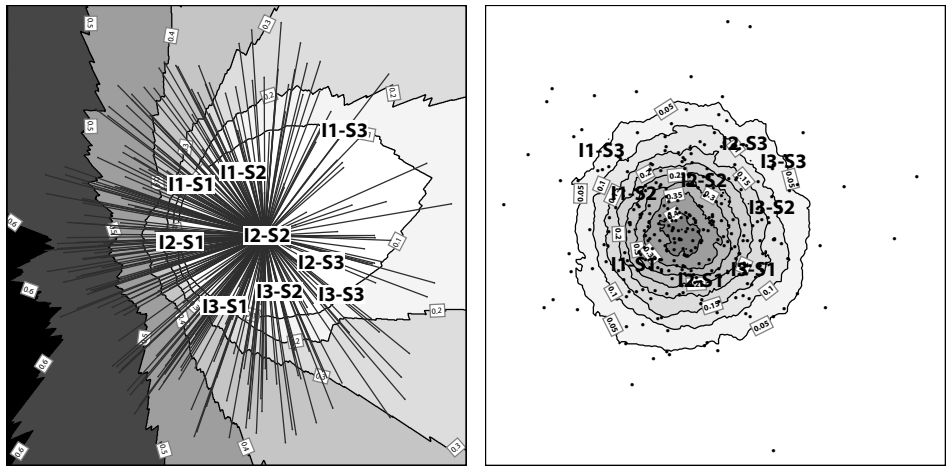
```
                 Code Start                                    Code Start
1  ...                                       1  ...
2  P = Y*X';                 % projections   2  for i = 1:nx              % x-loop
3  maxp = max (P);           % maximum       3    for j = 1:ny            % y-loop
4  for i = 1:nx              % x-loop        4       g = [x(i),y(j)];     % coordinate
5    for j = 1:ny            % y-loop        5       d = distance (g,X);  % distances
6       g = [x(i),y(j)];     % coordinate    6       Z(i,j) = sum (d < unit)/n;  % density
7       d = g*X';            % projections   7    end                     % end loop
8       Z(i,j) = sum (d > maxp)/n;  % proportion  8  end                      % end loop
9    end                     % end loop      9  contourf (x,y,Z');        % plot
10 end                       % end loop      10 ...
11 contourf (x,y,Z');        % plot                           Code End
12 ...
                 Code End
```

Figure 5.6   PCA solution (left-hand panel) and landscape segmentation analysis solution (right-hand panel) for the tomato soup data with first choice contours and corresponding MATLAB code.

is similar to the function for the contours with proportions of first choices (see code below Figure 5.4), replacing 'mind' with some fixed 'unit'.

The LSA solution shows almost the same solution for the product locations, although mirrored on both axes, but a quite different solution for the locations of the respondents. The highest density of respondents is situated in the center of the solution, surrounded by the 6 most preferred soups. The statistics (computed on optimally transformed data for a fair comparison and displayed in Table 5.4) indicate that the preference orders (RHO) are slightly worse recovered as compared to the restricted unfolding solution, and only half of the most preferred soups of the respondents (FIRST) are represented by the distance relations in the configuration (with 0.71 for the restricted unfolding solution). The sum-of-squares accounted for is identical to the sum-of-squares accounted for of the restricted unfolding model, while the variance accounted for is slightly worse. The fact that LSA is a probabilistic procedure and the

restricted unfolding method a deterministic procedure has not been taken into account.

The Euclidean Distance Ideal Point Mapping (EDIPM) method Meullenet et al. (2008) is an extension of MDPREF (Chang & Carroll, 1969) and identifies ideal points for the respondents in a PCA product configuration. Although EDIPM combines the vector model (to determine product locations) with the ideal point model (to determine respondent locations), the result is a true distance model. A related model, the VIPSCAL model (van Deun, 2005; van Deun, Groenen, & Delbeke, 2005), also combines both models but allows respondents to be represented by either vectors or points. The EDIPM method works as follows. First, PCA (equivalent with MDPREF on mean centered data) is used to obtain a product configuration. Then, a fine grid is searched for the optimal locations of each respondents ideal point. For this purpose, the correlation at each grid point is computed between the respondents preference scores and the distances between the grid point and the product locations. The grid point with the best correlation is chosen as ideal point for a respondent. This search is repeated for each respondent. The contours in Figure 5.7 are the overlaid acceptable regions of all respondents. An acceptable region for a single respondent consists of all grid points that are statistically non-significantly different ($\alpha = 0.10$) from the respondents ideal point concerning the correlations previously determined.

The product configuration from Figure 5.7 is identical to the configuration from the vector model (Figure 5.6, left-hand panel), obviously, since both configurations are determined by PCA. However, in contrast with the restricted unfolding model and the landscape segmentation analysis, most respondent locations are not only situated on the outside edge of the product configuration, but even on the outside edge of the specified grid (see Figure 5.7, left-hand panel). The grid, that was used for Figure 5.7, is about ten times the range of the product configuration, quite arbitrarily, although the extreme respondent positions are observed for a wide range of grid sizes. These extreme positions were to be expected, since the vector model is identical to an ideal point model with the ideal points at infinity (Carroll, 1972; DeSarbo & Rao, 1986; Borg & Groenen, 2005; van Deun, 2005). With the products located according to the

Table 5.4   Goodness-of-fit statistics ideal point methods.

| Statistic | Restricted Unfolding Model | Landscape Segmentation Analysis | Euclidean Distance Ideal Point Model |
|---|---|---|---|
| VAF | 0.82 | 0.73 | 0.45 |
| SSAF | 0.97 | 0.97 | 0.88 |
| RHO | 0.85 | 0.81 | 0.76 |
| FIRST | 0.71 | 0.49 | 0.64 |
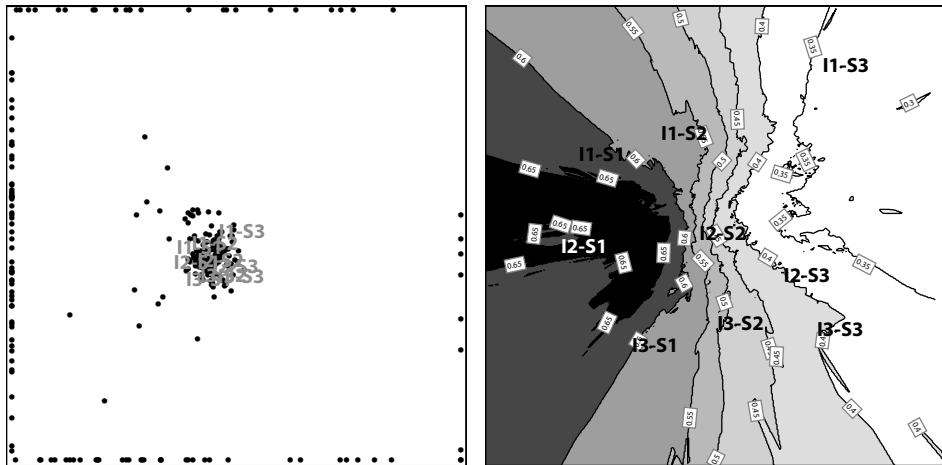
Figure 5.7 EDIPM solution for the tomato soup data (left-hand panel) with contours based on overlaid acceptable regions (right-hand panel). The right-hand panel depicts a detail of the left-hand panel, omitting the respondents and including contours.

vector model, with high preferred products on the outside of the configuration, as described before, respondent ideal points are prone to be at the periphery of the configuration, except for less well fitting respondents, who might reside in the vicinity of the products. The EDIPM merger of vector and ideal point models provides a contour plot with more or less parallel density paths (Figure 5.7, right-hand panel), similar to the PCA contour plot, instead of circular paths as with the 'real' ideal point models.

The goodness-of-fit statistics (see Table 5.4) for the EDIPM solution improves on those for PCA (VAF = 0.42), because badly fitting respondents in the vector model are better off in EDIPM, since they might find a better location in the middle of the products than on the periphery. The other statistics, again using optimally transformed data, are worse as compared to the landscape segmentations analysis (except for FIRST) and the restricted unfolding model, although the differences in configurations, particularly concerning the respondent locations, are more remarkable.

## 5.6 DISCUSSION

The restricted unfolding model, as proposed in this chapter, enables the researcher to represent both preference data and attributes data in one lower-dimensional configuration. Optimal transformations allow the use of both numerical and ordinal data and, in combination with the specification of active and passive variables, more or less restricted configurations are determined for

the respondents or products. The loss function compares preferences and true distances, not squared distances or projections. As a consequence, the interpretation of the resulting configuration is straightforward. The distance-based interpretation also extends to the additional analyses, such as the probabilistic d-clustering analysis and the grid search for both deterministic and probabilistic choice models.

The current model extends the GENFOLD-2 model (DeSarbo & Rao, 1984, 1986), which was designated as "the most advanced modelling of MDS-based componential segmentation" by P. E. Green and Krieger (1989), where componential segmentation refers to the use of external variables to restrict the configuration. Moreover, the serious problems related to the MDS-approach, as described by P. E. Green and Krieger (1989, p. 132), are currently under control: The difficulty of constructing joint spaces and ideal points, relating perceived dimensions to manipulable attributes, and problems of computational time (see also P. E. Green, Carroll, & Goldberg, 1981, p. 19). Difficulties in achieving global optima are reduced with the use of multiple (random) starts. Improvements with respect to the degeneracy problem (see Busing, Groenen, & Heiser, 2005), the addition of external variables, the extended choices in transformation functions, and the general availability in a major statistical package (IBM SPSS STATISTICS), assure a valuable enhancement of the ideal point unfolding model.

Future research might concentrate on the (prior) specification of variables, specifically which variables to include in the model. Although this problem is also observed in, for example, multiple linear regression (A. Miller, 2002), progress in other fields might create opportunities for the restricted unfolding model as well, although the various transformation and restriction options might complicate things (see van der Kooij, 2007). Another open research topic concerns back-transformation or calibration. Calibration, currently achieved by linear interpolation and extrapolation, might be improved for better prediction, especially when monotone transformation are concerned. It is hoped that further applications in consumer and marketing research will prove the value of the restricted unfolding model.

6

Unfolding creates configurations from preference information. In this chapter, it is argued that not all preference information needs to be collected and that good solutions are still obtained, even when more than half of the data is missing. Simulation studies are conducted to compare missing data treatments, sources of missing data, and designs for the specification of missing data. Guidelines are provided and used in actual practice.

## 6.1 INTRODUCTION

Multidimensional unfolding methods create perceptual spaces well-suited for consumers research (DeSarbo, Kim, Choi, & Spaulding, 2002) and marketing research (Balabanis & Diamantopoulos, 2004; DeSarbo et al., 1997). Unfolding represents both consumers and products as points in Euclidean space. The distance relation between consumers and products provides information about the preference structure of the consumers in such a way that consumers are closer to the products they prefer. The geometrical properties of the Euclidean space allow for simple and comprehensible interpretation of the relationships.

In consumer or marketing research, it would be more than convenient if consumer respondents only evaluate a subset of products. Respondents may be unable or unwilling to comply and fail to complete the evaluation of the full set of products. For example, in memory-based evaluations, respondents must have knowledge or at least be aware of the products under consideration to provide a useful evaluation. Without providing the respondents with additional information, which may be undesirable for several reasons, the evaluation set for each respondent might differ: Certain familiar products are evaluated more often than other products and some respondents evaluate more products than other respondents. Free-choice profiling (Arnold & Williams, 1986; Dijksterhuis & Gower, 1991) and the repertory grid method (Kelly, 1955; Rowe et al., 2005) also provide unequally distributed incomplete data, as respondents exploit different vocabularies. On the other hand, in studies involving tasting food products, sensory fatigue is a real issue which means that one wants to restrict the number of products tasted. All respondents evaluate an equally sized subset of all products under evaluation. The

expensive alternative is asking respondents to return on another occasion to complete the entire evaluation.

From a technical point of view, it is not at all necessary that the respondents evaluate all products. Most unfolding procedures allow for missing data without falling back on complete case analysis or listwise deletion. Consumer respondents may judge a subset of products and the unfolding procedures only use the valid, non-missing data, either by skipping the missing observations during computations (pairwise deletion) or by inserting 'valid' data at the missing observation, i.e., by imputation of missing data. Obtaining a good solution with incomplete data can help researchers get the most out of limited resources.

The focus of this chapter is (1) to investigate missing data designs for unfolding, (2) to determine key success factors for unfolding incomplete data, and (3) to provide guidelines for the proportion of non-missing data, required for a good correspondence with the results of a complete data analysis. In the following, we will first present unfolding, then briefly discuss the degeneracy problem that haunted this technique for decades, and discuss how this problem is currently solved. Then, incomplete data designs are presented that will be used in the Monte Carlo simulation study that aims to provide guidelines for researchers and data collectors, as to the amount of data that is still sufficient for proper solutions. An example with empirical data will be shown and we conclude with some general remarks.

6.2   UNFOLDING

Multidimensional unfolding is a technique that finds low-dimensional configurations for two sets of objects, the consumers and the products. The distances in the configuration between consumers and products should correspond as closely as possible with the preference ratings of the consumers for these products, in such a way that consumers are closest to the products they prefer the most. Unfolding in general consists of several different models. In this chapter, we use the model initiated by Coombs (1950) and generalized to the multidimensional case by Bennett and Hays (1960). In this model, $n$ consumers and $m$ products are represented as points in multidimensional space. The coordinate $x_i$ of a consumer is generally referred to as its ideal point; hence, this model is called the ideal-point model. The closer a product is to a consumers ideal point, the more this product is preferred by this consumer. Specific models have been suggested (external unfolding, weighted unfolding (Carroll, 1972)), but the most influential model was the nonmetric model, which joined nonmetric data and metric distances. Typical 'unfolding data' consist of rankings of products. These ranking data only contain ordinal information (i.e., no metric information) and the data are thus called nonmetric.

Shepard (1962a) showed that transformations, specifically ordinal transformations, can be used to shape this nonmetric relation in multidimensional scaling. Keeping the order relations of the original data intact, ordinal or nonmetric data are transformed into intermediate ratio data, which in turn are used to construct a metric Euclidean space (Kruskal, 1964a). Kruskal proposed to use the standardized residual sum-of-squares, abbreviated 'Stress', with STRESS-1 (Kruskal & Carroll, 1969) given as

$$\sigma_1(\mathbf{\Delta}, \mathbf{X}, \mathbf{Y}) = \|\mathbf{\gamma} - \mathbf{d}\|^2 / \|\mathbf{d}\|^2, \tag{6.1}$$

where $\mathbf{\Delta}$ is an $n \times m$ matrix with preferences and $\mathbf{X}$ and $\mathbf{Y}$ are the $n \times p$ and $m \times p$ coordinate matrices for consumers and products, respectively. In the unfolding case, $\|\mathbf{\gamma} - \mathbf{d}\|^2$ is the squared Euclidean norm $\|\cdot\|^2$ of the differences between some monotone transformation $f(\cdot)$ of the consumer's preferences $\mathbf{\Delta}$, with $\mathbf{\gamma} = f(\mathbf{\Delta})$, and the distances $\mathbf{d} = d(\mathbf{X}, \mathbf{Y})$, where $\mathbf{\gamma} = \mathrm{vec}(\mathbf{\Gamma})$ and $\mathbf{d} = \mathrm{vec}(\mathbf{D})$. The vec operator stacks the columns of its matrix argument. Standardization is regulated by $\|\mathbf{d}\|^2$, the sum-of-squares of the distances.

Nonmetric multidimensional scaling was one of the biggest breakthroughs in psychological research methods, but it ultimately caused unfolding's existential crisis: The freedom of the coordinates in space *and* the almost unrestricted transformations ensured that the thus weakly constrained unfolding model (Lingoes, 1977) was no longer identifiable (Busing, 2006). As a result, analyses tend to produce perfect (in terms of loss function) but meaningless (in terms of interpretation) configurations of points (Kruskal & Carroll, 1969; Roskam, 1968). Attempts to resolve the degeneracy problem often ended up in relatively unknown procedures or procedures with still uncertain results (see Borg & Groenen, 2005; Busing, Groenen, & Heiser, 2005, for an overview). Currently, there is a revival of attempts to set afloat unfolding with more prominent results (Kim et al., 1999; Busing, Groenen, & Heiser, 2005; Busing, 2006; van Deun, Groenen, & Delbeke, 2005; van Deun et al., 2007). All these attempts somehow restrict the model, either by restricting the transformations or by restricting the coordinates. An overview of the history of unfolding, the degeneracy problem, and currently available (computer) procedures can be found in van Deun (2005) and this monograph, Chapter 2.

To avoid the degeneracy problem, researchers often restrict themselves to a metric unfolding analysis. Although this chapter focusses on nonmetric unfolding analyses, a comparison is made between both types of analysis, but the results are relegated to Appendix 6.A. For the nonmetric unfolding analyses, we use the penalty approach of Busing, Groenen, and Heiser (2005) implemented in PREFSCAL (available in IBM SPSS STATISTICS), which avoids degeneracy by penalizing on the coefficient of variation $\upsilon(\cdot)$ (Pearson, 1896). Solutions with no or low variation in the transformed preferences and/or

distances, characteristics of degenerate solutions, are avoided by dividing (6.1) with a function of the variation coefficient $\upsilon(\boldsymbol{\gamma})$, i.e., the standard deviation of $\boldsymbol{\gamma}$ divided by its mean. The division causes the loss function to attain minimum values only in combination with a definite non-zero variation coefficient, that is, with sufficient variation in transformed preferences. PENALIZED STRESS is defined as

$$\sigma_{p}(\boldsymbol{\Delta}, \mathbf{X}, \mathbf{Y}, \omega, \lambda) = \sigma_{1}(\boldsymbol{\Delta}, \mathbf{X}, \mathbf{Y})/\mu(\boldsymbol{\Delta}, \omega, \lambda)$$

where $\mu(\boldsymbol{\Delta}, \omega, \lambda) = 1 + \omega\upsilon(\boldsymbol{\gamma})^{2/\lambda}$ with penalty parameters $\omega \geqslant 0$ and $0 < \lambda \leqslant 1$. Strong penalty parameters, with high values for $\omega$ and values for $\lambda$ closer to zero, tend to produce linear transformations, whereas weak penalty parameters, with $\omega$ close to zero and $\lambda$ closer to one, are prone to result in degenerate solutions. Details can be found in Chapter 4 or in Busing, Groenen, and Heiser (2005), although the function currently implemented in IBM SPSS PREFSCAL deviates slightly from the function presented therein: NORMALIZED RAW STRESS (normalization done with the sum-of-squares of the transformed preferences) is used instead of R-STRESS (no normalization) and an additional constant $(\upsilon(\boldsymbol{\delta})^{2})$ is used in combination with $\omega$. The default value for $\omega$ changed from 0.5 to 1.0 under the influence of this last addition (see Technical Appendix B).

In subsequent sections, the default settings of PREFSCAL are used: Classical scaling start with data imputation based on the triangle inequality (Heiser & de Leeuw, 1979a), row-conditional, ordinal (ties are kept tied) transformations, and default values for penalty parameters and convergence criteria, except for the maximum number of iterations, which was doubled to prevent imprecision due to premature termination of the iterative algorithm. Important in the present context is that PREFSCAL allows for a preference weight matrix with fixed non-negative weights. When this matrix is specified as an incidence matrix (a matrix with solely zeros and ones) it allows for the specification of missing data.

### 6.3 MISSING DATA

Missing data can be initiated by the researcher when only a subset of the products is presented to a respondent for evaluation. Proper factorial designs can be used to define subsets which in turn can be randomly assigned to respondents. On the other hand, missing data may be a consequence of the knowledge set of the respondent. In this case, missing data might be irregularly distributed over both respondents and products. Whatever the source of the missing data, unfolding needs to cope with the fact that some data is absent.

*Handling missing data*

In general, there are two common approaches for dealing with missing data: Deletion and imputation. The first approach simply excludes cases containing missing data, either for all computations (listwise deletion), or only for those computations where a missing for that case is involved (pairwise deletion). Either deletion scheme, listwise or pairwise, ignores possible systematic differences between complete and incomplete samples and produces unbiased estimates only if deleted records are a random sub-sample of the original sample. Data imputation, on the other hand, replaces missing data with 'valid' data through either single (deterministic) or multiple (stochastic) imputation and could lead to the minimization of bias. However, no imputation model is free of assumptions and the imputation results should hence be thoroughly checked for their statistical properties, such as distributional characteristics as well as heuristically for their meaningfulness (e.g., whether, for example, negative imputed values are possible). See Little and Rubin (1987) for well-documented drawbacks of either approach. We will now compare the two methods in a small simulation study.

*Simulation study: Imputation versus deletion*

The breakfast data (P. E. Green & Rao, 1972; DeSarbo et al., 1997; Borg & Groenen, 2005; Busing, Groenen, & Heiser, 2005; van Deun, 2005) for which 21 MBA students and their wives rank ordered 15 breakfast items, are used to compare the recovery of the complete data solution using three methods: Deletion (no imputation), respondent average imputation, and product average imputation. To create incomplete data, 5 out of 15 items per respondent were set missing by specifying an incidence matrix using a known balanced incomplete block design (see Table 6.1). Each method was replicated 1000

Table 6.1 Incomplete block design for v=15, k=5, r=14, and b=42, taken from Design Computing* specifying the column numbers with either missing or non-missing data for a 42 × 15 incidence matrix. The entries indicate the 5 column numbers, displayed in 3 blocks of 14.

| | | | | | |
|---|---|---|---|---|---|
| 12  3  9 13 14 | 9 13  1  4 12 | 8  5  1 12  6 | 14 10  3  6 12 | 12 14  2 10  7 | 15  4  1  2 13 |
| 6 11 15  4  3 | 1 10  9  2 15 | 11 12 13 14 15 | 13  7  5  9 15 | 3  8 13 10 11 | 3  1  4 10  5 |
| 8  1  3  9 15 | 12  9  2  5  6 | 15  4  5 14  8 | 13  3 12  5  2 | 14  7  2  5  1 | 3  6 13  9  7 |
| 11  2  5  7  3 | 4  9  7  8 12 | 14  7  1  9 11 | 5 10  6  9 15 | 1  6 13 14  8 | 11  1 13  6  5 |
| 1  4 12  3 10 | 7 15 11  8 12 | 4  7 13  2 10 | 13 15  8 10  2 | 7 13  8 10  6 | 11  6  9  4  2 |
| 6  7 11  1 10 | 9  4 10 14 11 | 6 12 15  7  4 | 1  7 15 14  3 | 3 11  8  9  2 | 5 10 12 15 11 |
| 2  6  3 14 15 | 2  4  8 14  6 | 7  4  8  5  3 | 10 14  8  9  5 | 11 12  8  1  2 | 11 13  4  5 14 |

*see http://www.designcomputing.net/gendex/bib/b4.html.

times with permutated rows and columns of the incidence matrix on each instance.

The quality of the equivalence between the distances of the unfolding analyses with and without missing data, i.e., the recovery of the unfolding solution, is quantified using Tucker's congruence coefficient $\phi_{xy}$ comparing both sets (respondents and products) and $\phi_y$ comparing the product sets only (Burt, 1948b; Tucker, 1951). By using $\phi$, a scale-independent similarity measure for ratio data, a Procrustes analysis to match configurations becomes superfluous, since the distances are independent of rotation and translation and the congruence coefficient is independent of dilation (see Technical Appendix G). At the individual level, the influence of missing data is determined with Kendall's rank order correlation $\tau_b$ (Kendall, 1948), comparing the rank ordered distances of the complete and incomplete data solutions for identical respondents, averaged over respondents. The comparison measures $\phi_{xy}$, $\phi_y$, and $\tau_b$, take all distances into account, also the distances associated with missing data.

A multivariate analysis of variance indicates a significant overall difference between the recovery capabilities of the three imputation methods (using Wilks' Lambda: $F(6, 5944) = 227.990$; $p < .001$; $\eta_p^2 = .187$). Table 6.2 provides descriptive statistics and the tests of the between-subject effects, including effect sizes, expressed as partial eta squared ($\eta_p^2$). For the simulation studies, emphasis is on the effect sizes as the number of replications can always be increased to obtain significant results. Here, all differences are significant, but the descriptive statistics and the effect sizes indicate that the differences in recovery are not very serious. According to Cohen (1988), a partial eta squared of .010 indicates a small effect, .059 a medium effect, and .138 a large effect. The deletion method is slightly better than the imputation methods for $\phi_y$ and $\tau_b$, but worse for $\phi_{xy}$.

The actual solutions from the incomplete data are superior for the deletion method. Table 6.3 shows STRESS-1, indicated by $\sigma_1^-$ based on the valid data

Table 6.2  Descriptive statistics (upper part, with means and standard deviations in parentheses) and MANOVA tests of between-subjects effects (lower part, with F-statistics, significance in parenthesis, and effect sizes on the second line) comparing recovery of unfolding solutions using deletion (no imputation), respondent average imputation, and product average imputation methods.

| Method | $\phi_{xy}$ | | $\phi_y$ | | $\tau_b$ | |
|---|---|---|---|---|---|---|
| Deletion | .957 | (.015) | .967 | (.019) | .661 | (.054) |
| Respondent Average Imputation | .964 | (.008) | .957 | (.025) | .645 | (.055) |
| Product Average Imputation | .965 | (.008) | .962 | (.023) | .658 | (.047) |
| Between-Subjects Effects | $\phi_{xy}$ | | $\phi_y$ | | $\tau_b$ | |
| F   (p) | 157.536 | (.000) | 48.601 | (.000) | 27.169 | (.000) |
| $\eta_p^2$ | .096 | | .032 | | .018 | |

only, and the rank order correlations ($\tau_b^-$). The overall difference is significant (using Wilk's Lambda: $F(4, 5946) = 6145.930; p < .001; \eta_p^2 = .805$) and the tests of the between-subject effects show large effects for all measures in favor of the deletion method. Where the deletion method improves considerably on STRESS-1 and rank order correlations as compared to the complete data solution (with $\sigma_1 = .241$ and $\tau_b = .701$, respectively), the imputation methods worsen (see descriptives from Table 6.3). The introduction of additional error by imputation, causes higher STRESS-1 values for the imputation methods. The deletion method uses its freedom to find a better solution, mainly in the transformation part of the loss function, but without deviating from the imputation methods concerning the recovery of the unfolding solution. In conclusion, due to the inconclusive recovery results, the better actual solutions for the deletion method, and the absence of assumptions concerning the missing data, the deletion method is preferred for further analysis.

*Missing data by researcher*

Researchers may only want to provide a subset of products to a respondent for evaluation. These planned missings both reduces the burden on respondents, improving the quality of their evaluations, and saves time and money. With the missing data under the control of the researcher, the missing completely at random (MCAR) assumptions apply, if the missings are properly randomized (Little & Rubin, 1987). To determine which subset of products is presented to a respondent, simple missing data designs can be considered, but since the relations between objects of different sets are in order, rather than just means, more complicated fractional block designs might be necessary. A balanced incomplete block design (BIBD) (Cochran & Cox, 1957) is such a sophisticated fractional block design. A BIBD is usually defined as an arrangement of $v$ distinct objects in $b$ blocks, such that each block contains $k$ distinct objects,

Table 6.3 Descriptive statistics (upper part, with means and standard deviations in parentheses) and MANOVA tests of between-subjects effects (lower part, with F-statistics, significance in parenthesis, and effect sizes on the second line) comparing fit of unfolding solutions using deletion (no imputation), respondent average imputation, and product average imputation methods.

| Method | $\sigma_1^-$ | | $\tau_b^-$ | |
|---|---|---|---|---|
| Deletion | .164 | (.025) | .770 | (.022) |
| Respondent Average Imputation | .298 | (.013) | .545 | (.023) |
| Product Average Imputation | .273 | (.015) | .618 | (.023) |
| Between-Subjects Effects | $\sigma_1^-$ | | $\tau_b^-$ | |
| F   (p) | 15160.740 | (.000) | 25362.813 | (.000) |
| $\eta_p^2$ | .911 | | .945 | |
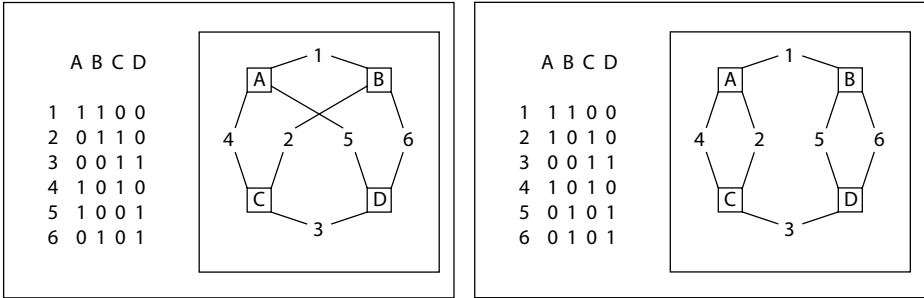
Figure 6.1 Example of a balanced incomplete block design (BIBD) (left-hand panel) and a row-balanced incomplete block design (row-BIBD) (right-hand panel), where valid data is represented by a connection (line) between respondents (numbers) and products (letters).

each object occurs in exactly r different blocks, and every two objects occur together in exactly $\lambda$ blocks (definition by Prestwich, 2001). Convenient for the current topic, with the same set of parameters, a BIBD can be defined in terms of an incidence matrix $\mathbf{I}$, which is then a binary matrix with $v$ rows and $b$ columns where each row sums to $r$ and each column sums to $k$ (see, for an example, Figure 6.1, left-hand panel). Any pair of distinct rows has scalar product $\lambda = \mathbf{I}_i' \mathbf{I}_j$ for all $i \neq j$. Since the parameters are not independent ($vr = bk$ and $\lambda(v-1) = r(k-1)$), a BIBD is commonly expressed in terms of $v$, $k$, and $\lambda$. The term 'incomplete' stems from the fact that $k < v$.

Although the description of a BIBD is relatively simple, the generation of a BIBD is a complex problem. BIBD's are only available for a limited series of particular parameter values (see for example Clatworthy's (1973) catalog) and solvable for small parameter values within an acceptable period of time (see Nguyen, 1994). These features pose a serious problem for a Monte Carlo study, which depends on fast problem solving, handling thousands of computational problems within a limited time period. To overcome this practical problem, a design is explored where each row sums to $r$, but where the requirement of sum $k$ per column and scalar product $\lambda$ between rows is relaxed. Still, $vr = bk$, but now $k$ is a random variable with mean $vr/b$ and some variation. This design is referred to as a row-balanced incomplete block design (ROW-BIBD), indicating that every respondent evaluates the same number ($r$) of products, but products might not be evaluated the same number ($k$) of times (an example is given in Figure 6.1, right-hand panel). Different $k$'s for different columns seems to be a minor problem, since the number of consumers is typically large compared to the number of products.

*Simulation study: BIBD versus ROW-BIBD*

The breakfast data are used to evaluate the difference between a BIBD and a ROW-BIBD concerning the recovery of unfolding solutions. A known BIBD, with $v = 15$, $k = 5$, and $\lambda = 8$ (Nguyen, 1993, 1994), is used that matched the breakfast data. The resulting incidence matrix is transposed and zero's and ones are interchanged to get a BIBD with $v = 42$, $k = 14$, and $\lambda = 18$ (see Table 6.1). The incomplete unfolding using the BIBD is replicated 1000 times, each time randomly interchanging rows and columns of the incidence matrix. For the analyses using the ROW-BIBD, 1000 runs are conducted creating another incidence structure on each instance. Both designs exclude 5 out of 15 products per respondent.

A multivariate analysis of variance indicates a significant overall difference between the recovery capabilities of both designs (using Wilks' Lambda: $F(3, 1996) = 2.840; p = .037$). The effect size, however, is very small ($\eta_p^2 = .004$), which is reflected in the tests of the between-subject effects provided in Table 6.4: None of the three statistics shows a significant result. The descriptive statistics also show that the differences are very small, which leads to the conclusion that both designs perform alike. Subsequently, the more flexible and faster ROW-BIBD is used to specify the missing data by researcher.

*Missing data by respondent*

In memory-based evaluations, only products that are known to the respondents are available for evaluation. If a researcher still offers all products to the respondents, the results for the unknown products will mostly be neutral, random, invalid, or missing. Shocker, Ben-Akiva, Boccara, and Nedungadi (1991) discuss a hierarchical chain of sets modeling decision-making. In their view, consumers use a universal set, which contains an awareness or knowledge set, which in turn contains a consideration set, which contains a choice set,

Table 6.4    Descriptive statistics (upper part, with means and standard deviations in parentheses) and MANOVA tests of between-subjects effects (lower part, with F-statistics, significance in parenthesis, and effect sizes on the second line) comparing recovery of unfolding solutions using balanced incomplete block designs (BIBD) and row-balanced incomplete block designs (row-BIBD).

| Design | $\phi_{xy}$ | | $\phi_y$ | | $\tau_b$ | |
|---|---|---|---|---|---|---|
| BIBD | .957 | (.014) | .967 | (.018) | .659 | (.055) |
| row-BIBD | .957 | (.016) | .968 | (.017) | .662 | (.054) |
| Between-Subjects Effects | $\phi_{xy}$ | | $\phi_y$ | | $\tau_b$ | |
| F    (p) | .155 | (.694) | 2.777 | (.096) | 2.583 | (.108) |
| $\eta_p^2$ | .000 | | .001 | | .001 | |
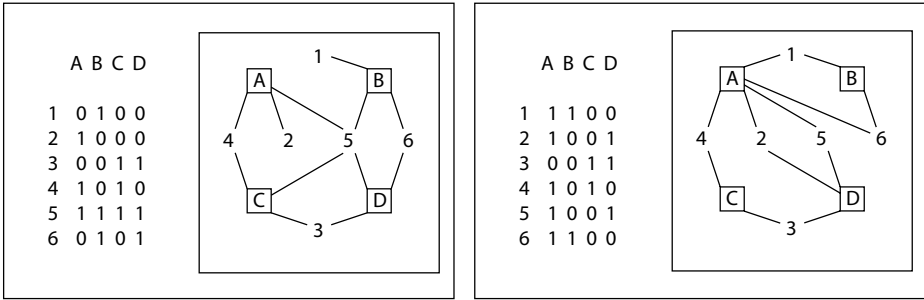
Figure 6.2   Example of a knowledge set design (left-hand panel) and a product familiarity design (right-hand panel), where valid data is represented by a connection (line) between respondents (numbers) and products (letters).

which finally contains the product of choice. Each set is smaller in number of products than or equal to the previous set. For the present research this means that the researcher offers a universal set for evaluation to all respondents, but respondents only evaluate the products they know, i.e., products from their knowledge set. For an even higher quality of their evaluations, the researcher might persuade respondents to use their consideration set or even their choice set. As a result of the use of knowledge sets, different respondents may evaluate a different number of products, simply because some respondents know more products than others. A simulation study is used to determine the consequences of the variation in the number of products per respondent for the recovery of the unfolding solutions. The products in the knowledge sets might be uniformly distributed over the entire product range, but this is not expected in practice. Some products are more familiar than other products, for example due to more advertising, longer existence, or wider availability. As a result, the knowledge sets are expected to be unequally distributed over the products. The impact of the unequal number of respondents per product on the recovery of unfolding solutions is investigated in another simulation study.

The missing data in the present case is related to the respondents (knowledge sets) or to the products (familiarity), hence the missing data is not completely at random (MCAR), but only missing at random (MAR). Since the MCAR assumptions no longer apply, it is imperative that additional analyses are performed to get a thorough insight in the distribution of the missing data over both respondents and products.

*Simulation study: Knowledge sets*

The breakfast data are used to determine the influence of knowledge sets on the recovery of unfolding solutions using incomplete data. To simulate knowledge

sets, the number of evaluations per respondent is varied. The variation is set by drawing the number of non-missings from a normal distribution with mean 10 (as in the previous simulation studies) and standard deviation $a$. The minimum number of non-missings is set to 2. An example of a *knowledge set design* is given in Figure 6.2 (left-hand panel), where different respondents know a different number of products, represented by a line between respondents (numbers) and products (letters). The levels of factor $a$ are set from 0 to 10 with steps of 1, where $a = 0$ specifies no variation, thus 10 evaluations per respondent exactly. This study uses 1000 replications of incomplete data per level of factor $a$.

A multivariate analysis of variance indicates a significant difference (using Wilks's Lambda: $F(30, 32247) = 304.186; p < .001; \eta_p^2 = .216$) between the variation levels $a$. The between-subject effects (Table 6.5, lower part) show significant differences with large effects and the descriptive statistics of all recovery measures give the same result: As the variation level $a$ increases, i.e.,, as the differences in number of evaluations per respondent increase, the recovery of the unfolding solutions worsens. Since the total number of missings is equal for all levels of $a$, the variation in number of missings definitely influences recovery. Especially the respondent points suffer from the variation, as can be concluded from the effect sizes and the differences in decrease of $\phi_{xy}$, $\phi_y$, and $\tau_b$, for increasing $a$.

Table 6.5   Descriptive statistics (upper part, with means and standard deviations in parentheses) and MANOVA tests of between-subjects effects (lower part, with F-statistics, significance in parenthesis, and effect sizes on the second line) comparing recovery of unfolding solutions using missing data designs with different levels of variation in number of products per respondent (a).

| Variation Level a | $\phi_{xy}$ | | $\phi_y$ | | $\tau_b$ | |
|---|---|---|---|---|---|---|
| 0 | .957 | (.015) | .966 | (.019) | .657 | (.055) |
| 1 | .952 | (.015) | .964 | (.019) | .649 | (.056) |
| 2 | .941 | (.019) | .960 | (.020) | .633 | (.054) |
| 3 | .923 | (.026) | .950 | (.022) | .604 | (.056) |
| 4 | .909 | (.028) | .941 | (.023) | .579 | (.054) |
| 5 | .892 | (.037) | .931 | (.025) | .559 | (.056) |
| 6 | .883 | (.040) | .928 | (.026) | .545 | (.058) |
| 7 | .875 | (.042) | .923 | (.027) | .537 | (.058) |
| 8 | .869 | (.046) | .921 | (.029) | .529 | (.060) |
| 9 | .866 | (.047) | .920 | (.029) | .527 | (.061) |
| 10 | .863 | (.049) | .918 | (.027) | .521 | (.060) |
| Between-Subjects Effects | $\phi_{xy}$ | | $\phi_y$ | | $\tau_b$ | |
| F   (p) | 1006.808 | (.000) | 587.935 | (.000) | 806.964 | (.000) |
| $\eta_p^2$ | .478 | | .349 | | .423 | |

*Simulation study: Product familiarity*

The breakfast data are used to determine the influence of product familiarity on the recovery of unfolding solutions using incomplete data. Product familiarity is reflected by increasing the chance that a product is chosen for evaluation, which differs from the approach taken by Chatterjee and DeSarbo (1992), where familiarity is linked with reliability and preferences require additional uncertainty information. For the current simulation study, the chance to choose the first 3 products (20%) for evaluation is b times greater than the chance to choose the remaining products, thus defining high familiar and low familiar products. Corresponding comparison measures $\phi_{xy}^{high}$ and $\phi_{xy}^{low}$ only use the distances between the respondents and the products under consideration, that is, the first 20% or the last 80% of the products. The levels of factor b are set from 1 to 10 with steps of 1, with equal chances for $b = 1$. An example of the missing data design is given in Figure 6.2 (right-hand panel). This study uses 1000 replications of incomplete data for each level of factor b.

A multivariate analysis of variance indicates significant differences (using Wilks's Lambda: $F(27, 29165) = 2.186; p < .001$) between the familiarity levels b, but with an effect size close to zero ($\eta_p^2 = .002$). The between-subject effects indicate that the differences are due to $\phi_y$ and $\tau_b$, but also with an effect sizes close to zero (see lower part of Table 6.6). Comparing the familiarity levels shows that only $b = 1$ is responsible for the differences, and not even with all other levels of b. It is nevertheless save to conclude that the familiarity level has no influence on the recovery of the unfolding solutions.

Table 6.6   Descriptive statistics (upper part, with means and standard deviations in parentheses) and MANOVA tests of between-subjects effects (lower part, with F-statistics, significance in parenthesis, and effect sizes on the second line) comparing recovery of unfolding solution using missing data designs with different familiarity level of the products (b).

| Familiarity Level b | $\phi_{xy}$ | | $\phi_y$ | | $\tau_b$ | |
|---|---|---|---|---|---|---|
| 1 | .957 | (.014) | .967 | (.018) | .659 | (.056) |
| 2 | .957 | (.015) | .969 | (.017) | .662 | (.052) |
| 3 | .957 | (.016) | .968 | (.017) | .665 | (.051) |
| 4 | .958 | (.014) | .970 | (.016) | .668 | (.048) |
| 5 | .957 | (.017) | .969 | (.016) | .665 | (.050) |
| 6 | .957 | (.017) | .968 | (.016) | .666 | (.049) |
| 7 | .957 | (.017) | .969 | (.016) | .668 | (.050) |
| 8 | .957 | (.017) | .969 | (.017) | .667 | (.051) |
| 9 | .957 | (.016) | .968 | (.017) | .667 | (.049) |
| 10 | .958 | (.014) | .969 | (.016) | .668 | (.048) |
| Between-Subjects Effects | $\phi_{xy}$ | | $\phi_y$ | | $\tau_b$ | |
| F   (p) | .846 | (.574) | 2.613 | (.005) | 3.061 | (.001) |
| $\eta_p^2$ | .001 | | .002 | | .003 | |

However, some of the products are more familiar than others and it is the difference between these two sets, high familiar and low familiar products, we are interested in. Table 6.7 (lower part) gives the results of a two-way analysis of variance with familiarity level and high-low familiarity as fixed factors. As for the multivariate analysis of variance, the familiarity level has a significant effect on recovery, but an effect size close to zero. The difference, however, between high familiar products and low familiar products is significant, together with a large effect size ($\eta_p^2 = .245$). It is therefore important to make a distinction between high familiar and low familiar products, whereas the familiarity level is a matter of secondary significance.

*Impossible missing data*

Unfolding is unable to compute a solution from an unconnected block design and it is therefore required that the incidence graph of any block design previously discussed is connected (i.e., that there exists a path joining any two of its vertices). In Figure 6.3, an example is shown of an unconnected design, as one small block, with respondent 4 and product C, is not connected with the large block of respondents (1, 2, 3, 5, and 6) and products (A, B, and D). Determining the positions of both blocks with respect to one another is impossible. Thus, we will ensure in the following that each design is connected.

Table 6.7   Average congruence coefficients for high familiar and low familiar products (upper part, with standard deviations in parentheses), and the univariate two-way analysis of variance (lower part) comparing recovery of high familiar and low familiar products in unfolding solutions using missing data designs with different familiarity levels of the products (b).

| Descriptive Statistics | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Familiarity Level b | $\phi_{xy}^{high}$ | | $\phi_{xy}^{low}$ | | Familiarity Level b | $\phi_{xy}^{high}$ | | $\phi_{xy}^{low}$ |
| 1 | .965 | (.014) | .953 | (.015) | 6 | .971 | (.015) | .952 | (.018) |
| 2 | .968 | (.013) | .953 | (.016) | 7 | .971 | (.015) | .952 | (.019) |
| 3 | .970 | (.014) | .953 | (.017) | 8 | .971 | (.014) | .952 | (.018) |
| 4 | .971 | (.012) | .953 | (.016) | 9 | .971 | (.014) | .952 | (.017) |
| 5 | .970 | (.015) | .952 | (.018) | 10 | .972 | (.012) | .953 | (.015) |

| Univariate Two-Way Analysis of Variance | | | | | | |
|---|---|---|---|---|---|---|
| Source | SS | df | MS | F | p | $\eta_p^2$ |
| Familiarity Level b | .016 | 9 | .002 | 7.468 | .000 | .003 |
| High-Low | 1.539 | 1 | 1.539 | 6485.668 | .000 | .245 |
| Interaction | .027 | 9 | .003 | 12.470 | .000 | .006 |
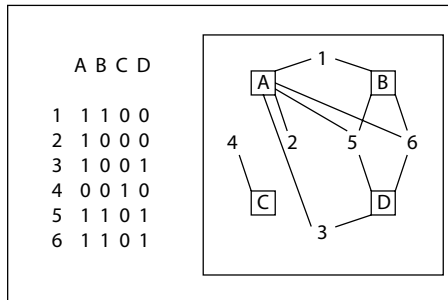
Figure 6.3  Example of an unconnected design, where valid data is represented by a connection (line) between respondents (numbers) and products (letters).

## 6.4  MONTE CARLO SIMULATION STUDY

A comparison is made between unfolding on a complete and an incomplete set of data, for which an incidence matrix is used to specify the incomplete set of data. The current Monte Carlo simulation study attempts to determine key success factors for unfolding with incomplete data and aims at providing guidelines for researchers and data collectors.

Data is generated according to the model of Wagenaar and Padmos (1971), that is, $\delta_{ij} = \|x_i - y_j\| \times \exp^{N(o,e)}$. After generating $i = 1, \ldots, n$ points for the respondents and $j = 1, \ldots, m$ points for the products in a p-dimensional space from a uniform distribution, 5% outliers are created in each set. Using the distances from the centroid of the configuration, points are shifted 1.5–3.0 times the interquartile range of the distances outside the maximum distance from the centroid to become an outlier. This choice is similar to the outlier definition in boxplots (see, for example, SPSS, 2006), when applied to the distances of points to the origin. Next, the distances between the sets are computed and perturbed by multiplying them with a log normal distribution ($\exp^{N(o,e)}$), generating a normally distributed error pattern $e$ on the distances. The levels of error are roughly equivalent to Kruskal's STRESS-1 values corresponding with a perfect to a very poor fit (Kruskal, 1964a), with slightly higher STRESS-1 values for the three-dimensional case. For each respondent, the (error-perturbed) distances are replaced with their rank number. The variation in the rank numbers, expressed in values of Kendall's rank order correlation $\tau_b$, average

Table 6.8  Summary of independent factors and accompanying levels for the simulation study.

| Factor | Description | # | Levels | Factor | Description | # | Levels |
|--------|-------------|---|--------|--------|-------------|---|--------|
| n | # Respondents | 5 | 10, 20, 40, 80, 160 | p | # Dimensions | 2 | 2, 3 |
| m | # Products | 4 | 5, 10, 20, 40 | e | Error Level | 3 | 0.00, 0.10, 0.25 |

0.87 and 0.70 for the error levels 0.10 and 0.25, respectively. The levels for the independent factors in the simulation study are summarized in Table 6.8.

For each generated data set, a complete unfolding as well as $m - p$ incomplete unfolding solutions are computed, as the number of inclusions ($i$, i.e., the number of non-missing products), starts at $p$ (the dimensionality) and ends at $m - 1$ (the total number of products minus one, i.e., with one missing per respondent). The factors from Table 6.8 are studied in a fully crossed factorial design with 1000 replications for each cell. Cases for which the incidence matrix is not connected or with insufficient free parameters are excluded from further analyses.

Based on the results of the simulation studies from the previous section, two types of incidence matrices are used to specify the incomplete data. The first type specifies missing data by researcher with a ROW-BIBD, where each respondent evaluates the same number of products and products are evaluated about the same number of times. The second type of incidence matrices specifies missing data by respondent, where the number of evaluations per respondent varies depending on the number of products ($a = m/4$) and 20% of the products (high familiar products) are evaluated $b = 10$ times more often than others (low familiar products).

*Guidelines for missing data by researcher*

The influence on recovery for the factors from Table 6.8 are determined with a multivariate analysis of covariance (main effects and 2-way interactions only), where the continuous variable inclusion proportion ($\text{prop}(i) = i/m$) is specified as a covariate. All multivariate tests are significant ($p < .001$), but with varying effect sizes. As indicated by the effect sizes of the multivariate effects (Table 6.9, second column), there is better recovery for data with fewer missings, more products ($m$), and more observations ($n \times m$). It is also beneficial to have data with a low level of error ($e$), while increasing the number of respondents ($n$) or changing the dimensionality ($p$) only has a marginal effect on recovery. The tests of the between-subject effects are also significant ($p < .001$) for all factors and for all recovery measures. Table 6.9 shows the effect sizes in the last three columns. These results lead to the same key success factors. Additionally to the multivariate effects, the number of respondents ($n$) does influence the recovery of the product configuration ($\phi_y$) and the rank order recovery per respondent ($\tau_b$), as $\eta_p^2 = .067$ and $\eta_p^2 = .068$, respectively. The number of observations ($n \times m$) has a large effect on the rank order correlation with $\eta_p^2 = .173$.

Figure 6.4 provides guidelines for applied research when the researcher is in control of the missing data. The panels show I-beams and markers for all factors of the Monte Carlo simulation study, except for dimensionality,

which has an insufficient effect on recovery to be included. We first explain the elements of such I-beam plots and then indicate how they should be read. The I-beam and markers, i.e., the high, low, and close in high-low graphs, indicate high, low and medium recovery. For the congruence coefficients $\phi$, these indicators correspond with the values .99, .95, and .98, respectively. Although Tucker (1951) employs .80 and Cureton and D'Agostino (1983) and Mulaik (1972) advocate .90 to identify congruent factors or component loadings, the relation between $\phi$ and $\sigma_1$ as discussed in Technical Appendix G combined with the rules-of-thumb by Kruskal (1964a) (although not specified for unfolding) called for much stricter values for $\phi$. For the rank order correlation $\tau_b$, values of .90, .70, and .80 are considered sufficiently high in actual practice, also considering the variation in rank order correlations for the different error levels. The actual values for the three recovery measures are reached with 95% accuracy, providing a common 5% type-I error.

