



Universiteit
Leiden
The Netherlands

Advances in multidimensional unfolding

Busing, F.M.T.A.

Citation

Busing, F. M. T. A. (2010, April 21). *Advances in multidimensional unfolding*. Retrieved from <https://hdl.handle.net/1887/15279>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/15279>

Note: To cite this publication please use the final published version (if applicable).

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 Bergemaier (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

This chapter is a revised version of 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.

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

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

where the γ_{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 <i>Stress-2</i> to avoid trivial solutions; Notification of importance of conditionality. | MINIRSA |
| 1969 | Kruskal Carroll | Development of <i>Stress-2</i> 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 <i>Stress-2</i> ; Classification of degeneracy types. | |
| 1984 | DeSarbo Rao | Fixed cell weights emphasizing certain cells to avoid degeneracies; Fast algorithm minimizing <i>Stress-2</i> . | 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 Lingo, 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,

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

with $i = 1, \dots, n$ the row entries (judges) and $j = 1, \dots, m$ 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 Lingoés in Michigan, he extensively compared the algorithms developed by Kruskal (*M-D-SCAL*, see Kruskal & Carmone, 1969) and by Guttman and Lingoés (*SSA*). This collaboration resulted in a monograph supplement of *Psychometrika* (Lingoés & Roskam, 1973) and in the *MINISSA* (Roskam & Lingoés, 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 Lingoés & 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 (Lingoés & 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/mini-rsa.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.

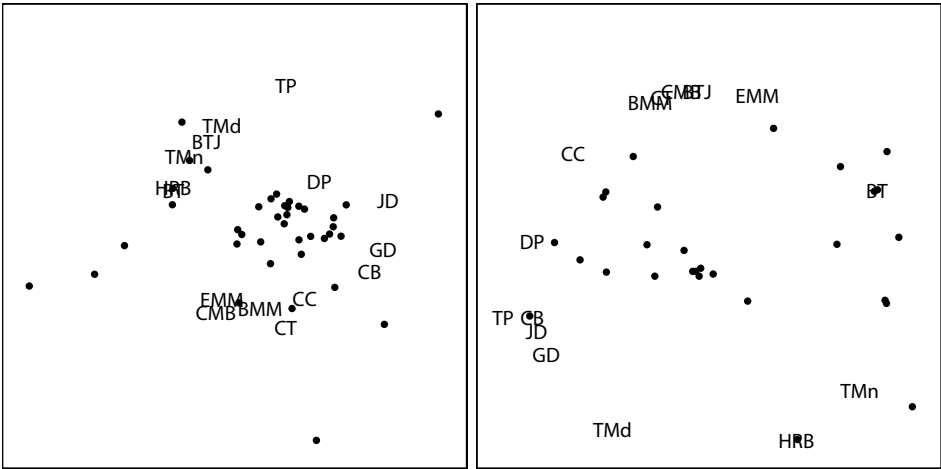


Figure 2.1 MINIRSA solution for the breakfast data (left-hand panel) and KYST unfolding solution of the breakfast data using Stress-2 and a rational start (right-hand panel).

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

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

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*

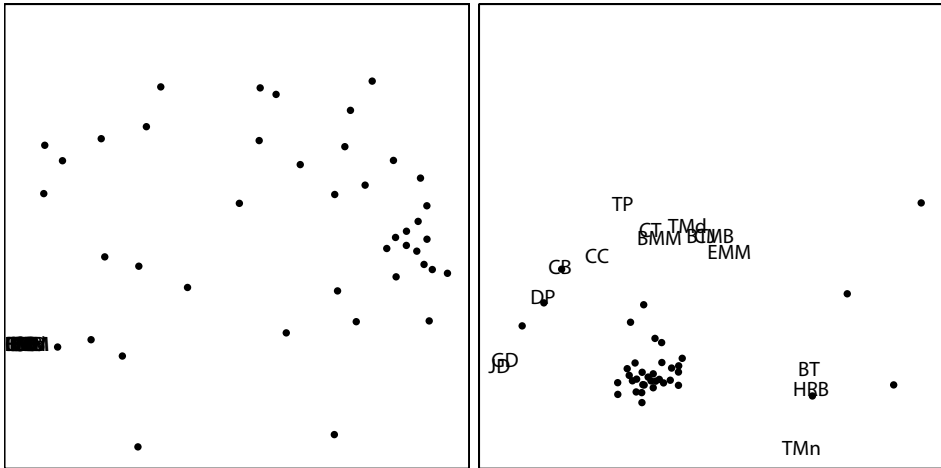


Figure 2.2 SSAP-II unfolding solution of the breakfast data (left-hand panel) with all breakfast items in a clutter of black ink on the left side and a mixed ordinal-interval unfolding solution of the breakfast data (right-hand panel).

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 were 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, particularly 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 (α) and upper bounds (β) for the transformed data (Γ), based on the raw data (Δ) with the smallest dissimilarity set to zero (Heiser, 1981, p. 223), such that $\beta(\delta_l - \delta_{l-1}) \geq \gamma_l - \gamma_{l-1} \geq \alpha(\delta_l - \delta_{l-1})$. 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 $0 \leq \alpha \leq 1$, making α smaller means a bigger cone, and introducing degeneracies for $\alpha \rightarrow 0$. To determine an optimal α , 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 α , 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

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

with $0 \leq \alpha \leq 1$. 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’Hôpital 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