Figure 6.4 can be read as follows. Suppose we have about 10 products, 20 respondents, and we expect almost errorfree data. Suppose we are interested in the rank order correlations for which we are satisfied with only $\tau_b = .70$ recovery. In this case, we use the upper left-hand panel for 10 products and 20 respondents and the left-hand side cluster for error level 0.0. The rank order correlation is on the right-hand side of the cluster, indicated with a square marker. The lower part of the I-beam provides the minimal $\tau_b = .70$ rank order correlation, which in this case allows for an inclusion proportion of .70. Thus, with a 95% chance that the rank orders corresponds at least

Table 6.9    Effect sizes for the main effects (Wilks' Lambda) and effect sizes for the tests of the between-subject effects of the multivariate covariance analysis comparing the recovery of unfolding solutions for different number of respondents (n), number of products (m), number of dimensions (p), and error levels (e), with inclusion proportion (prop(i)) as covariate.

| Source | Wilks' $\lambda$ | $\phi_{xy}$ | $\phi_y$ | $\tau_b$ |
|---|---|---|---|---|
| prop(i) | .551*** | .214*** | .172*** | .538*** |
| n | .041* | .004 | .067** | .068** |
| m | .071** | .105** | .024* | .134** |
| p | .039* | .018* | .008 | .002 |
| e | .061** | .066** | .027* | .096** |
| n $\times$ m | .072** | .083** | .106** | .173*** |
| n $\times$ p | .016* | .005 | .016* | .005 |
| m $\times$ p | .005 | .005 | .002 | .008 |
| n $\times$ e | .016* | .010* | .037* | .016* |
| m $\times$ e | .018* | .032* | .004 | .007 |
| p $\times$ e | .007 | .000 | .001 | .006 |
| R² | | .407 | .455 | .672 |

*, **, and *** indicate small, medium, and large effect sizes.
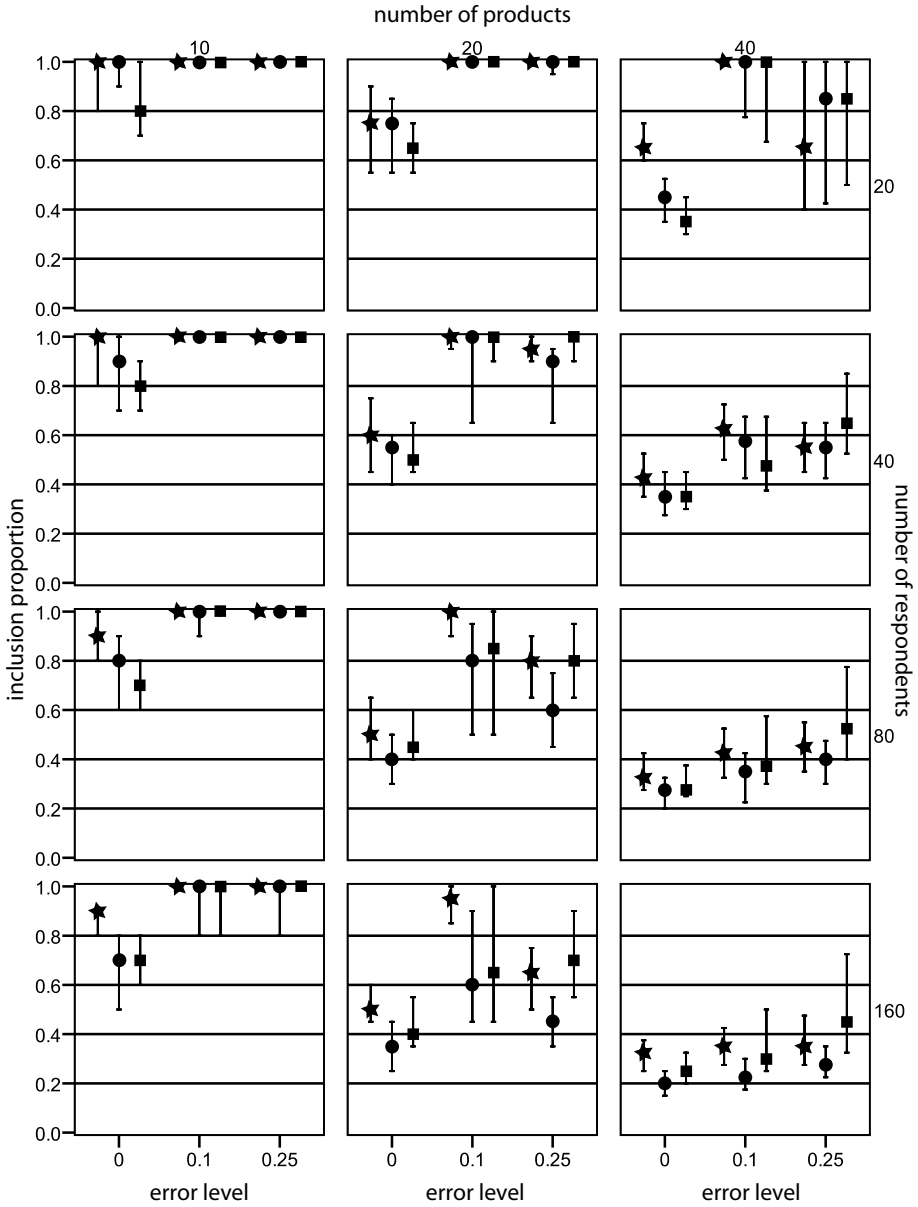
Figure 6.4    High-low graphs for inclusion proportions when data are missing by researcher, where the I-beams (low-close-high) indicate 95% chances on minimal values for $\phi_{xy}$ (.95–.98–.99), $\phi_y$ (.95–.98–.99), and $\tau_b$ (.70–.80–.90), indicated by stars, dots and squares, respectively.

$\tau_b = .70$ with the complete unfolding solution 3 products can be set missing per respondent.

The three different foundations ($\phi_{xy}$, $\phi_y$, and $\tau_b$) for the inclusion proportions in Figure 6.4 and the multivariate covariance analysis provide similar results: More products ($m$), more observations ($n \times m$), and less error ($e$) allow for lower inclusion proportions. Figure 6.4 lacks small sample sizes with $n = 10$ and $m = 5$, because for these cases the inclusion proportion is always equal to 1.0. The recovery of the product configuration, quantified with $\phi_y$, and situated in the middle of the clusters of three with the dot marker, allows for the lowest inclusion proportions. This is plausible considering the number of parameters to be estimated and the amount of data available. Notable is the fact that the high error levels often allows for a lower inclusion proportion as compared with the medium error levels, as can be seen in Figure 6.4 for $n = 20$ and $m = 40$ and for $n = 80$ and $m = 20$.

### Guidelines for missing data by respondent

The influence on recovery when the data are missing by respondent are determined with a multivariate covariance analysis. Recovery of the entire configuration ($\phi_{xy}$) is split up into the recovery of a high familiar set of products ($\phi_{xy}^{high}$) and the recovery of a low familiar set of products ($\phi_{xy}^{low}$).

All tests (multivariate and between-subjects) are significant ($p < .001$) and Table 6.10 shows the effect sizes only. The conclusions are similar to

Table 6.10    Main effects (Wilks' Lambda) and effect sizes for the tests of the between-subject effects of two multivariate covariance analyses, one for missing data by researcher and one for missing data by respondent, comparing the recovery of unfolding solutions for number of respondents (n), number of products (m), number of dimensions (p), and error level (e), with inclusion proportion (prop(i)) as covariate.

| Source | Wilks' $\lambda$ | $\phi_{xy}^{high}$ | $\phi_{xy}^{low}$ | $\phi_y$ | $\tau_b$ |
|---|---|---|---|---|---|
| prop(i) | .679*** | .424*** | .434*** | .261*** | .662*** |
| n | .036* | .032* | .028* | .027* | .048* |
| m | .069** | .086** | .088** | .004 | .099** |
| p | .037* | .019* | .021* | .004 | .001 |
| e | .057* | .003 | .003 | .009 | .095** |
| $n \times m$ | .063** | .028* | .036* | .071** | .204*** |
| $n \times p$ | .010* | .013* | .014* | .003 | .018* |
| $m \times p$ | .006 | .001 | .001 | .003 | .005 |
| $n \times e$ | .020* | .050* | .058* | .031* | .018* |
| $m \times e$ | .015* | .034* | .037* | .001 | .004 |
| $p \times e$ | .007 | .004 | .004 | .000 | .009 |
| R² | | .539 | .549 | .402 | .737 |

*, **, and *** indicate small, medium, and large effect sizes.
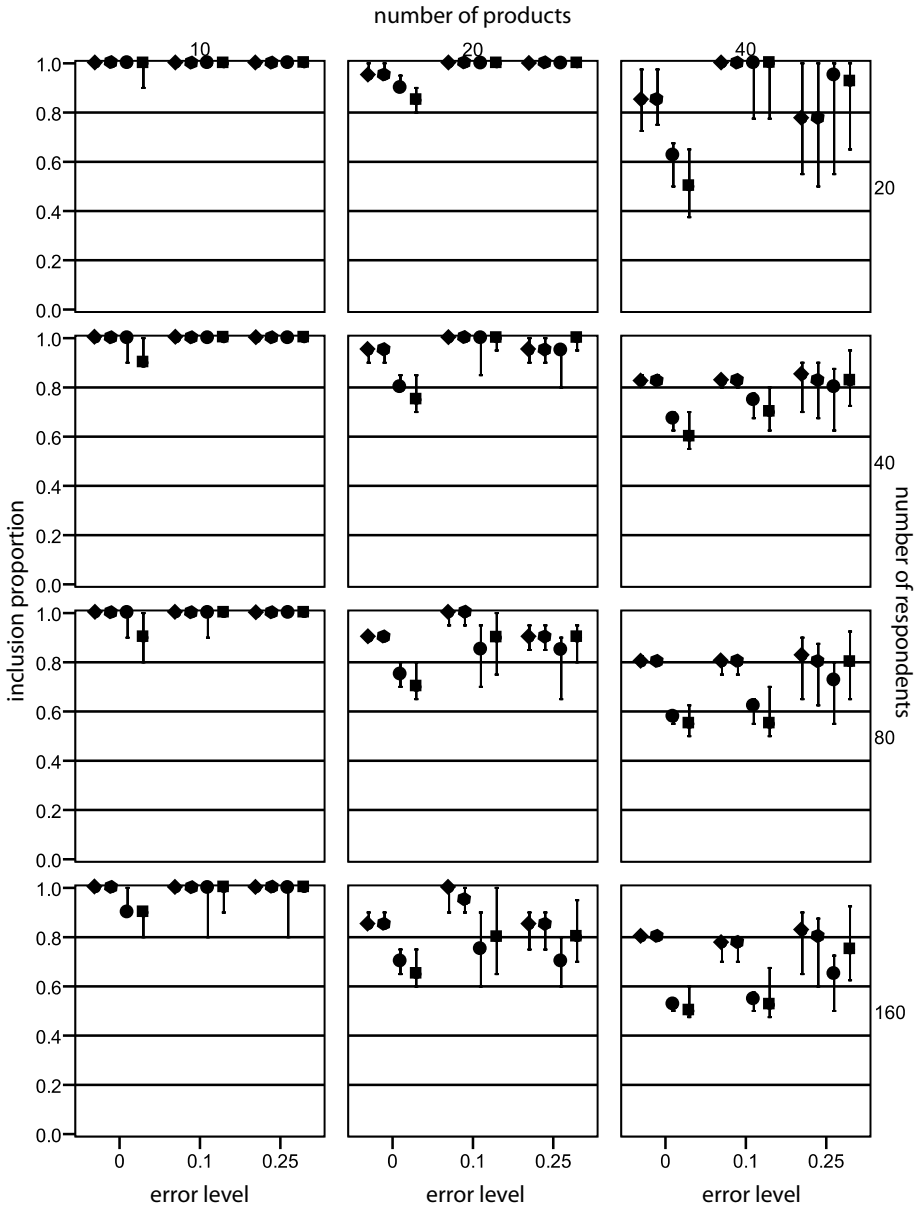
Figure 6.5    High-low graphs for inclusion proportions when data are missing by respondent, where the I-beams (low-close-high) indicate 95% chances on minimal values for $\phi_{xy}^{high}$ (.95–.98–.99), $\phi_{xy}^{low}$ (.95–.98–.99), $\phi_y$ (.95–.98–.99), and $\tau_b$ (.70–.80–.90), indicated by diamonds, polygons, dots and squares, respectively.

the conclusions from the missing data by researcher design, although less pronounced: Unfolding solutions are better recovered for data with fewer missings (prop($i$)), more products ($m$), less error($e$), and more observations ($n \times m$). There are small effects for the number of respondents ($n$), dimensionality ($p$), and some of the interactions. Considering the between-subject effects (last four columns of Table 6.10), the rank order correlation benefits exceptionally well from additional observations, but recovery of the correlation is also sensitive to error. Finally, it should be noted that although high familiar products are better recovered than low familiar products (significantly with very small effect sizes (not shown here)), the independent factors have similar effects on both sets of products.

Guidelines for applied research when the data is missing due to respondents are given in Figure 6.5. In general, the inclusion proportions are seriously higher than for the missing data by researcher design (Figure 6.4). Only for a large number of observations, and then even with a large number of products, the inclusion proportions approach 50%. Compare, for example, $n = 160$ and $m = 10$ with $n = 40$ and $m = 40$: Both samples have the same number of observations, but the latter, with more products, allows for more missing data.

## 6.5 EXAMPLE

The results of the Monte Carlo simulation study are used to determine the inclusion proportion for the breakfast data. The breakfast data consists of 42 respondents and 15 products (breakfast items) and the inclusion proportion is determined by taking the average between inclusion proportions of $m = 10$ and $m = 20$ for $n = 40$ and $e = .25$. In this case, the error level is known from the complete set of data, which is something to be guessed at in other circumstances. The number of missing preferences per respondent can be chosen, depending on the quality of recovery (low, medium, or high), on the primary interest of the researcher (the product configuration, the respondents rank orders, or the entire configuration), and on the missing data design (by researcher or by respondent). For the current illustration, we are interested in the product configuration and thus focuss on $\phi_y$. The inclusion proportions for low, medium, and high recoverability are .825, .95, and .975, for the missing data by researcher design and .90, .975, and .975 for the missing data by respondent design. With 15 products, this leads to 0–3 missing preferences per respondent. Since the complete set of data is available, multiple incomplete data analyses are possible and 1000 replications are used to create the configurations and boxplots.

Figure 6.6 shows the unit standard deviation confidence ellipses (Meulman & Heiser, 1983) or confidence regions for the incomplete data solutions after
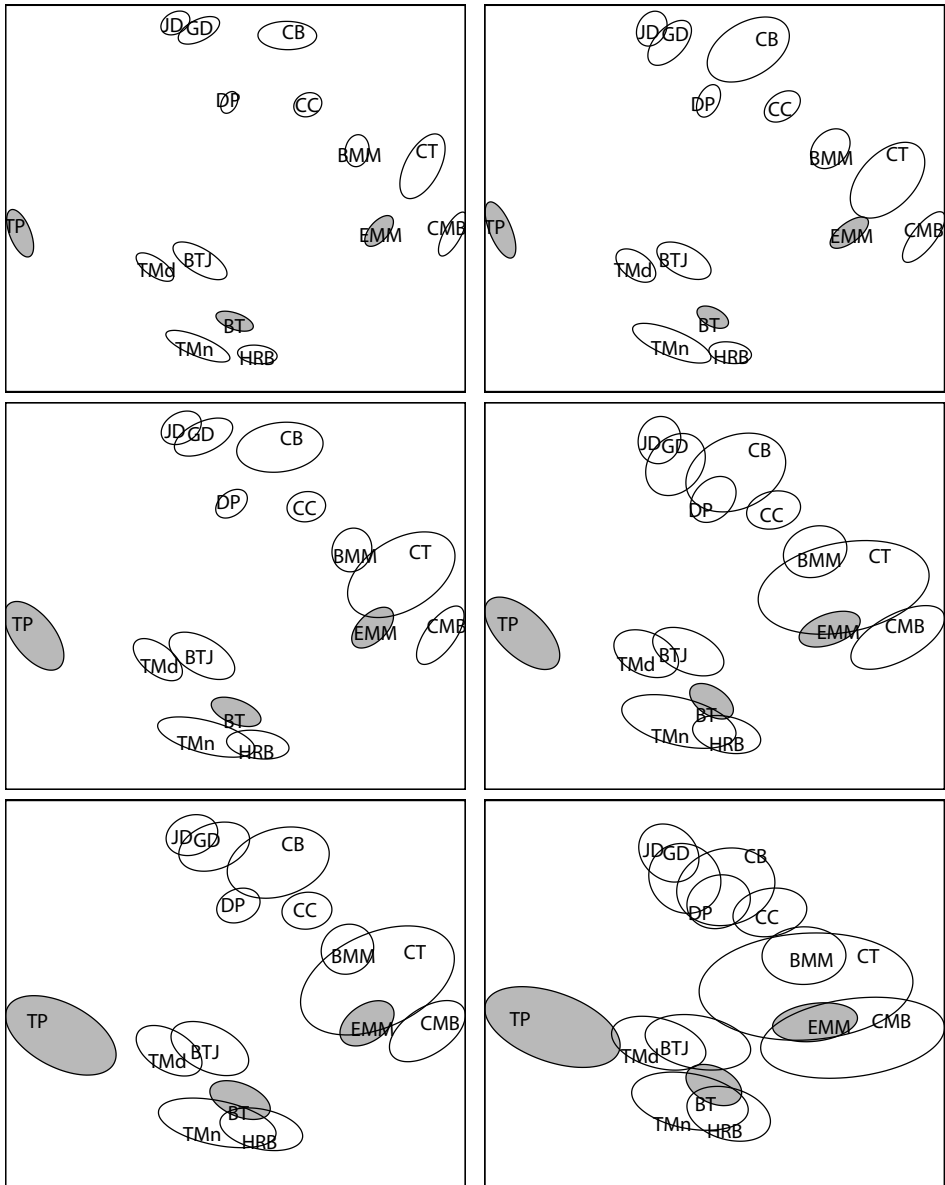
Figure 6.6   Configurations with unit standard deviation confidence ellipses for the incomplete breakfast data with 1, 2, and 3 missings (top-down) using two different designs for specifying missing data: Missing data by researcher design (left-hand panels) and missing data by respondent design (right-hand panels). The breakfast items (and plotting codes) are given in Table 2.1.

1000 replications. The incomplete data solutions are optimally rotated, translated, and dilated by orthogonal Procrustes analysis (Cliff, 1966) to match the complete unfolding solution. It is obvious, even by sight, that the solutions with fewer missing preferences per respondent and the solutions from the missing data by researcher design contain smaller regions. These solutions are more alike and provide better recovery of the complete data solution. Nevertheless, the three high familiar products in the missing data by respondent design, toast pop-up (TP), buttered toast (BT), and English muffin and margarine (EMM), indicated in the configurations by filled confidence regions, deviate from this general observation by maintaining their small regions, such that these products are comparable with the missing data by researcher design. Compare, for example, the confidence regions of CT and EMM, where the region of the latter remains small, while the region of the former increases considerably with each additional missing preference per respondent. In all cases, the true product points (indicated by the plotting codes) lie within the boundaries of their confidence region. This indicates that the incomplete data configurations are indeed very similar to the complete data configuration, although the variation of the coordinates from the incomplete data solutions increases for additional missing data.

The boxplots in Figure 6.7 display the distributions of the recovery measures. For the missing data by researcher design, nearly all congruence coefficients are greater than .98 (panel $a$ and $b$), and even greater than .99 considering only the product configuration (panel $b$). It seems that the guidelines from Table 6.4 are somewhat conservative, since $\phi_y \geqslant .99$ was expected for data without missings and $\phi_y \geqslant .98$ for data with only one missing per respondent. For the missing data by respondent design, the recovery is acceptable for one or two missing preferences per respondent, but recovery quickly worsens for additional missing data. High familiar products are better recovered than low familiar products (panel $d$), but extra missing data results in inferior configurations for the high familiar products too. However, returning to where we started from, the product configuration is recovered quite well, also for two and even three missing preferences, which is better than predicted from the Monte Carlo simulation study results.

## 6.6   CONCLUSION

An extensive study was performed that investigated the effects of incomplete data on the results of a multidimensional unfolding analysis. We focused on two research designs that are often utilized in consumer and marketing research. In the first, the missing data pattern is imposed by the researcher, while in the second design the respondent 'controls' the missing data pattern. The goal of the study was to propose guidelines to researchers about the
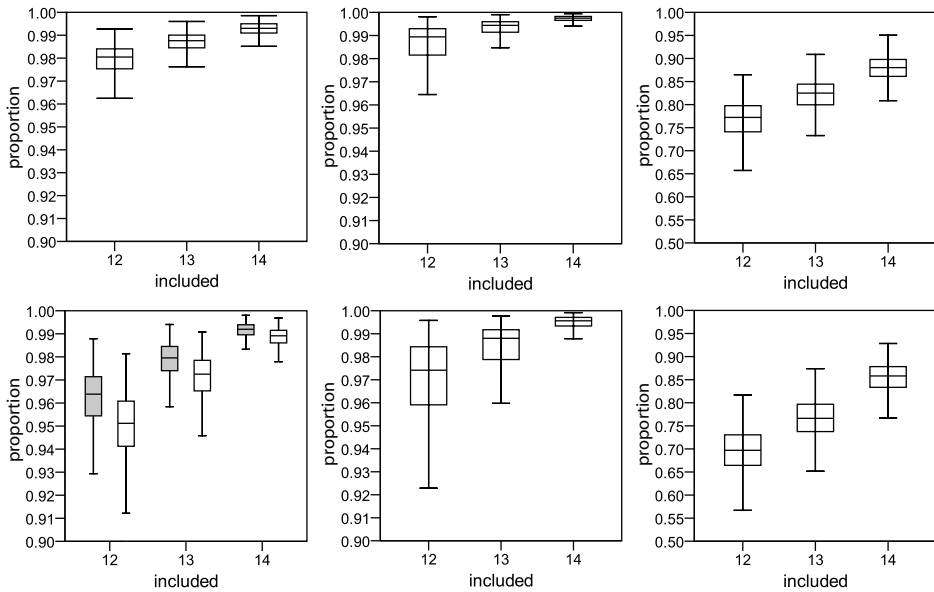
Figure 6.7  Boxplots of the recovery measures ($\phi_{xy}$, $\phi_y$, and $\tau_b$, for the left-hand, middle, and right-hand panels, respectively; lower left-hand panel includes both $\phi_{xy}^{high}$ (filled) and $\phi_{xy}^{low}$) for the incomplete breakfast data (boxes within a panel represent results for 1, 2, and 3 missings) using two different designs for specifying missing data: Missing data by researcher design (upper panels) and missing data by respondent design (lower panels).

amount of missing data that unfolding can handle without corrupting the results of the analysis. Therefore, we compared all incomplete data solutions with solutions obtained on complete data using two resemblance measures: Tucker's congruence coefficient ($\phi$) and Kendall's rank order correlation ($\tau_b$).

Unfolding analysis has the possibility to include a weight matrix. When this weight matrix is coded as a zero-one matrix, it can be used to handle missing data. This option is equal to the pairwise deletion scheme, as for a zero weight both the (missing) data and the corresponding distance are ignored in computations. Often, researchers choose to impute data for the missings. We compared the pairwise deletion scheme with two simple imputation methods, and it can be concluded that pairwise deletion works better (Tables 6.2 and 6.3). Of course, more elaborate imputation schemes could be thought off, but this is left for future research.

The first design, where the researcher controls the missing data, conforms to a situation where the data are missing completely at random (MCAR). In this case, often a balanced incomplete block design is utilized in order to

make the study as efficient as possible. For our simulation study, however, such a BIBD poses insurmountable problems and we proposed a row-balanced incomplete block design (ROW-BIBD). We investigated efficiency loss using a ROW-BIBD compared to a BIBD and it can be concluded that this loss is negligible (Table 6.4). All our further studies used the ROW-BIBD. In practical settings, where only a single analysis has to be performed, we advice to use a real BIBD (if available) to have a somewhat higher efficiency though.

The second design, where the respondent controls the missing data, conforms to a situation where the data are missing at random (MAR). Two factors determine the missing data pattern: The knowledge set of the respondent and product familiarity. The knowledge set corresponds to the set of products the respondent knows and so is able to judge. This knowledge set may differ over subjects in size and content. Product familiarity corresponds to the fact that some products are very well known (and thus in the knowledge set of every respondent) and others are less well known. We distinguished between high and low familiarity. We found that variance in knowledge set has a large influence on the recoverability (Table 6.5), while product familiarity on average has only a minor influence (Table 6.6). However, high familiar products are recovered better than low familiar products (Table 6.7).

Knowing this, we investigated recovery of the complete data unfolding solution using the two designs, where in the second design knowledge set variance and familiarity were used as additional factors. We varied the proportion of included data, the number of respondents, the number of products, the number of dimensions, and the error level. Key success factors in the recovery of the unfolding solution using incomplete data are (in order of importance): The proportion of non-missing data, the number of observations, the number of products, and the error in the data.

Figure 6.4 and 6.5 can be used as guidelines for researchers to choose on the amount of information to be collected. The first figure is for the case where the researcher can determine the missing data pattern. This case is the least sensitive to missing data and it was concluded that up to 80% of the data could be missing without real deterioration. In the second design, where respondents control the missing data pattern, unfolding is more sensitive. Researchers should be careful to include at least 50% of the data to have a good recovery. In both situation, when there are less respondents, or less products, the more the percentage of valid data should increase. In all cases, the researcher is advised to be careful in the research design to keep the error at the lowest possible level. The guidelines presented are in a sense conservative: Even for lower inclusion proportions than presented, and thus with lower recovery measures, the incomplete data solutions remain similar such that substantive conclusions would not change.

The guidelines were illustrated with an example on empirical data. In Figures 6.6 and 6.7, the effects can be seen of various amounts of missing data and the influence of both designs. It was concluded that the solutions are all very similar. Substantive conclusions of these different solutions will be equal.

The two set-ups studied correspond to situations where the data are missing completely at random or missing at random. A third case exists when data are not missing at random (NMAR). Such a situation would occur, for example, if the respondent is asked to indicate the top r ranked products (see, for example, DeSarbo et al., 1997). The possibilities to construct data in this manner are enormous and a thorough investigation would require (computer) time, (journal) space, and a serious (human) effort. It is beyond the scope of this chapter and left for future research.

The problem of local minima was not addressed in this chapter, although predecessors of the Monte Carlo simulation study used both random and close starts. Using random starts introduced unwanted variation in the somewhat more conservative results, while the close starts, using the results of the complete data solution, provided unrealistically good results. Results from these studies (not reported here) indicate that it is always better to use a good start, using available information about locations of products whenever possible. Random starts can not match the guideline results presented here, although PREFSCAL performed fairly well under both circumstances.

The results obtained are very promising. Although really small samples have low recoverability when data are missing for each respondent, for studies involving more products, less than half of the data has to be included without danger of changing the conclusions. This is of major importance for all consumer research: A lot of time and money can be saved.

APPENDIX 6.A    SIMULATION STUDY

In many cases, and certainly due to the huge problem of degenerate solutions in unfolding, nonmetric data are often analyzed metrically. In a simulation study using the breakfast data (P. E. Green & Rao, 1972), metric unfolding, only estimating a dilation parameter for each respondent, is compared with nonmetric unfolding, where the preferences of each respondent are transformed monotonically. The study was replicated 1000 times with a shuffled BIBD (see Table 6.1) on each instance. The metric and nonmetric unfolding solutions are both compared on their own merits and concerning their recovery capabilities.

The original metric unfolding solution for the complete data set is not a particular good solution with $\sigma_1^+ = .299$ and $\tau_b^+ = .608$, while the variation of the distances is quite good. The incomplete data solutions, with 33% less data to fit, only improves a little over the complete data solution with $\sigma_1^- = .272$

and $\tau_b^- = .647$ (see Table 6.11). The nonmetric unfolding solution for the complete data is definitely better than its metric counterpart with $\sigma_1^+ = .241$ and $\tau_b^+ = .701$, while the variation of both transformed preferences and distances are comparable. The nonmetric unfolding solution is certainly not degenerate. However, in contrast with the incomplete data solutions of the metric unfolding, the nonmetric unfolding solutions improve considerably with one third of the data missing, especially on STRESS, with $\sigma_1^- = .162$ and $\tau_b^- = .772$. A multivariate analysis of variance indicates a significant difference (using Wilks' Lambda: $F = 11868.735; p = .000; \eta_p^2 = .922$) between the metric and the nonmetric unfolding solutions with incomplete data. The between-subject effects for both $\sigma_1^-$ and $\tau_b^-$ are significant with large effect sizes (see Table 6.11).

For the discussion of the recovery of both unfolding methods, we have to keep in mind that different solutions need to be recovered. A multivariate analysis of variance indicates a significant difference (using Wilks' Lambda: $F = 630.381; p = .000; \eta_p^2 = .487$) between the recovery of the metric and the nonmetric unfolding solution. Although the differences in recovery are only minor for the whole configuration ($\phi_{xy}$) and the configuration of the products ($\phi_y$), there is a large effect for the differences in rank order recovery ($\tau_b$): The

Table 6.11 Descriptive statistics (upper part, with means and standard deviations in parentheses) and MANOVA tests of the between-subjects effects (lower part, with F-statistics, significance in parenthesis, and effect sizes on the second line) comparing the fit of metric and nonmetric unfolding solutions with incomplete data.

| Analysis | $\sigma_1^-$ | | $\tau_b^-$ | |
|---|---|---|---|---|
| Metric Unfolding | .272 | (.009) | .647 | (.019) |
| Nonmetric Unfolding | .162 | (.024) | .772 | (.020) |
| Between-Subjects Effects | $\sigma_1^-$ | | $\tau_b^-$ | |
| F   (p) | 18206.414 | (.000) | 19862.169 | (.000) |
| $\eta_p^2$ | .901 | | .909 | |

Table 6.12 Descriptive statistics (upper part, with means and standard deviations in parentheses) and the tests of the between-subjects effects (lower part, with F-statistics, significance in parenthesis, and effect sizes on the second line) comparing the recovery of metric and nonmetric unfolding solutions.

| Analysis | $\phi_{xy}$ | | $\phi_y$ | | $\tau_b$ | |
|---|---|---|---|---|---|---|
| Metric Unfolding | .953 | (.021) | .971 | (.019) | .714 | (.054) |
| Nonmetric Unfolding | .957 | (.015) | .968 | (.018) | .661 | (.055) |
| Between-Subjects Effects | $\phi_{xy}$ | | $\phi_y$ | | $\tau_b$ | |
| F   (p) | 27.180 | (.000) | 13.270 | (.000) | 478.766 | (.000) |
| $\eta_p^2$ | .013 | | .007 | | .193 | |

metric unfolding recovers the rank orders better than the nonmetric unfolding does (see Table 6.12, last column). This is noticeable, but appearances may be deceiving. The complete data metric unfolding has a worse fit for the rank orders ($\tau_b^+ = .608$) as mentioned before, but it recovers these less fitting rank orders better ($\tau_b = .714$) than the nonmetric unfolding does ($\tau_b = .661$) for its better fitting rank orders ($\tau_b^+ = .701$).

In conclusion, it is safe to state that the metric unfolding on incomplete data recovers the inferior complete data solution about equally well as the nonmetric unfolding does for the superior complete data solution.

CONCLUSION

This monograph has discussed some advances in multidimensional unfolding. The history of unfolding degeneracies, discussed in Chapter 2, made clear that little headway was made, especially since the technical conception of unfolding as a special form of multidimensional scaling. Chapter 3 discussed a small improvement as the intercept penalty allows metric unfolding to run without getting bogged down in degeneracy troubles. It is a simple procedure, which is applicable in almost any general computational software package. A more versatile approach to overcome the degeneracy problem was discussed in Chapter 4. A penalty on the variation of the transformed preferences, adjustable with two penalty parameters, provides an unfolding loss function that is available for all model options. With the degeneracy problem under control, it is now possible for multidimensional unfolding to attain its full development as a valuable data analysis technique. Examples of such developments were presented in subsequent chapters: Chapter 5 elaborated on a previously published model extension, restricting the coordinates to be linear combinations of independent variables, and Chapter 6 discussed the handling and possible extent of missing data in multidimensional unfolding.

The path from an idea to an ultimate publication, not to mention to the implementation of an idea in software or even to the application in other research areas, depends on many factors and takes a lot of time. During the research for this monograph, other, additional ideas came up. Although these ideas deserve a place in this monograph, they could not be inserted in the completed chapters, because these were published in journal articles. Therefore, we conclude with a short retrospect and somewhat longer prospect, combined per chapter, omitting the history chapter for obvious reasons.

## 7.1 THE INTERCEPT PENALTY

Chapter 3 established that degeneracy also occurs for unfolding with metric transformations of the preferences. By penalizing for an undesirable (large) intercept, the linear transformation is prevented from attaining a horizontal position, which consequently leads to variation in the transformed preferences. The loss function, finding a correspondence between the transformed preferences and the distances, then produces variation in the distances too.

A major drawback of the proposed procedure, besides the restricted set of transformations, is the addition of a penalty parameter. This parameter must

be provided in advance. Since the optimal value is unknown beforehand, the value is commonly determined by trial and error, thus leading to a series of analyses guided by subjective choices. To obviate this drawback, we propose to use resampling to help asses the optimal value or use a procedure that eliminates the penalty parameter from the loss function.

*Finding the optimal penalty parameter*

For the determination of the optimal value for the penalty parameter $\kappa$, which tunes the intercept penalty, a series of unfolding analyses can be performed, for example for $\kappa = 0, \ldots, K$, after which different fit, variation, and degeneracy measures can be compared. The simultaneous comparison of multiple measures undoubtedly leaves room for discussion. Using an automated procedure to determine the optimal parameter value, which circumvents this uncertainty, runs into the trouble of finding a single measure for degeneracy, probably a combination of (a subset of) existing measures. An attempt for such a measure will be discussed later (see page 129). Other procedures might prove a way out as well. Resampling techniques, for example, allow for the quantification of the stability of a solution by just repeating the analysis with slightly deviating data, and as such allow for the definition of a single measure to assess the quality of an unfolding solution.

From all the resampling techniques that are at our disposal, *cross-validation* seems to be an appropriate candidate. See Larson (1931) and Horst (1941) for early applications, Lachenbruch (1965, 1968), who developed the cross-validation criterion, Mosteller and Tukey (1968) for coining the term cross-validation, and Shao and Tu (1995) for a more recent reference. Cross-validation is applied as follows: Using a part of the data as training set and the remaining part of the data as test set, the mean squared error of prediction

$$\text{MSEP} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left( \delta_{ij} - \widehat{\delta}_{ij} \right)^2,$$

where $\widehat{\delta}_{ij} = f^{-1}(d_{ij}) = (d_{ij} - b_1)/b_2$ (with $b_1$ and $b_2$ as intercept and slope, respectively) is the predicted value of $\delta_{ij}$ (from the test set), can be used to assess the predictive validity of the model (Allen, 1974). The training set is specified by randomly selecting a specific number of cells from the data set. The training set is not required to be half of the original data set, which is known as two-fold cross-validation, or all data except one observation (hence the confusing relation with the leave-one-out jackknife, see Stone, 1974) but can be any integer division, such as, for example, 10 for a ten-fold cross-validation, defining a training set with 90% of the data and a test set consisting of the remaining 10%. The results from Chapter 6 might help to decide which

division is most appropriate for the data set at hand. The cross-validation procedure is repeated for all folds, making sure that all data belonged to the test set once. All folds are repeated $R \geqslant 1$ times.

It should be noted, however, that the degeneracy problem might interfere (again), since a degenerate solution might prove very stable indeed. The prediction error MSEP of a degenerate solution might be very small, since the almost constant distances $\mathbf{d}$ of a degenerate solution transform back $f^{-1}(d_{ij}) = (d_{ij} - b_1)/b_2$ into quite different estimated preferences $\widehat{\boldsymbol{\delta}}$ due to accompanying extreme regression parameters $b_1$ and $b_2$.

Instead of cross-validation where the data cells are used as training and test sets, whole rows can serve this purpose. Using $n - K$ rows of the data set as training set and the remaining $K$ rows as test set, opens up the possibility to use external unfolding to find the coordinates of the test set rows, using the column coordinates of the training set as fixed coordinates. This approach circumvents the problems with the extreme regression parameters as described above, since the external unfolding does not need the regression parameters from the training set. The results of small test runs with this approach are promising. Future research should allow us to determine the best approach to use cross-validation for unfolding.

*Eliminating the penalty parameter*

There are a few observations to make on how we can keep the metric transformations of the data under control and how this might lead to an improved (penalty parameter free) procedure.

The working principle for the intercept penalty relies on an explicit normalization of the loss function and, obviously, on a penalty for the intercept. For the normalization, the sum-of-squares of the transformed preferences are explicitly set equal to the number of preferences (cf. Equation 3.3). The thus defined loss function is penalized for an intercept deviating from zero. This concludes the first observation.

The second and third observation are concerned with smooth monotone regression (Heiser, 1989). Chapter 2 described that the step size for bounded monotone regression is restricted by a lower and an upper bound, both accompanied with a parameter to specify the relative size of the bounds (Heiser, 1981). Heiser immediately acknowledged the fact that the solution for the degeneracy problem "lacks the elegance of uniqueness", which he solved in 1989 by presenting the procedure of smooth monotone regression. This latter procedure omits both parameters and uses data-specific bounds based on the mean step of successive preferences.

However, due to the vast amount of inequality constraints, the procedure is very slow, even with the increase in computer speed over the last few decades.

Depending on the conditionality and the handling and number of ties, smooth monotone regression is considerable slower than ordinary monotone regression. For example, the breakfast data is unfolded (10 iterations only on a PENTIUM(R)D CPU 2.80 GHZ) with an unconditional ordinal transformation, untying ties, in less than a tenth of a second, while the smooth monotone transformation takes about 30 minutes, which is more than 18000 times slower.

The fact that a monotone spline transformation forms an intermediate transformation between a linear and a monotone transformation constitutes the final observation. On one extreme, a monotone spline transformation with linear polynomials and (only) two boundary knots is equal to a linear transformation including an intercept and a slope. On the other extreme, a monotone spline with a knot on each unique preference value amounts to a monotone (stepwise) transformation.

A combination of the above observations resulted in the following research in progress (Busing, Heiser, & Eilers, in preparation): Avoiding degeneracies in unfolding using *smooth monotone spline* (SMS) transformations, where a SMS transformation is defined as a monotone spline transformation (Ramsay, 1988) with smoothness restrictions on the knots. Typical features of the SMS transformation are that (1) the left most boundary (exterior) knot is linked to zero, corresponding to the first observation, and (2) the next consecutive steps (from knot to knot) are bounded by a mean step (cf. the second observation). Since the SMS transformation can be specified with fewer interior knots, the number of inequality constraints are decreased and consequently an increase in speed is realized (third observation). The SMS transformation function is more flexible (cf. fourth observation) and faster, and does not include a penalty parameter, which makes the procedure much simpler, a clear advantage, both theoretically and practically.

## 7.2 THE COEFFICIENT OF VARIATION PENALTY

In Chapter 4, the conditions for degeneracy were identified, insofar as degeneracy is defined as a solution with zero STRESS and constant distances. It was argued that the set of admissible transformations contains the cause for degeneracies. Using the coefficient of variation in a penalty function, a general badness-of-fit function was obtained that successfully avoids a degenerate solution in a wide range of circumstances. For this purpose, the penalty function was equipped with two penalty parameters to fine tune the penalty. The simulation study made clear that one of these parameters could be restricted to a constant value, whereas the other parameter was best chosen in a specific interval.

Despite the provision of default values, each analysis might require a different set of penalty parameters and it is left for the user to determine the

exact values. Specifying the penalty parameters too weak causes the solution to be(come) degenerate. Although goodness-of-fit will improve in this case, variation and degeneracy measures will indicate (signs of) degeneracy. On the other hand, when the penalty parameters are set too strong, this will not provide a degenerate solution but a solution with linearly transformed preferences, and with worse fit statistics.

*Finding the optimal penalty parameters*

It is not a trivial task to find a combination of optimal penalty parameters. This is mainly due to the difficult simultaneous comparison of multiple fit, variation, and degeneracy measures that are currently at our disposal. There are essentially two ways to proceed: Use existing measures and consolidate a selection into one single measure or find a new measure. Both ways would enable us to determine the 'best' unfolding solution with corresponding optimal penalty parameters

To start with the former way to proceed, there are several measures that seem suitable for the definition of a proper unfolding solution. Fit measures, however, are rather ambiguous: Both perfect (non-degenerate) and degenerate solutions have (near) perfect fit measures, thus making it impossible to distinguish between these two situations using a fit measure. A similar problem arises for degeneracy measures. For example, the intermixedness index (I-INDEX, see Chapter 4 and Technical Appendix G) measures intermixedness of the two sets of objects in the configuration. It might occur, however, that an otherwise normal solution with low STRESS and sufficient variation exhibits separated sets of objects, and thus an undeserved high intermixedness index. The inadequacy of these measures to distinguish automatically between good and bad unfolding solutions disqualify these measures as components of a single quality measure. Further, the most appropriate measure, the PENALIZED STRESS function value, is not an option, since its magnitude depends on the penalty parameters. Nevertheless, it is still possible to define a quality measure for unfolding solutions based on existing measures. Although this measure might fail at providing us with the optimal solution, it enables us at least to avoid solutions with unattractive characteristics.

Within this framework, attractive features of an unfolding solution can be specified as follows: Variation in both distances and transformed preferences (Busing, Groenen, & Heiser, 2005), preferably about equal; low STRESS values (Kruskal & Carroll, 1969); intermixed sets of objects (I-INDEX) (DeSarbo & Rao, 1986; Busing, Groenen, & Heiser, 2005); and a high number of sufficiently different values for both distances (Shepard, 1974) and transformed preferences (D-INDEX) (see Technical Appendix G for a description of these measures). To keep away from a too complex combination, I-INDEX and D-INDEX are dropped

due to the objections raised before and STRESS can be omitted because it is minimized by the least squares loss function. We therefore concentrate on the use of the variation coefficients for the transformed preferences and the distances.

The values for the coefficient of variation of the distances for normally or uniformly distributed coordinates are approximately equal to $t = .15 + (2p)^{-1}$ as determined by simulation, where $p$ is the dimensionality of the solution. This means that the following plausible rules might be applied: The coefficient of variation of the distances $\upsilon(\mathbf{D})$ must be equal to the target $(t)$ and the coefficients of variation of the distances $\upsilon(\mathbf{D})$ and the *conditional* coefficient of variation of the transformed proximities $\upsilon_c(\mathbf{\Gamma})$ must be equal. A single quality measure is then given as

$$q = |\ln\left(\frac{\upsilon(\mathbf{D})}{t}\right)| + |\ln\left(\frac{\upsilon(\mathbf{D})}{\upsilon_c(\mathbf{\Gamma})}\right)|,$$

where $q$ is equal to zero when we are dealing with a proper unfolding solution, since $\ln(1) = 0$. When the fractions deviate from 1, $q$ becomes larger than zero. To what extend $q$ may deviate from zero for proper unfolding solutions is still unknown. Further research should judge the validity of $q$ for comparing different unfolding solutions.

Instead of using existing measures, it might be feasible to determine a new measure. In the past, and in the previous section, researchers proposed to use resampling methods to assess the quality of multidimensional scaling solutions (Heiser & Meulman, 1983a; Weinberg, Carroll, & Cohen, 1984; de Leeuw & Meulman, 1986) or multidimensional unfolding solutions (Heiser & de Leeuw, 1979a; Heiser, 1981). The assessment consisted of crude non-parametric confidence regions (Heiser & de Leeuw, 1979a; Heiser, 1981) or variance estimates and accompanying confidence regions (based on multi-variate normal distributions) (Weinberg et al., 1984) for the coordinates, or actual stability measures (Heiser & Meulman, 1983a; de Leeuw & Meulman, 1986) and cross-validation and dispersion measures for the entire solution (de Leeuw & Meulman, 1986).

Currently, van de Velden, de Beuckelaer, Groenen, and Busing (2010) uses the bootstrap procedure to find stability measures for the coordinates of an unfolding solution. These stability measures, bias, variation, and mean squared error, are combined into a single stability measure, which is used to compare solutions with different values of the penalty parameters. Preliminary results indicate that the measure is at least capable of distinguishing proper from improper solutions and in most cases even indicates the 'best' solution. Improper solutions are often highly instable due to widely differing degenerate solutions, while proper of even the best solutions exhibit improved stability co-efficients. Even strongly penalized solutions, which are often relatively stable,

are distinguished from less penalized solutions with improved fit and trans-
formations. Although currently an entire grid of penalty parameter values is
searched, an automated procedure is being drawn up. Welcome side-products
of the bootstrap procedure are the stability measures for the coordinates (for
space weights, for regression coefficients, etc), allowing for nonparametric
adjusted parameter estimates and (nonparametric) confidence intervals. The
bootstrap procedure is further discussed in Technical Appendix F (page 196)

*An adjusted coefficient of variation penalty*

Apart from a single measure, it is also an option to continue the development
of the penalty function. A closer inspection of the results of the Monte Carlo
simulation study from Chapter 4 (Figure 4.3) reveals that in the conditional
case the coefficient of variation of the distances increases dramatically for small
values of $\omega$ and large values of $\lambda$ (see page 58), which defines a weak penalty.
In these cases, it seems that one column object isolates itself from the rest of
the objects, by which the distances with this object become relatively large,
as compared to the other distances. It is hypothesized that this phenomenon
arises when PENALIZED STRESS maximizes the coefficient of variation.

Lemma 1 *The coefficient of variation of the transformed preferences $\gamma$ given as*

$$\upsilon(\gamma) = \frac{\sqrt{(n-1)^{-1}\sum(\gamma_i - \overline{\gamma})^2}}{\overline{\gamma}}, \tag{7.1}$$

*where $\overline{\gamma} = n^{-1}\sum \gamma_i$ is the average of $\gamma$, is maximized for*

$$\gamma_i = \begin{cases} 0 & \text{for } i = 1, \dots, n-1 \\ \sqrt{n} & \text{otherwise} \end{cases}$$

*given the arbitrary normalization that $\sum \gamma_i^2 = n$.*

   *Proof.* [1] Given the arbitrary normalization that $\sum \gamma_i^2 = n$, the variation
coefficient of the transformed preferences $\gamma$, under the constraint that $\gamma_i \geqslant$

---

1   Thanks are due to E. Meijer for the formal proof, personal communication, January 15, 2009.

$0 \: \forall \: i = 1, \ldots, n$ and thus $\overline{\gamma} > 0$, can be rewritten as

$$v(\gamma) = \frac{\sqrt{(n-1)^{-1} \sum (\gamma_i - \overline{\gamma})^2}}{\overline{\gamma}}$$

$$= \left[ \frac{(n-1)^{-1} \sum \gamma_i^2 - (n-1)^{-1} n \overline{\gamma}^2}{\overline{\gamma}^2} \right]^{.5}$$

$$= \left[ \frac{n}{n-1} \left( \frac{1 - \overline{\gamma}^2}{\overline{\gamma}^2} \right) \right]^{.5}$$

$$= \left[ \frac{n}{n-1} \left( \frac{1}{\overline{\gamma}^2} - 1 \right) \right]^{.5},$$

which means that the coefficient of variation is maximized for a minimum average transformed preference. Given the constraints, $\max v(\gamma)$ or $\min \sum \gamma_i$, which is equivalent to $\min \overline{\gamma}$, is found as follows. For $\gamma_i$, there are two possibilities: either $\gamma_i = 0$ or $\gamma_i > 0$, so let the first $k$ elements of $\gamma$ be zero and the last $n - k$ elements be positive. The Lagrange function with the two restrictions is given as

$$L(\gamma, \lambda, \mu) = f(\gamma) + \eta \left[ h(\gamma) - c \right] + \mu \left[ g(\gamma) - d \right]$$

$$= \sum \gamma_i + \eta \left[ \sum \gamma_i^2 - n \right] + \left[ \sum \mu_i \gamma_i \right],$$

where $f(\gamma) = \sum \gamma_i$ is the function that is maximized, both $h(\gamma) = \sum \gamma_i^2 = n$ and $g(\gamma) = \sum \gamma_i \geqslant 0$ are the restrictions, and $\eta$ and $\mu$ are the Lagrange multipliers. For an optimum, the following conditions must apply

$$\frac{\partial L}{\partial \gamma_i} = \begin{cases} 1 + 2\eta\gamma_i + \mu_i = 1 + \mu_i = 0 & \text{for } i = 1, \ldots, k \ (\gamma_i = 0), \\ 1 + 2\eta\gamma_i + \mu_i = 1 + 2\eta\gamma_i = 0, & \text{for } i = k+1, \ldots, n \ (\gamma_i > 0), \end{cases}$$

(7.2)

and the partial derivatives $\partial L/\partial \eta = 0$, $\partial L/\partial \mu_i = 0$ for $\gamma_i = 0$ (obligatory restriction), and $\partial L/\partial \mu_i = 0$ for $\gamma_i > 0$ (non-obligatory restriction). It follows from (7.2) that $\gamma_i = -(2\eta)^{-1} \: \forall \: i > k$, for which all $\gamma_i$ are equal. Thus,

$$\gamma_i = \begin{cases} 0 & \text{for } i = 1, \ldots, k, \text{ and} \\ c & \text{for } i = k+1, \ldots, n, \end{cases}$$

where $c = -(2\eta)^{-1}$. Remains the determination of the values for $c$ and $k$. Since

$$\sum_{i=1}^{n} \gamma_i^2 = \sum_{i=1}^{k} \gamma_i^2 + \sum_{i=k+1}^{n} \gamma_i^2 = k \times 0 + (n-k) \times c^2 = n,$$

$c = \sqrt{\frac{n}{n-k}}$. Now, $k$ can be determined to minimize $\overline{\gamma}$ with $\gamma_i = 0 \; \forall \; i = 1,\dots,k$ and $\gamma_i = \sqrt{n}/\sqrt{n-k} \; \forall \; i = k+1,\dots,n$ as

$$\overline{\gamma} = \sum_{i=1}^{n} \frac{\gamma_i}{n} = \sum_{i=1}^{k} \frac{\gamma_i}{n} + \sum_{i=k+1}^{n} \frac{\gamma_i}{n} = \frac{k \times 0}{n} + \frac{(n-k)}{n} \sqrt{\frac{n}{n-k}} = \frac{\sqrt{n-k}}{\sqrt{n}}$$

and $\overline{\gamma}$ is thus minimized for a maximum $k$. Since $k$ is an integer value and $k$ cannot be equal to $n$, due to the $\sum \gamma_i^2 = n$ restriction, the minimum for $\overline{\gamma}$ is found for $k = n - 1$, which gives the solution as

$$\gamma_i = \begin{cases} 0 & \text{for } i = 1,\dots,n-1 \\ \sqrt{n} & \text{otherwise} \end{cases}$$

with $\overline{\gamma} = 1/\sqrt{n}$ and $\upsilon = \sqrt{n}$. $\qquad\qquad\qquad\qquad\qquad\qquad \Box$

A maximum coefficient of variation thus coincides with one large value and many small or zero values. This is identical to the observed phenomenon with one distant column object, but it is only effective for the row-conditional model: For each row object, all column objects are close, except for one column object, that is at a large distance. This allows the coefficient of variation to become maximal for each row and thus for the penalty as a whole.

Although the above argumentation indicates that the PENALIZED STRESS function (B.4) maximizes the coefficient of variation in such a case, this is not completely true, since the PENALIZED STRESS function does not exclusively consists of the penalty function and the penalty function itself does not consists only of the inverse of the variation coefficient. The penalty function (B.2) also consists of $1+$, a component of the penalty function that is easily overlooked. This component was used for the first time in Groenen (1993, pp. 54–55) to overcome the problem of 'attraction to the horizon'. Due to the $1+$, PENALIZED STRESS minimizes

$$\sigma_p^2(\gamma, d) = \sigma_n^2(\gamma, d) + \sigma_n^2(\gamma, d) \left( \frac{\omega}{\upsilon^2(\gamma)} \right)^{1/\lambda}, \qquad\qquad (7.3)$$

which shows that maximizing the coefficient of variation only minimizes half of the PENALIZED STRESS function, the second part on the right-hand side of (7.3), depending on the values for $\omega$ and $\lambda$. Maximization of the coefficient of variation must therefore also be advantageous for the STRESS part of the PENALIZED STRESS function, which is not the case when the distant single column object is not the least preferred by all row objects. If this is the case, however, it might be argued that we are dealing with an outlier, and that the object might be removed from the data for this reason.
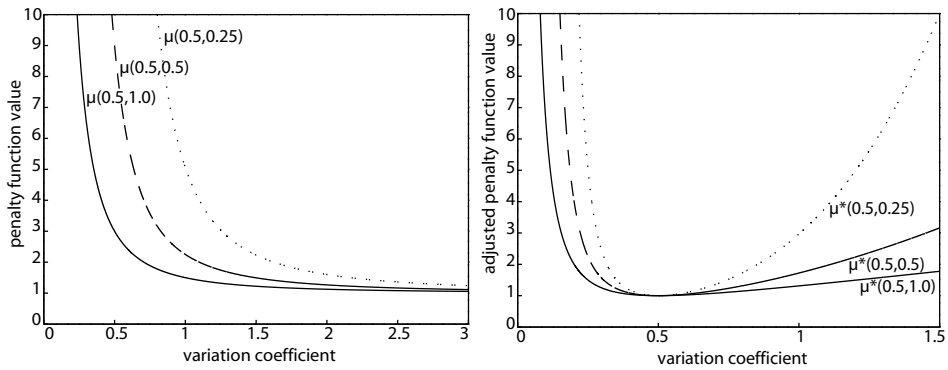
Figure 7.1   Function plots for the current penalty function (left-hand panel) and the suggested penalty function (right-hand panel).

Another way to proceed is that the penalty function can be adapted in order to avoid the undesirable side effects of maximization of the coefficient of variation. Let us recall the penalty function from Chapter 4. Figure 7.1 (left-hand panel) shows a plot of the function

$$\mu(\omega, \lambda) = \left(1.0 + \omega \upsilon^{-2}(\gamma)\right)^{1/\lambda} \tag{7.4}$$

with $\omega = 0.5$ for different values of $\upsilon(\gamma)$ and for $\lambda = 1.0, 0.5, 0.25$. It shows that an increase of the variation coefficient (horizontal axis) causes a decrease in penalty function values (vertical axis). Maximizing the variation coefficient thus minimizes the penalty function and consequently also PENALIZED STRESS, to a certain extend, as indicated above.

In order to avoid a continuous decrease in penalty function values for increasing variation coefficients, an adjusted penalty function should increase its values after a certain point, thus avoiding 'attraction to the horizon'. This requirement means that the adjusted penalty function will have a minimum, as the function increases in value for both smaller and larger variation coefficients. The minimum of the adjusted penalty function may conveniently be locked either to the variation coefficient of the original preferences or to the target variation coefficient. An adjusted penalty function might be specified as

$$\mu_a(\omega_a, \lambda_a) = \left(0.25 + 0.25\omega_a^2 \upsilon^{-2}(\gamma) + 0.5\omega_a^{-1}\upsilon(\gamma)\right)^{1/\lambda_a}, \tag{7.5}$$

where $\omega_a$ assumes the role of $\omega$ and $\lambda_a$ the role of $\lambda$, as compared to the original penalty function. The minimum of $\mu_a(\omega_a, \lambda_a)$ is found for $\omega_a$, which allows one to specify the minimum at either of the above suggested

values. Pre-specifying $\omega_a$ has the additional advantage of reducing the penalty parameter set with one parameter, leaving only $\lambda_a$ to be specified. The function value at the minimum $\omega_a$ is equal to 1.0 (due to the specific fractions used in (7.5)), irrespective of the values for $\omega_a$ and $\lambda_a$. In addition, this allows a fair comparison of adjusted PENALIZED STRESS values from different solutions. Figure 7.1 (right-hand panel) shows a plot of the adjusted penalty function $\mu_a(0.5, \lambda_a)$ for different values of $\upsilon(\gamma)$ and for $\lambda_a = 1.0, 0.5, 0.25$ with $\omega_a = 0.5$. The minimum for $\mu_a(0.5, \lambda_a)$ is attained for $\omega_a$ at 0.5, as discussed above, and moving in either direction causes the penalty function to increase. For smaller values of $\lambda_a$, $\mu_a(0.5, \lambda_a)$ shows a steeper increase in function values, while maintaining its minimum at $\omega_a = 0.5$. Further development and implementation of such an adjusted penalty function is left as plan for the future.

## 7.3    RESTRICTED UNFOLDING

The restricted unfolding model finds an optimal configuration of two sets of objects, where the coordinates of either one or both sets are restricted to be a linear combination of independent variables. The model further allows for optimal transformations of the variables. The merger of linear combinations and optimal transformations is equivalent with categorical regression analysis (CATREG, van der Kooij & Meulman, 2004). The restricted unfolding model is discussed in Chapter 5.

Most problems related to the restricted unfolding model, as described by P. E. Green and Krieger (1989, p. 132), have been resolved by the current unfolding approach, specifically the difficulty of constructing joint spaces and ideal points and relating perceived dimensions to manipulable attributes. Future research concentrates on the (prior) specification of variables (e.g., model or subset selection), improved prediction and interpolation, specifically with optimal transformations, and optimal graphical representations (Gower & Hand, 1996; Tufte, 2001). In the following, we will only touch upon the former problem, the selection of variables for the restricted unfolding model.

*Subset selection of variables in restricted unfolding*

There are different situations in which it is desired to use only a subset of a large number of variables. A. Miller (2002) is concerned with the situation in which the value of one variable (say a coordinate) is predicted from a number of other variables (say a number of independent variables) and uses subset selection to improve prediction. Another situation occurs when the number of variables exceeds the number of objects, in which case the (regression) model is not identified. At least two complications arise when selecting subsets for the

restricted unfolding model: The model employs (optimal) transformations for the variables and the actual regression procedure is only a subproblem of the entire model estimation. The latter might form a serious obstacle, for which possible solutions are discussed hereafter, whereas the former was recently addressed by van der Kooij (2007).

The number of variables can be reduced in advance by using a linear combination of independent variables, as suggested by DeSarbo and Rao (1986), who used a principal component analysis as a guard against multicollinearity. DeSarbo and Rao derived the principal component scores and replaced the variables with the scores. Instead of retaining all components, as in the case of DeSarbo and Rao, only the first few components can be used, components that correspond with the largest singular values of the matrix with independent variables. The disadvantages of this method are limited to the need for measuring all variables and the possible correlation of the predictant with low singular value scores. The method is not restricted to retaining orthogonal scores, and it is even possible to meet the categorical nature of the independent variables by using a categorical principal component analysis (CATPCA, Meulman et al., 2004). Whatever analysis is used, the interpretation of the model is not facilitated by the use of fewer components than variables. Direct relations between coordinates and variable categories are no longer present since additional reparametrizations (via scores and eigenvalues) are necessary to reestablish the original variable category scores.

Another procedure to reduce the number of variables is offered by the LASSO , the least absolute shrinkage and selection operator, as one example of a constrained version of ordinary least squares regression. The LASSO shrinks some coefficients and sets others to zero (Tibshirani, 1996), and as such it can be used for subset selection. The backfitting algorithm, already implemented to deal conveniently with the variable transformations, also ensures an easy implementation of the LASSO (cf. van der Kooij, 2007). Once the subset is identified, the regression weights can be computed without shrinkage. The advantages of the LASSO over a linear combination of variables are the use of the original variables and the possible optimal transformation thereof. For the interpretation, the original variables are used, transformed or not, although some variables are lost due to a coefficient equal to zero. A disadvantage of the LASSO, in the case of the restricted unfolding model, is the preliminary specification of the number of variables remaining in the model. Although the LASSO, as for example used in CATREG (van der Kooij, 2007), allows for the optimal determination of the shrinkage factor through bootstrap or cross-validation, it is premature to conclude that this will work for the restricted unfolding model. The variable restrictions are only a small subproblem, hidden deep in the unfolding algorithm (see Technical Appendix E), and all kinds of dependencies and time considerations will probably make the implementation

unfeasible. Only specifying the number of variables in advance seems to offer a practicable alternative.

Finally, it is also possible that the number of variables is reduced by finding an optimal subset of variables via forward selection, backward elimination, sequential replacement, branch-and-bound techniques, or exhaustive search (see, for example, A. Miller, 2002), but the iterative unfolding algorithm makes it very hard to combine one of these procedures with the variable restriction option of the unfolding model.

### 7.4    UNFOLDING INCOMPLETE DATA

It has been known, since Kruskal, 1964a, 1964b, that the least squares loss function for multidimensional scaling allows for missing data. Although this is also true for multidimensional unfolding, the extend to which data can be missing without changing the conclusions based on the results of the unfolding with incomplete data has been unknown. Some advances in this field were discussed in Chapter 6, Unfolding Incomplete Data. Research on incomplete data in least squares unfolding was initiated with the master thesis of Velderman (2005). The results from a subsequent publication (Busing & de Rooij, 2009), reproduced in Chapter 6, are promising: Moderate to large samples recover the original solution more than satisfactory with even half of the data missing. The method that was used to deal with the missing values is known as pairwise deletion, that is, the missing data was not replaced (imputed) but just ignored. A small comparison with imputed values was inconclusive and other considerations than recovery led to the choice for pairwise deletion (see Chapter 6, page 101).

For small samples, however, the deletion method performs less satisfactory, and, as was pointed out by one of the referees of Busing and de Rooij (2009), small samples are frequently observed in practical research with the potential use for unfolding. Research in progress by Busing (2010) now focusses on imputation techniques for small samples. Elaborating on the publication by Hedderley and Wakeling (1995), several imputation techniques are considered for comparison.

*Imputation techniques for unfolding incomplete small samples*

The simplest class of imputation methods is single imputation, of which the oldest method is probably *mean imputation* (presumably suggested by Wilks, 1932). The missing value is replaced by the mean based on (parts of) the remaining data, which might be the row mean $\bar{\delta}_i$ (the average of the preferences of a row object) or the column mean $\bar{\delta}_j$ (the average of the preferences for a column object), or a more sophisticated mean such as $\delta_{ij} = \bar{\delta}_i + \bar{\delta}_j - \bar{\delta}$, where $\bar{\delta}$ is

the overall mean (Bernaards & Sijtsma, 2000). This last method is commonly augmented with some random component, whether or not leading to multiple imputed values (see van Ginkel, van der Ark, & Sijtsma, 2007), and known as *two-way imputation*. Another remarkable simple imputation method was developed by Krzanowski (1988). The missing value is reconstructed based on the singular-value decompositions $\mathbf{\Delta} = \mathbf{UDV}'$ of two matrices, one matrix $\widetilde{\mathbf{\Delta}}$ omitting the row containing the missing value and one matrix $\overline{\mathbf{\Delta}}$ omitting the column containing the missing value, hence the name *row-column imputation*. The imputed value is computed as $\delta_{ij} = \sum [\widetilde{u}_{it}\widetilde{d}_t^{.5}][\overline{v}_{jt}\overline{d}_t^{.5}]$, where the summation is over t, the pre-specified dimensionality. A modification of this method, as described in Bergamo, dos Santos Dias, and Krzanowski (2008), leads to a multiple imputation method with differential weighting of the two singular values $\widetilde{d}$ and $\overline{d}$, although the advantages are rather unclear. For multiple missing values, the row-column imputation uses an iterative scheme to update the imputed values, which keeps iterating until the values stabilize.

An imputation method loosely based on the EM algorithm (Dempster, Laird, & Rubin, 1977; Little & Rubin, 1987), but utilizing the unfolding model, is the following. Starting with an initial guess (or starting with the deletion method), an unfolding solution is determined of which the distances are used to estimate the imputed values. The procedure is repeated until the solution stabilizes.

The most recent class of imputation methods concerns multiple imputation. Without regard to the possibility of adding random error in one of the methods described above, multivariate normal imputation randomly draws values from the conditional distribution of the missing values, given the observed preferences and the model parameters. The method is well-known, performs well, and is robust against departure from the multivariate normal model (Graham & Schafer, 1999), but, nevertheless, assumes a distribution, which can not be said from the other methods. The processing of the different imputed data sets can be handled in different ways. The simplest continuation is to unfold each imputed data set separately and combine the results to obtain point estimates (means) and interval estimates (variances), or display (non-parametric) confidence intervals in one final configuration (see Figure 7.2, left-hand panel). Another approach, graphically depicted in Figure 7.2 (right-hand panel), creates a third way stacking the imputed data sets and proceeds with a three-way unfolding analysis. This way, the point estimates are directly observed as the final coordinates and a decomposition of the mean squared error provides interval estimates or variances for the coordinates.

Further research should give answers to what method is preferred under which circumstances, circumstances that differ in data size, measurement level, transformation function, conditionality, error, and dimensionality.
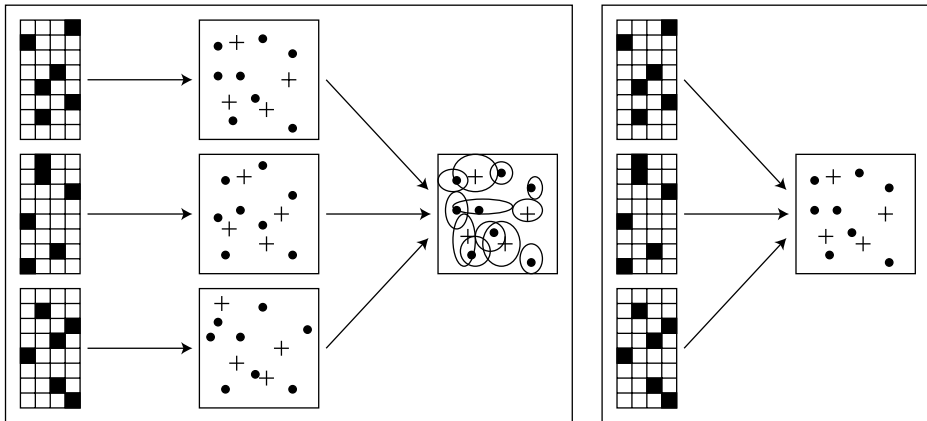
Figure 7.2    Processing of multiple imputed data sets with two-way unfolding (left-hand panel) and three-way unfolding (right-hand panel).

## 7.5    FINAL CONCLUSIONS

This monograph has tried to develop the unfolding technique into a more reliable and practical method for data analysis. It goes without saying that many advances are still needed some of which were indicated briefly in these conclusions. Other, seemingly less urgent, but definitely long-standing topics need to be addressed as well. For example, measures should be developed for obtaining rectangular matrices with data appropriate for unfolding analysis, assuring adequate use of different type of data, such as dichotomous data, paired-comparisons, frequencies, or abundances. The effectiveness of initial configurations needs to be evaluated and these procedures need to be properly matched with data characteristics and model options. Confirmatory analyses using resampling methods, such as the jackknife, bootstrap, cross-validation, and permutation analysis, should be implemented to help researchers make decisions, for example concerning the adequacy of transformation functions. Additional analyses, based on the unfolding outcomes, such as the analysis of angular variation, outlier analysis, cluster analysis, or latent class analysis, should be available as unfolding analysis options to facilitate the interpretation of the results. And finally, graphical output should be improved, with attention for the principles laid down by, for example, Tufte (2001) and colleagues. Research on these and previously described topics is only feasible after the creation of a firm basis. Least squares unfolding, as presented in this monograph, with its sound algorithm based on alternating least squares and iterative majorization, with its optimal transformations of the preferences, with its ability to handle missing data, and with its versatile restriction facilities, offers such a basis.

# NOTATION OVERVIEW

## A.1 NOTATION CONVENTIONS

Three-way arrays are denoted by underlined bold uppercase characters ($\underline{\mathbf{\Delta}}$), two-way arrays or matrices by bold uppercase characters ($\mathbf{\Delta}$), and one-way arrays or vectors by bold lowercase characters ($\boldsymbol{\delta}$). Scalars are denoted by lowercase characters ($\delta$) without discriminating between integers and floating point numbers. Functions are also denoted by lowercase characters, but these are followed by a set of parentheses containing the function parameters (e.g., $f(\delta)$). Vectorized arrays are denoted as $\boldsymbol{\delta} = \text{vec}\,\mathbf{\Delta}$, indicating a matrix $\mathbf{\Delta}$ with successive columns strung out below each other in a single vector $\boldsymbol{\delta}$. A matrix with $\boldsymbol{\delta}$ on the diagonal is denoted as $\mathbf{\Delta} = \text{diag}\,(\boldsymbol{\delta})$, whereas $\boldsymbol{\delta} = \text{diag}\,(\mathbf{\Delta})$ specifies a vector $\boldsymbol{\delta}$ containing the diagonal of matrix $\mathbf{\Delta}$. The functions $\text{tr}$, $\text{vec}$, and $\text{diag}$ only use a set of parentheses to avoid ambiguity.

The underlining for three-way arrays is described in Kiers (2000). Further notation originates from publications on multidimensional scaling and multidimensional unfolding.

## A.2 SYMBOLS

The following symbols, describing the two-way unfolding model, are used throughout this monograph, unless defined otherwise.

| | |
|---|---|
| $n$ | number of row objects |
| $m$ | number of column objects |
| $p_{min}$ | minimum dimensionality |
| $p_{max}$ | maximum dimensionality |
| $p$ | current dimensionality |
| $h$ | number of independent variables, either row or column |
| | |
| $\lambda$ | penalty parameter (strength) |
| $\omega$ | penalty parameter (range) |
| | |
| $\mathbf{0}$ | vector with zeros |
| $\mathbf{1}$ | vector with ones |
| $\mathbf{I}$ | identity matrix |
| $\mathbf{J}$ | centering matrix, where $\mathbf{J} = \mathbf{I} - \mathbf{11}'/\mathbf{1}'\mathbf{1}$ |

| | |
|---|---|
| $\boldsymbol{\Delta}$ | raw preferences |
| $\boldsymbol{\Gamma}$ | transformed preferences $\boldsymbol{\Gamma} = f(\boldsymbol{\Delta})$ |
| $\boldsymbol{\Gamma}^2$ | element-wise squared preferences $\boldsymbol{\Gamma}$ |
| $\boldsymbol{\Gamma}^+$ | $(n + m) \times (n + m)$ data matrix with $\boldsymbol{\Gamma}$ on off-diagonal |
| $\dot{\boldsymbol{\gamma}}$ | previous update of transformed preferences $\boldsymbol{\gamma} = \text{vec } \boldsymbol{\Gamma}$ |
| $\boldsymbol{\xi}$ | majorization vector for transformed preferences $\boldsymbol{\gamma}$ |
| | |
| $\boldsymbol{W}$ | preference weights |
| $\boldsymbol{R}$ | diagonal matrix with column sums of $\boldsymbol{W}$, $\boldsymbol{R} = \text{diag}\,(\boldsymbol{W}\boldsymbol{1})$ |
| $\boldsymbol{C}$ | diagonal matrix with row sums of $\boldsymbol{W}$, $\boldsymbol{C} = \text{diag}\,(\boldsymbol{1}'\boldsymbol{W})$ |
| | |
| $\boldsymbol{X}_o$ | initial coordinate matrix for row objects |
| $\boldsymbol{Y}_o$ | initial coordinate matrix for column objects |
| $\boldsymbol{X}$ | coordinate matrix for row objects |
| $\boldsymbol{Y}$ | coordinate matrix for column objects |
| $\boldsymbol{Z}$ | coordinate matrix for row *and* column objects $\boldsymbol{Z} = [\boldsymbol{X}', \boldsymbol{Y}']'$ |
| $\boldsymbol{X}^+$ | update for row coordinates $\boldsymbol{X}$ |
| $\boldsymbol{Y}^+$ | update for column coordinates $\boldsymbol{Y}$ |
| | |
| $\boldsymbol{D}$ | distances between rows of $\boldsymbol{X}$ and rows of $\boldsymbol{Y}$, $\boldsymbol{D} = d(\boldsymbol{X}, \boldsymbol{Y})$ |
| | |
| $\widetilde{\boldsymbol{X}}$ | preliminary update for $\boldsymbol{X}$, $\widetilde{\boldsymbol{X}} = \boldsymbol{P}\boldsymbol{X} - \boldsymbol{B}\boldsymbol{Y}$ |
| $\widetilde{\boldsymbol{Y}}$ | preliminary update for $\boldsymbol{Y}$, $\widetilde{\boldsymbol{Y}} = \boldsymbol{Q}\boldsymbol{Y} - \boldsymbol{B}'\boldsymbol{X}$ |
| $\dot{\widetilde{\boldsymbol{X}}}$ | preliminary update for $\boldsymbol{X}$ from previous iteration |
| $\dot{\widetilde{\boldsymbol{Y}}}$ | preliminary update for $\boldsymbol{Y}$ from previous iteration |
| | |
| $\boldsymbol{x}^i$ | individual space row object coordinates for row object $i$ |
| $\boldsymbol{Y}^i$ | individual space column objects coordinates for row object $i$ |
| $\underline{\boldsymbol{A}}$ | space weights, such that $\boldsymbol{x}^i = \boldsymbol{x}_i \boldsymbol{A}_i$ and $\boldsymbol{Y}^i = \boldsymbol{Y}\boldsymbol{A}_i$ |
| $\underline{\boldsymbol{A}}^+$ | update for space weights $\underline{\boldsymbol{A}}$ |
| | |
| $\boldsymbol{E}$ | independent variables for row or column objects |
| $\boldsymbol{Q}$ | transformed independent variables for row or column objects |
| $\boldsymbol{Q}^*$ | unrestricted update for transformed independent variables $\boldsymbol{Q}$ |
| $\boldsymbol{Q}^+$ | restricted update for transformed independent variables $\boldsymbol{Q}$ |
| $\boldsymbol{B}$ | matrix with regression coefficients for row or column objects |
| $\boldsymbol{B}^+$ | update for the regression coefficients $\boldsymbol{B}$ |
| $\boldsymbol{A}$ | matrix with direction coefficients for row or column objects |
| $\boldsymbol{P}$ | matrix with variable projections, $\boldsymbol{P} = \boldsymbol{X}\boldsymbol{A}$ or $\boldsymbol{P} = \boldsymbol{Y}\boldsymbol{A}$ |

Symbols used for the two-way model are generalized to the three-way model by adding an extra dimension to the appropriate array of the two-way model.

Specific and deviating symbols for the three-way model are provided here.

| | |
|---|---|
| s | number of sources, slices, matrices, or two-way arrays |
| $\underline{\Delta}$ | raw preferences |
| $\underline{W}$ | preference weights |
| $X_k$ | individual space row objects coordinates for source k |
| $Y_k$ | individual space column objects coordinates for source k |
| $\underline{A}$ | space weights, such that $X_k = XA_k$ and $Y_k = YA_k$ |

## A.3   FUNCTIONS

| | |
|---|---|
| $\sigma$ | badness-of-fit function STRESS |
| $\sigma_r$ | badness-of-fit function RAW STRESS |
| $\sigma_n$ | badness-of-fit function NORMALIZED RAW STRESS |
| $\sigma_1$ | badness-of-fit function STRESS-1 |
| $\sigma_2$ | badness-of-fit function STRESS-2 |
| $\sigma_{s1}$ | badness-of-fit function S-STRESS-1 |
| $\sigma_{s2}$ | badness-of-fit function S-STRESS-2 |
| $\sigma_p$ | badness-of-fit function PENALIZED STRESS |

| | |
|---|---|
| $\upsilon(\cdot)$ | coefficient of variation of the argument's elements |
| $\mu(\cdot)$ | penalty function in PENALIZED STRESS |

| | |
|---|---|
| $f(\cdot)$ | (transformation) function |
| $g(\cdot)$ | majorizing function |
| $d(\cdot)$ | Euclidean distance function |

| | |
|---|---|
| $\mathrm{argmin}\,(\cdot)$ | minimum value of the argument's elements |
| $\mathrm{argmax}\,(\cdot)$ | maximum value of the argument's elements |
| $\mathrm{med}\,(\cdot)$ | median of the argument's elements |

## A.4   ACRONYMS

*Function acronyms*

| | |
|---|---|
| DAF | goodness-of-fit function dispersion accounted for |
| D-INDEX | distinctness index |
| FIRST | percentage of first choices correctly recovered |
| I-INDEX | intermixedness index |
| KAPPA | goodness-of-fit function Cohen's κ |
| N-STRESS | badness-of-fit function NORMALIZED RAW STRESS |
| ORDERS | recovered preference orders |

| | |
|---|---|
| PHI | goodness-of-fit function Tuckers' congruence coefficient φ |
| P-STRESS | badness-of-fit function PENALIZED STRESS |
| RHO | Spearman's rank order correlation ρ |
| R-STRESS | badness-of-fit function RAW STRESS |
| SSAF | sum-of-squares accounted for |
| S-STRESS-1 | badness-of-fit function Young's STRESS formula one |
| S-STRESS-2 | badness-of-fit function Young's STRESS formula two |
| STRESS-1 | badness-of-fit function Kruskal's STRESS formula one |
| STRESS-2 | badness-of-fit function Kruskal's STRESS formula two |
| STRESS | (standardized) residual sum-of-squares |
| TAU | Kendall's rank order correlation τ |
| VAF | variance accounted for |

*Software acronyms*

| | |
|---|---|
| ALSCAL | alternating least squares MDS program (IBM SPSS procedure) |
| CATPCA | categorical principal component analysis (IBM SPSS procedure) |
| GENFOLD | general unfolding program, different variants |
| IBM | international business machines corporation |
| KYST | Kruskal-Young-Shepard-Torgerson MDS and MDU program |
| LSA | landscape segmentation analysis |
| MDPREF | multidimensional preference scaling |
| MDS(X) | multidimensional scaling program suite |
| MINIRSA | Michigan Israel Netherland integrated rectangular space analysis |
| MINISSA | Michigan Israel Netherland integrated smallest space analysis |
| NEWFOLD | new multidimensional unfolding program |
| PREFMAP | external unfolding program, different variants |
| PREFSCAL | preference scaling (IBM SPSS procedure) |
| PROFIT | property fitting |
| PROXSCAL | proximity scaling (IBM SPSS procedure) |
| SMACOF | scaling by majorizing a complicated function, different variants |
| SPSS | software package for the social sciences, an IBM company |
| SSA | smallest space analysis, different variants |
| TORSCA | Torgerson (classical) scaling |
| VIPSCAL | vector ideal point scaling |

*Other acronyms*

| | |
|---|---|
| ALS | alternating least squares minimization procedure |
| ANAVA | analysis of angular variation |
| BIBD | balanced incomplete block design |
| IM | iterative majorization |

| | |
|---|---|
| LASSO | least absolute shrinkage and selection operator |
| MAR | missing at random |
| MCAR | missing completely at random |
| MDS | multidimensional scaling |
| MDU | multidimensional unfolding |
| NMAR | not missing at random |
| PCA | principal components analysis |
| ROW-BIBD | row-balanced incomplete block design |

LEAST SQUARES UNFOLDING ALGORITHM

PREFSCAL minimizes a loss function called PENALIZED STRESS, of which the square is defined as the square of NORMALIZED RAW STRESS times a penalty function, i.e.,

$$\sigma_p^2(\boldsymbol{\gamma}, \mathbf{d}) = \sigma_n^{2\lambda}(\boldsymbol{\gamma}, \mathbf{d})\mu(\boldsymbol{\gamma}).$$

The square of NORMALIZED RAW STRESS, the same STRESS function that is minimized by IBM SPSS PROXSCAL, is given by

$$\sigma_n^2(\boldsymbol{\gamma}, \mathbf{d}) = \frac{\|\boldsymbol{\gamma} - \mathbf{d}\|_W^2}{\|\boldsymbol{\gamma}\|_W^2}, \tag{B.1}$$

where $\|\boldsymbol{\gamma} - \mathbf{d}\|_W^2$ is the weighted squared Euclidean norm of the differences between the transformed preferences $\boldsymbol{\gamma} = \mathrm{vec}\,(\boldsymbol{\Gamma})$ and the distances $\mathbf{d} = \mathrm{vec}\,\mathbf{D}$ in the metric $W$ and $\|\boldsymbol{\gamma}\|_W^2$ is the weighted sum-of-squares of $\boldsymbol{\gamma}$. Note that $W$ is somewhat confusingly defined as $\mathrm{diag}\,(\mathrm{vec}\,(\boldsymbol{W}))$, where the $\mathrm{vec}\,(\cdot)$ operation is performed over the original matrix with preference weights. The penalty function $\mu(\boldsymbol{\gamma})$ is defined as one plus the inverse of the squared coefficient of variation of the transformed preferences $\boldsymbol{\gamma}$, that is,

$$\mu(\boldsymbol{\gamma}) = 1 + \omega^* \frac{\upsilon^2(\boldsymbol{\delta})}{\upsilon^2(\boldsymbol{\gamma})}, \tag{B.2}$$

where $\upsilon^2(\boldsymbol{\gamma})$ is the squared variation coefficient which is defined as

$$\begin{aligned}
\upsilon^2(\boldsymbol{\gamma}) &= \frac{w_{++}^{-1}\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma} - w_{++}^{-2}\boldsymbol{\gamma}'\boldsymbol{w}\boldsymbol{w}'\boldsymbol{\gamma}}{w_{++}^{-2}\boldsymbol{\gamma}'\boldsymbol{w}\boldsymbol{w}'\boldsymbol{\gamma}} \\
&= \frac{\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma} - w_{++}^{-1}\boldsymbol{\gamma}'\boldsymbol{w}\boldsymbol{w}'\boldsymbol{\gamma}}{w_{++}^{-1}\boldsymbol{\gamma}'\boldsymbol{w}\boldsymbol{w}'\boldsymbol{\gamma}} \\
&= \frac{\|\boldsymbol{\gamma}\|_V^2}{\|\boldsymbol{\gamma}\|_M^2}, \tag{B.3}
\end{aligned}$$

where $w_{++} = \boldsymbol{w}'\boldsymbol{w}$, $\boldsymbol{M} = w_{++}^{-1}\boldsymbol{w}\boldsymbol{w}'$, and $\boldsymbol{V} = \boldsymbol{W} - \boldsymbol{M}$. Now, combining (B.1), (B.2), and (B.3) and re-introducing the two penalty parameters, $\lambda$ and $\omega = \omega^*\upsilon^2(\boldsymbol{\delta})$, expresses the square of PENALIZED STRESS as

$$\sigma_p^2(\boldsymbol{\gamma}, \mathbf{d}) = \left(\frac{\|\boldsymbol{\gamma} - \mathbf{d}\|_W^2}{\|\boldsymbol{\gamma}\|_W^2}\right)^\lambda \left(\frac{\|\boldsymbol{\gamma}\|_V^2 + \omega\|\boldsymbol{\gamma}\|_M^2}{\|\boldsymbol{\gamma}\|_V^2}\right), \tag{B.4}$$
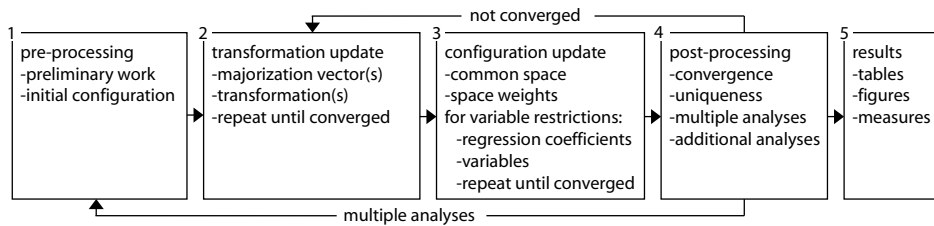
Figure B.1    PREFSCAL algorithm.

where $\lambda$, $\omega$, and $\omega^*$ are scalars, with $0 < \lambda \leqslant 1$, $\omega^* \geqslant 0$. Since $\upsilon^2(\delta) \geqslant 0$, $\omega \geqslant 0$. Both $\lambda$ and $\omega^*$ are specified by the user.

PENALIZED STRESS (B.4) is minimized by an alternating least squares (ALS) procedure, following the general strategy first used in MDS by Takane, Young, and de Leeuw (1977), in which we alternate between finding an update for the configuration given the current estimate of the transformed preferences and finding an update for the transformed preferences, given the current estimate of the configuration. In the present case, each of these two steps is carried out by iterative majorization (IM). A general discussion on iterative majorization, in the context of multidimensional scaling and multidimensional unfolding, can be found in de Leeuw (1977a), de Leeuw and Heiser (1980), Heiser (1981), Heiser (1987a), de Leeuw (1988), Heiser (1995), and Borg and Groenen (2005). More details and specific functions can be found in Heiser and Stoop (1986), Heiser (1991), Commandeur and Heiser (1993), Groenen (1993), Groenen, Mathar, and Heiser (1995), and Groenen and Heiser (1996).

The minimization of PENALIZED STRESS, which will be discussed in detail hereafter, and is schematically depicted in Figure B.1, consists of the following successive steps:

Step 1.    Choose an initial configuration and evaluate the function (Appendix C);
Step 2.    Compute an update for the transformed preferences (Appendix D);
Step 3.    Compute an update for the configuration (Appendix E);
Step 4.    Evaluate the function and if some predefined termination criterion is satisfied, continue; otherwise, go to Step 2 (Appendix F);
Step 5.    Compute tables, figures, and measures (Appendix G).

PENALIZED STRESS *considerations.*    The PENALIZED STRESS function implemented in PREFSCAL deviates from the function discussed in Busing, Groenen, and Heiser (2005) with respect to two parts of the loss function. First, initiated by the implementation of the three-way models, NORMALIZED RAW STRESS (B.1) is used instead of RAW STRESS, which adds another $\gamma$-based normalization

to the PENALIZED STRESS function. Secondly, $\omega$ is set dependent on the variation of the raw preferences, such that when the user specifies $\omega^* = 1$, it ensures a unit proportion between the variation of the original preferences and the transformed preferences. Busing, Groenen, and Heiser (2005) provide an extensive appendix describing the procedure to minimize the PENALIZED STRESS function with a user-provided $\omega$ and RAW STRESS.

*PREFSCAL considerations.* PREFSCAL actually exists as two related programs. The first PREFSCAL program is the beta version, a stand-alone console application developed at Leiden University to perform three-way three-mode multidimensional unfolding with restrictions through iterative majorization and alternating least squares of PENALIZED STRESS. Following the taxonomy of scaling methods according to Carroll and Arabie (1980), who separate data and model properties, the implementation contains for the following data properties: Two-way two-mode and three-way three-mode data, ordinal, interval, and ratio data, unconditional, row-conditional, and matrix-conditional data, and complete and incomplete data. The properties of the multidimensional measurement model include the Euclidean distance model for one (configuration) or two spaces (space weights), internal and external (one set of points fixed) solutions, and both coordinate and variable constraints on the coordinates whether or not combined. PREFSCAL beta is executed in a command window under MICROSOFT WINDOWS, uses simple text input, and provides simple text output (and one LATEX $2_\varepsilon$ figure). Since research is still being done to polish, improve, and add certain options to the program, beta testers are a prerequisite for helping us to deliver good software, which makes it inescapable to make the beta version freely available, under certain restrictions. The second PREFSCAL program is implemented in IBM SPSS STATISTICS. This version consists of a selection of options from the beta version, a selection that has been tested, polished, improved, tested, and polished again, and thus provides nice tables and smooth figures for output. The first implementation of PREFSCAL was in SPSS version 14.0. Since then, only a few bugs needed to be fixed.

C
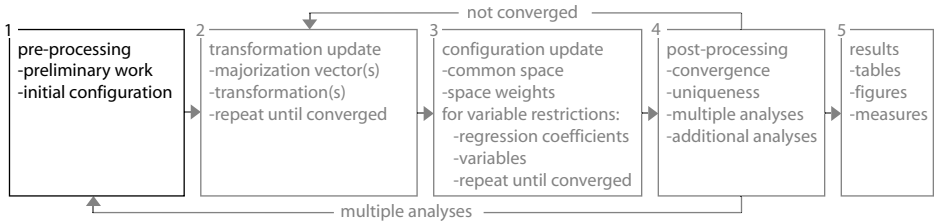
The pre-processing for PREFSCAL consists of several steps, mainly computations to ensure that the main algorithm is able to proceed without problems. For this purpose, the raw data is checked for missing elements and anomalies, initial transformations are computed for the raw data, as well as other preliminary computations are performed, for example, for the forthcoming computation of the optimally transformed preferences (Technical Appendix D) and the computation of the initial configuration.

Some computations, especially those linked with transformations, are executed per partition. A partition consists of a specific part of the data, mainly depending on the conditionality of the unfolding model. A partition therefore consists of a row for the row-conditional model, a matrix for the matrix-conditional model, and a partition consists of the complete data for the unconditional model.

## C.1    PRELIMINARY WORK

The raw preferences $\underline{\Delta}$ are checked for negative values. If negative values exist, these preferences, and the accompanying preference weights $\underline{W}$, are set to zero. If negative weights exist, these weights, and the accompanying preferences values, are also set to zero. When the values for either preferences or weights are missing (user-specified missing values), the respective preferences and weights are set to zero. Setting a preference weight to zero excludes the preference from subsequent calculations.

The preferences should correspond to the distances such that high preference values correspond to large distances and vice versa. Several initial transformations, replacing the original data, exist to establish this relationship. For example, for two-way models, for the exponential decay function $\delta_{ij} =$

$-\log \delta_{ij}$ (Shepard, 1957), for the Gaussian decay function $\delta_{ij} = \sqrt{-\log \delta_{ij}}$, or $\delta_{ij} = \sqrt{-2\log(f_{ij}/\max(f_{ij}))}$ and $\delta_{ij} = f_{i+}f_{+j}/f_{++}f_{ij}$ for frequencies, $\delta_{ij} = 1/(1+s_{ij})$ for similarities, and $\delta_{ij} = \sqrt{2(1-r_{ij})}$ for correlations, but only a similarity transformation is offered internally: When the preferences are similarities (s), the scale of the preferences is reversed, keeping both range and endpoints, i.e., $\delta_{ijk} = c_{ik} - s_{ijk}$. Depending on the conditionality of the model, the value for $c_{ik}$ is computed per partition as

$$
c_{ik} = \begin{cases} \mathrm{argmax}\left(\underline{\Delta}\right) + \mathrm{argmin}\left(\underline{\Delta}\right) & \text{if model = unconditional,} \\ \mathrm{argmax}\left(\Delta_k\right) + \mathrm{argmin}\left(\Delta_k\right) & \text{if model = matrix-conditional,} \\ \mathrm{argmax}\left(\delta_{ik}\right) + \mathrm{argmin}\left(\delta_{ik}\right) & \text{if model = row-conditional.} \end{cases}
$$

Note that an initial transformation differs from an optimal transformation as an initial transformation, specified by the user, is performed only once, at the pre-processing step, while an optimal transformation, although the type of transformation is also specified by the user, is repeatedly optimized during the minimization process carried out by the ALS-IM algorithm.

The scaling constant $\upsilon^2(\delta)$ that links $\omega$ with $\omega^*$ (see page 147) is computed per partition and remains constant for the rest of the process. Since the PENALIZED STRESS function is unable to cope with no variation, the algorithm terminates when $\upsilon(\delta) = 0$ for one or more partitions. This difficulty might be circumvented by removing the unwilling partitions, in this case rows, with the following SPSS syntax.

```
———————————————————————————— Code Start ————————————————————————————
1 COMPUTE novariation = SD(column_1,column_2,column_j,column_m) = 0.
2 COMPUTE removerow = (novariation = 0).
3 FILTER BY removerow.
4 EXECUTE.
———————————————————————————— Code End ————————————————————————————
```

For very low values of the variation coefficient $\upsilon(\delta)$, a warning is issued.

The weights are checked for rows or columns that add up to zero, since these rows or columns are clearly unavailable for coordinate estimation. Row and column sums of the preference weights are gathered in diagonal arrays for each $k$ as

$$
\mathbf{R}_k = \mathrm{diag}\left(\mathbf{W}_k \mathbf{1}\right), \ \mathbf{C}_k = \mathrm{diag}\left(\mathbf{1}'\mathbf{W}_k\right),
$$

and all weight related arrays are summed over the third way as

$$
\mathbf{W} = \sum_{k=1}^{s} \mathbf{W}_k, \ \mathbf{R} = \sum_{k=1}^{s} \mathbf{R}_k, \ \text{and} \ \mathbf{C} = \sum_{k=1}^{s} \mathbf{C}_k.
$$

For the initialization of the transformations, the number of non-missing elements and, analogously, the number and size of the tie-blocks are determined.

The preferences are sorted and an index array is kept for further reference. When the type of transformation is ordinal or smooth ordinal, the initially transformed preferences $\gamma$ are set to their respective rank numbers. For monotone spline transformation(s), which are essentially piecewise polynomials, the number of unique elements in each partition is determined. This number must be larger than the sum of the number of interior knots and the spline order, which define the degree and range of the polynomial pieces. The knot sequence is created and with this sequence the spline basis $S$ is computed (Ramsay, 1988). Note that at this point, the raw preferences $\delta$ are 'transported' to the transformed preferences $\gamma$, initially transformed or not, and that computations from here on use the transformed preferences $\gamma$. The raw preferences $\delta$ are kept for further reference and for the computation of the final results.

Before computing the initial configuration, and for the three-way models only, weighted average preferences and average weights are computed as

$$\gamma_{ij} = \frac{\frac{1}{s}\sum_{k=1}^{s} w_{ijk}\gamma_{ijk}}{\frac{1}{s}\sum_{k=1}^{s} w_{ijk}}, \text{ and } w_{ij} = \frac{1}{s}\sum_{k=1}^{s} w_{ijk},$$

since a two-way (rectangular) matrix suffices for the computation of the initial configuration. When it is not possible to determine a valid value for $\gamma_{ij}$, i.e., when $w_{ij} = 0$, $\gamma_{ij}$ is considered a missing value and is replaced (imputed) with

$$\gamma_{ij} = \begin{cases} c_j & \text{if initial imputation = column mean,} \\ r_i & \text{if initial imputation = row mean,} \\ c_j + r_i - t & \text{if initial imputation = total mean,} \end{cases}$$

where

$$c_j = \left(\sum_{i=1}^{n} w_{ij}\right)^{-1} \sum_{i=1}^{n} w_{ij}\gamma_{ij},$$

$$r_i = \left(\sum_{j=1}^{m} w_{ij}\right)^{-1} \sum_{j=1}^{m} w_{ij}\gamma_{ij}, \text{ and}$$

$$t = \left(\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}\right)^{-1} \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}\gamma_{ij}$$

are the weighted column mean, the weighted row mean, and the weighted total mean, respectively.

As might be deduced from the above calculations, for the pre-processing sofar, restrictions have not been taken into account, that is, the model is assumed to be the identity model and both fixed coordinates and independent

variables are assumed to be nonexistent. Nevertheless, the variables are initially transformed just like the preferences, depending on the user-specified transformations, and when there are more dimensions than variables, i.e., $p > h$, $l = p - h$ dummy variables are created such that for $l = h + 1, \ldots, p$

$$
q_{li} = \begin{cases} (l-h)/n - 1 & \text{for } i = 1, \ldots, l-h \\ (l-h)/n & \text{for } i = l-h+1, \ldots, n, \end{cases}
$$

which gives the first 3 dummy variables as

$$
\begin{bmatrix} \frac{1}{n} - 1 & \frac{2}{n} - 1 & \frac{3}{n} - 1 & \cdots \\ \frac{1}{n} & \frac{2}{n} - 1 & \frac{3}{n} - 1 & \cdots \\ \frac{1}{n} & \frac{2}{n} & \frac{3}{n} - 1 & \cdots \\ \frac{1}{n} & \frac{2}{n} & \frac{3}{n} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.
$$

All variables are centered and normalized, except when the variable in question is not transformed during iterations (transformation = none).

### C.2   INITIAL CONFIGURATION

There are several ways to compute the initial configuration. The two classes of rational ways, besides a user-provided configuration and a random start, are based on either the rectangular preference matrix or the augmented square symmetric super-matrix. After reading, sampling, or computing the initial configuration, the distances are optimally dilated to fit the preferences, since these quantities might be of a quite different scale, and all coordinates are adapted accordingly. In case of restrictions, the initial configuration is adapted to follow the restriction requirements.

*User-provided configuration*

The user is allowed to provide an initial configuration, either for the coordinates of the row objects, for the coordinates of the column objects, or for the coordinates of both sets of objects. If the coordinates of the row objects are omitted, an initial row objects configuration is determined by

$$
\mathbf{X}_o = -0.5\left(\mathbf{\Gamma}^2 - \mathbf{1}\mathbf{a}'\right)\mathbf{Y}_o\left(\mathbf{Y}_o'\mathbf{Y}_o\right)^{-1}, \tag{C.1}
$$

where $\mathbf{\Gamma}^2$ is equal to $\mathbf{\Gamma}$ with all elements squared, $\mathbf{a}$ is the vector with the row sums of $\mathbf{Y}_o\mathbf{Y}_o'$, and $\mathbf{1}$ a unit vector of appropriate length. In the opposite case, the initial column objects configuration is determined by

$$
\mathbf{Y}_o = -0.5\left(\mathbf{\Gamma}^2 - \mathbf{b}\mathbf{1}'\right)'\mathbf{X}_o\left(\mathbf{X}_o'\mathbf{X}_o\right)^{-1},
$$

where $\mathbf{b}$ is the vector with row sums of $\mathbf{X}_o\mathbf{X}_o'$.

*(multiple) Random configuration*

The random start option in PREFSCAL is provided for the use with multiple starts. Coordinates are drawn from an independent standard normal distribution. The Polar Box-Muller method is used to obtain standard normal deviates (see, for example, Dagpunar, 1988), with uniform pseudo-random numbers from the Mersenne Twister (Matsumoto & Nishimura, 1998).

*Rational configuration based on rectangular data*

*Ross and Cliff (1964).* The Ross-Cliff start is based on the article of Ross and Cliff (1964), but similar derivations can also be found in Schönemann (1970), Carroll (1980), Heiser (1981), and Greenacre and Browne (1986). First, the preferences are squared, double centered, and multiplied with $-0.5$ to obtain

$$\mathbf{B} = -0.5\mathbf{J}\mathbf{\Gamma}^2\mathbf{J}.$$

A singular-value decomposition $\mathbf{B} = \mathbf{P}\mathbf{\Phi}\mathbf{Q}'$ provides left and right singular vectors, $\mathbf{P}$ and $\mathbf{Q}$, respectively, and singular values $\mathbf{\Phi}$. The initial configuration is set as

$$\mathbf{X}_0 = \mathbf{P}$$
$$\mathbf{Y}_0 = \mathbf{Q}\mathbf{\Phi}.$$

Heiser (personal communication, October 24, 2005) suggested to use some additional scaling, such that the variances of both $\mathbf{X}_0$ and $\mathbf{Y}_0$ are equal on the first dimension by letting

$$\mathbf{X}_0 = \mathbf{P}\sqrt{s_1 n}$$
$$\mathbf{Y}_0 = \mathbf{Q}\mathbf{\Phi}^*\sqrt{s_1 m},$$

where $s_1$ is the largest (singular) value in $\mathbf{\Phi}$ and $\mathbf{\Phi}^* = s_1^{-1}\mathbf{\Phi}$. Recent publications (Nakanishi & Cooper, 2003; Kuga & Mayekawa, 2008) provide improvements over procedures suggested by Ross and Cliff (1964) and Schönemann (1970). These improvements are under review for implementation.

*Correspondence analysis.* Details on the Correspondence start can be found in Heiser (1981). The start is based on chi-square distances with row and column means removed and using a symmetrical normalization. A singular-value decomposition $\mathbf{B} = \mathbf{P}\mathbf{\Phi}\mathbf{Q}'$ is computed based on a matrix $\mathbf{B}$ with elements

$$b_{ij} = \frac{\gamma_{ij}^*}{\sqrt{\gamma_{i+}^*\gamma_{+j}^*}} - \frac{\sqrt{\gamma_{i+}^*\gamma_{+j}^*}}{\gamma_{++}^*},$$

where $\gamma_{ij}^* = \max(\boldsymbol{\Gamma}) - \gamma_{ij}$ and $\gamma_{i+}^*$, $\gamma_{+j}^*$, and $\gamma_{++}^*$ are the row totals, column totals, and overall total of $\boldsymbol{\Gamma}^*$, respectively. The initial configuration is set as

$$\mathbf{X}_o = \frac{\boldsymbol{\Phi}^{.5}\mathbf{P}}{\sqrt{\gamma_{i+}^*/\gamma_{++}^*}}$$

$$\mathbf{Y}_o = \frac{\boldsymbol{\Phi}^{.5}\mathbf{Q}}{\sqrt{\gamma_{+j}^*/\gamma_{++}^*}}.$$

Note that the Correspondence start with PREFSCAL without transformations is identical to the IBM SPSS CORRESPONDENCE procedure with the following syntax:

```
 ─────────────────────────── Code Start ───────────────────────────
1 CORRESPONDENCE
2    /TABLE         = ALL (n,m)
3    /MEASURE       = CHISQ
4    /STANDARDIZE   = RCMEAN
5    /DIMENSION     = p
6    /NORMALIZATION = SYMMETRICAL
7    /PRINT         = TABLE RPOINTS CPOINTS.
 ──────────────────────────── Code End ────────────────────────────
```

*Heiser and de Leeuw (1979).*    The centroid restriction was not meant as a restriction on the final configuration, according to Heiser (personal communication, 2005), but merely intended as a useful initial configuration. Suppose the column coordinates are in the centroid of the row coordinates, that is,

$$\mathbf{Y}_o = \mathbf{M}^{-1}\mathbf{E}\mathbf{X}_o,$$

where $\mathbf{E}$ is a $n \times m$ indicator matrix of first, second, third, etc choices and $\mathbf{M}$ is a $m \times m$ matrix containing the marginal frequencies of $\mathbf{E}$, i.e., $\mathbf{M} = \text{diag}(\mathbf{1E})$. The maximum number of choices in $\mathbf{E}$ is restricted to $n$ and the columns of $\mathbf{E}$ contain $1$ for a first, second, third, etc choice and $0$ otherwise. The number of choices are provided by the user. Using the classical scaling decomposition of the distances, the initial column coordinates $\mathbf{Y}_o$ can be found by an eigenvalue decomposition of $0.5\mathbf{M}^{-1}\mathbf{E}'\boldsymbol{\Gamma}^2 + 0.5(\mathbf{M}^{-1}\mathbf{E}'\boldsymbol{\Gamma}^2)'$. Coordinates for the row objects are estimated by least squares as in (C.1). Details on the Centroid start can be found in Heiser and de Leeuw (1979a).

*Rational configuration based on square symmetric data*

Classical scaling can be used to estimate coordinates from a square symmetric distance matrix. Suppose a distance matrix $\mathbf{D}$ is given, containing the distances between $n + m$ objects, in a $p$-dimensional Euclidean space. The squared distances are then given by

$$\mathbf{D}^2 = \mathbf{c1}' + \mathbf{1c}' - 2\mathbf{ZZ}',$$

where $\mathbf{c}$ are the diagonal elements of $\mathbf{ZZ}'$ and $\mathbf{Z}$ contains the coordinate values of the $n + m$ objects. Since distances are preserved under translation, we assume that $\mathbf{Z}$ has column means equal to zero. Double centering of $\mathbf{D}^2$ and multiplying by $-0.5$ gives

$$
\begin{aligned}
-0.5\mathbf{J}\mathbf{D}^2\mathbf{J} &= -0.5\mathbf{J}\big(\mathbf{c}\mathbf{1}' + \mathbf{1}\mathbf{c}' - 2\mathbf{ZZ}'\big)\mathbf{J} \\
&= -0.5\mathbf{J}\mathbf{c}\mathbf{1}'\mathbf{J} - 0.5\mathbf{J}\mathbf{1}\mathbf{c}'\mathbf{J} + 0.5\mathbf{J}\big(2\mathbf{ZZ}'\big)\mathbf{J} \\
&= -0.5\mathbf{J}\mathbf{c}\mathbf{0} - 0.5\mathbf{0}\mathbf{c}'\mathbf{J} + \mathbf{J}\mathbf{ZZ}'\mathbf{J} \\
&= \mathbf{ZZ}',
\end{aligned}
$$

since centering a vector of ones yields a vector of zeros and centering of $\mathbf{ZZ}'$ can be ignored due to the assumed column centering of $\mathbf{Z}$. Now, the eigenvalue decomposition $\mathbf{ZZ}' = \mathbf{V}\mathbf{\Lambda}\mathbf{V}' = (\mathbf{V}\mathbf{\Lambda}^{1/2})(\mathbf{V}\mathbf{\Lambda}^{1/2})'$ gives the coordinate values $\mathbf{Z} = \mathbf{V}\mathbf{\Lambda}^{1/2}$, where only the first $p$ ordered eigenvalues ($\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_{n+m}$) and eigenvectors are used. Because $\mathbf{ZZ}'$ is a symmetric, positive semi-definite matrix of rank $p$, it has $p$ non-negative eigenvalues. The $n + m - p$ zero or negative eigenvalues are ignored as error.

This technique, classical scaling, although initiated by G. Young and Householder (1938) and based on the Eckart-Young theorem (Eckart and Young, first issue of Psychometrika, 1936) is due to Torgerson (1952, 1958) and Gower (1966) and also known under the names Torgerson scaling, Torgerson-Gower scaling, Principal Coordinate Analysis, or even YoHoToGo scaling. Using classical scaling for unfolding requires, however, the estimation of the intra-set distances for both sets of objects. The resulting $(n+m) \times (n+m)$ super-matrix,

$$
\mathbf{\Gamma}^+ = \begin{bmatrix} \mathbf{\Gamma}^r & \mathbf{\Gamma}' \\ \mathbf{\Gamma} & \mathbf{\Gamma}^c \end{bmatrix},
$$

with row intra-set distances $\mathbf{\Gamma}^r$ and column intra-set distances $\mathbf{\Gamma}^c$, is decomposed to obtain initial coordinates for the row objects (first $n$ rows of $\mathbf{Z}$) and the column objects (last $m$ rows of $\mathbf{Z}$) (cf. Lingoes, 1970 as cited in Heiser, 1981).

It must be noted, however, that the data are mostly considered comparative distances, distances with an undetermined true zero point, i.e., a distance minus an unknown constant. Following Torgerson (1952), this constant is practically estimated with the 'triple equality' procedure from Carroll and Wish (1974), based upon Torgerson (1958), as

$$
c_0 = \max(\gamma_{ik}^+ - \gamma_{ij}^+ - \gamma_{jk}^+) \; \forall \, i, j, k,
$$

and also known as the triple equality difference (TED) test (see Coxon, 1982, p. 128). Suggestions concerning the additive constant problem by Saito (1978) or Cailliez (1983) have not been taken into consideration yet.

*Heiser and de Leeuw (1979).* For the Triangle start, the triangle inequality is used to compute the intra-set values. The triangle inequality states that the distance between two objects, given the distances of both objects with a third object, is bounded between the sum of the distances with the third object and the absolute difference between these two distances, i.e., the distance between objects i and j satisfies

$$\left| d_{ik} - d_{jk} \right| \leqslant d_{ij} \leqslant \left( d_{ik} + d_{jk} \right).$$

This is true for all objects k. Indeed, for multiple third objects ($k = 1, \ldots, K$), the distance between object i and object j lies in the intersection of K intervals:

$$\max_{k=1}^{K} \left| d_{ik} - d_{jk} \right| \leqslant d_{ij} \leqslant \min_{k=1}^{K} \left( d_{ik} + d_{jk} \right),$$

with equality if there is one object on the line passing through objects i and j. With unequal boundaries the distance between object i and object j lies in between the least upper bound and the greatest lower bound. Heiser and de Leeuw (1979a) suggests to use the midpoint of the least upper bound and the greatest lower bound, derived from the data, as

$$\gamma_{ij} = \frac{1}{2} \left( \max_{k=1}^{K} \left| \gamma_{ik} - \gamma_{jk} \right| + \min_{k=1}^{K} \left( \gamma_{ik} + \gamma_{jk} \right) \right),$$

which coincides with the average of the first rows (Level 1) of S (sorted sum) and D (sorted absolute difference) as given by Rabinowitz (1976, Table 1). The actual procedure of Rabinowitz is on the implementation nomination.

*Van Deun, Heiser, and Delbeke (2007).* The Spearman start computes the co-ordinates of both sets by classical scaling through the imputation of spearman distances between the row objects and the use of the preference sphere for the column objects. Details can be found in van Deun et al. (2007), although the current start deviates in the final stage from the procedure described in van Deun, Marchal, Heiser, Engelen, and van Mechelen (2008). The current procedure is given by the following steps. Add m additional rows $\Gamma^c$ to the preference data $\Gamma$ that represent the objects by tied ranks as

$$\gamma_{ij}^c = \begin{cases} \gamma_{ij}^c = 1 & \text{if } i = j, \\ \gamma_{ij}^c = 1 + m/2 & \text{if } i \neq j \end{cases}$$

and center the data with respect to $\mathbf{c} = [(m+1)/2]\mathbf{1}$. Now, equalize the norm for each row by dividing the row elements with $m^{-1}(\sum_j \gamma_{ij}^{c\,2})^{.5}$ and double center the result, which is now a $(n+m) \times m$ data matrix $[\Gamma', \Gamma^{c\,\prime}]'$. The result

of the singular-value decomposition $[\mathbf{\Gamma}', \mathbf{\Gamma}^{c\,'}]' = \mathbf{P\Phi Q}'$ is used to obtain first estimates of the coordinates for both sets of objects as

$$\left[\mathbf{X}_o', \mathbf{Y}_o'\right]' = \mathbf{P\Phi}^2,$$

where $\mathbf{\Phi}^2$ is a matrix with the squared elements of $\mathbf{\Phi}$. Finally, the configuration is centered on the column coordinates $\mathbf{Y}_o$, the norm for the row object coordinates is set to a maximum of 1, and the coordinates of the two sets are intermixed, based on the median distances from the origin, that is, the column object coordinates are scaled by multiplying with

$$s = \frac{\operatorname{med}\left(d(\mathbf{0}, \mathbf{X}_o)\right)}{\operatorname{med}\left(d(\mathbf{0}, \mathbf{Y}_o)\right)},$$

where $\operatorname{med}(\cdot)$ returns the median of its argument and $d(\mathbf{0}, \mathbf{X})$ and $d(\mathbf{0}, \mathbf{Y})$ are the distances from the origin, for the row and column coordinates, respectively.

One of the initial configurations that is still missing in PREFSCAL is the start discussed by Lingoes and Roskam (1973) and Guttman (1968), which is implemented in MINISSA. This start is on the implementation list for PREFSCAL.

*Restrictions*

*Fixed coordinates.*    Except for a random start, a Procrustes analysis is performed (named after Procrustes, the Greek innkeeper in Attica who always managed to fit his guests into his one-size beds by cutting or stretching their legs as necessary; Oreskovich, Klein, & Sutherland, 1991) to adapt the initial coordinates $\mathbf{Z} = [\mathbf{X}', \mathbf{Y}']'$ to the fixed coordinates $\overline{\mathbf{Z}}$. For this purpose, $\mathbf{Z}_f$ and $\overline{\mathbf{Z}}_f$ only contains the $n_f$ objects that are fixed in $\mathbf{Z}$ and $\overline{\mathbf{Z}}$, respectively. The Procrustes analysis computes a rotation matrix $\mathbf{R} = \mathbf{PQ}'$, where $\mathbf{P}$ and $\mathbf{Q}$ are the left and right singular values from the singular-value decomposition $\mathbf{Z}_f' \mathbf{J}\overline{\mathbf{Z}}_f = \mathbf{P\Phi Q}'$ and $\mathbf{J}$ is a centering matrix of appropriate size, a scaling factor $s = \operatorname{tr}(\mathbf{R}'\mathbf{Z}_f'\mathbf{J}\overline{\mathbf{Z}}_f)/\operatorname{tr}(\mathbf{Z}_f'\mathbf{J}\mathbf{Z}_f)$, and a translation vector $\mathbf{t} = \mathbf{1}'\mathbf{R}(\overline{\mathbf{Z}}_f - s\mathbf{Z}_f\mathbf{R})'/(n_f s)$. After applying the Procrustes transformation $\mathbf{Z}^+ = s(\mathbf{Z} - \mathbf{1}'\mathbf{t})\mathbf{R}$, the fixed coordinates $\overline{\mathbf{Z}}$ replace the corresponding coordinates in $\mathbf{Z}^+$ to get a 100% match. The above procedure, due to Gower (1975), requires all coordinates of a point to be present. It is also possible to fix only one coordinate and leaving the other $p - 1$ coordinates free. For this purpose, missing values should be allowed through the use of an intermediate (diagonal) weight matrix. Details on this procedure can be found in Commandeur (1991).

*Independent variables.* The unrestricted configuration is matched as closely as possible with the independent variables. For this purpose, and considering a restricting on the row objects only, the regression coefficients $\mathbf{B}$ are computed as $b_{hp} = (\mathbf{q}_h' \mathbf{W}_h \mathbf{q}_h)^{-1} \mathbf{q}_h' \mathbf{W}_h \mathbf{x}_p$, where $\mathbf{q}_h$ is independent variable $h$, $\mathbf{x}_p$ are the row coordinates for dimension $p$, and $\mathbf{W}_h$ is a diagonal matrix containing zero's and ones to account for missing values in $\mathbf{q}_h$. After normalizing the regression coefficients for identification, such that $p_{max} = \mathbf{b}_h' \mathbf{b}_h$ and adapting the variables accordingly, the restricted initial row configuration is computed as $\mathbf{X} = \mathbf{Q}\mathbf{B}$. The same procedure is applied for independent variables restricting the column coordinates $\mathbf{Y}$.

## TRANSFORMATION UPDATE
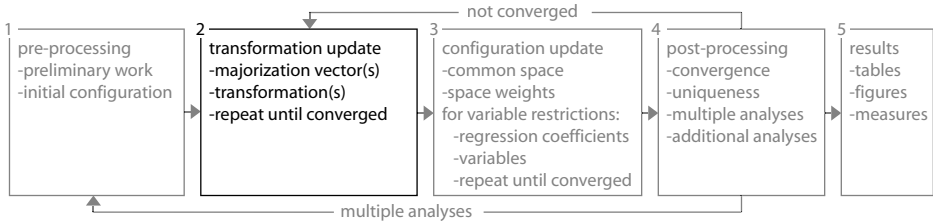


For the transformation update of PENALIZED STRESS ,

$$\sigma_p^2(\boldsymbol{\gamma}, \mathbf{d}) = \left( \frac{\|\boldsymbol{\gamma} - \mathbf{d}\|_W^2}{\|\boldsymbol{\gamma}\|_W^2} \right)^\lambda \left( \frac{\|\boldsymbol{\gamma}\|_V^2 + \omega\|\boldsymbol{\gamma}\|_M^2}{\|\boldsymbol{\gamma}\|_V^2} \right),$$

optimally transforming the preferences according to the conditionality and transformation function supplied, we developed a similar approach as Groenen and Heiser (1996). Given the current estimate of $\mathbf{d}$, we use iterative majorization to obtain a majorizing function in each iteration, that is,

$$c_m\|\boldsymbol{\gamma} - \boldsymbol{\xi}\|_W^2 + c_a, \tag{D.1}$$

which is in the metric $W$ and quadratic in $\boldsymbol{\gamma}$, and where $c_m$, $\boldsymbol{\xi}$ and $c_a$ do not depend on $\boldsymbol{\gamma}$. Function (D.1) can subsequently be minimized using standard least squares transformation routines, such as Kruskal's monotone regression.

### D.1 MAJORIZATION FUNCTIONS

For the theory of iterative majorization in MDS, we refer to de Leeuw (1977a) and Heiser (1995) as basic papers, and to Groenen (1993) and Groenen and Heiser (1996) for some extensions used in this appendix. The iterative majorization approach is briefly summarized as follows. Iterative majorization is a minimization method that uses an auxiliary function $g(a, b)$, the so-called majorizing function, as an aid in finding the minimum of the original function $f(a)$. For this purpose, a majorizing function has to fulfill the following requirements:

$$f(a) \leqslant g(a, b), \text{ for all } a, b \tag{D.2}$$

$$f(a_l) = g(a_l, a_l), \text{ for } l = 0, 1, \dots \tag{D.3}$$
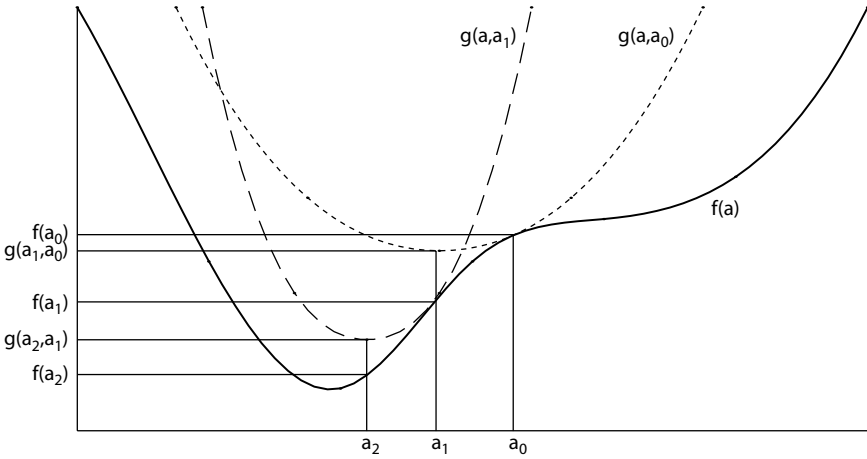
Figure D.1    Iterative majorization, where f(a) is the function to be minimized and $g(a,a_0)$ and $g(a,a_1)$ are majorizing functions that are located on or above the original function f(a).

where $b$ is a known estimate of $a$ and $l$ is an iteration counter. In words, the conditions above mean that the majorizing function always lies above the original function, or touches the function, but is never smaller than the original function (cf. (D.2)) and the majorizing function must touch the original function at the current estimate (cf. (D.3)), the so-called supporting point ($a_1$). Note that a sum of functions can be majorized by using a separate majorizing function for each function in the summation. Repeatedly minimizing consecutive majorizing functions produces a non-increasing sequence of objective function values. If the objective function is bounded from below, then this sequence has a limit. Furthermore, if consecutive function values are equal, that is, if iterative majorization fails to progress, then it has identified a stationary point of the objective function.

*Deriving the unconditional case*

Starting from the PENALIZED STRESS function (B.4) and using the following additional substitutions,

$$f_1 = \|\boldsymbol{\gamma} - \mathbf{d}\|_{\mathbf{W}}^2, \tag{D.4}$$

$$f_2 = \|\boldsymbol{\gamma}\|_{\mathbf{W}}^2, \tag{D.5}$$

$$f_3 = \|\boldsymbol{\gamma}\|_{\mathbf{V}}^2 + \omega\|\boldsymbol{\gamma}\|_{\mathbf{M}}^2, \text{ and} \tag{D.6}$$

$$f_4 = \|\boldsymbol{\gamma}\|_{\mathbf{V}}^2, \tag{D.7}$$

the square root of (B.4) can be written as

$$\frac{f_1^{\lambda/2} f_3^{1/2}}{f_2^{\lambda/2} f_4^{1/2}} = \frac{f_n}{f_d}, \tag{D.8}$$

with numerator $f_n = f_1^{\lambda/2} f_3^{1/2}$ and denominator $f_d = f_2^{\lambda/2} f_4^{1/2}$.

A majorizing function for (D.8) can be found by substitution of separate majorizing functions for (D.4) to (D.7) as long as (D.8) consists of a sum of functions. For this purpose, we first remove the fraction $f_n/f_d$, then split up the product $f_1^{\lambda/2} f_3^{1/2}$, as well as the product $f_2^{\lambda/2} f_4^{1/2}$, and remove $\lambda/2$ as a power in both $f_1^{\lambda/2}$ and $f_2^{\lambda/2}$. Then, substitution of the separate majorizing functions provides the final majorizing function for (B.4).

*Fractional programming the fraction.* Assume that we have $f = f_n/f_d \leqslant g_n/g_d = g$, where, considering $f$ in the current step, $g$ is the function value in the previous step of the iterative majorization algorithm. The reasoning of fractional programming (Dinkelbach, 1967) is that if $f_n/f_d \leqslant g_n/g_d$, then $f_n/f_d - g_n/g_d \leqslant 0$, and thus $f_n - f_d g_n/g_d \leqslant 0$, provided $f_d > 0$, which holds in our case. Applying fractional programming to (D.8) says that minimizing

$$f_1^{\lambda/2} f_3^{1/2} - f_2^{\lambda/2} f_4^{1/2} \frac{g_1^{\lambda/2} g_3^{1/2}}{g_2^{\lambda/2} g_4^{1/2}} = f_1^{\lambda/2} f_3^{1/2} - g f_2^{\lambda/2} f_4^{1/2}, \tag{D.9}$$

also reduces the value of the fraction in (D.8). The auxiliary function (D.9) is the product of two functions, $f_1^{\lambda/2}$ and $f_3^{1/2}$, minus the product of the previous function value $g$ (a constant in minimizing (D.9)) and the product of $f_2^{\lambda/2}$ and $f_4^{1/2}$.

*Majorizing the product.* Following Groenen (1993), the product of the two functions $f_1^{\lambda/2} f_3^{1/2}$ can be majorized as follows. Because the square of a real argument is non-negative, it is true that

$$\left( \frac{f_1^{\lambda/2}}{g_1^{\lambda/2}} - \frac{f_3^{1/2}}{g_3^{1/2}} \right)^2 \geqslant 0, \text{ and thus } f_1^{\lambda/2} f_3^{1/2} \leqslant \frac{1}{2} f_1^{\lambda} \frac{g_3^{1/2}}{g_1^{\lambda/2}} + \frac{1}{2} f_3 \frac{g_1^{\lambda/2}}{g_3^{1/2}}, \quad \text{(D.10)}$$

for $g_1^{\lambda/2} > 0$ and $g_3^{1/2} > 0$, which majorizes the product by a sum of functions, leaving us with only one obstacle: $\lambda$ in $f_1^{\lambda}$.

*Majorizing the λ root.* The first term on the right-hand side of (D.10) is a constant times $f_1^\lambda$. Because $0 < \lambda \leqslant 1$, $f_1^\lambda$ is concave and can be linearly majorized (see Groenen, 1993). Hardy et al. (1952) state that

$$x^\lambda \leqslant (1 - \lambda) + \lambda x,$$

for $0 \leqslant \lambda \leqslant 1$, and substituting $f_1/g_1$ for $x$ gives

$$f_1^\lambda \leqslant (1 - \lambda)g_1^\lambda + \lambda f_1 g_1^{\lambda - 1}, \tag{D.11}$$

which eliminates λ as a power in $f_1^\lambda$.

*Majorizing minus the product.* The second product on the right-hand side of (D.9), $-f_2^{\lambda/2} f_4^{1/2}$, can be majorized as follows (see Groenen, 2002). Because the square of the sum of two differences is non-negative, it is true that

$$\left[\left(f_2^{\lambda/2} - g_2^{\lambda/2}\right) + \left(f_4^{1/2} - g_4^{1/2}\right)\right]^2 \geqslant 0, \text{ and thus}$$

$$-f_2^{\lambda/2} f_4^{1/2} \leqslant \frac{1}{2}\left(f_2^\lambda + g_2^\lambda + f_4 + g_4\right) - \left(f_2^{\lambda/2} + f_4^{1/2}\right)\left(g_2^{\lambda/2} + g_4^{1/2}\right) + g_2^{\lambda/2} g_4^{1/2},$$

which majorizes minus the product by a sum of functions, leaving us with two final obstacles: λ in $f_2^\lambda$ and λ/2 in $-f_2^{\lambda/2}$.

*Majorizing the λ root again.* The majorizing function for $f_2^\lambda$ is identical to (D.11), the majorizing function for $f_1^\lambda$, and thus, for $0 \leqslant \lambda \leqslant 1$,

$$f_2^\lambda \leqslant (1 - \lambda)g_2^\lambda + \lambda f_2 g_2^{\lambda - 1},$$

which eliminates λ as a power in $f_2^\lambda$.

*Majorizing minus the λ root.* Kiers and Groenen (1996) provide the inequality for minus a root as

$$-x^\lambda \leqslant (1 - \lambda)x^2 + (\lambda - 2)x.$$

Replacing $x$ with $f_2^{1/2}/g_2^{1/2}$ gives the majorizing function for $-f_2^{\lambda/2}$ as

$$-f_2^{\lambda/2} \leqslant (1 - \lambda)\frac{f_2}{g_2^{1 - \lambda/2}} + (\lambda - 2)\frac{f_2^{1/2}}{g_2^{1/2 - \lambda/2}}, \tag{D.12}$$

where $0 < \lambda \leqslant 1$ and both $f_2$ and $g_2$ are positive, which is true in our case.

*Combining the majorization results sofar.* Combining (D.9) to (D.12) gives the majorizing function for the square root of (B.4) as a sum of the functions $f_1$, $f_2$, $-f_2^{1/2}$, $f_3$, $f_4$, and $-f_4^{1/2}$, i.e.,

$$\sigma_p(\gamma, d) \leqslant c + \alpha_2 f_1 + \alpha_1 f_3 + (\alpha_5 + \alpha_6)f_2 + \alpha_3 f_4 - \alpha_7 f_2^{1/2} - \alpha_4 f_4^{1/2} \tag{D.13}$$

with $\alpha_1$ to $\alpha_7$ defined as

$$\alpha_1 = (1/2)g_1^{\lambda/2} g_3^{-1/2}$$
$$\alpha_2 = (1/2)\lambda g_3^{1/2} g_1^{\lambda/2-1}$$
$$\alpha_3 = (1/2)g$$
$$\alpha_4 = g\left(g_2^{\lambda/2} + g_4^{1/2}\right)$$
$$\alpha_5 = g(1/2)\lambda g_2^{\lambda-1}$$
$$\alpha_6 = g\left(g_2^{\lambda/2} + g_4^{1/2}\right)(1-\lambda)g_2^{\lambda/2-1}$$
$$\alpha_7 = g\left(g_2^{\lambda/2} + g_4^{1/2}\right)(2-\lambda)g_2^{\lambda/2-1/2}$$

and

$$c = \frac{1}{2}(1-\lambda)g_1^{\lambda/2} g_3^{1/2} + g g_2^{\lambda/2} g_4^{1/2} + \frac{1}{2}g g_2^{\lambda} + \frac{1}{2}g g_4 + \frac{1}{2}(1-\lambda)g g_2^{\lambda}$$
$$= \frac{1}{2}(1-\lambda)g_1^{\lambda/2} g_3^{1/2} + \frac{1}{2}g\left(2 g_2^{\lambda/2} g_4^{1/2} + g_4 + (2-\lambda)g_2^{\lambda}\right),$$

which are all non-negative and constant in minimizing $\sigma_p(\gamma, d)$, and now only separate majorizing functions are needed for $f_1$, $f_2$, $-f_2^{1/2}$, $f_3$, $f_4$, and $-f_4^{1/2}$. In order to provide a majorizing function equivalent to (D.1), terms in the separate majorizing functions must be linear or quadratic in $\gamma$, all in the metric $W$, so that substitution in (D.13) will provide (D.1).

*Majorization of $f_1$.* Expanding $f_1$ shows that

$$\|\gamma - d\|_W^2 = \gamma' W \gamma - 2\gamma' W d + d' W d$$
$$= \beta_1 \gamma' W \gamma - 2\gamma' W b_1 + \beta_2, \tag{D.14}$$

where $\beta_1 = 1$, $b_1 = d$, and $\beta_2 = d' W d$. Equation (D.14) is both linear and quadratic in $\gamma$, in the metric $W$.

*Majorization of $f_2$.* Expanding $f_2$ gives

$$\|\gamma\|_W^2 = \beta_3 \gamma' W \gamma, \tag{D.15}$$

which is quadratic in $\gamma$, in the metric $W$, with $\beta_3 = 1$.

*Majorization of* $-f_2^{1/2}$. For the majorization of $-f_2^{1/2}$, it follows from the Cauchy-Schwarz inequality that

$$-\|\boldsymbol{\gamma}\|_W \leqslant -\frac{\boldsymbol{\gamma}'W\dot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}}\|_W} = -2\boldsymbol{\gamma}'W\mathbf{b}_2, \qquad (D.16)$$

with $\mathbf{b}_2 = (1/2)\|\dot{\boldsymbol{\gamma}}\|_W^{-1}\dot{\boldsymbol{\gamma}}$ and where $\dot{\boldsymbol{\gamma}}$ is the previous update of $\boldsymbol{\gamma}$.

*Majorization of* $f_3$. Expanding $f_3$ gives

$$\|\boldsymbol{\gamma}\|_V^2 + \omega\|\boldsymbol{\gamma}\|_M^2 = \boldsymbol{\gamma}'W\boldsymbol{\gamma} - \boldsymbol{\gamma}'M\boldsymbol{\gamma} + \omega\boldsymbol{\gamma}'M\boldsymbol{\gamma}. \qquad (D.17)$$

Because $M$ is a positive semi-definite matrix, it is true that $(\boldsymbol{\gamma}-\dot{\boldsymbol{\gamma}})'M(\boldsymbol{\gamma}-\dot{\boldsymbol{\gamma}}) \geqslant 0$ and the linear majorizing function of a minus quadratic function is written as

$$-\boldsymbol{\gamma}'M\boldsymbol{\gamma} \leqslant -2\boldsymbol{\gamma}'M\dot{\boldsymbol{\gamma}} + \dot{\boldsymbol{\gamma}}'M\dot{\boldsymbol{\gamma}}. \qquad (D.18)$$

For the last part of (D.17) quadratic majorization is used. Because $V = W-M$ is also a positive semi-definite matrix, it is also true that $(\boldsymbol{\gamma}-\dot{\boldsymbol{\gamma}})'V(\boldsymbol{\gamma}-\dot{\boldsymbol{\gamma}}) \geqslant 0$, and it follows that,

$$\boldsymbol{\gamma}'V\boldsymbol{\gamma} - 2\boldsymbol{\gamma}'V\dot{\boldsymbol{\gamma}} + \dot{\boldsymbol{\gamma}}'V\dot{\boldsymbol{\gamma}} \geqslant 0,$$
$$\boldsymbol{\gamma}'W\boldsymbol{\gamma} - \boldsymbol{\gamma}'M\boldsymbol{\gamma} - 2\boldsymbol{\gamma}'W\dot{\boldsymbol{\gamma}} + 2\boldsymbol{\gamma}'M\dot{\boldsymbol{\gamma}} + \dot{\boldsymbol{\gamma}}'W\dot{\boldsymbol{\gamma}} - \dot{\boldsymbol{\gamma}}'M\dot{\boldsymbol{\gamma}} \geqslant 0, \text{ and}$$
$$\boldsymbol{\gamma}'M\boldsymbol{\gamma} \leqslant \boldsymbol{\gamma}'W\boldsymbol{\gamma} - 2\boldsymbol{\gamma}'W\dot{\boldsymbol{\gamma}} + 2\boldsymbol{\gamma}'M\dot{\boldsymbol{\gamma}} + \dot{\boldsymbol{\gamma}}'W\dot{\boldsymbol{\gamma}} - \dot{\boldsymbol{\gamma}}'M\dot{\boldsymbol{\gamma}}. \qquad (D.19)$$

However, $2\boldsymbol{\gamma}'M\dot{\boldsymbol{\gamma}} = 2w_{++}^{-1}\boldsymbol{\gamma}'ww'\boldsymbol{\gamma}$ in (D.19) might lead to negative values for $\xi$. This is an undesired situation. To solve this problem, we need the following lemma.

**Lemma 2** *The quadratic majorizing function for* $\boldsymbol{\gamma}'W\mathbf{1}$ *is given by*

$$\boldsymbol{\gamma}'W\mathbf{1} \leqslant \tau_1\boldsymbol{\gamma}'W\boldsymbol{\gamma} - 2\boldsymbol{\gamma}'W\mathbf{t}_1 + \tau_2, \qquad (D.20)$$

*where* $\tau_1 = \min\{(2\gamma_{min})^{-1}, \epsilon^{-1}\}$ *with* $\gamma_{min} = \operatorname{argmin}_i(\dot{\boldsymbol{\gamma}})$ *and* $\epsilon$ *some small number,* $\mathbf{t}_1 = \tau_1\dot{\boldsymbol{\gamma}} - \frac{1}{2}\mathbf{1}$, *and* $\tau_2 = -\tau_1\dot{\boldsymbol{\gamma}}'W\dot{\boldsymbol{\gamma}} + 2\mathbf{t}_1'W\dot{\boldsymbol{\gamma}} + \dot{\boldsymbol{\gamma}}'w$.

*Proof.* The requirements (D.2) and (D.3) must hold for (D.20) to be a proper majorizing function. Substitution of $\mathbf{t}_1$ and $\tau_2$ in (D.20) shows that $\tau_1(\boldsymbol{\gamma} - \dot{\boldsymbol{\gamma}})'W(\boldsymbol{\gamma} - \dot{\boldsymbol{\gamma}}) \geqslant 0$, because $W$ is a positive semi-definite matrix and $\tau_1$ is a positive scalar. Equality occurs if $\boldsymbol{\gamma} = \dot{\boldsymbol{\gamma}}$, which proves (D.3). $\qquad\square$

The term $2\gamma'W\dot{\gamma}$ can be rewritten as $2\tau_3\gamma'W1$ with $\tau_3 = w_{++}^{-1}w'\dot{\gamma}$. Using Lemma 2, and substitution of the majorization results (D.18) to (D.20) in (D.17) gives

$$\|\gamma\|_V^2 + \omega\|\gamma\|_M^2 \leqslant \beta_4\gamma'W\gamma - 2\gamma'Wb_3 + \beta_5, \qquad (D.21)$$

where $\beta_4 = 1 + \omega + 2\omega\tau_3\tau_1$, $b_3 = \tau_3 1 + \omega\dot{\gamma} + 2\omega\tau_3 t_1$, and $\beta_5 = (1 - \omega)w_{++}^{-1}\dot{\gamma}'ww'\dot{\gamma} + 2\omega\tau_3\tau_2 + \omega\dot{\gamma}'W\dot{\gamma}$, providing the majorizing function for $f_3$.

*Majorization of* $f_4$.    Expanding $f_4$ gives

$$\|\gamma\|_V^2 = \gamma'W\gamma - \gamma'M\gamma.$$

Using the same substitutions as for the majorization of $f_3$, provides the majorizing function for $f_4$ as

$$\|\gamma\|_V^2 \leqslant \beta_6\gamma'W\gamma - 2\gamma'Wb_4 + \beta_7, \qquad (D.22)$$

with $\beta_6 = 1$, $b_4 = \tau_3 1$, and $\beta_7 = \dot{\gamma}'M\dot{\gamma}$.

*Majorization of* $-f_4^{1/2}$.    For the majorization of $-f_4^{1/2}$, it follows again from the Cauchy-Schwarz inequality that

$$\begin{aligned}
-\|\gamma\|_V &\leqslant -\frac{\gamma'V\dot{\gamma}}{\|\dot{\gamma}\|_V} \\
&= \frac{-\gamma'W\dot{\gamma} + \gamma'M\dot{\gamma}}{\|\dot{\gamma}\|_V} \\
&= -2\gamma'Wt_2 + \tau_4\gamma'W1, \qquad (D.23)
\end{aligned}$$

where $t_2 = (2\|\dot{\gamma}\|_V)^{-1}\dot{\gamma}$ and $\tau_4 = (\|\dot{\gamma}\|_V)^{-1}w_{++}^{-1}w'\dot{\gamma}$. Although (D.23) is a valid majorizing function for $-\|\gamma\|_V$, $\gamma'M\dot{\gamma}$ may again lead to negative values for $\xi$. Lemma 2 is used to avoid these negative values. Combining (D.20) and (D.23) gives

$$-\|\gamma\|_V \leqslant \beta_8\gamma'W\gamma - 2\gamma'Wb_5 + \beta_9, \qquad (D.24)$$

where $\beta_8 = \tau_4\tau_1$, $b_5 = t_2 + \tau_4 t_1$, and $\beta_9 = \tau_4\tau_2$, providing the majorizing function for $-f_4^{1/2}$.

*Combining majorization results.* The majorizing functions for $f_1$, $f_2$, $-f_2^{1/2}$, $f_3$, $f_4$, and $-f_4^{1/2}$, all containing terms linear and quadratic in $\gamma$, can now be combined to provide the overall majorizing function for PENALIZED STRESS. Substitution of (D.14), (D.15), (D.16), (D.21), (D.22), and (D.24) in (D.13) gives the majorizing function for PENALIZED STRESS as

$$\sigma_p(\gamma, d) \leqslant c + \alpha_2(\beta_1 \gamma' W \gamma - 2\gamma' W b_1 + \beta_2)$$
$$+ \alpha_1(\beta_4 \gamma' W \gamma - 2\gamma' W b_3 + \beta_5)$$
$$+ (\alpha_5 + \alpha_6)(\beta_3 \gamma' W \gamma)$$
$$+ \alpha_3(\beta_6 \gamma' W \gamma - 2\gamma' W b_4 + \beta_7)$$
$$+ \alpha_7(-2\gamma' W b_2)$$
$$+ \alpha_4(\beta_8 \gamma' W \gamma - 2\gamma' W b_5 + \beta_9).$$

This leads to the following expression for $\xi$ in (D.1),

$$\xi = c_m^{-1}(\alpha_2 b_1 + \alpha_1 b_3 + \alpha_3 b_4 + \alpha_7 b_2 + \alpha_4 b_5),$$

where $c_m = \alpha_2 \beta_1 + \alpha_1 \beta_4 + (\alpha_5 + \alpha_6)\beta_3 + \alpha_3 \beta_6 + \alpha_4 \beta_8$. As indicated before, $\xi$ is a vector with non-negative values and $\xi$ can be used in standard minimization routines to find optimally transformed values of $\gamma$.

*Deriving the conditional case*

The variation of the transformed preferences in row $i$ is denoted as $\upsilon(\gamma_i)$, yielding the penalty function

$$\mu_c(\Gamma) = 1 + \omega^* \frac{1}{\left(\frac{1}{n}\sum_{i=1}^n \upsilon^2(\delta_i)\upsilon^{-2}(\gamma_i)\right)^{-1}} \tag{D.25}$$

for the conditional case. The row-conditional PENALIZED STRESS function can, considering (D.25), be written as

$$\sigma_p(\gamma, d) = \left[\left(\frac{1}{n}\sum_{i=1}^n \frac{\|\gamma_i - d_i\|_{W_i}^2}{\|\gamma_i\|_{W_i}^2}\right)^\lambda \left(\frac{1}{n}\sum_{i=1}^n \frac{\|\gamma_i\|_{V_i}^2 + \omega\|\gamma_i\|_{M_i}^2}{\|\gamma_i\|_{V_i}^2}\right)\right]^{1/2}, \tag{D.26}$$

where the subscript $i$ indicates that it only concerns elements corresponding to row $i$.

*Block relaxation.* For the minimization of (D.26) we use block relaxation, which updates one row at the time, while keeping the other rows fixed. Consider one single row $i$, while keeping the other $n-1$ rows fixed. Now, (D.26)

is written as

$$\sigma_p(\gamma_i, d_i) = \left[ \left( \frac{nc_{1i}\|\gamma_i\|_{W_i}^2 + \|\gamma_i - d_i\|_{W_i}^2}{n\|\gamma_i\|_{W_i}^2} \right)^\lambda \right.$$
$$\left. \left( \frac{(1 + nc_{2i})\|\gamma_i\|_{V_i}^2 + \omega\|\gamma_i\|_{M_i}^2}{n\|\gamma_i\|_{V_i}^2} \right) \right]^{1/2},$$

where

$$c_{1i} = \sum_{j \neq i}^n \frac{\|\gamma_j - d_j\|_{W_j}^2}{n\|\gamma_j\|_{W_j}^2} \text{ and } c_{2i} = \sum_{j \neq i}^n \frac{\|\gamma_j\|_{V_j}^2 + \omega\|\gamma_j\|_{M_j}^2}{n\|\gamma_j\|_{V_j}^2}.$$

*Conditional derivations.* For the row-conditional case, the substitutions described in (D.4) to (D.7) are,

$$f_{1i} = nc_{1i}\|\gamma_i\|_{W_i}^2 + \|\gamma_i - d_i\|_{W_i}^2,$$
$$f_{2i} = n\|\gamma_i\|_{W_i}^2,$$
$$f_{3i} = (1 + nc_{2i})\|\gamma_i\|_{V_i}^2 + \omega\|\gamma_i\|_{M_i}^2, \text{ and}$$
$$f_{4i} = n\|\gamma_i\|_{V_i}^2.$$

Since the initial derivations (D.9) to (D.12) are equivalent for the conditional case, only the majorizing functions for $f_{1i}$, $f_{2i}$, $f_{3i}$, and $f_{4i}$ need to be adapted and only in those cases where the constants $c_{1i}$, $c_{2i}$, and $n$ are concerned. For the quadratic part, the following constants are adapted for the conditional case as:

$$\beta_{1i} = 1 + nc_{1i},$$
$$\beta_{3i} = n,$$
$$\beta_{4i} = 1 + nc_{2i} + \omega + 2\omega\tau_3\tau_1,$$
$$\beta_{6i} = n, \text{ and}$$
$$\beta_{8i} = n^{1/2}\tau_4\tau_1,$$

whereas for the linear part, the following vectors are adapted:

$$b_{2i} = (1/2)n^{1/2}\|\dot{\gamma}_i\|_{W_i}^{-1}\dot{\gamma}_i,$$
$$b_{3i} = \tau_3 1 + \tau_3 nc_{2i} 1 + \omega\dot{\gamma}_i + 2\omega\tau_3 t_1,$$
$$b_{4i} = n\tau_3 1, \text{ and}$$
$$b_{5i} = n^{1/2}t_2 + n^{1/2}\tau_4 t_1.$$

Remaining constants and vectors are taken from the unconditional derivations, considering only row $i$.

*Combining majorization results.* Combining the results for the row-conditional case gives the majorizing function for the row-conditional PENALIZED STRESS as

$$
\begin{aligned}
\sigma_p(\boldsymbol{\gamma}_i, \mathbf{d}_i) \leqslant \; & c_i + \alpha_{2i}(\beta_{1i}\boldsymbol{\gamma}_i'\boldsymbol{W}_i\boldsymbol{\gamma}_i - 2\boldsymbol{\gamma}_i'\boldsymbol{W}_i\mathbf{b}_{1i} + \beta_{2i}) \\
& + \alpha_{1i}(\beta_{4i}\boldsymbol{\gamma}_i'\boldsymbol{W}_i\boldsymbol{\gamma}_i - 2\boldsymbol{\gamma}_i'\boldsymbol{W}_i\mathbf{b}_{3i} + \beta_{5i}) \\
& + (\alpha_{5i} + \alpha_{6i})(\beta_{3i}\boldsymbol{\gamma}_i'\boldsymbol{W}_i\boldsymbol{\gamma}_i) \\
& + \alpha_{3i}(\beta_{6i}\boldsymbol{\gamma}_i'\boldsymbol{W}_i\boldsymbol{\gamma}_i - 2\boldsymbol{\gamma}_i'\boldsymbol{W}_i\mathbf{b}_{4i} + \beta_{7i}) \\
& + \alpha_{7i}(-2\boldsymbol{\gamma}_i'\boldsymbol{W}_i\mathbf{b}_{2i}) \\
& + \alpha_{4i}(\beta_{8i}\boldsymbol{\gamma}_i'\boldsymbol{W}_i\boldsymbol{\gamma}_i - 2\boldsymbol{\gamma}_i'\boldsymbol{W}_i\mathbf{b}_{5i} + \beta_{9i}),
\end{aligned}
$$

which provides the following expression for $\boldsymbol{\xi}_i$ in (D.1),

$$
\boldsymbol{\xi}_i = c_{mi}^{-1}\left(\alpha_{2i}\mathbf{b}_{1i} + \alpha_{1i}\mathbf{b}_{3i} + \alpha_{3i}\mathbf{b}_{4i} + \alpha_{7i}\mathbf{b}_{2i} + \alpha_{4i}\mathbf{b}_{5i}\right),
$$

where $c_{mi} = \alpha_{2i}\beta_{1i} + \alpha_{1i}\beta_{4i} + (\alpha_{5i} + \alpha_{6i})\beta_{3i} + \alpha_{3i}\beta_{6i} + \alpha_{4i}\beta_{8i}$. As indicated before, again, $\boldsymbol{\xi}_i$ is a vector with non-negative values and $\boldsymbol{\xi}_i$ can be used in standard minimization routines to find optimally transformed values of $\boldsymbol{\gamma}_i$.

### D.2 TRANSFORMATION FUNCTIONS

Transformations are performed over different parts of the data, depending on the conditionality of the data, as specified by the user. For the discussion of the following transformation functions, it is assumed that the transformation is computed over a single vector, without reference to a row (in case of row-conditional transformations), source (in case of matrix-conditional transformations), or complete data array (in case of unconditional transformations).

*Linear transformations*

The linear transformation is defined as

$$
\boldsymbol{\gamma} = b_1\mathbf{1} + b_2\boldsymbol{\delta} \tag{D.27}
$$

with $b_2 \geqslant 0$ to ensure increasing values of $\boldsymbol{\gamma}$ for increasing values of $\boldsymbol{\delta}$ and $\boldsymbol{\gamma} = b_1\mathbf{1} + b_2\boldsymbol{\delta} \geqslant \mathbf{0}$ for all elements of $\boldsymbol{\gamma}$ to ensure non-negative transformed preferences. By subtracting $\operatorname{argmin}(\boldsymbol{\delta})$ from $\boldsymbol{\delta}$ the latter constraint becomes $b1 \geqslant 0$. This shift by $\operatorname{argmin}(\boldsymbol{\delta})$ is restored afterwards. Now, a non-negative least squares procedure can be used to find estimates for $b_1$ and $b_2$ under the non-negativity restrictions, minimizing (D.1) given $\boldsymbol{\delta}$ and $\boldsymbol{\xi}$. Since only two parameters need to be estimated, a limited number of steps suffice. Expanding

(D.1), substituting (D.27) for $\gamma$, and adding a diagonal matrix with preference weights $W$, gives two equations with two unknowns as

$$b_1 \mathbf{1}'W\mathbf{1} + b_2 \mathbf{1}'W\delta = \mathbf{1}'W\xi, \qquad \text{(D.28)}$$

$$b_1 \delta'W\mathbf{1} + b_2 \delta'W\delta = \delta'W\xi. \qquad \text{(D.29)}$$

Substitution of $b_1$ from (D.28) in (D.29) provides an estimate for $b_2$ as

$$\widehat{b_2} = \frac{\mathbf{1}'W\mathbf{1}\delta'W\xi - \delta'W\mathbf{1}\mathbf{1}'W\xi}{\mathbf{1}'W\mathbf{1}\delta'W\delta - \delta'W\mathbf{1}\mathbf{1}'W\delta}.$$

If $\widehat{b_2}$ is smaller than zero, then $\widehat{b_2}$ is set to zero. Substitution of $\widehat{b_2}$ in (D.28) provides the estimate for $b_1$ as

$$\widehat{b_1} = \frac{\xi'W\mathbf{1} - \widehat{b_2}\delta'W\mathbf{1}}{\mathbf{1}'W\mathbf{1}}.$$

If, however, $\widehat{b_1}$ is smaller than zero, then $\widehat{b_1}$ is set to zero and $b_2$ is re-estimated, using either (D.28) or (D.29), with $b_1 = 0$.

Special cases of the linear transformation (D.27) exist for $b_1 = 0$, a slope-only or ratio transformation, $b_2 = 1$, an intercept-only transformation, and both $b_1 = 0$ and $b_2 = 1$, which implies no transformation, obviously.

*Monotone transformations*

*Monotone splines.*   For the monotone spline transformations (Winsberg & Ramsay, 1983; Ramsay, 1988), smooth nondecreasing piecewise polynomial transformations, additional pre-processing is required, which is described in Technical Appendix C, PRE-PROCESSING, page 153. A weighted regression $\xi = SB$ with spline basis $S$ is computed in each iteration of the main algorithm. The regression coefficients $B$ are restricted to be non-negative and computed using the non-negative least squares procedure of Lawson and Hanson (1974).

Note that monotone spline transformations establish a link between linear transformations and monotone regression (discussed hereafter). Without interior knots, thus only using two boundary knots, and with a linear polynomial, the monotone spline transformation is identical to the linear transformation, estimating an intercept (boundary knot) and a slope (linear polynomial). When, on the other hand, there exists a knot for every preference value, the monotone spline transformation conforms to a monotone regression transformation, that is, to a stepwise transformation with nondecreasing steps (due to the non-negativity constraint on $B$) on each knot.

*Monotone regression.* Kruskal (1964a) discusses monotone regression in his famous article discussing nonmetric MDS. By itself, monotone regression was already known from Ayer, Brunk, Ewing, Reid, and Silverman (1955), Barton and Mallows (1961), Bartholomew (1959, 1961), and Miles (1959). Other references (in the context of MDS) are: Kruskal (1971), proving existence and continuity of the gradient, de Leeuw (1977b), proving correctness in case of ties, and van Waning (1976), finding an optimal computer procedure. A good discussion under the name *isotone regression* can be found in Barlow, Bartholomew, Bremner, and Brunk (1972).

Monotone regression, in our case, finds a vector $\gamma$ that minimizes (D.1) under the inequality restrictions $\gamma_1 \leqslant \gamma_2 \leqslant \ldots \leqslant \gamma_n$ for an ordered vector $\xi$. The index order of $\xi$ is identical to the index order of the ascendingly sorted original preferences $\delta$. In case of ties, there are two approaches implemented: Ties are allowed to be untied or ties must be kept tied. The first (primary) approach to ties uses the entire transformation vector $\xi$ and sorts the vector according to the original preferences $\delta$. The pool-adjacent-violators algorithm (Ayer et al., 1955), or equivalently the up-and-down blocks algorithm (Kruskal, 1964b), now re-establishes the correct monotone order in $\xi$, whereas within tie-blocks the values are set in optimal (least squares) order. The second (secondary) approach to ties requires ties to be kept tied. In this case, the preferences within tie-blocks are first contracted (and expanded afterwards), that is, tie-blocks are represented in the sorted vector by their means, leading to a transformation vector $\xi^*$ that is number-of-tie-blocks large. This vector is, in turn, fed to the up-and-down blocks algorithm to re-establish monotonicity. A third (tertiary) approach is discussed by de Leeuw (1977b).

The master thesis of van Waning (1976) discusses several algorithms for monotone regression in terms of speed and memory requirements. A simple and fast version of the up-and-down blocks algorithm, using a minimum of memory, is used in PREFSCAL.

*Smooth monotone regression.* An alternative algorithm for monotone regression is non-negative least squares. Using an ordered vector $\xi$ and a special coefficient matrix

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix},$$

specifying the constraints (here for four elements only), leads to a monotone regression procedure similar to

$$\min_{\beta \geqslant 0} \| \mathbf{E}\beta - \xi \|_W^2, \tag{D.30}$$

providing the solution as $\mathbf{E}\boldsymbol{\beta}$. With monotonicity as a condition on the first order differences $\theta_l^{(1)} = \xi_l - \xi_{l-1} \geqslant 0$, smoothness is defined as a condition on the second order differences $\theta_l^{(2)} = |\theta_l^{(1)} - \theta_{l-1}^{(1)}| \leqslant \overline{\theta}^{(1)}$, where $\overline{\theta}^{(1)}$ is the *mean step*, which leads to a somewhat more complicated coefficient matrix $\mathbf{E}$, but to the same non-negative least squares minimization procedure. Heiser (1985, 1986, 1989) discusses an alternating least squares algorithm to find a solution for (D.30), but mainly due to the huge number of inequalities, the algorithm is very slow for only a limited number of elements in $\boldsymbol{\xi}$. Nevertheless, smooth monotone regression functions adequately and is able to avoid degenerate solutions.

## CONFIGURATION UPDATE
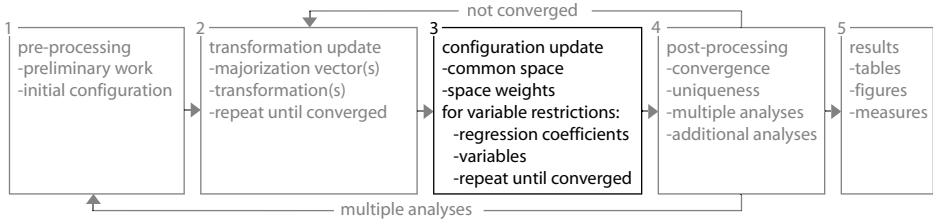


For the configuration update of PENALIZED STRESS ,

$$\sigma_p^2(\boldsymbol{\gamma}, \mathbf{d}) = \left( \frac{\|\boldsymbol{\gamma} - \mathbf{d}\|_W^2}{\|\boldsymbol{\gamma}\|_W^2} \right)^\lambda \left( \frac{\|\boldsymbol{\gamma}\|_V^2 + \omega \|\boldsymbol{\gamma}\|_M^2}{\|\boldsymbol{\gamma}\|_V^2} \right), \qquad \text{(E.1)}$$

we developed an approach first discussed in the context of unfolding by Heiser (1987a). Minimizing PENALIZED STRESS reduces to the minimization of NORMALIZED RAW STRESS , as the penalty function (B.2), i.e., the second part on the right-hand side of (E.1), and the penalty parameters $\lambda$ and $\omega$ are constants with respect to the configuration. Re-defining $W$ as $\|\boldsymbol{\gamma}\|_W^{-2} W$, i.e., dividing the weights with the weighted sum-of-squares of the transformed preferences, yields RAW STRESS

$$\sigma_r^2(\mathbf{d}) = \|\boldsymbol{\gamma} - \mathbf{d}\|_W^2 \qquad \text{(E.2)}$$

as the objective function for the configuration update, which is minimized using iterative majorization (see Section D.1 for the theory of iterative majorization and Heiser, 1987a for unfolding specific derivations) and alternating least squares (see Takane et al., 1977; Greenacre & Browne, 1986; Heiser, 1987a). Depending on the individual differences model and the restrictions on the configuration, updates are determined for the row coordinates $\mathbf{X}$, the column coordinates $\mathbf{Y}$, the space weights $\underline{\mathbf{A}}$, and the independent variables $\mathbf{Q}$ and corresponding regression coefficients $\mathbf{B}$ (see de Leeuw & Heiser, 1980; DeSarbo & Rao, 1984; DeSarbo & Carroll, 1985; DeSarbo & Rao, 1986).

The general strategy to find an update is first to rewrite the function such that only relevant quantities remain. The non-relevant quantities, i.e., variables that are constant in the minimization, are gathered in a variable c, using a subscript indicating the (non-)fixed content. Then, the function is rewritten, enabling us the take the derivative, which is set equal to zero to provide us with the minimum and thus the update of the variable in question.

The basic two-way unfolding loss function (E.2), defined as a weighted least squares function, can be written as

$$
\begin{aligned}
f_2 &= \|\gamma - d\|_W^2 \\
&= c_2 + \mathrm{tr}\left(X'RX + Y'CY - 2X'WY - 2X'\widetilde{X} - 2Y'\widetilde{Y}\right),
\end{aligned}
\qquad (E.3)
$$

where $c_2 = \mathrm{tr}\,\gamma'W\gamma$, $R = \mathrm{diag}\,(W1)$, $C = \mathrm{diag}\,(1'W)$, $\widetilde{X} = PX - BY$, and $\widetilde{Y} = QY - B'X$, where $B$ is defined with elements

$$
b_{ij} = \begin{cases} w_{ij}\gamma_{ij}/d(x_i, y_j) & \text{if } d(x_i, y_j) > 0.0 \\ 0.0 & \text{if } d(x_i, y_j) = 0.0 \end{cases}
$$

and $P = \mathrm{diag}\,(B1)$ and $Q = \mathrm{diag}\,(1'B)$. Manipulation of the Cauchy-Schwarz inequality

$$
d(x_i, y_j)d(\dot{x}_i, \dot{y}_j) \geqslant (x_i - y_j)'(\dot{x}_i - \dot{y}_j),
$$

where $\dot{x}$ and $\dot{y}$ are other instances of $x$ and $y$, respectively, shows that

$$
-\mathrm{tr}\,X'\widetilde{X} - \mathrm{tr}\,Y'\widetilde{Y} \leqslant -\mathrm{tr}\,X'\dot{\widetilde{X}} - \mathrm{tr}\,Y'\dot{\widetilde{Y}},
\qquad (E.4)
$$

where $\dot{\widetilde{X}}$ and $\dot{\widetilde{Y}}$ are the previous updates of $\widetilde{X}$ and $\widetilde{Y}$, respectively. Combining (E.3) and (E.4) provides the majorizing function for $f_2$ as

$$
g_2 = c_2 + \mathrm{tr}\left(X'RX + Y'CY - 2X'WY - 2X'\dot{\widetilde{X}} - 2Y'\dot{\widetilde{Y}}\right).
\qquad (E.5)
$$

After some rearrangement of terms, it is clear that (E.5) can be minimized by alternating least squares, i.e., by alternating between minimizing $g_2$ for $X$ and minimizing $g_2$ for $Y$, keeping the other terms fixed. The solutions for $X$ and $Y$ are then given by

$$
X^+ = R^{-1}\left(\dot{\widetilde{X}} + WY\right) \text{ and}
\qquad (E.6)
$$

$$
Y^+ = C^{-1}\left(\dot{\widetilde{Y}} + W'X^+\right).
\qquad (E.7)
$$

Repeatedly updating $X$ and $Y$ until convergence provides at least a local minimum of the loss function.

In 1972, Carroll introduced additional models which allows for different, individual points of view, such that an individual (row object) is allowed to view the common space in a certain way. These two-way individual differences models are specified as:

$$x^i = x_i A_i,$$
$$Y^i = Y A_i, \tag{E.8}$$

where $x_i$ specifies a row vector with common space coordinates for row object $i$ and $x^i$ specifies a row vector with the individual point of view. The models allow for different weighting and/or rotation of the common space $X, Y$ for each individual. The different *forms* of the space weights matrices $A_i$ result in the following individual differences models:

- $A_i = I \, \forall \, i = 1, \ldots, n$; Identity model, where $I$ is an identity matrix, specifying individual spaces $x^i, Y^i$ identical to the common space $x_i, Y$;

- $A_i = D_i$; Diagonal or weighted Euclidean model, with diagonal matrices $A_i$, resulting in individual dimension weighting of the common space for each individual space;

- $A_i = F_i$; Full or generalized Euclidean model, with full space weight matrices $A_i$, which can be divided into individual rotation and dimension weighting matrices for each individual space;

- $A_i = R_i$; Rectangular or reduced rank model, with rectangular space weight matrices $A_i$, which can be divided into individual rotation, dimension weighting, and dimension reduction matrices for each individual space, where the rank of an individual space is smaller than the rank of the common space.

Incorporating the space weights $\underline{A}$ into the two-way loss function (E.5) gives the general two-way individual differences model as

$$g_2' = c_2 + \frac{1}{n} \sum_{i=1}^{n} \mathrm{tr} \left( A_i' x_i' r_{ii} x_i A_i + A_i' Y' \mathrm{diag}\left(w_i\right) Y A_i - 2 A_i' x_i' w_i Y A_i \right.$$
$$\left. - 2 A_i' x_i' \dot{\tilde{x}}_i - 2 A_i' Y' \dot{\tilde{Y}}_i \right) \tag{E.9}$$

which is simply an average over all $n$ row objects.

Note that unfolding IRT models need the diagonal or weighted Euclidean model *for each column* rather than each row, so that each item can have different diagonal discrimination matrices (J. S. Roberts, personal communication, March 24, 2009).

*Row coordinates update.* (E.9) is minimized with respect to the row coordinates $\mathbf{X}$ for each row $i$ separately, while keeping the other $n-1$ rows fixed. For this purpose, (E.9) is written as

$$g_2' = c_{\mathbf{x}_i} + \operatorname{tr} \mathbf{A}_i\mathbf{A}_i'\mathbf{x}_i'r_{ii}\mathbf{x}_i - 2\operatorname{tr}\mathbf{x}_i'\mathbf{w}_i\mathbf{Y}\mathbf{A}_i\mathbf{A}_i' - 2\operatorname{tr}\mathbf{x}_i'\dot{\tilde{\mathbf{x}}}_i\mathbf{A}_i'. \qquad \text{(E.10)}$$

Using the equalities

$$\begin{aligned}
\operatorname{tr}\mathbf{ABCD} &= \big(\operatorname{vec}\mathbf{D}\big)'\big(\mathbf{A}\otimes\mathbf{C}'\big)\operatorname{vec}\mathbf{B}' \text{ and}\\
\operatorname{tr}\mathbf{A}'\mathbf{B} &= \big(\operatorname{vec}\mathbf{A}\big)'\operatorname{vec}\mathbf{B}
\end{aligned} \qquad \text{(E.11)}$$

(see, for example, Magnus & Neudecker, 1988), (E.10) can be written as

$$\begin{aligned}
g_2' = c_{\mathbf{x}_i} &+ \big(\operatorname{vec}\mathbf{x}_i\big)'\big(\mathbf{A}_i\mathbf{A}_i'\otimes r_{ii}'\big)\operatorname{vec}\mathbf{x}_i'\\
&- 2\big(\operatorname{vec}\mathbf{x}_i\big)'\operatorname{vec}\mathbf{w}_i\mathbf{Y}\mathbf{A}_i\mathbf{A}_i' - 2\big(\operatorname{vec}\mathbf{x}_i\big)'\operatorname{vec}\dot{\tilde{\mathbf{x}}}_i\mathbf{A}_i',
\end{aligned} \qquad \text{(E.12)}$$

which provides the update for $\mathbf{x}_i$ as

$$\operatorname{vec}\mathbf{x}_i^+ = \Big[\mathbf{A}_i\mathbf{A}_i'\otimes r_{ii}'\Big]^{-1}\Big[\operatorname{vec}\mathbf{w}_i\mathbf{Y}\mathbf{A}_i\mathbf{A}_i' + \operatorname{vec}\dot{\tilde{\mathbf{x}}}_i\mathbf{A}_i'\Big],$$

which reduces to (E.6) for the identity model with $\mathbf{A}_i = \mathbf{I}\ \forall\ i = 1,\ldots,n$. Writing vec in front of $\mathbf{x}_i$ is redundant, but left as it is for clarity and for correspondence with matrix arguments.

*Column coordinates update.* The minimization of (E.9) with respect to $\mathbf{Y}$ cannot be performed per row or column. In this case, $g_2'$ is written as

$$\begin{aligned}
g_2' = c_{\mathbf{Y}} &+ \frac{1}{n}\sum_{i=1}^{n}\operatorname{tr}\mathbf{A}_i\mathbf{A}_i'\mathbf{Y}'\operatorname{diag}\big(\mathbf{w}_i\big)\mathbf{Y}\\
&- 2\frac{1}{n}\sum_{i=1}^{n}\operatorname{tr}\mathbf{A}_i\mathbf{A}_i'\mathbf{x}_i'\mathbf{w}_i\mathbf{Y} - 2\frac{1}{n}\sum_{i=1}^{n}\operatorname{tr}\mathbf{Y}'\dot{\tilde{\mathbf{Y}}}_i\mathbf{A}_i'
\end{aligned} \qquad \text{(E.13)}$$

and after using the equalities (E.11) the function becomes

$$\begin{aligned}
g_2' = c_{\mathbf{Y}} &+ \big(\operatorname{vec}\mathbf{Y}\big)'\bigg(\frac{1}{n}\sum_{i=1}^{n}\mathbf{A}_i\mathbf{A}_i'\otimes\operatorname{diag}\mathbf{w}_i\bigg)\operatorname{vec}\mathbf{Y}\\
&- 2\big(\operatorname{vec}\mathbf{Y}\big)'\operatorname{vec}\frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i'\mathbf{x}_i\mathbf{A}_i\mathbf{A}_i' - 2\big(\operatorname{vec}\mathbf{Y}\big)'\operatorname{vec}\frac{1}{n}\sum_{i=1}^{n}\dot{\tilde{\mathbf{Y}}}_i\mathbf{A}_i'. \quad \text{(E.14)}
\end{aligned}$$

The update is then given as

$$\operatorname{vec}\mathbf{Y}^+ = \bigg[\frac{1}{n}\sum_{i=1}^{n}\mathbf{A}_i\mathbf{A}_i'\otimes\operatorname{diag}\mathbf{w}_i\bigg]^{-1}\bigg[\operatorname{vec}\frac{1}{n}\sum_{i=1}^{n}\big(\mathbf{w}_i'\mathbf{x}_i\mathbf{A}_i\mathbf{A}_i' + \dot{\tilde{\mathbf{Y}}}_i\mathbf{A}_i'\big)\bigg],$$

which reduces to (E.7) when $\mathbf{A}_i = \mathbf{I}\ \forall\ i = 1,\ldots,n$ as $\mathbf{A}_i\mathbf{A}_i' = \mathbf{I}$ and $\mathbf{C} = \sum_i \operatorname{diag}\mathbf{w}_i$.

*Space weights update.*    For the minimization of (E.9) with respect to $\underline{\mathbf{A}}$, we consider one $\mathbf{A}_i$ at the time. The equalities (E.11) need not be applied, since $\mathbf{A}_i$ is on the outside edge of terms and the update for $\mathbf{A}_i$ is given as

$$\mathbf{A}_i^+ = \left( \mathbf{x}_i' \dot{\tilde{\mathbf{x}}}_i + \mathbf{Y}' \dot{\tilde{\mathbf{Y}}}_i \right) \left( \mathbf{x}_i' r_{ii} \mathbf{x}_i + \mathbf{Y}' \mathrm{diag}\left( \mathbf{w}_i \right) \mathbf{Y} - \mathbf{x}_i' \mathbf{w}_i \mathbf{Y} - \mathbf{Y}' \mathbf{w}_i' \mathbf{x}_i \right)^{-1}.$$

For the diagonal model, with diagonal $\mathbf{A}_i$'s, only the diagonals are used of the numerator and denominator, and for the rectangular model, the best rank $r$ approximation is used. For this purpose the singular-value decomposition $\mathbf{A}_i = \mathbf{P}\boldsymbol{\Phi}\mathbf{Q}'$ is used with the singular values in nondecreasing order on the diagonal of $\boldsymbol{\Phi}$ and specify $\mathbf{A}_i$ for the rectangular model as $\mathbf{P}\boldsymbol{\Phi}$ using only the first $r$ (reduced rank) columns of $\mathbf{P}$ and $\boldsymbol{\Phi}$. After updating $\underline{\mathbf{A}}$, the individual spaces are computed according to (E.8).

### E.3    THREE-WAY UNFOLDING MODELS

Allowing for multiple sources, i.e., multiple rectangular $n \times m$ matrices, the relationship between the common space $\mathbf{X}, \mathbf{Y}$ and the $s$ individual spaces $\mathbf{X}_k, \mathbf{Y}_k$ ($k = 1, \ldots, s$) is specified by the following model:

$$\begin{aligned} \mathbf{X}_k &= \mathbf{X}\mathbf{A}_k \\ \mathbf{Y}_k &= \mathbf{Y}\mathbf{A}_k. \end{aligned} \tag{E.15}$$

The different forms of the space weights matrices $\mathbf{A}_k$ ($k = 1, \ldots, s$) result in the same individual differences models as described for the two-way model (see page 177). The two-way loss function (E.5) is extended with an additional way, using the average over $s$ sources, which provides the three-way individual differences model as

$$\begin{aligned} g_3' &= c_3 + \frac{1}{s} \sum_{k=1}^s \mathrm{tr} \left( \mathbf{X}_k' \mathbf{R}_k \mathbf{X}_k + \mathbf{Y}_k' \mathbf{C}_k \mathbf{Y}_k - 2\mathbf{X}_k' \mathbf{W}_k \mathbf{Y}_k - 2\mathbf{X}_k' \dot{\tilde{\mathbf{X}}}_k - 2\mathbf{Y}_k' \dot{\tilde{\mathbf{Y}}}_k \right) \\ &= c_3 + \frac{1}{s} \sum_{k=1}^s \mathrm{tr} \left( \mathbf{A}_k' \mathbf{X}' \mathbf{R}_k \mathbf{X}\mathbf{A}_k + \mathbf{A}_k' \mathbf{Y}' \mathbf{C}_k \mathbf{Y}\mathbf{A}_k - 2\mathbf{A}_k' \mathbf{X}' \mathbf{W}_k \mathbf{Y}\mathbf{A}_k \right. \\ &\qquad \left. - 2\mathbf{A}_k' \mathbf{X}' \dot{\tilde{\mathbf{X}}}_k - 2\mathbf{A}_k' \mathbf{Y}' \dot{\tilde{\mathbf{Y}}}_k \right). \end{aligned} \tag{E.16}$$

The individual differences in the three-way model thus refer to differences between the additional third way matrices or sources and not between the rows as in the two-way model.

*Row coordinates update.* Using the equalities (E.11), and moving the summation over s inside each term, such that the sum over traces becomes the trace of sums, (E.16) is written as

$$
g_3' = c_\mathbf{X} + \frac{1}{s} \sum_{k=1}^{s} \mathrm{tr}\left(\mathbf{A}_k\mathbf{A}_k'\mathbf{X}'\mathbf{R}_k\mathbf{X} - 2\mathbf{X}'\mathbf{W}_k\mathbf{Y}\mathbf{A}_k\mathbf{A}_k' - 2\mathbf{A}_k'\mathbf{X}'\dot{\tilde{\mathbf{X}}}_k\right)
$$

$$
= c_\mathbf{X} + \left(\mathrm{vec}\,\mathbf{X}\right)'\left(\frac{1}{s} \sum_{k=1}^{s} \mathbf{A}_k\mathbf{A}_k' \otimes \mathbf{R}_k'\right)\mathrm{vec}\,\mathbf{X}
$$

$$
- 2\left(\mathrm{vec}\,\mathbf{X}\right)'\mathrm{vec}\,\frac{1}{s} \sum_{k=1}^{s} \mathbf{W}_k\mathbf{Y}\mathbf{A}_k\mathbf{A}_k' - 2\left(\mathrm{vec}\,\mathbf{X}\right)'\mathrm{vec}\,\frac{1}{s} \sum_{k=1}^{s} \dot{\tilde{\mathbf{X}}}_k\mathbf{A}_k'.
$$

$$(\text{E.17})$$

The update for $\mathbf{X}$ is now given as

$$
\mathrm{vec}\,\mathbf{X}^+ = \left[\frac{1}{s} \sum_{k=1}^{s} \mathbf{A}_k\mathbf{A}_k' \otimes \mathbf{R}_k'\right]^{-1}\left[\mathrm{vec}\,\frac{1}{s} \sum_{k=1}^{s} \left(\mathbf{W}_k\mathbf{Y}\mathbf{A}_k\mathbf{A}_k' + \dot{\tilde{\mathbf{X}}}_k\mathbf{A}_k'\right)\right].
$$

*Computational considerations.* For the identity model, where $\mathbf{A}_k = \mathbf{I} \; \forall \; k = 1,\ldots,s$, all $\mathbf{A}_k\mathbf{A}_k'$ are also equal to $\mathbf{I}$, and the update for $\mathbf{X}$ is reduced to

$$
\mathbf{X}^+ = \left[\frac{1}{s} \sum_{k=1}^{s} \mathbf{R}_k\right]^{-1}\left[\frac{1}{s} \sum_{k=1}^{s} \mathbf{W}_k\mathbf{Y} + \frac{1}{s} \sum_{k=1}^{s} \dot{\tilde{\mathbf{X}}}_k\right],
$$

where the inverse is very simple, due to the diagonal matrices $\mathbf{R}_k$ (see page 152). For the diagonal model, $\sum \mathbf{A}_k\mathbf{A}_k' \otimes \mathbf{R}_k'$ is diagonal and simple divisions also suffice for the update. The inverse for the full model is somewhat more difficult, since $\sum \mathbf{A}_k\mathbf{A}_k' \otimes \mathbf{R}_k'$ is a *block* diagonal matrix as the $\mathbf{A}_k$'s are full matrices and the $\mathbf{R}_k$'s are diagonal matrices. Still, the inverse of such a matrix is computed very fast by taking the corresponding diagonal elements of each block, combining them into one matrix, taking the proper inverse and transferring the elements back to the original position in the matrix. This means taking $n$ times the inverse of a $p \times p$ matrix for the complete update of $\mathbf{X}$. However, this is not the update strategy for this or any other model in PREFSCAL, because updating the coordinates one by one even eliminates the computation of the inverses as described above (see Greenacre & Browne, 1986). The one-by-one procedure is described in the section concerning the coordinate restrictions.

*Column coordinates update.* The update for the column coordinates follows the same steps as for the update of the row coordinates.

*Space weights update.* Minimizing (E.16) with respect to the space weights $\underline{\mathbf{A}}$ can be solved for one $\mathbf{A}_k$ at the time, keeping the other $s-1$ fixed. As with the two-way models, rewriting (E.16) in vector notation using (E.11) is not necessary. The derivative with respect to $\mathbf{A}_k$ is straightforward and provides the update for $\mathbf{A}_k$ as

$$\mathbf{A}_k^+ = \left( \mathbf{X}'\dot{\tilde{\mathbf{X}}}_k + \mathbf{Y}'\dot{\tilde{\mathbf{Y}}}_k \right) \left( \mathbf{X}'\mathbf{R}_k\mathbf{X} + \mathbf{Y}'\mathbf{C}_k\mathbf{Y} - \mathbf{X}'\mathbf{W}_k\mathbf{Y} - \mathbf{Y}'\mathbf{W}_k'\mathbf{X} \right)^{-1}.$$

Note that because $-2\mathbf{A}_k\mathbf{X}'\mathbf{W}_k\mathbf{Y}\mathbf{A}_k$ is not symmetric, the derivative is taken from $-\mathbf{A}_k'\mathbf{X}'\mathbf{W}_k\mathbf{Y}\mathbf{A}_k - \mathbf{A}_k'\mathbf{Y}'\mathbf{W}_k'\mathbf{X}\mathbf{A}_k$ instead. The identity model does not need any update for $\mathbf{A}_k$, since $\mathbf{A}_k = \mathbf{I} \ \forall \ k = 1, \ldots, s$, whereas for the diagonal model, only the diagonals are needed, that is,

$$\mathbf{A}_k^+ = \mathrm{diag}\left( \mathbf{X}'\dot{\tilde{\mathbf{X}}}_k + \mathbf{Y}'\dot{\tilde{\mathbf{Y}}}_k \right) \mathrm{diag}\left( \mathbf{X}'\mathbf{R}_k\mathbf{X} + \mathbf{Y}'\mathbf{C}_k\mathbf{Y} - \mathbf{X}'\mathbf{W}_k\mathbf{Y} - \mathbf{Y}'\mathbf{W}_k'\mathbf{X} \right)^{-1},$$

which is a simple division instead of a multiplication with an inverse. For the rectangular model, the best rank $r$ approximation is used. To this end, we use the singular-value decomposition $\mathbf{A}_k = \mathbf{P}\mathbf{\Phi}\mathbf{Q}'$ with the singular values in nondecreasing order on the diagonal of $\mathbf{\Phi}$ and specify $\mathbf{A}_k$ for the rectangular model as $\mathbf{P}\mathbf{\Phi}$ using only the first $r$ columns of $\mathbf{P}$ and $\mathbf{\Phi}$. After updating $\underline{\mathbf{A}}$, the individual spaces are computed according to (E.15).

## E.4 COORDINATE RESTRICTIONS

Restricting coordinates are easily incorporated when the coordinates are computed, not only point after point, but coordinate after coordinate. For this purpose (E.9) for the two-way models and (E.16) for the three-way models are rewritten such that the updates can be computed per coordinate. Fixed coordinates are simply skipped while updating the coordinates.

*Two-way model, row coordinate update.* The vector with coordinates for row $i$ can be written as $\mathbf{x}_i = \mathbf{x}^- + \mathbf{e}x$, where $\mathbf{x}^-$ is equal to $\mathbf{x}_i$, except for one element, $x$, which is set equal to zero in $\mathbf{x}^-$, where $\mathbf{e}$ has a one and zero's elsewhere. Using this decomposition, (E.12) can be written as

$$\begin{aligned} g_2' &= c_{\mathbf{x}_i} + \left( \mathbf{x}^- + \mathbf{e}x \right)' \left( \mathbf{A}_i\mathbf{A}_i' \otimes \mathbf{r}_{ii}' \right) \left( \mathbf{x}^- + \mathbf{e}x \right) \\ &\quad - 2\left( \mathbf{x}^- + \mathbf{e}x \right)' \mathrm{vec}\, \mathbf{w}_i\mathbf{Y}\mathbf{A}_i\mathbf{A}_i' - 2\left( \mathbf{x}^- + \mathbf{e}x \right)' \mathrm{vec}\, \dot{\tilde{\mathbf{x}}}_i\mathbf{A}_i' \\ &= c_x + 2x\mathbf{e}'\left( \mathbf{A}_i\mathbf{A}_i' \otimes \mathbf{r}_{ii}' \right)\mathbf{x}^- + x\mathbf{e}'\left( \mathbf{A}_i\mathbf{A}_i' \otimes \mathbf{r}_{ii}' \right)\mathbf{e}x \\ &\quad - 2x\mathbf{e}'\mathrm{vec}\, \mathbf{w}_i\mathbf{Y}\mathbf{A}_i\mathbf{A}_i' - 2x\mathbf{e}'\mathrm{vec}\, \dot{\tilde{\mathbf{x}}}_i\mathbf{A}_i'. \end{aligned}$$

The update for $x$ is now given as

$$x^+ = \left[ e'\left(A_iA_i'\otimes r_{ii}'\right)e\right]^{-1}\left[e'\text{vec}\left(w_iYA_iA_i' + \dot{\tilde{x}}_iA_i'\right) - e'\left(A_iA_i'\otimes r_{ii}'\right)x^-\right],$$

which is repeated for all $np$ coordinates.

*Two-way model, column coordinate update.* The vector with column coordinates can be written as $\text{vec}\,Y = y^- + ey$, where $y^-$ is equal to $\text{vec}\,Y$, except for one element, $y$, which is set equal to zero in $y^-$, where $e$ has a one and zero's elsewhere. Substitution is (E.14) gives

$$g_2' = c_Y + (y^- + ey)'\left(\frac{1}{n}\sum_{i=1}^{n} A_iA_i' \otimes \text{diag}\,w_i\right)(y^- + ey)$$

$$- 2(y^- + ey)'\text{vec}\,\frac{1}{n}\sum_{i=1}^{n} w_i'x_iA_iA_i' - 2(y^- + ey)'\text{vec}\,\frac{1}{n}\sum_{i=1}^{n}\dot{\tilde{Y}}_iA_i'$$

$$= c_y + 2ye'\left(\frac{1}{n}\sum_{i=1}^{n} A_iA_i' \otimes \text{diag}\,w_i\right)y^-$$

$$+ ye'\left(\frac{1}{n}\sum_{i=1}^{n} A_iA_i' \otimes \text{diag}\,w_i\right)ey$$

$$- 2ye'\text{vec}\,\frac{1}{n}\sum_{i=1}^{n} w_i'x_iA_iA_i' - 2ye'\text{vec}\,\frac{1}{n}\sum_{i=1}^{n}\dot{\tilde{Y}}_iA_i'.$$

The update for $y$ is now given as

$$y^+ = \left[e'\left(\frac{1}{n}\sum_{i=1}^{n} A_iA_i' \otimes \text{diag}\,w_i\right)e\right]^{-1}$$

$$\left[e'\text{vec}\,\frac{1}{n}\sum_{i=1}^{n}\left(w_i'x_iA_iA_i' + \dot{\tilde{Y}}_iA_i'\right) - e'\left(\frac{1}{n}\sum_{i=1}^{n} A_iA_i' \otimes \text{diag}\,w_i\right)y^-\right],$$

which is repeated for all $mp$ coordinates.

*Three-way model, row coordinate update.* The vector with coordinates can be written as $\text{vec}\,X = x^- + ex$, where $x^-$ is equal to $\text{vec}\,X$, except for one element, $x$, which is set equal to zero in $x^-$, where $e$ has a one and zero's elsewhere.

Combining this decomposition with (E.17) gives

$$
\begin{aligned}
g'_3 &= c_X + (x^- + ex)' \left( \frac{1}{s} \sum_{k=1}^{s} A_k A'_k \otimes R'_k \right) (x^- + ex) \\
&\quad - 2(x^- + ex)' \mathrm{vec}\, \frac{1}{s} \sum_{k=1}^{s} W_k Y A_k A'_k - 2(x^- + ex)' \mathrm{vec}\, \frac{1}{s} \sum_{k=1}^{s} \dot{\tilde{X}}_k A'_k \\
&= c_x + 2xe' \left( \frac{1}{s} \sum_{k=1}^{s} A_k A'_k \otimes R'_k \right) x^- + xe' \left( \frac{1}{s} \sum_{k=1}^{s} A_k A'_k \otimes R'_k \right) ex \\
&\quad - 2xe' \mathrm{vec}\, \frac{1}{s} \sum_{k=1}^{s} W_k Y A_k A'_k - 2xe' \mathrm{vec}\, \frac{1}{s} \sum_{k=1}^{s} \dot{\tilde{X}}_k A'_k .
\end{aligned}
$$

The update for each coordinate $x$ is given as

$$
\begin{aligned}
x^+ &= \left[ e' \left( \frac{1}{s} \sum_{k=1}^{s} A_k A'_k \otimes R'_k \right) e \right]^{-1} \\
&\quad \left[ e' \mathrm{vec}\, \frac{1}{s} \sum_{k=1}^{s} \left( W_k Y A_k A'_k + \dot{\tilde{X}}_k A'_k \right) - e' \left( \frac{1}{s} \sum_{k=1}^{s} A_k A'_k \otimes R'_k \right) x^- \right].
\end{aligned}
$$

*Three-way model, column coordinate update.* The update for a column co-ordinate follows the same steps as for the update of a three-way model row coordinate.

*Computational considerations.* Note that the vector $e$ contains only zero's and a single one, so after multiplication with another vector (inner product) only a scalar remains. The Kronecker product is even pre- and post-multiplied with $e$, thus providing one diagonal element of the product, a scalar, which reduces the inverse to a simple division. For both identity and diagonal models, the terms $e'(\sum \ldots \otimes \ldots) x^-$ and $e'(\sum \ldots \otimes \ldots) y^-$ are equal to zero, due to the special form of $e$, $x^-$, and $y^-$. For the full model, these terms are not equal to zero, but consist of a summation over $p - 1$ off-diagonal products, where $p$ is the dimensionality of the model.

### E.5 VARIABLE RESTRICTIONS

The coordinates of the row objects of common space can be restricted to be a linear combination of independent variables, i.e., $X = QB$, where $X$ is a $n \times p$ matrix with common space row coordinates, $Q$ is a $n \times h$ matrix of $h$ optimally transformed independent variables, and $B$ is a $h \times p$ matrix with regression coefficients. Instead of updating $X$, updates are computed for the optimally

transformed variables $\mathbf{Q}$ (if needed) and for the regression coefficients $\mathbf{B}$. These updates are computed iteratively per variable until converged. The same restrictions can be applied to the column objects of the common space. To avoid too many subscripts, superscripts, or other scripts, we use the same characters for both row and column regression coefficients and variables, although their sizes and contents differ, of course.

*Two-way model, row regression coefficients update.* For the update of $\mathbf{B}$, we substitute $\mathbf{q}_i\mathbf{B}$ for $\mathbf{x}_i$ in (E.10) and rearrange terms yielding

$$g_2' = c_{\mathbf{B}} + \frac{1}{n}\sum_{i=1}^{n} \operatorname{tr}\left(\mathbf{A}_i'\mathbf{B}\mathbf{q}_i'r_{ii}\mathbf{q}_i\mathbf{B}\mathbf{A}_i - 2\mathbf{A}_i'\mathbf{B}'\mathbf{q}_i'w_i\mathbf{Y}\mathbf{A}_i - 2\mathbf{A}_i'\mathbf{B}'\mathbf{q}_i'\overset{\bullet}{\tilde{\mathbf{x}}_i}\right). \tag{E.18}$$

Letting

$$\mathbf{q}_i\mathbf{B} = \sum_{t=1}^{h} q_{it}\mathbf{b}_t' = q_{il}\mathbf{b}_l' + \sum_{t\neq l}^{h} q_{it}\mathbf{b}_t' = \mathbf{U}_l + q_{il}\mathbf{b}_l' \tag{E.19}$$

allows us to find an update for $\mathbf{B}$ one row at the time. Substituting (E.19) in (E.18) and writing the equation for one row $l$ only, omitting the subscript $l$ for $\mathbf{q}$, $\mathbf{b}$, and $\mathbf{U}$, gives

$$g_2' = c_{\mathbf{b}} + \frac{1}{n}\sum_{i=1}^{n} \operatorname{tr}\left(\mathbf{A}_i\mathbf{A}_i'\mathbf{b}q_i'r_{ii}q_i\mathbf{b}' - 2\mathbf{b}q_i'r_{ii}\mathbf{U}\mathbf{A}_i\mathbf{A}_i'\right.$$
$$\left. - 2\mathbf{b}q_i'w_i\mathbf{Y}\mathbf{A}_i\mathbf{A}_i' - 2\mathbf{b}q_i'\overset{\bullet}{\tilde{\mathbf{x}}_i}\mathbf{A}_i'\right)$$

and using the equalities from (E.11) and the fact that $\operatorname{vec}\mathbf{b} = \mathbf{b}$ gives

$$g_2' = c_{\mathbf{b}} + \mathbf{b}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{A}_i\mathbf{A}_i' \otimes q_i'r_{ii}q_i\right)\mathbf{b}'$$
$$+ 2\mathbf{b}\left(\operatorname{vec}\frac{1}{n}\sum_{i=1}^{n} q_i'r_{ii}\mathbf{U}\mathbf{A}_i\mathbf{A}_i' - \operatorname{vec}\frac{1}{n}\sum_{i=1}^{n} q_i'w_i\mathbf{Y}\mathbf{A}_i\mathbf{A}_i'\right.$$
$$\left. - \operatorname{vec}\frac{1}{n}\sum_{i=1}^{n} q_i'\overset{\bullet}{\tilde{\mathbf{x}}_i}\mathbf{A}_i'\right).$$

Now, using the same strategy as with the coordinates restriction, that is, update one b at the time by setting $\mathbf{b} = b\boldsymbol{e}' + \mathbf{b}^-$, such that

$$
g_2' = c_b + b\boldsymbol{e}'\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{A}_i\mathbf{A}_i' \otimes q_i'r_{ii}q_i\right)\boldsymbol{e}b
$$

$$
+ 2b\boldsymbol{e}'\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{A}_i\mathbf{A}_i' \otimes q_i'r_{ii}q_i\right)\mathbf{b}^- + 2b\boldsymbol{e}'\left(\text{vec } \frac{1}{n}\sum_{i=1}^{n} q_i'r_{ii}\mathbf{U}\mathbf{A}_i\mathbf{A}_i'\right.
$$

$$
\left. - \text{vec } \frac{1}{n}\sum_{i=1}^{n} q_i'w_i\mathbf{Y}\mathbf{A}_i\mathbf{A}_i' - \text{vec } \frac{1}{n}\sum_{i=1}^{n} q_i'\overset{\bullet}{\tilde{x}}_i\mathbf{A}_i'\right),
$$

which provides the update for b as

$$
\mathbf{b}^+ = \left[\boldsymbol{e}'\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{A}_i\mathbf{A}_i' \otimes q_i'r_{ii}q_i\right)\boldsymbol{e}\right]^{-1}
$$

$$
\boldsymbol{e}'\left[\text{vec } \frac{1}{n}\sum_{i=1}^{n} q_i'w_i\mathbf{Y}\mathbf{A}_i\mathbf{A}_i' + \text{vec } \frac{1}{n}\sum_{i=1}^{n} q_i'\overset{\bullet}{\tilde{x}}_i\mathbf{A}_i'\right.
$$

$$
\left. - \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{A}_i\mathbf{A}_i' \otimes q_i'r_{ii}q_i\right)\mathbf{b}^- - \text{vec } \frac{1}{n}\sum_{i=1}^{n} q_i'r_{ii}\mathbf{U}\mathbf{A}_i\mathbf{A}_i'\right].
$$

This update is repeated for all hp regression coefficients.

*Two-way model, row variable update.* For a variable, two updates are necessary: An unrestricted and a restricted update. The latter update restricts the variables conform the user-specified transformations of the original independent variables $\mathbf{E}$.

For the unrestricted update, using (E.19) and updating $\mathbf{q}$ for one row $i$ at the time, (E.10) is written as

$$
g_2' = c_{q_i} + \text{tr}\left(\mathbf{b}'\mathbf{A}_i\mathbf{A}_i'\mathbf{b}q_i'r_{ii}q_i + 2q_i'r_{ii}\mathbf{U}\mathbf{A}_i\mathbf{A}_i'\mathbf{b}\right.
$$

$$
\left. - 2q_i'w_i\mathbf{Y}\mathbf{A}_i\mathbf{A}_i'\mathbf{b} - 2q_i'\overset{\bullet}{\tilde{x}}_i\mathbf{A}_i'\mathbf{b}\right).
$$

Applying the equalities from (E.11) gives

$$
g_2' = c_{q_i} + q_i'\left(\mathbf{b}'\mathbf{A}_i\mathbf{A}_i'\mathbf{b} \otimes r_{ii}'\right)q_i + 2q_i'\text{vec } r_{ii}\mathbf{U}\mathbf{A}_i\mathbf{A}_i'\mathbf{b}
$$

$$
- 2q_i'\text{vec } w_i\mathbf{Y}\mathbf{A}_i\mathbf{A}_i'\mathbf{b} - 2q_i'\text{vec } \overset{\bullet}{\tilde{x}}_i\mathbf{A}_i'\mathbf{b}, \tag{E.20}
$$

which provides the unrestricted update for $q_i$ as

$$
q_i^* = \left[r_{ii}\mathbf{b}'\mathbf{A}_i\mathbf{A}_i'\mathbf{b}\right]^{-1}\left[w_i\mathbf{Y}\mathbf{A}_i\mathbf{A}_i'\mathbf{b} + \overset{\bullet}{\tilde{x}}_i\mathbf{A}_i'\mathbf{b} - r_{ii}\mathbf{U}\mathbf{A}_i\mathbf{A}_i'\mathbf{b}\right], \tag{E.21}
$$

which is repeated $nh$ times.

For the restricted update, first rewrite (E.20) for one variable $h$ as

$$g_2' = c_q + q'Vq - 2q'Tb, \tag{E.22}$$

where $V$ is an $n \times n$ matrix with diagonal elements $v_{ii} = r_{ii}b'A_iA_i'b$ and $T$ is an $n \times p$ matrix with row elements $t_i = w_iYA_iA_i' + \dot{\tilde{x}}_iA_i' - r_{ii}UA_iA_i'$. The general solution for an unrestricted variable $q$ is then given as $q^* = V^{-1}Tb$, as can be deduced from (E.21), and thus

$$Vq^* = Tb. \tag{E.23}$$

Combining (E.22) and (E.23) gives the objective function for a restricted variable $q$ as

$$\begin{aligned} g_2' &= c_q + q'Vq - 2q'Vq^* \\ &= c_q + (q - q^*)'V(q - q^*), \end{aligned}$$

which provides the update for a restricted variable $q$ as

$$q^+ = (q'Vq)^{-1}q'Vq^*. \tag{E.24}$$

Since (E.24) is a weighted least squares function with weights $V$, standard weighted least squares procedures suffice to find a solution for the restricted variable $q^+$.

*Two-way model, column regression coefficients update.* Starting from (E.13) and substituting $QB$ for $Y$ gives

$$\begin{aligned} g_2' = c_B + \frac{1}{n}\sum_{i=1}^{n} \text{tr}\Big( &A_i'B'Q'\text{diag}\,(w_i)QBA_i \\ &- 2A_i'B'Q'w_i'x_iA_i - 2A_i'B'Q'\dot{\tilde{Y}}_i \Big). \end{aligned}$$

Now, replacing $Y$ with the sum of rank-one matrices

$$QB = \sum_{t=1}^{h} q_tb_t' = q_lb_l' + \sum_{t \neq l}^{h} q_tb_t' = q_lb_l' + U_l \tag{E.25}$$

and changing to vectors using (E.11), the objective function becomes

$$
g_2' = c_b + \frac{1}{n} \sum_{i=1}^{n} \left( A_i A_i' b q' \mathrm{diag}\,(w_i)\, q b' + 2 b q' \mathrm{diag}\,(w_i)\, U A_i A_i' \right.
$$

$$
\left. - 2 b q' w_i' x_i A_i A_i' - 2 b q' \dot{\widetilde{Y}}_i A_i' \right)
$$

$$
= c_b + b' \left( \frac{1}{n} \sum_{i=1}^{n} A_i A_i' \otimes q' \mathrm{diag}\,(w_i)\, q \right) b
$$

$$
+ 2 b' \mathrm{vec}\, \frac{1}{n} \sum_{i=1}^{n} \left( q' \mathrm{diag}\,(w_i)\, U A_i A_i' - q' w_i' x_i A_i A_i' - q' \dot{\widetilde{Y}}_i A_i' \right),
$$

where the subscript $l$ is omitted for $b$, $q$, and $U$. By setting $b = b e' + b^-$, the function is written in terms of only one regression coefficient $b$ as

$$
g_2' = c_b + b e' \left( \frac{1}{n} \sum_{i=1}^{n} A_i A_i' \otimes q' \mathrm{diag}\,(w_i)\, q \right) e b
$$

$$
+ 2 b e' \left( \frac{1}{n} \sum_{i=1}^{n} A_i A_i' \otimes q' \mathrm{diag}\,(w_i)\, q \right) b^-
$$

$$
+ 2 b e' \mathrm{vec}\, \frac{1}{n} \sum_{i=1}^{n} \left( q' \mathrm{diag}\,(w_i)\, U A_i A_i' - q' w_i' x_i A_i A_i' - q' \dot{\widetilde{Y}}_i A_i' \right),
$$

which gives the update for $b$ as

$$
b^+ = \left[ e' \left( \frac{1}{n} \sum_{i=1}^{n} A_i A_i' \otimes q' \mathrm{diag}\,(w_i)\, q \right) e \right]^{-1}
$$

$$
e' \left[ \mathrm{vec}\, \frac{1}{n} \sum_{i=1}^{n} \left( q' w_i' x_i A_i A_i' + q' \dot{\widetilde{Y}}_i A_i' - q' \mathrm{diag}\,(w_i)\, U A_i A_i' \right) \right.
$$

$$
\left. - \left( \frac{1}{n} \sum_{i=1}^{n} A_i A_i' \otimes q' \mathrm{diag}\,(w_i)\, q \right) b^- \right].
$$

*Two-way model, column variable update.* For the unrestricted update, we rewrite (E.13) for one variable $\mathbf{q}$ using (E.25) as

$$
\begin{aligned}
g_2' = c_{\mathbf{q}} &+ \frac{1}{n} \sum_{i=1}^{n} \operatorname{tr} \left( \mathbf{b}' \mathbf{A}_i \mathbf{A}_i' \mathbf{b} \mathbf{q}' \operatorname{diag}(\mathbf{w}_i) \mathbf{q} \right. \\
&+ 2\mathbf{q}' \operatorname{diag}(\mathbf{w}_i) \mathbf{U} \mathbf{A}_i \mathbf{A}_i' \mathbf{b} - 2\mathbf{q}' \mathbf{w}_i' x_i \mathbf{A}_i \mathbf{A}_i' \mathbf{b} - 2\mathbf{q}' \dot{\tilde{\mathbf{Y}}}_i \mathbf{A}_i' \mathbf{b} \Big) \\
= c_{\mathbf{q}} &+ \mathbf{q}' \left( \frac{1}{n} \sum_{i=1}^{n} \mathbf{b}' \mathbf{A}_i \mathbf{A}_i' \mathbf{b} \otimes \operatorname{diag} \mathbf{w}_i \right) \mathbf{q} \\
&+ 2\mathbf{q}' \operatorname{vec} \frac{1}{n} \sum_{i=1}^{n} \left( \operatorname{diag}(\mathbf{w}_i) \mathbf{U} \mathbf{A}_i \mathbf{A}_i' \mathbf{b} - \mathbf{w}_i' x_i \mathbf{A}_i \mathbf{A}_i' \mathbf{b} - \dot{\tilde{\mathbf{Y}}}_i \mathbf{A}_i' \mathbf{b} \right),
\end{aligned}
$$

which provide the unrestricted update for $\mathbf{q}$ as

$$
\begin{aligned}
\mathbf{q}^* = & \left[ \frac{1}{n} \sum_{i=1}^{n} \mathbf{b}' \mathbf{A}_i \mathbf{A}_i' \mathbf{b} \otimes \operatorname{diag} \mathbf{w}_i \right]^{-1} \\
& \left[ \operatorname{vec} \frac{1}{n} \sum_{i=1}^{n} \left( \mathbf{w}_i' x_i \mathbf{A}_i \mathbf{A}_i' \mathbf{b} + \dot{\tilde{\mathbf{Y}}}_i \mathbf{A}_i' \mathbf{b} - \operatorname{diag}(\mathbf{w}_i) \mathbf{U} \mathbf{A}_i \mathbf{A}_i' \mathbf{b} \right) \right].
\end{aligned}
$$

The restricted update for a column variable follows the same steps as for the restricted update of a row variable (see page 186).

*Three-way model, row regression coefficients update.* The three-way model (E.16) substituting $\mathbf{QB}$ for $\mathbf{X}$ is written as

$$
g_3' = c_{\mathbf{QB}} + \frac{1}{s} \sum_{k=1}^{s} \operatorname{tr} \left( \mathbf{A}_k' \mathbf{B}' \mathbf{Q}' \mathbf{R}_k \mathbf{Q} \mathbf{B} \mathbf{A}_k - 2 \mathbf{A}_k' \mathbf{B}' \mathbf{Q}' \mathbf{W}_k \mathbf{Y} \mathbf{A}_k \right.
$$
$$
\left. - 2 \mathbf{A}_k' \mathbf{B}' \mathbf{Q}' \dot{\tilde{\mathbf{X}}}_k \right),
$$

which in turn is rewritten as

$$
g_3' = c_{\mathbf{q}\mathbf{b}'} + \frac{1}{s} \sum_{k=1}^{s} \operatorname{tr} \left( \mathbf{q}' \mathbf{R}_k \mathbf{q} \mathbf{b}' \mathbf{A}_k \mathbf{A}_k' \mathbf{b} + 2 \mathbf{q}' \mathbf{R}_k \mathbf{U} \mathbf{A}_k \mathbf{A}_k' \mathbf{b} \right.
$$
$$
\left. - 2 \mathbf{q}' \mathbf{W}_k \mathbf{Y} \mathbf{A}_k \mathbf{A}_k' \mathbf{b} - 2 \mathbf{q}' \dot{\tilde{\mathbf{X}}}_k \mathbf{A}_k' \mathbf{b} \right), \qquad \text{(E.26)}
$$

where $\mathbf{QB}$ is replaced with the sum of rank-one matrices as given in (E.25) with the index $l$ omitted. By using the equalities in (E.11) first and then setting

$b = be' + b^-$, the function is written in terms of one b only as

$$g_3' = c_b + b'\left(\frac{1}{s}\sum_{k=1}^{s} A_k A_k' \otimes q'R_k q\right)b + 2b'\text{vec}\frac{1}{s}\sum_{k=1}^{s} q'R_k U A_k A_k'$$

$$- 2b'\text{vec}\frac{1}{s}\sum_{k=1}^{s} q'W_k Y A_k A_k' - 2b'\text{vec}\frac{1}{s}\sum_{k=1}^{s} q'\dot{\tilde{X}}_k A_k'$$

$$= c_b + be'\left(\frac{1}{s}\sum_{k=1}^{s} A_k A_k' \otimes q'R_k q\right)eb + 2be'\left(\frac{1}{s}\sum_{k=1}^{s} A_k A_k' \otimes q'R_k q\right)b^-$$

$$+ 2be'\left(\frac{1}{s}\sum_{k=1}^{s} q'R_k U A_k A_k'\right) - 2be'\left(\frac{1}{s}\sum_{k=1}^{s} q'W_k Y A_k A_k'\right)$$

$$- 2be'\left(\frac{1}{s}\sum_{k=1}^{s} q'\dot{\tilde{X}}_k A_k'\right),$$

which provides the update for b as

$$b^+ = \left[e'\left(\frac{1}{s}\sum_{k=1}^{s} A_k A_k' \otimes q'R_k q\right)e\right]^{-1}$$

$$\left[e'\left(\frac{1}{s}\sum_{k=1}^{s} q'W_k Y A_k A_k'\right) + e'\left(\frac{1}{s}\sum_{k=1}^{s} q'\dot{\tilde{X}}_k A_k'\right)\right.$$

$$\left. - e'\left(\frac{1}{s}\sum_{k=1}^{s} A_k A_k' \otimes q'R_k q\right)b^- - e'\left(\frac{1}{s}\sum_{k=1}^{s} q'R_k U A_k A_k'\right)\right].$$

Note that $e'(\ldots)b^-$ is zero, and thus drops out of the equation, for diagonal $A_k$'s, that is, for the identity models and the diagonal models.

*Three-way model, row variable update.* The unrestricted update for a row variable q is found by first rewriting (E.26) as

$$g_3' = c_{qb'} + q'\left(\frac{1}{s}\sum_{k=1}^{s} R_k \otimes b'A_k A_k' b\right)q + 2q'\text{vec}\frac{1}{s}\sum_{k=1}^{s} R_k U A_k A_k' b$$

$$- 2q'\text{vec}\frac{1}{s}\sum_{k=1}^{s} W_k Y A_k A_k' b - 2q'\text{vec}\frac{1}{s}\sum_{k=1}^{s} \dot{\tilde{X}}_k A_k' b,$$

which gives the unrestricted update of q as

$$q^* = \left[\frac{1}{s}\sum_{k=1}^{s} R_k \otimes b'A_k A_k' b\right]^{-1}$$

$$\left[\text{vec}\frac{1}{s}\sum_{k=1}^{s}\left(W_k Y A_k A_k' b + \dot{\tilde{X}}_k A_k' b - R_k U A_k A_k' b\right)\right].$$

Note that $\mathbf{b}'\mathbf{A}_k\mathbf{A}'_k\mathbf{b}$ is a scalar, so, irrespective of the model, the Kronecker product $\mathbf{R}_k \otimes \mathbf{b}'\mathbf{A}_k\mathbf{A}'_k\mathbf{b}$ results in a diagonal matrix for which the inverse consists of simple divisions. Then, for an update of the restricted $\mathbf{q}$, we use $g'_3 = c_\mathbf{q} + \mathbf{q}'\mathbf{V}\mathbf{q} - 2\mathbf{q}'\mathbf{T}\mathbf{b} = c_\mathbf{q} + \mathbf{q}'\mathbf{V}\mathbf{q} - 2\mathbf{q}'\mathbf{V}\mathbf{q}^*$ with

$$\mathbf{V} = \frac{1}{s}\sum_{k=1}^{s} \mathbf{R}_k \otimes \mathbf{b}'\mathbf{A}_k\mathbf{A}'_k\mathbf{b} \ \text{ and}$$

$$\mathbf{T} = \text{vec}\,\frac{1}{s}\sum_{k=1}^{s} \left(\mathbf{W}_k\mathbf{Y}\mathbf{A}_k\mathbf{A}'_k + \dot{\tilde{\mathbf{X}}}_k\mathbf{A}'_k - \mathbf{R}_k\mathbf{U}\mathbf{A}_k\mathbf{A}'_k\right),$$

which gives the update satisfying the constraints as (E.24). Since $\mathbf{V}$ is a diagonal matrix, standard weighted least squares procedures suffice for a transformation of an unrestricted variable $\mathbf{q}^*$ to obtain a restricted variable $\mathbf{q}^+$.

*Computational considerations.* Updates for the variables $\mathbf{Q}$ and for the regression coefficients $\mathbf{B}$ are computed per variable $h$, iterating over updates $\mathbf{q}^*$, $\mathbf{q}^+$, and $\mathbf{b}^+$ until convergence is reached. For this purpose, $\sum \mathbf{W}_k\mathbf{Y}\mathbf{A}_k\mathbf{A}'_k$ and $\sum \dot{\tilde{\mathbf{X}}}_k\mathbf{A}'_k$ can be computed in advance. Since $\mathbf{U}$ does not change while working on variable $h$, $\sum \mathbf{R}_k\mathbf{U}\mathbf{A}_k\mathbf{A}'_k$ is computed before iterations start, where $\mathbf{U}$ is computed as $\mathbf{X} - \mathbf{q}\mathbf{b}'$. Finally, after convergence, an update for $\mathbf{X}$ is computed as $\mathbf{X}^+ = \mathbf{U} + \mathbf{q}^+\mathbf{b}^{+\prime}$. It might happen that a monotonically transformed variable becomes constant, i.e., $\mathbf{q}^+ = c\mathbf{1}$, in which case the direction of the variable $\mathbf{q}^*$ as well as other related quantities (for example, rank-indices and tie-blocks) are reversed, and new updates for the restricted variable $\mathbf{q}$ and the regression coefficients $\mathbf{b}$ are computed. Optimal procedures, besides constant transformed variables, such as minimizing the number of ties, maximizing variable variation, minimizing loss, or maximizing the correlation between the original and the transformed variables, are still under investigation.

*Three-way model, column regression coefficients update.* The update for a column regression coefficient follows the same steps as for the update of a row regression coefficient discussed on page 188.

*Three-way model, column variable update.* The update for a column variable follows the same steps as for the update of a row variable (see page 189).

*Identification considerations.* In contrast with PROXSCAL, for PREFSCAL it is not allowed to center the variables after updating $\mathbf{Q}$ and $\mathbf{B}$ as centering in PREFSCAL changes the distances $\mathbf{D}$. But since $\mathbf{Q}\mathbf{B} = \mathbf{Q}\mathbf{T}\mathbf{T}^{-1}\mathbf{B} = \mathbf{Q}^*\mathbf{B}^*$, with $\mathbf{Q}^* = \mathbf{Q}\mathbf{T}$ and $\mathbf{B}^* = \mathbf{T}^{-1}\mathbf{B}$, for an arbitrary matrix $\mathbf{T}$ for which the inverse exists, identification is needed for $\mathbf{Q}\mathbf{B}$ and satisfied by letting $\mathbf{b}'\mathbf{b} = p$ for each row of $\mathbf{B}$. The corresponding columns of $\mathbf{Q}$ are adapted accordingly, so that $\mathbf{Q}\mathbf{B}$ remains the same.

Other restrictions, besides coordinate and variable restrictions, are possible for the configuration. These restrictions are on the implementation list, but currently not available in PREFSCAL.

The restriction that follows logically from the centroid start is the centroid restriction. Objects from one set can be restricted to be located in the centroid of objects from the other set, i.e., $Y = M^{-1}EX$, where $M$ and $E$ are defined as before (see page 156). For example, product coordinates can be placed in the centroid of the respondents that ranked the product highest. Using only the closest respondents conforms to an indicator matrix $E$ with first choices only. Also using second or third placed respondents expands the indicator matrix with second and third choices. Obviously, the number of choices must be smaller than the number of products, because all products will otherwise be placed in the same centroid, the centroid of all respondents. It is possible to combine the centroid restriction for one set with another restriction, coordinate or variable restriction, on the other set. The centroid restriction is similar to the approach taken by DeSarbo and Rao to avoid degenerate solutions.

Another restriction is the orthogonality restriction. One of the (dis)advantages of principal component analysis are the uncorrelated dimensions. This is not true for multidimensional unfolding. To meet with this feature, an orthogonality restriction can be placed on the configuration. Currently it is unknown whether such a restriction will also lead to nested models, that is, whether the two dimensions of a two-dimensional solution are identical to the first two dimensions of a three-dimensional solution. Future research should reveal the practical value of such a restriction.

## POST-PROCESSING
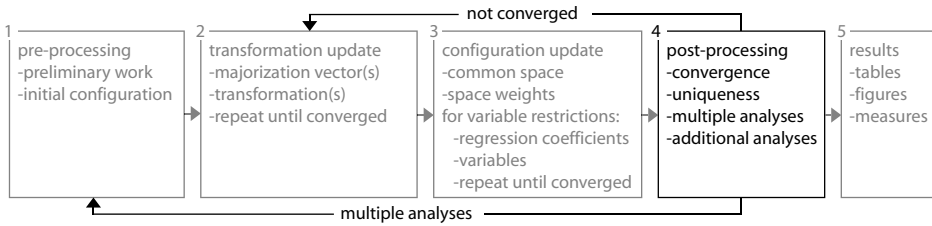


After the iterative process has stopped, unique solutions are computed for the common space, with the space weights adapted accordingly. When multiple unfolding analyses, for example multiple random starts, are requested, the process returns to the pre-processing step (Technical Appendix C). Before computing the final results (Technical Appendix G), additional analyses, for example cluster analysis, can be performed on request to enhance the results of the multidimensional unfolding analysis.

### F.1    ALGORITHM TERMINATION

The algorithm is terminated if any of the three following criteria is satisfied (cf. Dennis & Schnabel, 1983, pp. 159–161). If the current P-STRESS value is close to zero, i.e., $\sigma_p < \epsilon$, the lower bound of the function, the problem is solved and iterations are no longer continued. Further, if the algorithm has converged or if it has stalled, i.e., $\sigma_p^{old} - \sigma_p^{new} \leqslant \Delta_{\sigma_p}(\sigma_p^{old} + \sigma_p^{new} + \epsilon)/2$, or if the current P-STRESS value is sufficiently small, i.e., $\sigma_p < \sigma_p^{min}$, iterations are stopped. Finally, to avoid an unlimited amount of computer time, an iteration limit is imposed on the algorithm. Once the number of iterations exceeds a predefined maximum, i.e., $l > l^{max}$, the iterations are terminated. Default specifications for closeness ($\sigma_p^{min} = 0.0001$), change criterion ($\Delta_{\sigma_p} = 0.000001$), and maximum number of iterations ($l^{max} = 5000$) can be modified by the user.

### F.2    UNIQUENESS

An unrestricted common space has the freedom of both dilation and rotation. For the different models, a unique orientation of the common space is given in the next subsections. This transformation, discussed for the three-way models

only, is needed to overcome an indeterminacy in $\mathbf{Z}\mathbf{A}_k$, where $\mathbf{Z} = [\mathbf{X}', \mathbf{Y}']'$, since there always exists a non-singular matrix $\mathbf{T}$, such that $\mathbf{Z}\mathbf{T}^{-1}\mathbf{T}\mathbf{A}_k$, thus providing an equally good solution with $\mathbf{Z}^* = \mathbf{Z}\mathbf{T}^{-1}$ and $\mathbf{A}_k^* = \mathbf{T}\mathbf{A}_k$. The unique orientations are also optimal for changing from a higher dimensionality to a lower dimensionality, since the maximum variation in coordinates is retained in the lower dimensionality. Once restrictions are involved, rotation and dilation of the common space are not allowed, since the orientation is then determined by the restrictions.

*Identity model.* The common space of the identity model may be rotated and uniformly dilated freely. For a unique solution, the common space is rotated to principal axes. First, an eigenvalue decomposition is computed on the common space,

$$\mathbf{Z}'\mathbf{Z} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}',$$

where $\mathbf{V}'\mathbf{V} = \mathbf{V}\mathbf{V}' = \mathbf{I}$, and $\mathbf{\Lambda}$ is a diagonal matrix with descending elements on the diagonal. Then, the rotated common space is computed as

$$\mathbf{X}^+ = \mathbf{X}\mathbf{V}$$
$$\mathbf{Y}^+ = \mathbf{Y}\mathbf{V},$$

such that $\mathbf{Z}^{+\prime}\mathbf{Z}^+ = \mathbf{V}'\mathbf{Z}'\mathbf{Z}\mathbf{V} = \mathbf{V}'\mathbf{V}\mathbf{\Lambda}\mathbf{V}'\mathbf{V} = \mathbf{\Lambda}$.

*Diagonal model.* The three-way diagonal model has a unique orientation, except for the uniform dilation part. Carroll and Chang (1970) suggest two normalizations: Normalize each individual's sum of squared scalar products, or normalize the variance of the projections of the points on each axis. Since PREFSCAL does not use scalar products, the first normalization can not be used. Instead, the normalization used in PROXSCAL (Commandeur & Heiser, 1993; Commandeur, 1994) is applied, where the dilated common space is computed as

$$\mathbf{Z}^+ = (n + m)^{.5}\mathbf{Z}(\text{diag}\,\mathbf{Z}'\mathbf{Z})^{-.5},$$

meaning that $\text{diag}(\mathbf{Z}^{+\prime}\mathbf{Z}^+) = (n + m)\mathbf{I}$. The inverse transformation is applied to the space weights. Further, the dimensions of the common space are permutated depending on the sum of squared space weights per dimension, where the largest sum-of-squares is related to the first dimension and so on, which eventually corresponds to the second normalization suggested by Carroll and Chang and also suggested by Gower and Hand (1996, p. 219). These latter authors also suggest the parameterization $s^{-1} \sum_{k=1}^{s} \mathbf{A}_k = \mathbf{I}$.

*Full and rectangular model.*    The three-way full model and the rectangular model allow for rotation and non-uniform dilation of the common space, as long as the space weights are rotated and dilated inversely. For a unique orientation, again, PROXSCAL is followed (see Commandeur & Heiser, 1993; Commandeur, 1994). Using the Cholesky decomposition $\mathbf{Z}'\mathbf{Z} = \mathbf{L}\mathbf{L}'$, the common space is normalized by adapting the common space and the space weights as

$$\mathbf{Z}^{+} = (n + m)^{.5}\mathbf{Z}(\mathbf{L}')^{-1}$$
$$\mathbf{A}_k^{+} = \mathbf{L}'\mathbf{A}_k \ \forall \, k = 1, \ldots, s,$$

such that $\mathbf{Z}^{+\prime}\mathbf{Z}^{+} = (n + m)\mathbf{I}$.

## F.3    MULTIPLE ANALYSES

It is possible to perform multiple unfolding analyses to address issues like local minima or stability. Besides multiple random starts and dimensionality reduction, which both address the issue of local minima, resampling methods are used to evaluate the stability of the parameters of the unfolding model.

*Dimensionality reduction.*    PREFSCAL allows the user to specify different values for the minimum and maximum dimensionality, such that $p_{min} < p_{max}$. The reduced rank dimensionality must in turn be smaller than $p_{min}$, whenever the rectangular model is specified. When the dimensionality of the common space is lowered, the transformed preferences need to be re-computed based on the lower dimensionality. Since the configurations are optimal in some sense, as described above (page 193), the first $p - 1$ (most important) dimensions are used for this purpose. The initial configuration and the space weights for the $p - 1$ analysis also use these optimal dimensions, which illustrates the way local minima are avoided. The results of the lowest dimensionality specified are provided afterwards. Dimensionality reduction in combination with restrictions on the common space is not allowed. If dimensionality reduction is desired anyway, it can be performed manually.

*Multiple random starts.*    The user can specify the number of random starts, as well as the seed for the random number generator (see Technical Appendix C, PRE-PROCESSING, page 155). The seed can be used to replicate the analyses with exactly the same numbers. From all random starts, the best solution is chosen, and replicated for output. The user specifies one of the following criteria by which the best solution is determined: P-STRESS, DAF, VAF, RHO, and KAPPA. These criteria are discussed in Technical Appendix G, RESULTS. Analyzing the statistics of multiple random starts provides insight in the reliability of the results, especially concerning the existence of local minima.

*Bootstrap.*    The bootstrap (Efron, 1982) is used to determine the reliability or stability of the unfolding solution. Heiser and Meulman (1983a) and Weinberg et al. (1984) already discuss the advantages of the bootstrap in case of an unknown distribution function and a complicated loss function as in the case of MDS and MDU models. Assuming a distribution function, as in probabilistic models, often provides a too optimistic stability picture (Weinberg et al., 1984). Using a bootstrap procedure, the stability of the solution is measured in terms of mean squared error, bias, and variance of the parameters. Following Heiser and de Leeuw (1979a) and van de Velden et al. (2010), the bootstrap procedure consists of the following steps:

Step 1.    Compute unfolding solution and keep results for further reference;
Step 2.    Draw bootstrap sample of $n$ rows from original preferences $\underline{\Delta}$;
Step 3.    Compute unfolding solution for bootstrap sample;
Step 4.    Use Procrustes to match bootstrap to original configuration;
Step 5.    Save intermediate, augmented bootstrap sample results;
Step 6.    Repeat Step 2–5 $R$ times;
Step 7.    Compute measures based on bootstrap sample results.

The unfolding analysis from Step 3 can start with any initial configuration, but in order to avoid local minima it is started with the final configuration from Step 1 (see Weinberg et al., 1984, p. 483). The bootstrap used here is a balanced bootstrap (see, for example, Davison, Hinkley, & Schechtman, 1986; Hinkley, 1988). This means that although the rows are randomly drawn without replacement, and within one bootstrap sample the rows are probably unequally distributed, after drawing $R$ bootstrap samples, each row is drawn $R$ times. Obviously, each column is also used (but not drawn) $R$ times. The Procrustes analysis (see page 159) from Step 4 only matches the column objects of the bootstrap and original configuration, since these points are always present in each bootstrap sample analysis. The number of bootstrap replications must be large ($R > 500$), due to the interest in variation and the tail of the distribution. The jackknife (another resampling technique) uses much less computation, but relies on a linear approximation of the statistic for the estimation of the accompanying standard error (Efron & Gong, 1983). The results of the $R$ bootstrap analyses are used to compute three measures to assess stability: Bias, variance, and mean squared error. All measures are computed for the row and column coordinates separately. The squared bias for row object $i$ is defined as

$$\text{BIAS}^2_{\mathbf{x}_i} = \left(\mathbf{x}_i - \overline{\mathbf{x}}^b_i\right)'\left(\mathbf{x}_i - \overline{\mathbf{x}}^b_i\right), \tag{F.1}$$

where $\mathbf{x}_i$ is the original coordinate of row object $i$ and $\overline{\mathbf{x}}^b_i = R^{-1}\sum_r \mathbf{x}^b_{ir}$ is the average over $R$ bootstrap replications of the coordinate of row object $i$. The squared bias is thus equal to the squared distance between the original

coordinate $x_i$ and the average of the bootstrap coordinates $\bar{x}_i^b$. The variance (although divided by $R$ instead of $R - 1$) for row object $i$ is defined as

$$\text{VAR}_{x_i} = \frac{1}{R} \sum_{r=1}^{R} \left(x_{ir}^b - \bar{x}_i^b\right)' \left(x_{ir}^b - \bar{x}_i^b\right), \tag{F.2}$$

which is the average of the squared distances between the bootstrap coordinates $x_{ir}^b$ and the bootstrap average $\bar{x}_i^b$. The sum of (F.1) and (F.2) defines the mean squared error for row object $i$ as

$$\text{MSE}_{x_i} = \text{BIAS}_{x_i}^2 + \text{VAR}_{x_i}$$

$$= \left(x_i - \bar{x}_i^b\right)' \left(x_i - \bar{x}_i^b\right) + \frac{1}{R} \sum_{r=1}^{R} \left(x_{ir}^b - \bar{x}_i^b\right)' \left(x_{ir}^b - \bar{x}_i^b\right)$$

$$= \frac{1}{R} \sum_{r=1}^{R} \left(x_{ir}^b - x_i\right)' \left(x_{ir}^b - x_i\right), \tag{F.3}$$

which is equal to the average of the squared distances between the bootstrap coordinates $x_{ir}^b$ and the original coordinate $x_i$.

Heiser and Meulman (1983a) compare MDS models on overall stability using the within and total ratio sum-of-squares, which is defined as

$$\text{STAB} = \frac{\sum_i \sum_r \left(x_{ir}^b - \bar{x}_i^b\right)' \left(x_{ir}^b - \bar{x}_i^b\right)}{\sum_i \sum_r \left(x_{ir}^b - \bar{x}^b\right)' \left(x_{ir}^b - \bar{x}^b\right)},$$

where $\bar{x}^b = (nR)^{-1} \sum_i \sum_r x_{ir}^b$. In Heiser and Meulman (1983a), $\bar{x}^b$ is assumed to be zero. Also in that publication, a clever procedure is used to obtain a plot of the confidence regions. The bootstrap distances are first subjected to an INDSCAL analysis (Pruzansky, 1975), of which the results are used as a target for an orthogonal Procrustes rotation (Cliff, 1966).

Instead, van de Velden et al., working on MDU models, use the average of (F.3) over row coordinates and column coordinates separately to ensure that both sets have an equal share in the assessment of stability, thus avoiding that the row objects, which are often in the majority, dominate the MSE measure. The normalization establishes a comparison with the variation of the original configuration. The stability measure of van de Velden et al. is now given as the relative mean squared error,

$$\text{REL-MSE} = \frac{1}{2} \frac{\sum_{i=1}^{n} \sum_{r=1}^{R} \left(x_{ir}^b - x_i\right)' \left(x_{ir}^b - x_i\right)}{R \sum_{i=1}^{n} \left(x_i - \bar{x}\right)' \left(x_i - \bar{x}\right)}$$

$$+ \frac{1}{2} \frac{\sum_{j=1}^{m} \sum_{r=1}^{R} \left(y_{jr}^b - y_j\right)' \left(y_{jr}^b - y_j\right)}{R \sum_{j=1}^{m} \left(y_j - \bar{y}\right)' \left(y_j - \bar{y}\right)}, \tag{F.4}$$

where $\overline{x} = n^{-1} \sum_i x_i$ and $\overline{y} = m^{-1} \sum_j y_j$, which are assumed zero in van de Velden et al. (2010). The REL-MSE measure equals 1.0 when the bootstrap variation around the unfolding solution is equal to the variation of the original solution. Smaller values are obviously more desirable, whereas solutions with values greater than 1.0 demonstrate larger bias and/or higher variation.

An optimal solution, concerning the penalty parameters, can be found by alternatingly running the bootstrap procedure for different values of one parameter, while fixing the other parameter, using the optimal parameter values in subsequent runs. Subsequent runs might employ smaller intervals for the trial values. Usually, after 2-4 cycles the procedure converges. Experiments show that, starting with $\lambda = 0.5$ and $\omega^* = 1.0$ and varying $\omega^*$ by doubling or halving its value repeatedly (1.0, 2.0, 0.5, 4.0, 0.25, ...), this procedure provides encouraging results.

Other measures, comparing the inter-set variability of the bootstrap coordinates, are currently being developed. An adequate unfolding solution should after all exhibit *non*-overlapping *intra-set* bootstrap clouds (a cloud consists of R bootstrap points for one object only) to display individual differences for row and column objects and overlapping *inter-set* bootstrap clouds to mark the relations between row and columns objects.

*Permutation test.* Following Manly (1991), ter Braak (1992b), ter Braak and Prentice (1988), ter Braak and Šmilauer (1998), and Lepš and Šmilauer (1999), the permutation test is used to determine the significance level of the correlation $r_{qp}$ between the independent variable $q$ and the projection of the coordinates (onto the variable's direction) $p$. The permutation procedure consists of the following steps:

Step 1.  Compute restricted unfolding solution and save results;
Step 2.  Permutate rows of independent variables matrix $E$;
Step 3.  Compute restricted unfolding solution with permutated $E$;
Step 4.  Save intermediate results, i.e., correlations $r_{qp}^p$;
Step 5.  Repeat Step 2–4 R times;
Step 6.  Compute significance level $p$ of $r_{qp}$ by using (F.5).

The procedure preserves the correlational structure of the variables, the preferences remain unchanged, but the relation between the variables and the preferences is altered. The null hypothesis is that the coordinates are independent of the variables. If that is the case, the distribution of the correlation structure remain the same in the original data and in the permutated data. The significance level $p$ from Step 6 is determined by

$$p = \frac{1 + \#\left(r_{qp}^p > r_{qp}\right)}{1 + R} \text{ or } p = \frac{1 + \#\left(|r_{qp}^p| > |r_{qp}|\right)}{1 + R}, \qquad (F.5)$$

for a one-tailed test or a two-tailed test respectively. The estimate of the p-value is improved by adding one extra result more extreme than the observed results. Other evaluations are also possible through a randomization or permutation test (see, for example, Wakeling, Raats, & MacFie, 1992). Fit statistics, for example, can be tested against the randomized data statistics using a Student's t-test (Wemelsfelder, Hunter, Mendl, & Lawrence, 2000). Note that exact permutation tests require all possible permutations, which for the current case entails $n! - 1$ possibilities. Random permutation tests use only a small portion of the possible replications, but require a mere 500 re-arrangements for a reliable estimate of the significance level, and even more than 10000 replications for accurate results.

## F.4    ADDITIONAL ANALYSES

After termination of the algorithm, determination of a unique solution, and, possibly, multiple analyses, additional analyses can be performed on request of the user for clarification and easier interpretation of the basic results.

*Property fitting.*    Independent variables can be used to facilitate the interpretation of the unfolding solution, i.e., the common space. In the following, the variables are discussed in relation to the row objects, but the same reasoning can be applied to the variables in relation to the column objects. A procedure based on linear regression (J. E. Miller, Shepard, & Chang, 1964) and a procedure based on optimizing an index of nonlinear correlation (Carroll & Chang, 1964a, 1964b) are implemented in the computer program PROFIT (Chang & Carroll, 1968) to perform *property fitting*. PREFSCAL only uses the first procedure that minimizes

$$f(\mathbf{Q}, \mathbf{A}) = \text{tr}\,(\mathbf{Q} - \mathbf{XA})'(\mathbf{Q} - \mathbf{XA}), \tag{F.6}$$

where $\mathbf{Q}$ is a matrix containing $h$ independent variables, the regression coefficients $\mathbf{A}$ are the so-called direction cosines (Chang & Carroll, 1968), and $\mathbf{X}$ is defined as before. The projection of the coordinates onto the directions of the variables are thus defined by $\mathbf{P} = \mathbf{XA}$. A solution for the *direction coefficients* $\mathbf{A}$ is given as

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}, \tag{F.7}$$

which is computed in alteration with a transformation update of the variables $\mathbf{Q}$. This also specifies the difference with PROFIT, which does not allow transformations of the independent variables.

By now, there are four (two times two) options for the independent variables of either set (rows or columns): The variables can be used to restrict the common space (direct, as compared to direct gradient analysis, see ter Braak

& Prentice, 1988) or the variables can be used afterwards, after the common space has been determined and is considered fixed (indirect, as compared to indirect gradient analysis, see ter Braak & Prentice, 1988). The other options are determined by what is regressed on what.

On the one hand, the variables used to restrict one set of coordinates as described in Technical Appendix E, CONFIGURATION UPDATE, can be used to find *optimal locations* for the objects and thus the coordinates $\mathbf{X}$ are regressed on the variables $\mathbf{Q}$, which essentially minimizes

$$f(\mathbf{Q}, \mathbf{B}) = \operatorname{tr}(\mathbf{X} - \mathbf{QB})'(\mathbf{X} - \mathbf{QB}),$$

providing a solution for the *regression coefficients* $\mathbf{B}$ as

$$\mathbf{B} = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{X}. \tag{F.8}$$

Gower and Hand (1996) refer to this process of finding a position in the display as *interpolation*.

On the other hand, *optimal quantifications* for the original variables are found by regressing the variables $\mathbf{Q}$ on the coordinates $\mathbf{X}$ as in (F.6) with the solution for the direction coefficients $\mathbf{A}$ given in (F.7), a process called *prediction* by Gower and Hand (1996). A summary of the options is given in Table F.1.

Restricting the coordinates $\mathbf{X}$ as $\mathbf{QB}$ and computing the direction coefficients $\mathbf{A}$ with fixed $\mathbf{X}$ and $\mathbf{Q}$ shows that the relation between $\mathbf{A}$ and $\mathbf{B}$ is given as $\mathbf{AB} = \mathbf{I}$ since

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q} = (\mathbf{B}'\mathbf{Q}'\mathbf{QB})^{-1}\mathbf{B}'\mathbf{Q}'\mathbf{Q}.$$

Depending on the researcher's intentions, one should use either $\mathbf{A}$ or $\mathbf{B}$ with corresponding optimally transformed variables $\mathbf{Q}$. Suppose a location $x_{n+1}$ is

Table F.1    Independent variables options summary[*].

| Goal | Variables | |
| --- | --- | --- |
| | Direct (Restriction) | Indirect |
| Optimal Locations (Interpolation) | estimate location given quantification(s) $\min \operatorname{tr}(\mathbf{X} - \mathbf{QB})'(\mathbf{X} - \mathbf{QB})$ set $\mathbf{X} = \mathbf{QB}$ use $\mathbf{B} = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{X}$ | estimate location given quantification(s) $\min \operatorname{tr}(\mathbf{X} - \mathbf{QB})'(\mathbf{X} - \mathbf{QB})$ use $\mathbf{B} = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{X}$ |
| Optimal Quantifications (Prediction) | estimate quantification(s) given location $\min \operatorname{tr}(\mathbf{X} - \mathbf{QB})'(\mathbf{X} - \mathbf{QB})$ set $\mathbf{X} = \mathbf{QB}$ use $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}$ (see Meulman & Heiser, 1984) | estimate quantification(s) given location $\min \operatorname{tr}(\mathbf{Q} - \mathbf{XA})'(\mathbf{Q} - \mathbf{XA})$ use $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}$ |

[*]Options are provided for the row objects only. Transfer to the columns objects is straightforward.
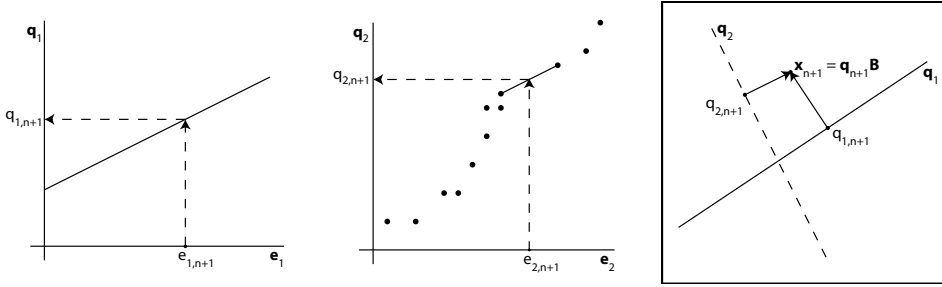
Figure F.1   Interpolation: Finding location $x_{n+1}$ based on variables values $e_{n+1}$.
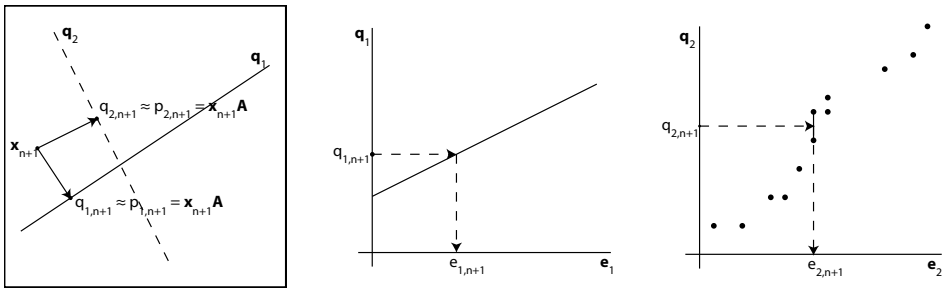


Figure F.2   Prediction: Finding quantifications $e_{n+1}$ based on location $x_{n+1}$.

searched for based on additional independent variables values $\boldsymbol{e}_{n+1}$ (interpolation). In this case, the transformed quantifications $\mathbf{q}_{n+1}$ are determined first, based on the known transformation functions $\mathbf{q} = f(\boldsymbol{e})$, as shown in Figure F.1 (left-hand panels). We then use the corresponding regression coefficients $\mathbf{B}$ to compute the location as $\boldsymbol{x}_{n+1} = \mathbf{q}_{n+1}\mathbf{B}$ (see Figure F.1, right-hand panel, the interpolative biplot, Gower & Hand, 1996). Whether $\mathbf{Q}$ and $\mathbf{B}$ are the result of a restricted or an unrestricted procedure is a matter of secondary concern. Instead, suppose we want to find quantifications $\boldsymbol{e}_{n+1}$ based on an additional location $\boldsymbol{x}_{n+1}$ (prediction). Then, we first project the coordinates $\boldsymbol{x}_{n+1}$ onto the direction of the independent variables $\mathbf{Q}$ to obtain the projected values $\mathbf{p}_{n+1}$ (Figure F.2, left-hand panel, the predictive biplot, Gower & Hand, 1996). Setting $\mathbf{q}_{n+1} = \mathbf{p}_{n+1}$ and using the known transformation functions $\mathbf{q} = f(\boldsymbol{e})$, original quantifications $\boldsymbol{e}_{n+1}$ are found by back-transformation (see Figure F.2, right-hand panels), i.e., $\boldsymbol{e} = f^{-1}(\mathbf{q})$ (see Gower & Hand, 1996; Gower et al., 1999).

Another method to find quantifications $\boldsymbol{e}_{n+1}$ based on (optimal) location $\boldsymbol{x}_{n+1}$ is given by Danzart, Sieffermann, and Delarue (2004) and Blumenthal (2004). Estimating $\boldsymbol{x}_{n+1}$ based on three close and surrounding points $\boldsymbol{x}_i$,

$x_j$, and $x_k$ as $x_{n+1} = w_i x_i + w_j x_j + w_k x_k$, where $w_i + w_j + w_k = 1$, provides weights $w$ that can be used to determine the weighted average $e_{n+1} = w_i e_i + w_j e_j + w_k e_k$. A solution for the weights is given by $w = (G'G)^{-1}G'h$, where $G = [(x_i, x_j, x_k), 1]'$ and $h = [x_{n+1}, 1]'$. This procedure, the *barycenter method*, is especially useful when products (locations) do not consist of actual sensory properties (quantifications), but the (optimal) product is composed of a mixture of products, such as, for example, blends of whiskies (Lee, Paterson, Piggott, & Richardson, 2001) or Muscadine juices (Meullenet et al., 2008).

References on the topics of projection, prediction, interpolation, and the interpretation of variables in configurations are Gower and Hand (1996) and Gower et al. (1999) in the context of biplots and de Leeuw and Heiser (1980), Heiser (1981), de Leeuw and Heiser (1982), Meulman and Heiser (1984), and ter Braak (1992a) in the context of multidimensional scaling. Especially, Meulman and Heiser (1984) is interesting in our context, since they show that despite optimizing the relation $X = QB$, the relation $Q = XA$ improves simultaneously. The argumentation is as follows. When $\text{tr}\,(U - QB)'(U - QB)$ is minimized with respect to $B$, where $U$ is the unrestricted configuration, this results in the restricted configuration $R = QB$ with projector $P = Q(Q'Q)^{-1}Q'$ (symmetric and idempotent) linking the restricted and the unrestricted configurations as $R = PU$. Meulman and Heiser now state that for the restricted configuration $R$, there are projections on some directions $A$ that are a better approximations of the independent variables than the projections of the unrestricted configuration on some directions $A$, as, with respect to $A$,

$$\min \text{tr}\,(Q - UA)'(Q - UA) \geqslant \min \text{tr}\,(Q - UA)'P(Q - UA)$$
$$= \min \text{tr}\,(Q - PUA)'(Q - PUA)$$
$$= \min \text{tr}\,(Q - RA)'(Q - RA),$$

which means that although restricted unfolding restricts the configuration to be a linear combination of variables as $X = QB$, at the same time the configuration holds improved projections of points onto the variables directions as $Q = XA$.

*Probabilistic distance clustering.* A probabilistic distance clustering analysis is performed on the row coordinates $X$ to facilitate the interpretation of the configuration. The following is based on Ben-Israel and Iyigun (2007) and Iyigun and Ben-Israel (2008). After computing initial cluster centers $c_k \,\forall\, k = 1, \ldots, K$, equally spaced and diagonal in the configuration as $c_{kp} = \text{argmin}\,(x_p) + k(\text{argmax}\,(x_p) - \text{argmin}\,(x_p))/(k+1)$, four steps are performed until convergence:

Step 1.   Compute distances $d(x_i, c_k)$ for all $i, k$;

Step 2. Update cluster sizes $q_k^+$ according to (F.9);

Step 3. Update the cluster centers $c_k^+$ according to (F.11);

Step 4. Check for convergence and if some predefined termination criterion is satisfied, continue; otherwise, go to Step 1.

It might occur that a row object coincides with a cluster center, in which case the distance between the row object and the cluster center is set to an arbitrary small value $\epsilon$. The cluster sizes are updated in Step 2 as

$$q_k^+ = n \left( \sum_{i=1}^{n} d(x_i, c_k) p_k^2(x_i) \right)^{.5} \left[ \sum_{l=1}^{K} \left( \sum_{i=1}^{n} d(x_i, c_l) p_l^2(x_i) \right)^{.5} \right]^{-1}, \quad (F.9)$$

where the membership probability for cluster $k$ of row object $i$ is defined as

$$p_k(x_i) = \left[ \sum_{j=1}^{K} \frac{d(x_i, c_k)/q_k}{d(x_i, c_j)/q_j} \right]^{-1}, \quad (F.10)$$

with $q_k$ and $q_j$ specifying the current size of cluster $k$ and $j$, respectively. When the cluster sizes are fixed, the cluster sizes remain untouched, obviously. The cluster centers $c_k$ from Step 3 are updated by

$$c_k^+ = \sum_{i=1}^{n} \frac{p_k^2(x_i)/d(x_i, c_k)}{\sum_{t=1}^{n} p_k^2(x_t)/d(x_t, c_k)} x_i, \quad (F.11)$$

with updated membership probabilities $p_k(x_i)$ and $p_k(x_t)$. Convergence is established when either the sum of the distances between cluster centers of two consecutive iterations is smaller than some critical value, or when the number of iterations exceeds some maximum.

An adjusted Calinsky-Harabasz index (original index from Calinski & Harabasz, 1974) is used to determine the optimal number of clusters, although substantive interpretation of the clusters must not be overlooked. The adjusted index is computed as

$$CH_{adjusted} = \frac{(n - K) \sum_{k=1}^{K} d(c_k, c)}{(K - 1) \sum_{i=1}^{n} d(x_i, c_{max(i)})}, \quad (F.12)$$

where $c$ is the median center of the cluster centers $c_k$ and $c_{max(i)}$ specifies the cluster with maximum probability for row object $i$.

*Analysis of angular variation.* The dimension weights of the non-identity models are often misinterpreted (Takane et al., 1977; MacCallum, 1977). The actual values of the weights must not be interpreted as points in space, but

as endpoints of vectors, of which the angles are interpretable. Directional statistics are thus needed to further (correctly) analyze the dimension weights (Heiser & Meulman, 1983a).

An analysis of angular variation (ANAVA) splits the total variation of the dimension weights (diagonal, full, and rectangular models only) in a between groups and a within groups part. The resulting F-statistic is too optimistic because the weights are not independent, but only conditional independent upon the given (unchanging) configuration. Nevertheless, an ANAVA is a good starting point to facilitate the interpretation of the unfolding results, as advised by Shepard (1972). Following Schiffman et al. (1981), the following steps are taken to perform an ANAVA in case of a three-way non-identity model. Transfer to the two-way model is straightforward and therefore omitted. First, the dimension weights are normalized such that the weights have equal length, that is,

$$w_{kp} = \frac{a_{kpp}}{\left(\sum_{t=1}^{p_{min}} a_{ktt}^2\right)^{.5}}, \tag{F.13}$$

with $a_{kpp}$ as the dimension weight for source $k$ and dimension $p$. Now, the weights have length 1 from the origin for each $k$. The mean direction per dimension is given as the average, $m_p = \frac{1}{s}\sum_{k=1}^{s} w_{kp}$, and the *mean resultant length* is given as $\overline{R} = (\sum_{t=1}^{p_{min}} m_t^2)^{1/2}$. A statistic that indicates the degree of variation in dimension weights, the *index of angular variation* (Mardia, 1972), is the counterpart of the mean resultant length, i.e., $\overline{S} = 1 - \overline{R}$. The index of angular variation can be divided into

$$s\overline{S} = \left(s - \sum_{t=1}^{s} g_t R_t\right) + \left(\sum_{t=1}^{s} g_t R_t - R\right),$$

where $R = s\overline{R}$ and $g_t$ specifies group membership (0/1). The analysis of angular variation results are given in Table F.2.

The index of angular variation ranges from 0 to 1, but for non-negative weights, which is the case here, the maximum is $1 - (1/p_{min})^{1/2}$. Computing $S^* = \overline{S}/(1 - (1/p_{min})^{1/2})$ sets the range for $S^*$ to $[0, 1]$, with values close to 0 indicating no or very little variation, and values close to 1 indication much variation in dimension weight directions. Mardia (1972) proposes to transform

Table F.2  Analysis of angular variation table.

| Source | SS | DF | MS | F |
|---|---|---|---|---|
| Between-Groups | $\sum_{t=1}^{s} g_t R_t - R$ | $(p_{min} - 1)(q - 1)$ | $SS_b/df_b$ | $MS_b/MS_w$ |
| Within-Groups | $s - \sum_{t=1}^{s} g_t R_t$ | $(p_{min} - 1)(s - q)$ | $SS_w/df_w$ | |
| Total | $m - R$ | $(p_{min} - 1)(s - 1)$ | | |

$S^*$ such that the range becomes $[0, \infty]$, like an ordinary standard deviation. The resulting *circular standard deviation index* is given by

$$s_w = \sqrt{-2 \log(1 - S^*)}.$$

Additionally, the *simplicity index* is computed, analogous to the characteristic of a simple structure matrix in factor analysis (Thurstone, 1947, p. 335), and named after Kaiser (1958), as he mentions the definition of *the simplicity of factorial composition* of a test or multiple tests. The simplicity index is identical to the weirdness index of F. W. Young (1982) (see also F. W. Young & Harris, 1997). The simplicity index has a value of 0 when a vector of dimension weights is proportional to the average vector, and a value of 1 for an extremely deviating vector. The value of 1 is only reached when one of the dimension weights is zero, and thus coincides with a simple structure as described by Thurstone (1947). The simplicity index is computed as follows: First, the weights are normalized so that the average vector has a 45° angle, $w'_{kp} = a_{kpp} / \sum_k a_{kpp}$, after which the normalized weights are set to unit length as in (F.13). The difference in angles can be defined as $\cos^{-1} p_{min}^{-.5} \sum_p w_{kp}$ and finally, the simplicity index is given as a value between 0 and 1 by

$$\widetilde{w}_k = \frac{\cos^{-1}\left(p_{min}^{-.5} \sum_{p=1}^{p_{min}} w_{kp}\right)}{\cos^{-1} p_{min}^{-.5}}. \tag{F.14}$$

Finally, *dimension importance* is determined as the explained sum-of-squares of the dimension weights,

$$I_p = \frac{\sum_{k=1}^{s} a_{kpp}^2}{\sum_{k=1}^{s} \sum_{t=1}^{p_{min}} a_{ktt}^2}. \tag{F.15}$$

It is also possible to perform a Watson-Williams test (Watson & Williams, 1956), which corresponds with an analysis of angular variation (Schiffman et al., 1981), although nonparametric tests (Wheeler & Watson, 1964, only for two dimensions) seem more appropriate, due to the problems with the distributional assumptions (dependence among angles) in the analysis of angular variation. Jones (1983) states that linear statistics can also be used, whereas Coxon and Jones (1974, 1978) use the logs of the within-subject ratio of the dimension weights for this purpose. Other references concerning this topic are Batschelet (1981), Bijleveld and Commandeur (1987), and van der Kloot, Bijleveld, and Commandeur (1990).

G
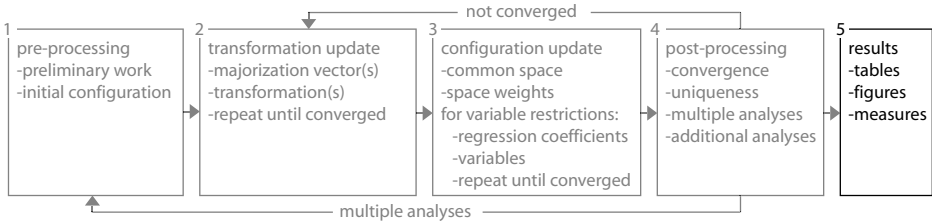


The results of an unfolding analysis consists of data and measures presented in tables and figures. These results are discussed here, with special attention for fit and degeneracy measures.

### G.1  TABLE OUTPUT

*Initial results.*     Input data, such as the original preferences $\underline{\Delta}$ and the preference weights $\underline{W}$, fixed coordinates and independent variables, are given for verification. The results of the pre-processing, discussed in Technical Appendix C, PRE-PROCESSING, which determine the input for the main algorithm, are: The initially transformed preferences $\underline{\Gamma}$ and the initial configurations $X_o$ and $Y_o$. Additionally, the eigenvalues resulting from initial configuration computations are given, as well as the additive constant when computed in advance of a classical scaling, that is, for the triangle and the spearman starts, and the dilation factor which adapts the configuration to match the preferences.

*Intermediate results.*     The main algorithm iterates over updates of the transformed preferences $\underline{\Gamma}$ and the configurations $X$ and $Y$, restricted or not, and the space weights $\underline{A}$. After each iteration, the iteration history is updated, which displays: Iteration number, number of inner iterations (transformation update), P-STRESS function value, difference with previous P-STRESS function value, and the decomposition of P-STRESS in (B.1) and (B.2), $\sigma_n^2(\gamma, \mathbf{d})$ and $\mu(\gamma)$, respectively.

For the option multiple random starts, the following results of each analysis are displayed: The number of the start, the initial seed for the start, and some fit and variation measures. Additionally, the initial seed (for the random number generator) is given, as well as the multiple random start criterion, which can either be P-STRESS, DAF, VAF, RHO, or KAPPA, all to be discussed hereafter.

*Fit-, variation-, and degeneracy measures.*    Several measures have been suggested to assess the quality of the unfolding solution. Every unfolding program specifies its own measures. For PREFSCAL, we attempt to provide all available and sensible measures. These measures will be discussed after the discussion of the figure output. PENALIZED STRESS and RAW STRESS decompositions per observation or per partition are discussed here.

Depending on the conditionality of the model, PENALIZED STRESS is decomposed in a STRESS and a penalty part. Both parts, (B.1) and (B.2), are given for each partition, that is, per row (row-conditional), per source (matrix-conditional), or for all observations simultaneously (unconditional). Additionally, RAW STRESS is decomposed for each cell of $\underline{\Gamma}$ and $\underline{D}$. Per row and per column, the averages and standard deviations are given.

Fit measures per partition are also provided, one for each measurement (transformation) level: SSAF for ratio transformed preferences, VAF for interval transformed preferences, and TAU for monotonically transformed preferences. Finally, a simple outlier analysis provides z-scores and Mahalanobis distances per row.

*Configurations.*    The main result of the unfolding analysis is the common space $\mathbf{X}, \mathbf{Y}$. The correlations between the dimensions are displayed for each set, since the dimensions of the common space are not necessarily orthogonal. For models not equal to the identity model, the individual spaces, $\mathbf{x}^i, \mathbf{Y}^i$ for the two-way models or $\mathbf{X}_k, \mathbf{Y}_k$ for the three-way models, and the space weights $\underline{\mathbf{A}}$ are provided in the output. The distances $d(\mathbf{X}, \mathbf{Y})$ and $d(\mathbf{x}^i, \mathbf{Y}^i)$ or $d(\mathbf{X}_k, \mathbf{Y}_k)$ are given for all configurations, upon request.

*Space weights.*    Irrespective of the orientation or size of the common space and the space weights, the space weights are additionally decomposed for the full and rectangular models for both two-way and three-way models. Hereafter, only the decompositions for the three-way model are given for one source $k$ at the time. Transfer to the two-way model is straightforward and therefore omitted.

For the decomposition of the space weights, the approach of de Leeuw and Heiser is adapted, as described in Commandeur (1994). After computing the singular-value decomposition $\mathbf{A}_k = \mathbf{P}_k \boldsymbol{\Phi}_k \mathbf{Q}_k'$, the space weights are divided into a rotation matrix $\mathbf{P}_k$ and a dimension weights matrix $\boldsymbol{\Phi}_k$. Since the individual coordinate matrices $\mathbf{X}_k$ and $\mathbf{Y}_k$ are post-multiplied with $\mathbf{Q}_k$, the common space remains unaltered as

$$\mathbf{Z}_k \mathbf{Q}_k = \mathbf{Z} \mathbf{A}_k \mathbf{Q}_k = \mathbf{Z} \mathbf{P}_k \boldsymbol{\Phi}_k \mathbf{Q}_k' \mathbf{Q}_k = \mathbf{Z} \mathbf{P}_k \boldsymbol{\Phi}_k,$$

where the complete individual configuration is specified as $\mathbf{Z}_k = [\mathbf{X}_k', \mathbf{Y}_k']'$ and the adapted individual spaces and space weights are defined as $\mathbf{X}_k^* = \mathbf{X}_k \mathbf{Q}_k$, $\mathbf{Y}_k^* = \mathbf{Y}_k \mathbf{Q}_k$, and $\mathbf{A}_k^* = \mathbf{P}_k \boldsymbol{\Phi}_k$.

Tucker (1972) and Harshman (1972) use the decomposition $A_k = S_k C_k S'_k$, where $S_k$ is a diagonal matrix and $C_k$ is a symmetric matrix with the diagonal elements equal to unity. In this way, $S_k$ can be interpreted as a matrix with standard deviations and $C_k$ as a correlation matrix. Computations are as follows for each source k: First compute $H_k = A_k A'_k$ and set the variances as $S_k^2 = \text{diag } H_k$. The correlation matrix is now set as $R_k = S_k^{-1} H_k S_k^{-1}$.

The difference between the two procedures is that the procedure of de Leeuw and Heiser first rotates (orthogonally) and then stretches the dimensions, while the procedure of Tucker and Harshman first stretches and then rotates (obliquely) the dimensions (Borg & Groenen, 2005). Furthermore, the decomposition of de Leeuw and Heiser coincides with the decomposition suggested by Carroll and Chang (1972) and described in Carroll and Wish (1974).

*Analysis of angular variation.*    For the dimension weights, we provide an analysis of angular variation table as described on page 204, together with the mean resultant length, the index of angular variation, and the circular standard deviation index. Finally, the simplicity index (F.14) is given per row (two-way model) or per source (three-way model), and the dimension importance (F.15) per dimension.

*Transformations.*    The optimally transformed preferences $\underline{\Gamma}$ are given and, in case of linear transformations, the regression coefficients are provided too.

*Independent variables.*    For both sets of objects, row and column objects, independent variables can be specified, either as a restriction (direct) or not (indirect). The output is the same for both sets and consists of all statistics described hereafter.

The optimally transformed variables $Q$ are displayed in combination with corresponding regressions coefficients $B$. These latter coefficients are the unstandardized regression coefficients (F.8), also called the canonical coefficients (ter Braak, 1986). The standardized regression coefficients are equal to the correlations between the variables $Q$ and the axes of the common space, i.e., the columns of $X$, also known as intra-set correlations $r(Q, X)$. These correlations are displayed as directions in the configuration by ter Braak (1992a) due to the strength interpretation of the values, but ter Braak agrees that the direction coefficients (F.7) leading to optimal projected values makes theoretical more sense to display.

The direction coefficients $A$ are given in combination with the projections $P = XA$, and the transformed independent variables $Q$. The squared correlation between the projection $p$ and the variable $q$ can be interpreted as the variance accounted for (VAF) by this variable. The average over direct and indirect variables separately indicates the fit achieved by the respective (group of) variables.

*Clustering.* The additional analysis distance clustering gives information about the number of clusters and the corresponding adjusted Calinsky-Harabasz index (F.12), as well as the resulting optimal number of clusters.

For the optimal number of clusters, or for the user-specified number of clusters, the distances $d(x_i, c_k)$ are given, and the probabilities per row and per cluster $p_k(x_i)$, according to (F.10), as well as the share of choices based on the distances and based on the probabilities and the average probabilities and distances for each cluster. The share of choices gives the proportion of row objects that have either the highest probability $p_k(x_i)$ for cluster k or the smallest distance $d(x_i, c_k)$ to cluster center $c_k$ for each cluster k. The average probabilities $\overline{p}_k$ and distances $\overline{d}(c_k)$ for each cluster k are computed as

$$\overline{p}_k = n^{-1} \sum_{i=1}^{n} p_k(x_i)$$

$$\overline{d}_k = n_k^{-1} \sum_{i=1}^{n} w_{ik} d(x_i, c_k),$$

where $n_k$ is the number of row objects with maximum probability for cluster k and $w_{ik} = 1$ for the cluster with maximum probability for object i and $w_{ik} = 0$ otherwise. We also provide the cluster centers $c_k$ as given by (F.11) for each cluster k, the average scores of the column objects per cluster,

$$\overline{\delta}_{jk} = \frac{\sum_{i=1}^{n} w_{ik} \delta_{ij}}{\overline{p}_k},$$

and the share proportion of the column objects per cluster,

$$\overline{s}_{jk} = \frac{\sum_{i=1}^{n} s_{ijk}}{\overline{p}_k},$$

where $s_{jk} = 1$ for $k = \max(i)$ and $j = \min(i)$, $s_{jk} = 0$ otherwise. Here, $j = \min(i)$ specifies the column object that is closest for object i. The values for $\overline{s}_{jk}$ provide information about which product coincides with which cluster.

*Bootstrap.* It is commonly known that resampling methods in general, and bootstrap procedures in particular, provide a profusion of information which must be carefully diminished to result in conveniently arranged tables and measures. After running the bootstrap procedure in PREFSCAL, tables are provided for row objects and column object separately, for squared bias (F.1), variance (F.2), and mean squared error (F.3). The aggregated measures are also broken down for the row objects and the column objects separately. Finally, stability measures are provided as described in Technical Appendix F, POST-PROCESSING, page 197.

*Permutation.*    The permutation procedure determines the significance levels of the correlations $r_{QP}$. The output is arranged per variable and consists of the variable number, the true (non-permutated) value of the correlation, the minimum, mean, and maximum value over permutations, and the p-value, following (F.5).

### G.2    FIGURE OUTPUT

Whether it was Fred R. Barnard who used it in an article to promote the use of images in advertisements "One Look is Worth A Thousand Words" (Barnard, 1921) or "One Picture is Worth Ten Thousand Words" (Barnard, 1927) and labeled it a Chinese proverb, soon to be attributed to Confucius, or whether it was the Russian writer Ivan Turgenef, who wrote: "A picture shows me at a glance what it takes dozens of pages of a book to expound" (Turgenef, 1867) or whether it was Napoleon Bonaparte who said: "Un bon croquis vaut mieux qu'un long discours" ("a good sketch is better than a long speech") (source: wikipedia), fact remains that a configuration or a transformation plot is appealing, comprehensive, and easily interpretable, even for non-mathematicians.

*Scree plot.*    A scree plot displays functions values, or derivatives thereof, against the dimensionality of the model. This plot can only be produced when the maximum dimensionality differs from the minimum dimensionality, that is, when $p_{max} > p_{min}$. The scree plot allows one to visually inspect the optimal model dimensionality. The optimal dimensionality occurs at the number of dimensions indicated by the elbow. This criterion differs from the criterion used in factor analysis, where the number *before* the elbow indicates the number of latent factors (Cattell, 1966).

*Configurations.*    A configuration is a scatter plot with markers and/or colors indicating the row objects, the column objects, or both row- and columns objects. Besides the common space, a plot representing the row object coordinates $X$ and the column object coordinates $Y$, configurations are plotted for the initial configuration $X_o, Y_o$, the individual spaces for the two-way models $x^i, Y^i$ and for the three-way models $X_k, Y_k$ and a plot for the dimensions weights. Plots for the individual spaces and the dimension weights are omitted for the identity model. Note that it is of the utmost importance that the aspect ratio of these configurations is equal to 1.0. Changing the aspect ratio changes the distances and a different (non-optimal) solution remains. Unfortunate examples can be found in the results of IBM SPSS ALSCAL. The independent variables, when present, are plotted as lines, solid lines for numerical variables and dashed lines for categorical variables, with originally valued markers calibrated on the lines.
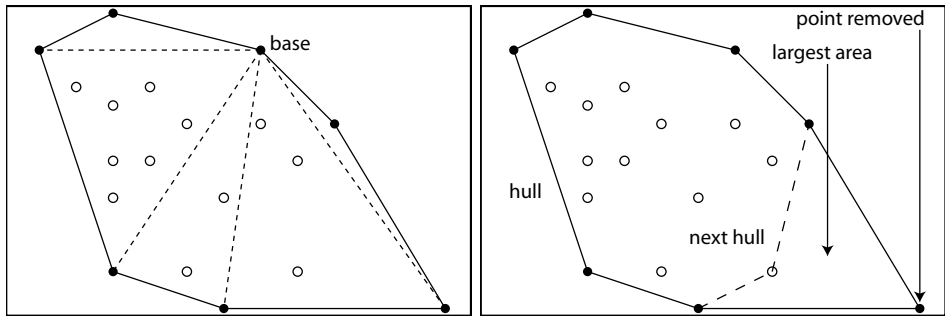
Figure G.1    Peeling the hull based on iterative polygon area minimization.

*Transformation plots.*    The transformation plots give the original preferences on the horizontal axes and the optimally transformed preferences on the vertical axes. Depending on the conditionality of the unfolding model, separate plots are provided for the rows or sources (matrices).

*Fit and Residual plots.*    The fit plots and the residual plots provide the same information but in a different form. A fit plot gives the optimally transformed preferences on the horizontal axis and the distances on the vertical axis. A good fit is thus provided by a 45° line from the origin in the lower left-hand corner to the upper right-hand corner. A residual plot gives the optimally transformed preferences on the horizontal axis and the *difference* between the optimally transformed preferences and the distances on the vertical axis. Here, a good fit is recognized by a horizontal line starting from the origin.

*Independent variable plots.*    For the independent variables, transformation and regression plots are produced for each variable. A transformation plot gives the original independent variable $e$ on the horizontal axis and the optimally transformed independent variable $q$ on the vertical axis. A regression plot gives either $X$ or $Q$ on the horizontal axis and $X = QB$ or $P = XA$, respectively, on the vertical axis. The former regression plot is only provided for the unrestricted case.

*Nonparametric confidence intervals.*    Nonparametric (emphasizing the distribution free least squares approach) confidence intervals are plotted to facilitate the interpretation of the unfolding configuration. The computation of a convex hull is based on the procedure suggested by P. J. Green and Silverman (1979) and executed with the envelope code by A. Miller (1987). Peeling the hull one point at the time is based on the size of the area of the polygon, computed by adding all triangle areas from one base point (see Figure G.1, left-hand panel)

The point removed in subsequent steps is the point that minimizes the area of the polygon (see Figure G.1, right-hand panel). The process is repeated until the desired percentage of points is reached. The confidence intervals are plotted on request of the user, after providing one or more vectors with zero's (not included in the hull computations) and ones (included in the hull computations) to indicate the group of row objects to be taken into account (for example, husbands or wives in the breakfast data), or the intervals are plotted to show the clusters after the distance clustering analysis. The perunage (percentage/100) or proportion (default equals one) for the interval can be provided by the user.

## G.3   FIT MEASURES

*Fit measures comparing $\gamma$ and $\mathbf{d}$*

PREFSCAL optimizes a loss function that compares the transformed preferences $\gamma$ and the distances $\mathbf{d}$. Other loss functions are minimized by other programs and PREFSCAL provides as many loss functions as possible. These functions are averaged over partitions (Kruskal & Carroll, 1969), so for clarity only the unconditional (two-way model) version is provided.

For an objective comparison between $\gamma$ and $\mathbf{d}$, where $\gamma$ might also be the distance vector of another configuration when comparing configurations, or $\mathbf{d}$ is the result of a fixed configuration from another analysis, note that the optimal dilation factor $\alpha$ for the distances is found by minimizing the loss function with respect to $\alpha$. The optimal values for $\alpha$ are provided for each loss function.

*R-STRESS.*   RAW STRESS is defined as the weighted sum of squared differences between $\gamma$ and $\mathbf{d}$:

$$\sigma_r = \|\gamma - \mathbf{d}\|_W^2 = \gamma' W \gamma + \mathbf{d}' W \mathbf{d} - 2\gamma' W \mathbf{d} \qquad \text{(G.1)}$$

Despite the optimal dilation factor $\alpha = (\gamma' W \mathbf{d})/(\mathbf{d}' W \mathbf{d})$ and taking the average over all difference, the magnitude of R-STRESS is scale-dependent and therefore not shown in the output, but mentioned here because of the important position among other STRESS functions.

*N-STRESS.*   NORMALIZED RAW STRESS is defined as RAW STRESS divided by (normalized with) the weighted sum of squared transformed preferences $\gamma$:

$$\sigma_n = \frac{\|\gamma - \mathbf{d}\|_W^2}{\|\gamma\|_W^2} = \frac{\gamma' W \gamma + \mathbf{d}' W \mathbf{d} - 2\gamma' W \mathbf{d}}{\gamma' W \gamma}.$$

N-STRESS is the function that is used by IBM SPSS PROXSCAL (Commandeur & Heiser, 1993; Busing, Commandeur, & Heiser, 1997; Meulman et al., 1999). Due to explicit normalization in PROXSCAL by letting $\gamma'W\gamma = 1$, N-STRESS is equivalent to R-STRESS. In 1977, de Leeuw and Heiser proved, that for a local minimum of N-STRESS, it holds that $d'Wd = \gamma'Wd$. Therefore, after convergence, $\alpha = (\gamma'Wd)/(d'Wd) = 1$. Substituting $\alpha$ (R-STRESS) in (G.1) and dividing both sides by $\gamma'W\gamma$ gives

$$\frac{\sigma_r(\alpha)}{\gamma'W\gamma} + \frac{(\gamma'Wd)^2}{(d'Wd)(\gamma'W\gamma)} = 1, \tag{G.2}$$

which links NORMALIZED RAW STRESS with dispersion accounted for (DAF) (after Commandeur & Heiser, 1993, p. 70). Using rank images, N-STRESS is equivalent to the alienation coefficient K (see PHI).

*STRESS-1.* The square of Kruskal's STRESS-1 (Kruskal & Carroll, 1969) is defined as the sum of squared differences between $\gamma$ and $d$, divided by the sum of squared distances $d$,

$$\sigma_1^2 = \frac{\|\gamma - d\|_W^2}{\|d\|_W^2} = \frac{\gamma'W\gamma + d'Wd - 2\gamma'Wd}{d'Wd}. \tag{G.3}$$

Substituting the optimal dilation factor $\alpha = (\gamma'W\gamma)/(\gamma'Wd)$ in (G.3) and using (G.2) shows that, in a local minimum, the square of Kruskal's STRESS-1 is equal to NORMALIZED RAW STRESS as

$$\sigma_1^2 = \frac{\gamma'W\gamma + (\gamma'Wd)^{-2}(\gamma'W\gamma)^2(d'Wd) - 2(\gamma'Wd)^{-1}(\gamma'W\gamma)(\gamma'Wd)}{(\gamma'Wd)^{-2}(\gamma'W\gamma)^2(d'Wd)}$$

$$= 1 - \frac{(\gamma'Wd)^2}{(d'Wd)(\gamma'W\gamma)}$$

$$= \frac{\sigma_r}{\gamma'W\gamma}$$

$$= \sigma_n.$$

*STRESS-2.* The square of Kruskal's STRESS-2 (Kruskal & Carroll, 1969; Kruskal et al., 1978) is defined as the sum of squared differences between $\gamma$ and $d$, divided by the variance of the distances. Defining $\overline{d} = (nm)^{-1}\sum_i\sum_j d_{ij}$ and $\overline{d} = 1\overline{d}$ gives STRESS-2 as

$$\sigma_2^2 = \frac{\|\gamma - d\|_W^2}{\|d - \overline{d}\|_W^2} = \frac{\gamma'W\gamma + d'Wd - 2\gamma'Wd}{d'Wd + \overline{d}'W\overline{d} - 2d'W\overline{d}}.$$

The optimal dilation factor $\alpha$ is equal to $(\gamma'W\gamma)/(\gamma'Wd)$.

*S-STRESS-1.* Young's s-STRESS-1 (Takane et al., 1977) uses *squared* preferences and *squared* distances instead of preferences and distances. For the rest, s-STRESS-1 is equal to STRESS-1, so with $\alpha^2$ specifying that each element of vector $\alpha$ is squared, s-STRESS-1 is given by

$$\sigma_{S_1}^2 = \frac{\|\gamma^2 - d^2\|_W^2}{\|d^2\|_W^2} = \frac{\gamma^{2\prime}W\gamma^2 + d^{2\prime}Wd^2 - 2\gamma^{2\prime}Wd^2}{d^{2\prime}Wd^2}.$$

The optimal dilation factor $\alpha = (\gamma^{2\prime}W\gamma^2)/(\gamma^{2\prime}Wd^2)$ is the reverse of the one mentioned in Takane et al. (1977), because it is applied to the preferences instead of the distances as in our case.

s-STRESS-1 is minimized by IBM SPSS ALSCAL . Note that minimizing s-STRESS-1 emphasizes large preferences and large distances. This can be illustrated as follows. Assume that s-STRESS is given as $\|\delta^2 - d^2\|^2$. Then,

$$
\begin{aligned}
\|\delta^2 - d^2\|^2 &= \sum_i \sum_j \left(\delta_{ij}^2 - d_{ij}^2\right)^2 \\
&= \sum_i \sum_j \left(\delta_{ij} + d_{ij}\right)^2 \left(\delta_{ij} - d_{ij}\right)^2 \\
&= \sum_i \sum_j w_{ij}\left(\delta_{ij} - d_{ij}\right)^2,
\end{aligned}
$$

where $w_{ij} = (\delta_{ij} + d_{ij})^2$. Assuming relatively small residuals, i.e., $\delta_{ij} \approx d_{ij}$, it shows that s-STRESS is similar to STRESS with weights equal to $4\delta_{ij}^2$ (Heiser, 1988; Borg & Groenen, 2005), thus illustrating the (over-)emphasis of large preferences. Weinberg and Menil (1993) state that this is a reason for not using ALSCAL: Since the error variances of dissimilarities tend to be positively correlated with their means, large dissimilarities should be down-weighted, if anything, relative to small dissimilarities.

*S-STRESS-2.* Young's s-STRESS-2 also uses squared variants of the preferences and the distances. Otherwise, the function is equal to Kruskal's STRESS-2 and given as

$$\sigma_{S_2}^2 = \frac{\|\gamma^2 - d^2\|_W^2}{\|d^2 - \overline{d}^2\|_W^2} = \frac{\gamma^{2\prime}W\gamma^2 + d^{2\prime}Wd^2 - 2\gamma^{2\prime}Wd^2}{d^{2\prime}Wd^2 + \overline{d}^{2\prime}W\overline{d}^2 - 2d^{2\prime}W\overline{d}^2},$$

with the optimal dilation factor $\alpha$ is equal to $(\gamma^{2\prime}W\gamma^2)/(\gamma^{2\prime}Wd^2)$.

*DAF.* Dispersion accounted for (DAF) is a goodness-of-fit statistic, dividing the co-sum-of-squares by the product of the sum-of-squares (compare co-variance and standard deviations), which is equal to one minus N-STRESS . As before, defining R-STRESS as

$$\sigma_r = \gamma'W\gamma + d'Wd - 2\gamma'Wd,$$

dilating the distances with $\alpha$, normalizing with $\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}$, and re-arranging terms gives (G.2). Since in a local minimum $\boldsymbol{d}'\boldsymbol{W}\boldsymbol{d} = \boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{d}$,

$$1 = \frac{\sigma_r}{\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}} + \frac{(\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{d})^2}{(\boldsymbol{d}'\boldsymbol{W}\boldsymbol{d})(\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma})} = \frac{\sigma_r}{\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}} + \frac{\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{d}}{\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}} = \sigma_n + \frac{\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{d}}{\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}},$$

which divides one in a badness-of-fit part, N-STRESS, and a goodness-of-fit part,

$$\text{DAF} = \frac{(\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{d})^2}{(\boldsymbol{d}'\boldsymbol{W}\boldsymbol{d})(\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma})} = \frac{\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{d}}{\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}}. \tag{G.4}$$

Dispersion accounted for (Heiser & Groenen, 1997) is also referred to as sum-of-squares accounted for (SSAF) by DeSarbo and Carroll (1985), although DeSarbo and Carroll match the preferences with *squared* distances and always compute SSAF unconditionally. SSAF is used as one of the convergence criteria in GENFOLD (DeSarbo & Rao, 1984; DeSarbo & Carroll, 1985).

*PHI.* The square root of DAF (or SSAF) is equal to Tucker's congruence coefficient $\phi$ (Burt, 1948a; Tucker, 1951; Wrigley & Neuhaus, 1955),

$$\text{PHI} = \sqrt{\frac{\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{d}}{\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}}},$$

which makes the *squared* congruence coefficient the counterpart of N-STRESS too, i.e., $\text{PHI}^2 = 1 - \text{N-STRESS}$. Using distances and rank images, PHI is equivalent to Guttman's (1981) monotonicity coefficient $\mu$ and related to N-STRESS, STRESS-1, and alienation coefficient K as $K = (1 - \phi^2)^{1/2} = (1 - (1 - \sigma_n))^{1/2} = \sigma_n^{1/2} = \sigma_1$ (see Borg & Groenen, 2005). Due to the normalization factors (the denominators), the function values of N-STRESS, STRESS-1, DAF, SSAF, and PHI are insensitive to differences in scale and/or sample size and thus well-suited to objectively compare distances of different multidimensional scaling configurations.

*VAF.* Variance accounted for (VAF) is defined as the squared correlation between $\boldsymbol{\gamma}$ and $\boldsymbol{d}$. Pearson's correlation $r$ is a measure for direction and strength of a linear relation between two variables of (at least) the interval measurement level. The correlation coefficient can be computed as

$$r = \frac{1}{nm-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \left( \frac{\gamma_{ij} - \overline{\gamma}}{s_\gamma} \right) \left( \frac{d_{ij} - \overline{d}}{s_d} \right), \tag{G.5}$$

but in PREFSCAL a weighted version is computed which defines variance accounted for as

$$\text{VAF} = \frac{\left( (\boldsymbol{1}'\boldsymbol{W}\boldsymbol{1})(\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{d}) - (\boldsymbol{1}'\boldsymbol{W}\boldsymbol{\gamma})(\boldsymbol{1}'\boldsymbol{W}\boldsymbol{d}) \right)^2}{(\boldsymbol{1}'\boldsymbol{W}\boldsymbol{1})(\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}) - (\boldsymbol{1}'\boldsymbol{W}\boldsymbol{\gamma})^2(\boldsymbol{1}'\boldsymbol{W}\boldsymbol{1})(\boldsymbol{d}'\boldsymbol{W}\boldsymbol{d}) - (\boldsymbol{1}'\boldsymbol{W}\boldsymbol{d})^2}.$$

Just like for the combination of N-STRESS and DAF, STRESS-2 and VAF are counterparts as

$$\text{VAF} = 1 - \frac{\|\boldsymbol{\gamma} - \mathbf{d}^2\|_W^2}{\|\boldsymbol{\gamma} - \overline{\boldsymbol{\gamma}}\|_W^2},$$

as described by DeSarbo and Carroll (1985, p. 296), where it is used as one of the convergence criteria for GENFOLD , especially for interval scaled data. Note that GENFOLD minimizes squared distances, always computes VAF unconditionally, and that the STRESS-2 function uses $\boldsymbol{\gamma}$ in the denominator for the normalization.

*Fit measures comparing rank($\boldsymbol{\delta}$) and rank($\mathbf{d}$)*

There are several functions that measure the recovery of the preferences by the distances. Since the transformed preferences might form an disturbing intermediate entity, we can use the rank orders of both $\boldsymbol{\delta}$ and $\mathbf{d}$ for the comparison, thus circumventing the transformations.

FIRST.    The statistic 'first choices' (FIRST) only takes the first choice of a row (usually a respondent) into consideration. This is especially important for discrete or deterministic choice processes. When there are tied first choices, all these choices are considered the first choice. When there are tied smallest distances, something that is not likely to happen (much), all distances are considered smallest. FIRST is defined as

$$\text{FIRST} = (sn)^{-1} \sum_{k=1}^{s} \sum_{i=1}^{n} \#\{\operatorname{argmin}(\delta_{i1k}, \ldots, \delta_{imk}) = \operatorname{argmin}(d_{i1k}, \ldots, d_{imk})\}$$

FIRST is computed per row, averaged over rows, and results in values between zero and one.

ORDERS.    The recovered preference orders provides the percentage of pairwise compared preferences recovered by the distances. For this purpose, a concordant pair is defined as $\operatorname{sign}(\delta_i - \delta_j) = \operatorname{sign}(d_i - d_j)$, where $\operatorname{sign}(x) = 1$ for $x \geqslant 0$ and $\operatorname{sign}(x) = -1$ for $x \leqslant 0$. Ties contribute positively to the result. The statistic ORDERS is now defined as the number of concordant pairs $n_c$ relative to the total number of pairs (total),

$$\text{ORDERS} = \frac{n_c}{\text{total}}.$$

ORDERS is computed per partition, averaged over partitions, and possible values for ORDERS range from zero to one.

*RHO.* Spearman's rank order correlation RHO ($\rho$) is commonly used as a nonparametric measure of correlation between two ordinal variables (see Kendall, 1948). For all non-missing cases, the values of original preferences and the distances are ranked from smallest to largest and the Pearson's correlation coefficient r (G.5) is computed on the ranks. RHO is computed per partition and averaged over partitions. The sign of the coefficient indicates the direction of the relationship, and its absolute value indicates the strength, with larger absolute values indicating stronger relationships. Possible values range from minus one to plus one. Significance of RHO is determined by $z = \rho \sqrt{n-1}$ for $n > 2000$ using standard normal probabilities and by $t = \rho \sqrt{n-2}/\sqrt{1-\rho^2}$ with $df = n - 2$ and the critical values of a t-distribution otherwise.

*TAU.* Kendall's rank order correlation TAU ($\tau$) is a nonparametric measure of association for ordinal or ranked variables that takes ties into account ($\tau_b$ for square tables and $\tau_c$ for rectangular tables) (Kendall, 1938, 1948). PREFSCAL displays TAU ($\tau_b$) as

$$\text{TAU} = \frac{n_c - n_d}{n(n-1)/2} \tag{G.6}$$

where $n_c$ is the number of concordant pairs, $n_d$ is the number of discordant pairs, and $n(n-1)/2$ is the total number of pairs ($n$ depends on the chosen conditionality). A concordant pair is defined as $\text{sign}\,(\delta_i - \delta_j) = \text{sign}\,(d_i - d_j)$ and a discordant pair as $\text{sign}\,(\delta_i - \delta_j) = -\text{sign}\,(d_i - d_j)$, where $\text{sign}\,(x) = -1$ for $x < 0$, $\text{sign}\,(x) = 1$ for $x > 0$, and $\text{sign}\,(x) = 0$ for $x = 0$. For ties, the denominator is corrected accordingly. TAU is computed per partition and averaged over partitions. The sign of the coefficient indicates the direction of the relationship, and its absolute value indicates the strength, with larger absolute values indicating stronger relationships. Possible values range from minus one to plus one, but a value of minus one or plus one can only be obtained from square tables. Significance for TAU is determined by $z = 3\tau \sqrt{n(n-1)}/\sqrt{2(2n+5)}$ using standard normal probabilities.

*Fit measures comparing $\delta$ and $\widehat{\delta}$*

The original preferences are transformed to obtain the transformed preferences. The transformed preferences are least squares approximated by the distances. Recovery of the original preferences can thus be determined by reversing the optimization process: The distances are back-transformed (or inversely transformed) to the scale of the original preferences and the results are used to compute a recovery or fit measure, which depends on the scale type of the original preferences.

The computation of the inverse transformation depends on the transformation function. For a linear transformation (see page 170), the inverse transformation is given by

$$\widehat{\delta}_{ijk} = \frac{d_{ijk} - b_{1ik}}{b_{2ik}},$$

where the estimation of the original preference for row object $i$, column object $j$, and source $k$ depends on the distance between row object $i$ and column object $j$ for source $k$, and the linear transformation parameters $b_{1ik}$ and $b_{2ik}$. Obviously, the linear transformation parameters depend on the conditionality of the unfolding model, such that for the unconditional model $b_{1ik} = b_1$ and $b_{2ik} = b_2$ and for the matrix-conditional model $b_{1ik} = b_{1k}$ and $b_{2ik} = b_{2k}$.

For an ordinal transformation, the distances are compared to the results of the monotone regression procedure. Linear interpolation and extrapolation are used to compute the estimated value $\widehat{\delta}_{ijk}$ for $\delta_{ijk}$, when the distance does not coincide with one of the existing values. Interpolation is used when the distance falls in between two values and extrapolation is used when the distance falls outside the range of values.

*SSAF.* For a numerical scale type, the sum-of-squares accounted for (SSAF) is computed between the original preferences $\underline{\Delta}$ and the estimated preferences $\widehat{\underline{\Delta}}$, taken into account the conditionality. For the unconditional model,

$$\text{SSAF} = 1 - \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{s} \left( \delta_{ijk} - \widehat{\delta}_{ijk} \right)^2}{\sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{s} \delta_{ijk}^2},$$

which is identical to (G.4). SSAF is averaged over partitions.

*TAU.* For an ordinal scale type, Kendall's rank order correlation TAU ($\tau$) is computed, comparing the ranks of the original preferences $\underline{\Delta}$ and the estimated preferences $\widehat{\underline{\Delta}}$ per partition. TAU is computed according to (G.6), where $\underline{D}$ is replaced with $\widehat{\underline{\Delta}}$.

*KAPPA.* Cohen's kappa coefficient $\kappa$ measures the agreement between the original preferences and the estimated preferences when the preferences are categorical data (with a limited number of categories), taking into account the agreement occurring by chance. PREFSCAL gives the weighted version of $\kappa$, which is defined as

$$\text{KAPPA} = \frac{\sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} p_{ij} - \sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} p_{i+} p_{+j}}{1 - \sum_{i=1}^{c} \sum_{j=1}^{c} w_{ij} p_{i+} p_{+j}},$$

where c is the number of categories, $p_{ij}$ is the proportion observed in cell $i, j$, $p_{i+}$ and $p_{+j}$ are the row and column totals, respectively, and $W$ is defined with elements $w_{ij} = 1 - (i - j)^2 / (k - 1)^2$. Key references are: Galton (1892), Cohen (1960, 1968), and Fleiss and Cohen (1973).

## G.4  VARIATION MEASURES

For the original preferences $\underline{\Delta}$, the transformed preferences $\underline{\Gamma}$, and the distances $\underline{D}$, variances and variation coefficients are reported. For $\underline{\Delta}$ and $\underline{\Gamma}$, the computation of the variation coefficients depend on the conditionality of the model.

*Variance.*   The variance or the standard deviation are variation measures that depend on the magnitude of the data. When the data is scaled up by a factor 3, the variance is scaled up by a factor $3^2$. Since PREFSCAL is implicitly normalized, the scales of the configurations are not constant. Variances and standard deviation should thus be interpreted with care. Whether the variance is computed over the complete data array $\underline{\Gamma}$, or averaged over partitions, is indicated in the output.

*Variation.*   The coefficient of variation (Pearson, 1896) is defined as the standard deviation divided by the mean. For the transformed preferences, the coefficient of variation is computed per partition. Since the variation of the preferences is observed per partition, the variation coefficient is computed per partition and averaged over partitions. In order to stay close to the penalty of the PENALIZED STRESS function, the harmonic mean is used for this average. Note that the harmonic mean emphasizes small values, so a very small variation in one of the partitions is not likely to be compensated by larger variations in other partitions. Since the distances are observed in the context of a configuration, the coefficient of variation for the distances are *not* computed per partition, but for the distance array ($\mathbf{D}$ or $\underline{D}$) as a whole.

## G.5  DEGENERACY INDICES

*D-INDEX.*   Shepard (1974) describes "a rough index of the *non*-degeneracy of a solution" as the ratio of the number of distinct distances to the total number of distances. Distinctness was difficult to define and a slightly modified version as compared to Shepard (1974) is used in PREFSCAL. Distinctness is defined as

$$v_{ij}(\mathbf{d}) = \begin{cases} 1 & \text{if } (d_i - d_j)/(d_i + d_j) > 0.1, \\ 0 & \text{if } (d_i - d_j)/(d_i + d_j) \leqslant 0.1. \end{cases}$$

Each distance is compared to every other distance and considered distinct whenever the ratio of the difference and the sum of the two distances exceeds 0.1. The sum of distinct distances is divided by the total number of comparisons to obtain the distinctness index

$$\text{D-INDEX} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \nu_{ij}(\mathbf{d})}{nm(nm-1)/2}$$

D-INDEX is computed per partition (row, matrix, or complete array), depending on the conditionality of the model, and then averaged over the partitions. The distinctness index is also computed for the original preferences $\underline{\Delta}$ and the transformed preferences $\underline{\Gamma}$.

I-INDEX.  DeSarbo, Young, and Rangaswamy (1997) introduce three indices that measure the intermixedness of two sets of points. The indices stem from the observation that degenerate solutions often show separated sets of row and column points, or, alternatively, show large differences in average within-set and/or between-set distances. The average within-set distances are defined as

$$\overline{d}_x = \frac{1}{n(n-1)/2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left( \sum_{p=1}^{p_{min}} (x_{ip} - x_{jp})^2 \right)^{.5}$$

$$\overline{d}_y = \frac{1}{m(m-1)/2} \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} \left( \sum_{p=1}^{p_{min}} (y_{jp} - y_{ip})^2 \right)^{.5},$$

whereas the average between-set distance is defined as

$$\overline{d}_{xy} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} d_{ij}.$$

Now, DeSarbo et al. define three indices to flag common types of degeneracies, based on these average distances, as

$$I_1 = \log \frac{\overline{d}_x}{\overline{d}_{xy}}, \quad I_2 = \log \frac{\overline{d}_y}{\overline{d}_{xy}}, \text{ and } I_3 = \log \frac{\overline{d}_x}{\overline{d}_y}.$$

These indices can be used as a diagnostic aid and aim to indicate a well-interspersed configuration when all three indices are near zero, or, equivalently, when the sum-of-squares of the indices is close to zero (Busing, Groenen, & Heiser, 2005). The sum-of-squares of the indices is reported by PREFSCAL as the *intermixedness index*, I-INDEX $= I_1^2 + I_2^2 + I_3^2$.

This glossary contains terms directly related with degenerate solutions, which, in general, are zero STRESS solutions that fail to preserve the structure that is contained in the data. Other terms can be found through the subject index.

*Trivial solution* — A perfect but meaningless solution having zero STRESS. Often, this solution can be avoided by using a proper normalization of the STRESS function. The most renowned trivial solutions are the two-point or object-sphere solution for unconditional unfolding using RAW STRESS, the two-plus-two-point solution using STRESS-1, and the object-point solution for row-conditional unfolding using STRESS-2 with an improper summation for the normalization factor.

*Absolutely or completely degenerate solution* — A flawless but purposeless zero STRESS solution without variation in the transformed preferences or the distances that *cannot* be avoided by using a proper normalization of the STRESS function. An absolutely degenerate solution resembles a trivial solution, such as the object-sphere solution, but differs in computational respect to that solution.

*Partially degenerate solution* — An impeccable but ineffectual *row-conditional* zero STRESS solution without variation in the rows of the transformed preferences or the distances that *cannot* be avoided by using a proper normalization of the STRESS function. A partially degenerate solution resembles a trivial solution, for example an object-point solution, but deviates in view of computational aspects.

*Extremely, near(ly), quasi, strongly, or approximately degenerate solution* — A solution that tends to a absolutely or partially degenerated solution, but deviates from this solution due to non-convergence, local minima, or other anomalies. Such a solution is difficult to identify, because only part of the solution is absolutely or partially degenerate. Commonly, this solution shows conspicuously low variation in either the transformed preferences or the distances and has a strikingly low STRESS value.

For the following solutions, subjects (rows) are represented by a small dot (one subject) or a large dot (multiple subject) and objects (columns) are represented by a plus sign (one object) or a cross (multiple objects).

*Two-point solution* — This zero STRESS solution holds only one distance value for all subject-object distances. This equal distance solution is the result of minimizing RAW STRESS without a normalization factor. The transformation of the preferences results in one value for all transformed preferences, i.e., $\gamma = c$, and with distances identical to this value, i.e., $\mathbf{d} = c$, R-STRESS $\|\gamma - \mathbf{d}\|^2$ is equal to zero.
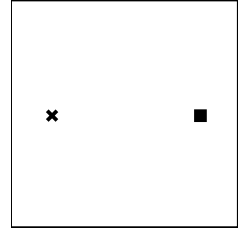


*Two-plus-two-point solution* — This zero STRESS solution contains one distance value for all subject-object distances, except for one subject-object distance, which has a different value. This solution is the result of minimizing STRESS-1 with the sum-of-squares of the distances as normalization factor. Due to the one differing distance value, and corresponding transformed preferences, the normalization factor $\|\mathbf{d}\|^2$ remains unequal to zero, while STRESS-1 $\|\gamma - \mathbf{d}\|^2/\|\mathbf{d}\|^2$ is equal to zero.
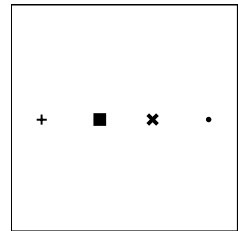


*Object-sphere solution* — This zero STRESS solution forms an extension of the two-point solution as there only exist one value for the distances. Whether the subjects are on the edge of the circle with the object in the center or the objects are on the fringe with the subjects in the middle, is of subordinate significance, since only subject-object distances matter for the loss function.
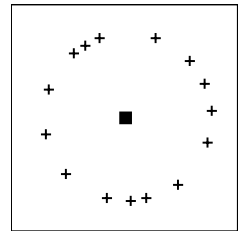


*Object-point solution* — This zero STRESS solution concerns a specific row-conditional solution with an improper normalization factor. Each subject has one value for the subject-object distances, but different values may exist for different subjects. A normalization factor, even a variance-based normalization factor as used in STRESS-2 or S-STRESS-2, that considers all distances simultaneously (i.e., improper summation), remains unequal to zero, while the RAW STRESS part of the loss function is equal to zero.
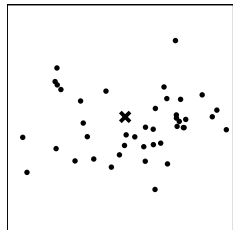
## REFERENCES

Albers, S., & Brockhoff, K. (1977). A procedure for new product positioning in an attribute space. *European Journal of Operational Research*, *1*, 230–238.

Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, *16*, 125–127.

Andrich, D. (1988). The application of an unfolding model of the PIRT type to the measurement of attitude. *Applied Psychological Measurement*, *12*, 33–51.

Andrich, D. (1989). A probabilistic IRT model for unfolding preference data. *Applied Psychological Measurement*, *13*(2), 193–216.

Arditti, S. (1997). Preference mapping: A case study. *Food Quality and Preference*, *8*(5/6), 323–327.

Arnold, G. M., & Williams, A. A. (1986). The use of Generalized Procrustes analysis in sensory analysis. In J. R. Piggott (Ed.), *Statistical procedures in food research* (pp. 233–253). London: Elsevier Applied Science.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., & Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, *26*, 641–647.

Baier, D., & Gaul, W. (1999). Optimal product positioning based on paired comparison data. *Journal of Econometrics*, *89*, 365–392.

Balabanis, G., & Diamantopoulos, A. (2004). Domestic country bias, country-of-origin effects, and consumer ethnocentrism: A multidimensional unfolding approach. *Journal of the Academy of Marketing Science*, *32*(1), 80–95.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., & Brunk, H. D. (1972). *Statistical inference under order restrictions.* New York: Wiley.

Barnard, F. R. (1921, December 8). *One look is worth a thousand words.* Advertisement in Printers' Ink.

Barnard, F. R. (1927, March 10). *Make a cake for bobby.* Advertisement in Printers' Ink. (Chinese Proverb: One Picture is Worth Ten Thousand Words)

Bartholomew, D. J. (1959). A test of homogeneity for ordered alternatives. *Biometrika*, *45*, 36–48.

Bartholomew, D. J. (1961). A test of homogeneity of means under restricted alternatives (with discussion). *Journal of the Royal Statistics Society, Series B*, *23*, 239–281.

Barton, D. E., & Mallows, C. L. (1961). The randomization bases of the amalgamation of weighted means. *Journal of the Royal Statistics Society, Series B*, 23, 423–433.

Batschelet, E. (1981). *Circular statistics in biology*. New York: Academic Press.

Bedall, F. K., & Zimmermann, H. (1979). Algorithm AS143: The mediancentre. *Applied Statistics*, 28, 325–328.

Ben-Israel, A., & Iyigun, C. (2007). Probabilistic d-clustering. *Journal of Classification*, 25, 5–26.

Bennett, J. F., & Hays, W. L. (1960). Multidimensional unfolding: Determining the dimensionality of ranked preference data. *Psychometrika*, 25, 27–43.

Bergamo, G. C., dos Santos Dias, C. T., & Krzanowski, W. J. (2008, July/August). Distribution-free multiple imputation in an interaction matrix through singular value decomposition. *Scientia Agricola*, 65(4), 422–427.

Bernaards, C. A., & Sijtsma, K. (2000). Influence of imputation and EM methods on factor analysis when item nonresponse in questionnaire data is nonignorable. *Multivariate Behavioral Research*, 35, 321–364.

Bezembinder, T. (1997). In memoriam Eddy Roskam (1932-1997). *Nederlands Tijdschrift voor de Psychologie*, 52, 149-150.

Bijleveld, C. C. J. H., & Commandeur, J. J. F. (1987). *The extension of analysis of angular variation to m-way designs* (Leiden Psychological Reports No. PRM-8703). Leiden: Department of Psychology, Leiden University.

Blumenthal, D. (2004). How to obtain the sensory scores of the optimal product according to external preference mapping. In *Proceedings of the seventh sensometrics meeting*. Davis, California: The Sensometric Society.

Borg, I., & Bergermaier, R. (1982). Degenerationsprobleme im Unfolding und Ihre Lösung. *Zeitschrift für Sozialpsychologie*, 13, 287–299.

Borg, I., & Groenen, P. J. F. (1997). *Modern multidimensional scaling: Theory and applications*. New York: Springer-Verlag.

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications* (second ed.). New York: Springer-Verlag.

Borg, I., & Lingoes, J. C. (1987). *Multidimensional similarity structure analysis*. Berlin: Springer.

Bringhurst, R. (2005). *The elements of typographic style* (3rd ed.). Vancouver, Canada: Hartley and Marks.

Brown, B. M. (1985). Spatial median. In S. Kotz, C. B. Read, & D. Banks (Eds.), *Encyclopedia of statistical sciences* (Vol. 8, pp. 574–575). New York: Wiley.

Burt, C. (1948a). Factor analysis and canonical correlations. *British Journal of Psychology (Statistical Section)*, 1, 95–106.

Burt, C. (1948b). The factorial study of temperamental traits. *British Journal of Psychology (Statistical Section)*, 1, 178–203.

Busing, F. M. T. A. (2006). Avoiding degeneracy in metric unfolding by penalizing the intercept. *British Journal of Mathemetical and Statistical Psychology*, *59*, 419–427.

Busing, F. M. T. A. (2010). *Advances in multidimensional unfolding.* Unpublished doctoral dissertation, Leiden University, Leiden, the Netherlands.

Busing, F. M. T. A., Commandeur, J. J. F., & Heiser, W. J. (1997). PROXSCAL: A multidimensional scaling program for individual differences scaling with constraints. In W. Bandilla & F. Faulbaum (Eds.), *Softstat '97 advances in statistical software* (pp. 237–258). Stuttgart, Germany: Lucius.

Busing, F. M. T. A., & de Rooij, M. (2009). Unfolding incomplete data: Guidelines for unfolding row-conditional rank order data with random missings. *Journal of Classification*, *26*, 329–360.

Busing, F. M. T. A., Groenen, P. J. F., & Heiser, W. J. (2005). Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika*, *70*(1), 71–98.

Busing, F. M. T. A., Heiser, W. J., & Eilers, P. (in preparation). *Avoiding degeneracies in unfolding using smooth monotone spline transformations.*

Busing, F. M. T. A., Heiser, W. J., Neufeglise, P., & Meulman, J. J. (2005). *PREFSCAL. Program for metric and nonmetric multidimensional unfolding, including individual differences modeling and fixed coordinates.* SPSS Inc. Chicago, IL. (Version 14.0)

Cailliez, F. (1983). The analytical solution of the additive constant problem. *Psychometrika*, *48*(2), 305–308.

Calinski, R. B., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*, 1–27.

Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. 1, pp. 105–155). New York: Seminar Press.

Carroll, J. D. (1980). Models and methods for multidimensional analysis of preferential choice (or other dominance) data. In E. Lantermann & H. Feger (Eds.), *Similarity and choice* (pp. 234–289). Bern: Hans Huber.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology*, *31*, 607–649.

Carroll, J. D., & Chang, J. J. (1964a). A general index of nonlinear correlation and its application to the problem of relating physical and psychological dimensions. *American Psychologist*, *19*(7), 540. (Abstract)

Carroll, J. D., & Chang, J. J. (1964b). *A general index of non-linear correlation and its application to the problem of relating physical and psychological dimensions* (Tech. Rep.). Murray Hill, NJ: Bell Telephone Laboratories. (Unpublished Manuscript)

Carroll, J. D., & Chang, J. J. (1967). *Relating preference data to multidimensional scaling solutions via a generalization of Coombs' unfolding model.* Bell Telephone Laboratories. (Unpublished Manuscript)

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*, 283–319.

Carroll, J. D., & Chang, J. J. (1972). *IDIOSCAL (individual differences in orientation scaling): a generalization of INDSCAL allowing idiosyncratic reference systems.* Paper presented at Psychometric Meeting, Princeton, NJ.

Carroll, J. D., & Wish, M. (1974). Models and methods for three-way multidimensional scaling. In D. H. Krantz, R. L. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2). San Francisco: W. H. Freeman.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276.

Chang, J. J., & Carroll, J. D. (1968). *How to use PROFIT, a computer program for property fitting by optimizing nonlinear or linear correlation* (Tech. Rep.). Murray Hill, NJ: Bell Telephone Laboratories. (Unpublished Manuscript)

Chang, J. J., & Carroll, J. D. (1969). *How to use MDPREF, a computer program for multidimensional analysis of preference data* (Tech. Rep.). Murray Hill, NJ: Bell Telephone Laboratories.

Chatterjee, R., & DeSarbo, W. S. (1992). Accommodating the effects of brand unfamiliarity in the multidimensional scaling of preference data. *Marketing Letters*, *3*(1), 85–99.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, *91*, 862–872.

Clatworthy, H. W. (1973). Tables of two-associate-class partially balanced designs. In *Applied mathematics series 63*. Washington: National Bureau of Standards.

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, *31*, 33–42.

Cochran, W. G., & Cox, G. M. (1957). *Experimental designs* (2nd ed.). New York: Wily.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*, 213–220.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (second ed.). New York: Academic Press.

Commandeur, J. J. F. (1991). *Matching configurations.* Unpublished doctoral dissertation, Leiden University, Department of Psychometrics and Research Methodology, Leiden University.

Commandeur, J. J. F. (1994, december). *The PROXSCAL program book.* (Implementation note for J. J. F. Commandeur and W. J. Heiser (1993).)

Commandeur, J. J. F., & Heiser, W. J. (1993). *Mathematical derivations in the proximity scaling (PROXSCAL) of symmetric data matrices* (Tech. Rep. No. RR-93-04). Leiden, The Netherlands: Leiden University, Department of Data Theory.

Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, *57*, 148–158.

Coombs, C. H. (1964). *A theory of data.* New York: Wiley.

Coombs, C. H., & Kao, R. C. (1960). On a connection between factor analysis and multidimensional unfolding. *Psychometrika*, *25*, 219–231.

Coxon, A. P. M. (1982). *The user's guide to multidimensional scaling. with special reference to the mds(x) library of computer programs.* London: Heinemann Educational Books. (hardcover)

Coxon, A. P. M., & Jones, C. L. (1974). Occupational similarities: subjective aspects of social stratification. *Quality and Quantity*, *8*, 139–157.

Coxon, A. P. M., & Jones, C. L. (1978). *The images of occupational prestige.* London: Macmillan.

Cureton, E. E., & D'Agostino, R. B. (1983). *Factor analysis: An applied approach.* Hillsdale, NJ: Lawrence Erlbaum.

Dagpunar, A. (1988). *Principles of random variate generation.* Oxford: Clarendon Press.

Daillant-Spinnler, B., MacFie, H. J. H., Beyts, P. K., & Hedderley, D. (1996). Relationships between percieved sensory properties and major preference directions of 12 variaties of apples from the southern hemisphere. *Food Quality and Preference*, *7*(2), 113–126.

Danzart, M., Sieffermann, J.-M., & Delarue, J. (2004). New developments in preference mapping techniques: Finding out a consumer optimal product, its sensory profile and the key sensory attributes. In *Proceedings of the seventh sensometrics meeting.* Davis, California: The Sensometric Society.

Davison, A. C., Hinkley, D. V., & Schechtman, E. (1986). Efficient bootstrap simulation. *Biometrika*, *73*(3), 555–566.

de Leeuw, J. (1977a). Applications of convex analysis to multidimensional scaling. In J. R. Barra, F. Brodeau, G. Romier, & B. van Cutsem (Eds.), *Recent developments in statistics* (pp. 133–145). Amsterdam, The Netherlands: North-Holland.

de Leeuw, J. (1977b). Correctness of Kruskal's algorithms for monotone regression with ties. *Psychometrika*, *42*(1), 141–144.

de Leeuw, J. (1983). *On degenerate nonmetric unfolding solutions* (Tech. Rep.). Leiden, The Netherlands: Department of Data Theory, FSW/RUL.

de Leeuw, J. (1988). Convergence of the majorization method for multidimensional scaling. *Journal of Classification*, *5*, 163–180.

de Leeuw, J., & Heiser, W. J. (1977). Convergence of correction-matrix algorithms for multidimensional scaling. In J. C. Lingoes, E. E. C. I. Roskam, & I. Borg (Eds.), *Geometric representations of relational data* (pp. 735–752). Ann Arbor, MI: Mathesis Press.

de Leeuw, J., & Heiser, W. J. (1980). Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (Vol. 5, pp. 501–522). Amsterdam, The Netherlands: North-Holland Publishing Company.

de Leeuw, J., & Heiser, W. J. (1982). Theory of multidimensional scaling. In L. Kanal & P. Krishnaiah (Eds.), *Handbook of statistics* (Vol. II, pp. 285–317). Amsterdam: North-Holland.

de Leeuw, J., & Meulman, J. J. (1986). A special jackknife for multidimensional scaling. *Journal of Classification*, *3*, 97–112.

de Soete, G., Carroll, J. D., & DeSarbo, W. S. (1986). The wandering ideal point model: A probabilistic multidimensional unfolding model for paired comparison data. *Journal of Mathematical Psychology*, *30*(1), 28–41.

de Soete, G., & Heiser, W. J. (1993). A latent class unfolding model for analyzing single stimulus preference ratings. *Psychometrika*, *58*, 545–565.

Delbeke, L. (1968). *Construction of preference spaces: an investigation into the applicability of multidimensional scaling models.* Unpublished doctoral dissertation, Leuven. (University Press)

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, *B39*, 1–38.

Dennis, J. E., & Schnabel, R. B. (1983). *Numerical methods for unconstrained optimization and nonlinear equations.* Englewood Cliffs, New Jersey: Prentice-Hall. (Republished by SIAM, Philadelphia, in 1996 as Volume 16 of Classics in Applied Mathematics)

Derbaix, C., & Sjöberg, L. (1994). Movie stars in space: A comparison of preferences and similarity judgements. *International Journal of Research in Marketing*, *11*, 261–274.

DeSarbo, W. S. (1978). *Three-way unfolding and situational dependence in consumer preference analysis.* Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.

DeSarbo, W. S., & Carroll, J. D. (1980). Three-way unfolding and situational dependence in consumer preference analysis. In K. Bernhardt, I. Dolich, M. Etzel, T. Kinnear, W. Perreault, & R. Roering (Eds.), *The changing marketing environment: New theories and applications* (pp. 321–325). Chicago: American Marketing Association.

DeSarbo, W. S., & Carroll, J. D. (1983). *Three-way unfolding via weighted least-squares.* (Unpublished Memorandum, AT&T Bell Laboratories: Murray Hill, N.J.)

DeSarbo, W. S., & Carroll, J. D. (1985). Three-way metric unfolding via alternating weighted least squares. *Psychometrika*, *50*(3), 275–300.

DeSarbo, W. S., & Cho, J. (1989, Mar). A stochastic multidimensional scaling vector threshold model for the spatial representation of 'pick any/n' data. *Psychometrika*, *54*(1), 105–129.

DeSarbo, W. S., de Soete, G., & Eliashberg, J. (1987). A new stochastic multi-dimensional unfolding model for the investigation of paired comparison consumer preference/choice data. *Journal of Economic Psychology*, *8*, 357–384.

DeSarbo, W. S., & Hoffman, D. L. (1987, Feb). Constructing MDS joint spaces from binary choice data: A multidimensional unfolding threshold model for marketing research. *Journal of Marketing Research*, *24*(1), 40–54.

DeSarbo, W. S., Kim, J., Choi, S. C., & Spaulding, M. (2002). A gravity-based multidimensional scaling model for deriving spatial structures underlying consumer preference/choice judgments. *Journal of Consumer Research*, *29*, 91–100.

DeSarbo, W. S., & Rao, V. R. (1984). GENFOLD2: A set of models and algo-rithms for the GENeral unFOLDing analysis of preference/dominance data. *Journal of Classification*, *1*, 147–186.

DeSarbo, W. S., & Rao, V. R. (1986). A constrained unfolding methodology for product positioning. *Marketing Science*, *5*, 1–19.

DeSarbo, W. S., Young, M. R., & Rangaswamy, A. (1997). A parametric multidimensional unfolding procedure for incomplete nonmetric prefer-ence/choice set data in marketing research. *Journal of Marketing Research*, *34*(4), 499–516.

Dijksterhuis, G. B., & Gower, J. C. (1991). The interpretation of generalized procrustes analysis and allied methods. *Food Quality and Preference*, *3*, 67–87.

Dinkelbach, W. (1967). On nonlinear fractional programming. *Management Science*, *13*, 492–498.

Eckart, C., & Young, G. (1936). Approximation of one matrix by another of lower rank. *Psychometrika*(1), 211–218.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia: SIAM.

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, *37*, 36–48.

Elam, K. (2007). *Typographic systems*. New York: Princeton Architectural Press.

Ennis, D. M. (1993). A single multidimensional model for discrimination, identification, and preferential choice. *Acta Psychologica*, *84*, 17–27.

Ennis, D. M. (1999). Multivariate preference mapping. *IFPress*, *2*(2), 2–3.

Ennis, D. M. (2005). Analytic approaches to accounting for individual ideal points. *IFPress*, *8*(2), 2–3.

Ennis, D. M., Palen, J., & Mullen, K. (1988). A multidimensional stochastic theory of similarity. *Journal of Mathematical Psychology*, *32*, 449–465.

Ennis, D. M., & Rousseau, B. (2004). Motivations for product consumption: Application of a probabilistic model to adolescent smoking. *Journal of Sensory Studies*, *19*, 107–117.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measure of reliability. *Educational and Psychological Measurement*, *33*(3), 613–619.

Galton, F. (1892). *Finger prints*. London: Macmillan.

Gleason, T. C. (1967). *A general model for nonmetric multidimensional scaling* (Tech. Rep.). Michigan: University of Michigan: MMPP 67-3.

Gold, M. (1958). Power in the classroom. *Sociometry*, *21*, 50–60.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, *53*, 325–338.

Gower, J. C. (1974). Algorithm AS78: The mediancentre. *Applied Statistics*, *23*, 466–470. (Corrigendum in Applied Statistics, 24(3), 390)

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, *40*(1), 33–51.

Gower, J. C., & Hand, D. J. (1996). *Biplots*. London: Chapman and Hall.

Gower, J. C., Meulman, J. J., & Arnold, G. M. (1999). Nonmetric linear biplots. *Journal of Classifcation*, *16*, 181–196.

Graham, J. W., & Schafer, J. L. (1999). On the performance of multiple imputation for multivariate data with small sample size. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 1–29). Thousand Oaks, CA: Sage.

Green, P. E., Carroll, J. D., & Goldberg, S. M. (1981). A general approach to product design optimization via conjoint analysis. *Journal of Marketing*, *45*, 17–37.

Green, P. E., & Krieger, A. M. (1989). Recent contributions to optimal product positioning and buyer segmentation. *European Journal of Operational Research*, *41*, 127–141.

Green, P. E., & Rao, V. R. (1972). *Applied multidimensional scaling*. Hinsdale, IL: Dryden Press.

Green, P. J., & Silverman, B. W. (1979). Constructing the convex hull of a set of points in the plane. *The Computer Journal*, *22*(3), 262–266.

Greenacre, M. J., & Browne, M. W. (1986, June). An efficient alternating least-squares algorithm to perform multidimensional unfolding. *Psychometrika*, *51*(2), 241–250.

Groenen, P. J. F. (1993). *The majorization approach to multidimensional scaling: Some problems and extensions*. Leiden, The Netherlands: DSWO Press.

Groenen, P. J. F. (2002, July, 16–19). *Iterative majorization algorithms for minimizing loss functions in classification*. Krakau.

Groenen, P. J. F., & Heiser, W. J. (1996). The tunneling method for global optimization in multidimensional scaling. *Psychometrika*, *61*, 529–550.

Groenen, P. J. F., Mathar, R., & Heiser, W. J. (1995). The majorization approach to multidimensional scaling for Minkowski distances. *Journal of Classification*, *12*, 3–19.

Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, *9*, 139–150.

Guttman, L. (1946). An approach for quantifying paired comparisons and rank order. *Annals of mathematical statistics*, *17*, 144–163.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, *33*, 469–506.

Guttman, L. (1981). Multidimensional data representations: When and why. In I. Borg (Ed.), (pp. 1–10). Ann Arbor, MI: Mathesis.

Hardy, G. H., Littlewood, J. E., & Polya, G. (1952). *Inequalities*. Cambridge: Cambridge University Press.

Harshman, R. A. (1972). *Determination and proof of minimum uniqueness conditions for PARAFAC-1* (UCLA Working Papers in Phonetics). Los Angeles: UCLA.

Hayford, J. F. (1902). What is the center of an area, or the center of a population? *Journal of the American Statistical Association*, *8*, 47–58.

Hays, W. L., & Bennett, J. F. (1961). Multidimensional unfolding: Determining configuration from complete rank order preference data. *Psychometrika*, *26*, 221–238.

Hedderley, D., & Wakeling, I. N. (1995). A comparison of imputation techniques for internal preference mapping, using monte carlo simulation. *Food Quality and Preference*, *6*, 281–297.

Heiser, W. J. (1981). *Unfolding analysis of proximity data*. Unpublished doctoral dissertation, Leiden University.

Heiser, W. J. (1985). *Multidimensional scaling by optimizing goodness-of-fit to a smoothness hypothesis* (Tech. Rep. No. RR-85-07). Leiden: Leiden University, Department of Data Theory.

Heiser, W. J. (1986). *Order invariant unfolding analysis under smoothness restrictions* (Tech. Rep. No. RR-86-07). Leiden: Leiden University, Department of Data Theory.

Heiser, W. J. (1987a). Joint ordination of species and sites: The unfolding technique. In P. Legendre & L. Legendre (Eds.), *Developments in numerical ecology* (pp. 189–221). Berlin, Heidelberg: Springer-Verlag.

Heiser, W. J. (1987b). Reihenfolgeninvariante entfaltungsanalyse unter glättebedingungen. *Zeitschrift für Sozialpsychologie*, *18*, 220–235.

Heiser, W. J. (1988). Multidimensional scaling with least absolute residuals. In H.-H. Bock (Ed.), *Classification and related methods of data analysis* (pp. 455–462). Amsterdam: North-Holland.

Heiser, W. J. (1989). Order invariant unfolding analysis under smoothness restrictions. In G. De Soete, H. Feger, & K. C. Klauer (Eds.), *New developments in psychological choice modeling* (pp. 3–31). Amsterdam: Elsevier Science Publisher B.V. (North-Holland).

Heiser, W. J. (1991). A generalized majorization method for least squares multidimensional scaling of pseudodistances that may be negative. *Psychometrika*, *55*, 7–27.

Heiser, W. J. (1995). Convergent computation by iterative majorization: Theory and applications in multidimensional data analysis. In W. J. Krzanowski (Ed.), *Recent advances in descriptive multivariate analysis* (pp. 157–189). Oxford: Oxford University Press.

Heiser, W. J., & Busing, F. M. T. A. (2004). Multidimensional scaling and unfolding of symmetric and asymmetric proximity relations. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 25–48). Thousand Oaks, CA: Sage Publications, Inc.

Heiser, W. J., & de Leeuw, J. (1979a). *How to use SMACOF-III: A program for metric multidimensional unfolding* (Tech. Rep.). Leiden, The Netherlands: Leiden University, Department of Data Theory.

Heiser, W. J., & de Leeuw, J. (1979b). Metric multidimensional unfolding. *Methoden en Data Nieuwsbrief van de Sociaal Wetenschappelijke Sectie van de Vereniging voor Statistiek*, *4*, 26–50.

Heiser, W. J., & Groenen, P. J. F. (1997). Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. *Psychometrika*, *62*, 63–83.

Heiser, W. J., & Meulman, J. J. (1983a). Constrained multidimensional scaling, including confirmation. *Applied Psychological Measurement*, *7*(4), 381–404.

Heiser, W. J., & Meulman, J. J. (1983b, July). *Smoothed monotone regression.* Handout. (Workshop on Nonmetric Data Analysis, Paris)

Heiser, W. J., & Stoop, I. (1986). *Explicit smacof algorithms for individual differences scaling* (Tech. Rep. No. RR-86-14). Leiden: Leiden University, Department of Data Theory. (PROXSCAL Progress Report)

Hinich, M. J. (2005). A new method for statistical multidimensional unfolding. *Communications in Statistics: Theory and Methods*, *34*(12), 2299–2310.

Hinkley, D. V. (1988). Bootstrap methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, *50*(3), 321–337.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(3), 55–67.

Hojo, H. (1997). A marginalization model for the multidimensional unfolding analysis of ranking data. *Japanese Psychological Research*, *39*(1), 33–42.

Horst, P. (1941). *Prediction of personal adjustment* (Bulletin No. 48). New York: Social Science Research Council.

Iyigun, C. (2007). *Probabilistic distance clustering adjusted for cluster size*. Unpublished doctoral dissertation, Rutgers University, New Brunswick, NJ.

Iyigun, C., & Ben-Israel, A. (2008). Probabilistic distance clustering adjusted for cluster size. *Probability in the Engineering and Informational Sciences*, *22*, 603–621.

Jeffrey, A. (2007). *Fontinst* [Type Software]. `http://www.tug.org/application/fontinst/`.

Jones, C. L. (1983). A note on the use of directional statistics in weighted Euclidean distances multidimensional scaling models. *Psychometrika*, *48*(3), 473–476.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187–200.

Katsnelson, J., & Kotz, S. (1957). On the upper limits of some measures of variability. *Archiv. f. Meteor., Geophys. u. Bioklimat. (B)*, *8*, 103.

Kelly, G. A. (1955). *The psychology of personal constructs: theory of personality*. New York: Norton.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*, 81–89.

Kendall, M. G. (1948). *Rank correlation methods*. London: Charles Griffin and Company Limited.

Kiers, H. A. L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics*, *14*(3), 105-122.

Kiers, H. A. L., & Groenen, P. J. F. (1996). A monotonically convergent algorithm for orthogonal congruence rotation. *Psychometrika*, *61*, 375–389.

Kim, C. (1990). *NEWFOLD: A new unfolding methodology*. Unpublished doctoral dissertation, Department of Marketing, The Wharton School, University of Pennsylvania.

Kim, C., Rangaswamy, A., & DeSarbo, W. S. (1999). A quasi-metric approach to multidimensional unfolding for reducing the occurrence of degenerate solutions. *Multivariate Behavioral Research*, *34*, 143–180.

Kohler, E. (2008). *LCDF typetools* [Type Software].

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. *Psychometrika*, *29*, 1–27.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*, 115–129.

Kruskal, J. B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society B*, *27*, 251–563.

Kruskal, J. B. (1971). Monotone regression: Continuity and differentiability properties. *Psychometrika*, *36*(1), 57–62.

Kruskal, J. B. (1977). Multidimensional scaling and other methods for discovering structure. In K. Enslein, A. Ralston, & H. S. Wilf (Eds.), *Mathematical methods for digital computers* (Vol. 2, pp. 296–339). New York: Wiley.

Kruskal, J. B., & Carmone, F. J. (1969). *How to use M-D-SCAL (Version 5M) and other useful information* (Tech. Rep.). Murray Hill, N. J.: Bell Telephone Laboratories.

Kruskal, J. B., & Carroll, J. D. (1969). Geometrical models and badness-of-fit functions. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (Vol. 2, pp. 639–671). New York: Academic Press.

Kruskal, J. B., & Shepard, R. N. (1974). A nonmetric variety of linear factor analysis. *Psychometrika*, *39*, 123-157.

Kruskal, J. B., Young, F. W., & Seery, J. B. (1978). *How to use KYST, a very flexible program to do multidimensional scaling and unfolding* (Tech. Rep.). Murray Hill, NJ: Bell Laboratories.

Krzanowski, W. J. (1988). Missing value imputation in multivariate data using the singular value decomposition of a matrix. *Biometrical Letters*, *25*, 31–39.

Kuga, N., & Mayekawa, S. (2008). New analytic solution to metric unfolding. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.), *New trends in psychometrics* (pp. 189–198). Tokyo: Universal Academic Press.

Lachenbruch, P. A. (1965). *Estimation of error rates in discriminant analysis*. PhD. Dissertation, University of California, Los Angeles.

Lachenbruch, P. A. (1968). Estimation of error rates in discriminant analysis. *Technometrics*, *10*(1), 1–11.

Larson, S. C. (1931). The shrinkage of the coefficient of multiple correlation. *Journal of Educational Psychology*, *22*, 45–55.

Lawson, C. L., & Hanson, R. J. (1974). *Solving least squares problems*. Englewoods Cliffs, NJ: Prentice Hall.

L'Ecuyer, P. (1999). Tables of maximally equidistributed combined LFSR generators. *Mathematics of Computing*, *68*, 261–269.

Lee, K.-Y. M., Paterson, A., Piggott, J. R., & Richardson, G. D. (2001). Sensory discrimination of blended scotch whiskies of different product categories. *Food Quality and Preference*, *12*, 109–117.

Lepš, J., & Šmilauer, P. (1999). Multivariate analysis of ecological data [Computer software manual]. Ceské Budejovice. (Course textbook)

Lingoes, J. C. (1966). An IBM-7090 program for Guttman-Lingoes smallest space Analysis-RII. *Behavioral Science*, *11*, 322.

Lingoes, J. C. (1970). A general nonparametric model for representing objects and attributes in a joint metric space. In J. C. Gardin (Ed.), *Archéologie et calculateurs.* Paris: Centre National de la Recherche Scientifique.

Lingoes, J. C. (1977). A general nonparametric model for representing objects and attributes in a joint metric space. In J. C. Lingoes (Ed.), *Geometric representations of relational data* (pp. 475–496). Ann Arbor, Michigan: Mathesis Press.

Lingoes, J. C., & Roskam, E. E. Ch. I. (1973). A mathematical and empirical analysis of two multidimensional scaling algorithms. *Psychometrika*, *38*. (Monograph Supplement)

Linting, M., Meulman, J. J., Groenen, P. J. F., & van der Kooij, A. J. (2007). Nonlinear principal components analysis: Introduction and application. *Psychological Methods*, *12*(3), 336–358.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: J. Wiley & Sons.

MacCallum, R. C. (1977). Effects of conditionality on INDSCAL and ALSCAL weights. *Psychometrika*, *42*, 297–305.

MacKay, D. B. (2001). Probabilistic unfolding models for sensory data. *Food Quality and Preference*, *12*, 427–436.

MacKay, D. B., & Zinnes, J. L. (1995). Probabilistic multidimensional unfolding: An anisotropic model for preference ratio judgements. *Journal of mathematical Psychology*, *39*, 99–111.

MacKay, D. B., & Zinnes, J. L. (2008, May). *PROSCAL: A program for probabilistic scaling* (5.0 ed.) [Computer software manual]. Kelley School of Business, Indiana University.

Magnus, J. R., & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. New York: Wiley.

Manly, B. E. J. (1991). *Randomization and monte carlo methods in biology*. London: Chapman and Hall.

Marden, J. I. (1995). *Analyzing and modeling rank data*. London: Chapman and Hall.

Mardia, K. V. (1972). *Statistics of directional data*. New York: Academic Press.

Matsumoto, M., & Nishimura, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, *8*(1), 3–30.

McClelland, G. H., & Coombs, C. H. (1975). ORDMET: A general algorithm for constructing all numerical solutions to ordered metric solutions. *Psychometrika*, *40*, 269–290.

McEwan, J. A., & Thomson, D. M. H. (1989). The repertory grid method and preference mapping in market research: A case study on chocolate confectionery. *Food Quality and Preference*, *1*(2), 59–68.

McGee, V. C. (1968). Multidimensional scaling of n sets of similarity measures: A nonmetric individual differences approach. *Multivariate Behavioral Research*, *3*, 233–248.

Meullenet, J.-F., Lovely, C., Threlfall, R., Morris, J. R., & Striegler, R. K. (2008). An ideal point density plot method for determining an optimal sensory profile for Muscadine grape juice. *Food Quality and Preference*, *19*, 210–219.

Meulman, J. J., & Heiser, W. J. (1983). *The display of bootstrap solutions in multidimensional scaling* (Tech. Rep.). Leiden, The Netherlands: Leiden University, Department of Data Theory.

Meulman, J. J., & Heiser, W. J. (1984). Constrained multidimensional scaling: More directions than dimensions. In *Compstat 1984, proceedings in computational statistics* (pp. 137–142). Vienna: Physica Verlag.

Meulman, J. J., Heiser, W. J., & Carroll, J. D. (1986). PREFMAP-3 user's guide [Computer software manual].

Meulman, J. J., Heiser, W. J., & SPSS. (1999). *SPSS Categories 10.0*. Chicago, IL: SPSS Inc.

Meulman, J. J., Heiser, W. J., & SPSS. (2005). *SPSS Categories 14.0*. Chicago, IL: SPSS Inc.

Meulman, J. J., van der Kooij, A. J., & Heiser, W. J. (2004). Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 49–70). Thousand Oaks, CA: Sage Publications, Inc.

Meyers, J. H., & Shocker, A. D. (1981). The nature of product-related attributes. *Research in Marketing*, *5*, 211–236.

Miles, R. E. (1959). The complete amalgamation into blocks, by weighted means, of a finite set of real numbers. *Biometrika*, *45*, 317–327.

Miller, A. (1987). *A simple but efficient routine for finding 2d convex hulls, i.e. in finding the minimum polygon to enclose a set of points*. Internet publication.

Miller, A. (2002). *Subset selection in regression* (second ed.). Boca Roton, Florida: Chapman & Hall.

Miller, J. E., Shepard, R. N., & Chang, J. J. (1964). An analytical approach to the interpretation of multidimensional scaling solutions. *American Psychologist*, *19*(7), 579–580. (abstract)

Mosteller, F., & Tukey, J. W. (1968). Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 2). Reading, Massachusetts: Addison-Wesley.

Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.

Mullen, K., & Ennis, D. M. (1991). A simple multivariate probabilistic model for preferential and triadic choices. *Psychometrika*, *56*, 69–75.

Murray, J. M., & Delahunty, C. M. (2000). Mapping consumer preference for the sensory and packaging attributes of Cheddar cheese. *Food Quality and Preference*, *11*, 419–435.

Nakanishi, M., & Cooper, L. G. (2003). *Metric unfolding revisited: Straight answers to basic questions* (eScholarship Repository No. 2003010112). Los Angeles, CA: University of California.

Nguyen, N. (1993). An algorithm for constructing optimal resolvable incomplete block designs. *Communications in Statistics, Simulation & Computation*, *22*, 911–923.

Nguyen, N. (1994). Construction of optimal block design by computer. *Technometrics*, *36*, 300–307.

Oreskovich, D. C., Klein, B. P., & Sutherland, J. W. (1991). Procrustes analysis and its applications to free-choice and other sensory profiling. In H. T. Lawless & B. P. Klein (Eds.), *Sensory science: Theory and applications in foods* (pp. 353–393). New York: Marcel Dekker.

Pearson, K. (1896). Regression, heridity, and panmixia. *Philosophical Transactions of the Royal Society of London, Ser. A*, *187*, 253–318.

Prestwich, S. D. (2001). Balanced incomplete block design as satisfiability. In D. O'Donoghue (Ed.), *Proceedings of the 12th irish conference on ai and cognitive science (AICS 2001)*. NUI Maynooth, Ireland: AICS.

Price, R. H., & Bouffard, D. L. (1974). Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, *30*, 579–586.

Pruzansky, S. (1975). How to use SINDSCAL: A computer program for individual differences in multidimensional scaling [Computer software manual]. Murray Hill, NJ.

Rabinowitz, G. (1976). A procedure for ordering object pairs consistent with the multidimensional unfolding model. *Psychometrika*, *41*(3), 349–373.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science*, *3*(4), 425–441.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement*, *24*(1), 3–32.

Roberts, J. S., & Laughlin, J. E. (1996). A unidimensional item response model for unfolding responses from a graded disagree–agree response scale. *Applied Psychological Measurement*, *20*, 231–255.

Roskam, E. E. Ch. I. (1968). *Metric analysis of ordinal data*. Voorschoten: VAM.

Roskam, E. E. Ch. I., & Lingoes, J. C. (1970). MINISSA-I: a FORTRAN IV (G)

program for the smallest space analysis of square symmetric matrices. *Behavioral Sciences*, *15*, 204–205.

Ross, J., & Cliff, N. (1964). A generalization of the interpoint distance model. *Psychometrika*, *29*, 167–176.

Rousseau, B., & Ennis, D. M. (2008). An application of landscape segmentation analysis to blind and branded data. *IFPress*, *11*(3), 2–3.

Rowe, G., Lambert, N., Bowling, A., Ebrahim, S., Wakeling, I. N., & Thomson, R. (2005). Assessing patients' preferences for treatments for angina using a modified repertory grid method. *Social Science & Medicine*, *60*, 2585–2595.

Saito, T. (1978). The problem of the additive constant and eigenvalues in metric multidimensional scaling. *Psychometrika*, *43*(2).

Schiffman, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to multidimensional scaling*. New York: Academic Press Inc.

Schönemann, P. H. (1970). On metric multidimensional unfolding. *Psychometrika*, *35*, 349–366.

Shao, J., & Tu, D. (1995). *The jackknife and bootstrap*. New York: Springer-Verlag.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model for relating generalization to distance in psychological space. *Psychometrika*, *22*, 325–345.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, *27*, 125–140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, *27*, 219–246.

Shepard, R. N. (1972). A taxonomy of some principal types of data and of multidimensional methods for their analysis. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. 1, pp. 21–47). New York: Seminar Press.

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika*, *39*(4), 373–421.

Sherif, C. W., Sherif, M., & Nebergall, R. E. (1965). *Attitude and attitude change; the social judgement-involvement approach*. Philadelphia: W. B. Sounders Company.

Shocker, A. D., Ben-Akiva, M., Boccara, B., & Nedungadi, P. (1991). Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing Letters*, *2*(3), 181–197.

Shocker, A. D., & Srinivasan, V. (1974). A consumer-based methodology for the identification of new product ideas. *Management Science*, *20*(6), 921–937.

Sixtl, F. (1973). Probabilistic unfolding. *Psychometrika*, *38*(2), 235–248.

SPSS. (2006). *SPSS for WINDOWS, Release 15.0* [Computer Software]. Chicago, IL.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.

Steverink, M. H. M., Heiser, W. J., & van der Kloot, W. A. (2002). *Avoiding degenerate solutions in multidimensional unfolding by using additional distance information* (Tech. Rep. No. RR-02-01). Leiden, The Netherlands: Leiden University, Department of Psychology.

Stone, M. (1974). Cross-validation choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B*, *36*, 111–147.

Sviatlovsky, E. E., & Eells, W. C. (1937). The centrographical method and regional analysis. *Geographical review*, *27*(2), 240–254.

Takane, Y., Bozdogan, H., & Shibayama, T. (1987). Ideal point discriminant analysis. *Psychometrika*, *52*(3), 371–392.

Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences MDS: An alternating least squares method with optimal scaling features. *Psychometrika*, *42*, 7–67.

ter Braak, C. J. F. (1986). Canonical correspondence analysis: a new eigenvecor technique for multivariate direct gradient analysis. *Ecology*, *67*, 1167–1179.

ter Braak, C. J. F. (1992a). Multidimensional scaling and regression. *Statistica Applicata (Italian Journal of Applied Statistics)*, *4*, 577–586.

ter Braak, C. J. F. (1992b). Permutation vs. bootstrap significance tests in multiple regression and ANOVA. In K. H. Jockel, G. Rothe, & W. Sendler (Eds.), *Bootstrapping and related techniques* (pp. 79–86). Berlin, Germany: Springer.

ter Braak, C. J. F., & Prentice, I. C. (1988). A theory of gradient analysis. *Advances in Ecological Research*, *18*, 271–319.

ter Braak, C. J. F., & Šmilauer, P. (1998). Canoco reference manual and user's guide to canoco for windows: Software for canonical community ordination (4th ed.) [Computer software manual]. Ithaca, NY.

Thompson, J. L., Drake, M. A., Lopetcharat, K., & Yates, M. D. (2004). Preference mapping of commercial chocolate milks. *Journal of Food Science*, *69*(9), 406–413.

Thurstone, L. L. (1947). *Multiple-factor analysis.* Chicago: University Chicago Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288.

Tikhonov, A. N., & Arsenin, V. Y. (1977). *Solutions of ill-posed problems.* Washington: Winston.

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

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

Trosset, M. W. (1998). A new formulation of the nonmetric STRAIN problem in multidimensional scaling. *Journal of Classification*, *15*, 15–35.

Tschichold, J. (1991). *The form of the book: Essays on the morality of good design* (R. Bringhurst, Ed.). London: Lund Humphries. (Translated by Hajo Hadeler. Originally published as Ausgewählte Aufsätze über Fragen der Gestalt des Buches und der Typographie. Basel: Birkhäuser Verlag, 1975)

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* (Tech. Rep.). Washington, D. C.: Department of the Army.

Tucker, L. R. (1972). Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika*, *37*, 3–27.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.

Turgenef, I. S. (1867). *Fathers and sons*. New York: Leypoldt & Holt. (Translated from the Russian, with approval of the author, by E. Schuyler, Ph. D.)

van Blokland-Vogelesang, A. W. (1989). Unfolding and consensus ranking: A prestige ladder for technical occupations. In G. De Soete, H. Feger, & K. C. Klauer (Eds.), *New developments in psychological choice modeling* (pp. 237–258). Amsterdam, The Netherlands: North-Holland.

van Blokland-Vogelesang, A. W. (1993). A nonparametric distance model for unidimensional unfolding. In M. A. Fligner & J. S. Verducci (Eds.), *Probability models and statistical analyses for ranking data* (pp. 241–276). New York: Springer-Verlag.

van de Graaf, J. A. (1946). Nieuwe berekening voor de vormgeving. *Tété*, 95–100.

van de Velden, M., de Beuckelaer, A., Groenen, P. J. F., & Busing, F. M. T. A. (2010). Visualizing preferences using penalized nonmetric unfolding: Stability and parameter selection. *Journal of Marketing Research*. (Manuscript submitted for publication)

van der Kloot, W. A., Bijleveld, C. C. J. H., & Commandeur, J. J. F. (1990). *Two approaches to the extension of analysis of angular variation to multi-way classifications* (Leiden Psychological Reports No. PRM 01-90). Leiden: Psychometrics and Research Methodology, Department of psychology, University of Leiden.

van der Kooij, A. J. (2007). *Prediction accuracy and stability of regression with optimal scaling transformations*. Unpublished doctoral dissertation, Leiden University.

van der Kooij, A. J., & Meulman, J. J. (2004). Regression with optimal scaling. In J. J. Meulman, W. J. Heiser, & SPSS (Eds.), *SPSS categories 13.0* (pp. 1–10, 107–157). Chicago, IL: SPSS Inc.

van Deun, K. (2005). *Degeneracies in multidimensional unfolding*. Unpublished doctoral dissertation, Catholic University Leuven.

van Deun, K., Groenen, P. J. F., & Delbeke, L. (2005). *VIPSCAL: A combined vector ideal point model for preference data* (Econometric Institute Report No. EI 2005-03). Rotterdam: Erasmus University Rotterdam.

van Deun, K., Groenen, P. J. F., & Delbeke, L. (2006). *VIPSCAL: A combined vector ideal point model for preference data.* Manuscript submitted for publication.

van Deun, K., Groenen, P. J. F., Heiser, W. J., Busing, F. M. T. A., & Delbeke, L. (2005). Interpreting degenerate solutions in unfolding by use of the vector model and the compensatory distance model. *Psychometrika*, *70*(1), 23–47.

van Deun, K., Heiser, W. J., & Delbeke, L. (2007). Multidimensional unfolding by nonmetric multidimensional scaling of Spearman distances in the extended permutation polytope. *Multivariate Behavioral Research*, *42*(1), 103–132.

van Deun, K., Marchal, K., Heiser, W. J., Engelen, K., & van Mechelen, I. (2008). Joint mapping of genes and conditions via multidimensional unfolding analysis. *BMC Bioinformatics*, *8*(1), 181.

van Ginkel, J. R., van der Ark, L. A., & Sijtsma, K. (2007). Multiple imputation of item scores in test and questionnaire data, and influence on psychometric results. *Multivariate Behavioral Research*, *42*(2), 387-414.

van Kleef, E., van Trijp, H., & Luning, P. (2006). Internal versus external preference analysis: An exploratory study on end-user evaluation. *Food Quality and Preference*, *17*, 387–399.

van Waning, E. (1976). *A set of programs to perform a Kruskal-type monotone regression.* Unpublished master's thesis, Leiden University.

Velderman, M. J. (2005). *Unfolding with incomplete data.* Unpublished master's thesis, Leiden University.

Wagenaar, W. A., & Padmos, P. (1971). Quantitative interpretation of stress in Kruskal's method multidimensional scaling technique. *British Journal of Mathematical and Statistical Psychology*, *24*, 101–110.

Wakeling, I. N., Raats, M. M., & MacFie, H. J. H. (1992). A new significance test for consensus in generalized procrustes analysis. *Journal of Sensory Studies*, *7*(2), 91–96.

Watson, G. S., & Williams, E. J. (1956). On the construction of significance tests on the circle and the sphere. *Biometrika*, *43*, 344–352.

Weber, A. (1909). *Über den standort der industrien (English translation by C. J. Freidrich (1929). Alfred Weber's Theory of Location of Industries).* Chicago: University of Chicago Press.

Weeks, D. G., & Bentler, P. M. (1982). Restricted multidimensional scaling models for asymmetric matrices. *Psychometrika*, *47*, 201–208.

Weinberg, S. L., Carroll, J. D., & Cohen, H. S. (1984). Confidence regions for

INDSCAL using the jackknife and bootstrap techniques. *Psychometrika*, *49*(4), 475–491.

Weinberg, S. L., & Menil, V. C. (1993). The recovery of structure in linear and ordinal data: INDSCAL versus ALSCAL. *Multivariate Behavioral Research*, *28*(2), 215–233.

Weiszfeld, E. (1937). Sur le point par lequel la somme des distances de n points donneès est minimum. *Tohoku Mathematical Journal*, *43*, 355–386.

Wemelsfelder, F., Hunter, E. A., Mendl, M. T., & Lawrence, A. B. (2000). The spontaneous qualitative assessment of behavioural expressions in pigs: first explorations of a novel methodology for integrative animal welfare measurement. *Applied Animal Behaviour Science*, *67*, 193–215.

Wheeler, S., & Watson, G. S. (1964). A distribution-free two sample test on a circle. *Biometrika*, *51*, 256–257.

Wilkinson, L. (1999). *Systat 9.0 for windows.* Chicago, IL: SPSS Inc.

Wilks, S. S. (1932). Moments and distribution of estimates of population parameters from fragmentary samples. *Annals of Mathematical Statistics*, *3*, 163–195.

Wilson, P. (2008). *The memoir class for configurable typesetting; user guide* [LaTeX $2_\varepsilon$ Document Class]. The Herries Press, Normandy Park, WA.

Winsberg, S., & Carroll, J. D. (1989). A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. *Psychometrika*, *54*, 217–229.

Winsberg, S., & Ramsay, J. O. (1983). Monotone spline transformations for dimension reduction. *Psychometrika*, *48*(4), 575–595.

Wrigley, C., & Neuhaus, J. O. (1955). *The matching of two sets of factors* (Tech. Rep.). Urbana: University of Illinois Press.

Young, F. W. (1972). A model for polynomial conjoint analysis algorithms. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling, theory* (Vol. I, pp. 69–104). New York: Seminar Press.

Young, F. W. (1982). Enhancements in ALSCAL-82. In *Proceedings of the seventh annual SAS users group* (pp. 633–642). Cary, NC: The SAS Institute.

Young, F. W. (1987). *Multidimensional scaling: History, theory, and applications* (R. M. Hamer, Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Young, F. W., & Harris, D. F. (1997). Multidimensional scaling. In M. J. Norŭsis (Ed.), *SPSS professional statistics 7.5* (pp. 155–222). Chicago, IL: SPSS Inc.

Young, F. W., & Lewyckyj, R. (1979). *ALSCAL-4 USERS GUIDE, 2nd edition.* Data Analysis and Theory Associates.

Young, F. W., & Torgerson, W. S. (1967). TORSCA, a FORTRAN IV program for Shepard-Kruskal multidimensional scaling analysis. *Behavioral Science*, *12*, 498.

Young, G., & Householder, A. S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, *3*(1), 19–22.

Zhang, B. (2000). *Generalized k-harmonic means – boosting in unsupervised learning* (Hewlett Packard Technical Report No. HPL-2000-137). Palo Alto, CA: Hewlett Packard Laboratories.

Zhang, B., Hsu, M., & Dayal, U. (1999). *K-harmonic means – a data clustering algorithm* (Hewlett Packard Technical Report No. HPL-1999-124). Palo Alto, CA: Hewlett Packard Laboratories.

Zinnes, J. L., & Griggs, R. A. (1974). Probabilistic, multidimensional unfolding analysis. *Psychometrika*, *39*(3), 327–350.

Als ik even, nogmaals, de samenvatting mag geven
want anders komen we natuurlijk niet verder.

— Rutger Kopland

Meerdimensionale ontvouwing is een analyse techniek die afbeeldingen maakt
van twee sets van objecten, bijvoorbeeld van personen en producten, geba-
seerd op de voorkeuren van de personen voor die producten. De afstanden
tussen de personen en de producten in de afbeelding dienen zo goed mogelijk
te corresponderen met deze voorkeuren en wel zo dat een kleine afstand over-
eenkomt met een grote voorkeur, terwijl een grote afstand correspondeert
met een geringe voorkeur.

De voorkeuren voor de 15 broodjes zijn door de personen aangegeven met
de rangnummers 1 tot en met 15, met op 1 het meest geprefereerde broodje,
ongeacht of de persoon in kwestie van broodjes houdt of niet. Dus zowel een
persoon die in het algemeen van broodjes houdt als een persoon die in het
algemeen broodjes verafschuwt, hebben beiden een broodje op plaats 1 staan.
Om met deze absolute verschillen in waardering om te kunnen gaan, mogen
de rangnummers (de getallen 1 tot en met 15) worden veranderd voor iedere
persoon afzonderlijk en wel zodanig dat de getalswaarden mogen veranderen,
maar de volgorde van de (nieuwe) getallen, genaamd *pseudo-afstanden*, in
stand moet blijven. De persoon die van broodjes houdt, krijgt kleinere waar-
den (bijvoorbeeld 0.1, 0.2, . . . , 1.4, 1.5), terwijl hogere waarden (bijvoorbeeld
13.6, 13.7, 13.8, . . . , 14.9, 15.0) beter passen bij de persoon die niet van broodjes
houdt. Deze zogenaamde persoonsgebonden *monotone transformatie* van de
rangnummers wordt door het ontvouwingsprogramma (optimaal) bepaald.

Bijvoorbeeld, voor het boek van Green en Rao (1972) hebben 21 studenten
en hun partners hun voorkeur aangegeven voor 15 ontbijtbroodjes. Meerdi-
mensionale ontvouwing maakt van deze persoonlijke voorkeuren een afbeel-
ding met personen en broodjes als punten in de ruimte, zoals te zien is in
Figuur 1 op de volgende bladzijde, waarbij de personen (zwarte stippen) het
dichtst liggen bij hun meest geprefereerde broodje. Hoe verder de broodjes
verwijderd liggen van een persoonspunt, hoe minder de broodjes worden
gewaardeerd door die persoon.

Figuur 1   PREFSCAL ontvouwingsoplossing voor de ontbijtbroodjes gegevens (Green en Rao, 1972) met 42 personen (stippen) en 15 broodjes (namen).

Het ontvouwingsalgoritme gaat nu als volgt: Eerst worden de rangnummers omgezet in de zogenaamde pseudo-afstanden, waarna de pseudo-afstanden zo goed mogelijk worden weergegeven als afstanden in de afbeelding. De broodjeshater komt dus ver van de broodjes te liggen (13.6 − 15.0), terwijl de broodjesverorberaar er vlak bij zal liggen (0.1 − 1.5) Als zodanig levert ontvouwing dus een metrische oplossing (afstanden) op basis van louter niet-metrische gegevens (voorkeuren of rangnummers).

Ondanks dit unieke gegeven is ontvouwing tot op de dag van vandaag geen populaire techniek: "Toepassingen van meerdimensionale ontvouwing

blijven ernstig achter, ongetwijfeld door de vele technische problemen die een hardnekkig obstakel vormen voor succesvolle data analyse ..." (Heiser en Busing, 2004, p. 27, vertaling FB). Het serieuze obstakel betreft *gedegenereerde oplossingen*: Ontvouwing geeft vaak oplossingen die perfect zijn in termen van de verliesfunctie (de afstanden geven de voorkeuren perfect weer), maar die volstrekt onbruikbaar zijn in termen van interpretatie (de perfecte weergave is nietszeggend). Het is een probleem dat volkomen versmolten is geraakt met ontvouwing. De vrijheid van de monotone transformatie staat (bijna) gelijke waarden toe voor de pseudo-afstanden. Wanneer de afstanden hieraan worden gelijkgesteld, levert dat een perfecte oplossing op in termen van 'overeenkomst', maar een waardeloze oplossing in termen van 'interpretatie'. De broodjes liggen in zo'n geval allemaal op dezelfde afstand van een persoon, hetgeen eenzelfde voorkeur voor alle broodjes impliceert. We weten echter dat dit niet het geval is, daar de personen hun voorkeuren hebben aangegeven middels de rangnummer 1 tot en met 15. Het meerdimensionale ontvouwingsmodel is als zodanig niet geïdentificeerd, daar ongeacht welke gegevens geanalyseerd worden, het immer eenzelfde soort, niet te interpreteren afbeelding oplevert.

Gedegenereerde oplossingen ontstonden gelijktijdig met de eerste algo-ritmen voor ontvouwing. Het conceptuele idee van ontvouwing dat daaraan vooraf ging, is afkomstig van Coombs en collega's. De term ontvouwing komt van de volgende metafoor: Stel je een kralenketting voor met zwarte en witte kralen, waarbij de zwarte kralen de personen voorstellen en de witte kralen de broodjes. Pak één van de zwarte kralen tussen duim en wijsvinger en laat de ketting hangen, zodat beide zijden van de ketting nu naast elkaar hangen. De broodjes, eerst nog aan twee kanten van de persoon, vallen nu samen en wel zo dat de positie van de persoon zich bovenaan de ketting bevindt en de broodjes eronder hangen. De volgorde van de broodjes op de gevouwen ketting, geme-ten vanaf de zwarte persoonskraal, correspondeert met de voorkeuren van de persoon. Ontvouwen is de omgekeerde operatie, waarbij de individuele voorkeuren van de personen (gevouwen kettingen) gebruikt worden om één lange kralenketting met alle zwarte en witte kralen te rijgen.

Het ontvouwingsidee is uitgebreid naar meer dimensies door Bennett en Hayes (1960) en Hayes en Bennett (1961). In plaats van een één-dimensionale kralenketting hebben we nu bijvoorbeeld een twee-dimensionaal gehaakte sprei met zwarte en witte kralen. Het basisprincipe blijft echter gelijk: De afstanden tussen de persoonspunten (zwarte kralen) en de broodjespunten (witte kralen) dienen overeen te komen met de eventueel getransformeerde voorkeuren.

Coombs' werk had een enorme impact op het conceptuele niveau, maar technisch gezien stelde het teleur. Hiervoor moest gewacht worden op een spin-off uit de hoek van de meerdimensionale schaling. Met name Shepard (1962a, 1962b) en Kruskal (1964a, 1964b) lieten zien dat rangnummers (niet-

metrische gegevens) voldoende informatie in zich hadden voor een unieke en bovenal kwantitatieve (metrische) oplossing en zorgden voor een expliciete formulering van de verliesfunctie. Negentienhonderdvierenzestig was het jaar van de niet-metrische doorbraak.

Kruskal zelf maakte in 1969 het probleem van de gedegenereerde oplossingen wereldkundig, terwijl Gleason (1967) en Roskam (1968) zich er al eerder mee bezig hielden. Gleason benadrukte de noodzaak voor persoonsgebonden transformaties en Roskam stelde een aangepaste verliesfunctie voor, net als Kruskal zelf overigens. Zowel Roskam (1968) als Kruskal en Carroll (1969) rapporteerden teleurstellende resultaten. Hoewel de aangepaste verliesfunctie triviale oplossingen vermeed, leken de oplossingen erg op oplossingen zonder de noodzakelijk geachte aanpassingen. In 1983 bewees de Leeuw dat de aangepaste verliesfunctie geen garantie biedt tegen degeneraties. Lingoes (1977) stelde voor om de schalingsvariant aan te houden. Daar schaling geen degeneraties kent, althans meestal niet, kunnen deze in ontvouwing worden voorkomen door ontvouwing te definiëren als een schalingsprobleem met ontbrekende diagonaalblokken. Heiser (1981), Borg en Bergermaier (1982) en Heiser (1989) probeerden het degeneratieprobleem te voorkomen door de transformaties minder vrijheid te geven. Heiser (1981) stelde grenzen aan de verschillen tussen opeenvolgende pseudo-afstanden, eerst met hard gecodeerde grenzen, later met intern bepaalde grenzen, terwijl Borg en Bergermaier (1982) het zochten in een (gewogen) combinatie van monotone en lineaire transformaties. DeSarbo en Rao (1984) legden de schuld van de degeneraties bij de meetfouten in de gegevens en gebruiken gewichten voor de gegevens om deze meetfouten af te zwakken en zo degeneraties te voorkomen. En passant kwamen ze met een snel algoritme voor het minimaliseren van de verliesfunctie. Net voor het sluiten van de twintigste eeuw opperden Kim, Rangaswamy en DeSarbo (1999) een algoritme waarbij vooraf één monotone transformatie wordt gedaan, maar verder een metrische ontvouwing wordt uitgevoerd, zonder monotone transformatie.

Om verschillende redenen hebben geen van bovenstaande aanpakken geleid tot een definitieve oplossing voor het degeneratieprobleem. De zoektocht naar niet-gedegenereerde oplossingen stopte echter niet bij de eeuwwisseling. Recente ontwikkelingen (Steverink, Heiser en van der Kloot, 2002; Borg en Groenen, 2005; van Deun, Groenen, Heiser, Busing en Delbeke, 2005; van Deun, Groenen en Delbeke, 2006; van Deun, Heiser en Delbeke, 2007) staan echter wel in het teken van de ontwikkelingen zoals hierboven geschetst.

Uit het voorgaande mag de indruk gewekt zijn dat het degeneratieprobleem alleen maar voor zou komen bij het ontvouwen van volgorde of niet-metrische gegevens. Niets is minder waar. Degeneraties komen ook voor bij ontvouwing met lineaire transformaties. Een eenvoudige oplossing voor dit specifieke probleem is het onder controle houden van een ongewenst hoog intercept,

terwijl de slope ongelijk aan nul wordt gehouden. Op deze manier blijven de pseudo-afstanden ongelijk en wordt metrische ontvouwing uit de degeneratieproblemen gehouden. De methode is eenvoudig toepasbaar in algemene rekensoftware (MATLAB of R) of statistische software (SAS of SPSS).

Een algemener toepasbare oplossing voor het degeneratieprobleem is gevonden in een substantiële aanpassing van de aanpak voorgesteld door Roskam (1968) en Kruskal en Carroll (1969). Om gelijke pseudo-afstanden te vermijden, wordt de verliesfunctie gedeeld door de variatie van de pseudo-afstanden. Wanneer deze variatie steeds kleiner wordt, zal de waarde van de verliesfunctie steeds groter worden: een onaantrekkelijke situatie. De deling door de variatie, gemeten middels de *coefficient of variation* (Pearson, 1896), ontmoedigt aldus persoonsgebonden pseudo-afstanden met weinig variatie en vermijdt zo ook oplossingen met weinig variatie in de afstanden. Twee penalty parameters zorgen voor voldoende power (een omissie in eerdere aanpassingen) en fine tuning. Een algoritme voor het minimaliseren van de aangepaste verliesfunctie, gebaseerd op *iterative majorization* (IM) en *alternating least squares* (ALS), staat beschreven in de technische appendix. De resultaten van een simulatiestudie en de analyse van empirische gegevens verschaffen advieswaarden voor de twee penalty parameters en laten zien dat de voorgestelde aanpassingen werken.

Met de controle over het degeneratieprobleem is de weg vrij gemaakt om het ontvouwingsmodel verder te ontwikkelen. Eén mogelijke uitbreiding van het model is het toevoegen van verklarende variabelen, niet alleen om de interpretatie te vereenvoudigen, maar ook om voorspellingen te kunnen doen. Afhankelijk van de toegevoegde informatie, gebruikt het zogenaamde *gerestricteerde ontvouwingsmodel* persoonskenmerken om de persoonspunten te restricteren of broodjeskenmerken om de broodjespunten te restricteren. De leeftijd of het geslacht van een persoon of het aantal calorieën of de hardheid van een broodje kunnen zorgen voor een bepaalde indeling van de ruimte. Deze variabelen helpen vervolgens om de ruimte eenvoudiger te kunnen interpreteren. Aangezien de coördinaten gelijk zijn aan een lineaire combinatie van de respectievelijke variabelen, kunnen zowel de variabelen gebruikt worden om locaties te voorspellen als locaties om waarden van variabelen te voorspellen. Transformaties van de variabelen maken het mogelijk de coördinaten in meer of mindere mate te restricteren. Dit is mede afhankelijk van het meetniveau van de variabelen.

Een andere ontwikkeling in dit proefschrift vormt het onderzoek naar de mate waarin gegevens mogen ontbreken zonder een doorslaggevende invloed te hebben op de eindoplossing, de afbeelding. Het blijkt dat met name voor relatief grote data sets zelfs bijna de helft van deze gegevens mogen ontbreken zonder de conclusies te verstoren. Proefpersonen hoeven bijvoorbeeld maar de helft van de items te beoordelen, hetgeen nauwkeuriger gegevens oplevert

(minder vermoeidheid en personen beoordelen broodjes die ze kennen) in minder tijd (goedkoper, ook door minder uitval).

Dit proefschrift heeft ontvouwing verder ontwikkeld in de richting van een meer betrouwbare en praktische methode voor gegevensanalyse. Het moge duidelijk zijn dat deze ontwikkeling nog niet tot stilstand is gekomen: Onderzoek naar locale minima, start configuraties, stabiliteit, aanvullende analyses, en onderzoek naar verbeterde grafische weergaven van de resultaten, zijn maar enkele voorbeelden van mogelijke onderzoeksgebieden. Deze ontwikkelingen zijn alleen mogelijk na het creëren van een stevige basis. Ontvouwing, zoals gepresenteerd in dit proefschrift, met z'n flexibele algoritme gebaseerd op alternating least squares en iterative majorization, met z'n mogelijkheden voor optimale transformaties van de gegevens en het omgaan met ontbrekende gegevens, en met z'n uitgebreide mogelijkheden voor allerhande restricties, heeft deze basis gelegd.

## CURRICULUM VITAE

Frank Busing was born on December 28, 1963 in Amsterdam as Franciscus (after his uncle) Martinus (after his father's father) Theodorus (after his mother's father) Antonius (after his father) Busing. In 1983 he graduated from the Fioretti College in Lisse (VWO-A). Before receiving his master's degree (1993) in Psychology, Frank studied Law, as well as Child and Education Studies (propedeuse 1987), all at Leiden University. In 2000 he received his qualification to teach at primary schools from the Hogeschool IPABO in Amsterdam.

During his studies, Frank was employed by Leiden University as research assistant and by LEVOB Assurance Company in Amersfoort as researcher and programmer. From 1993 on, he works for Leiden University as researcher and statistical programmer. His research topics are estimation and resampling procedures in multilevel analysis and multidimensional scaling. During this period, Frank developed computer programs for 'automatiseren t/m 20' (AUTO), multilevel analysis (MLA), multidimensional scaling (IBM SPSS PROXSCAL), and multidimensional unfolding (IBM SPSS PREFSCAL). From 2000 to 2007, Frank was a part time teacher at the primary school Oostelijke Eilanden in Amsterdam.

Frank is co-author of publications on resampling in multilevel analysis, multidimensional scaling, and multidimensional unfolding. Currently, Frank is employed as researcher and teacher to the Section for Psychometrics and Research Methodology, Institute of Psychology, Faculty of Social and Behavioural Sciences, Leiden University.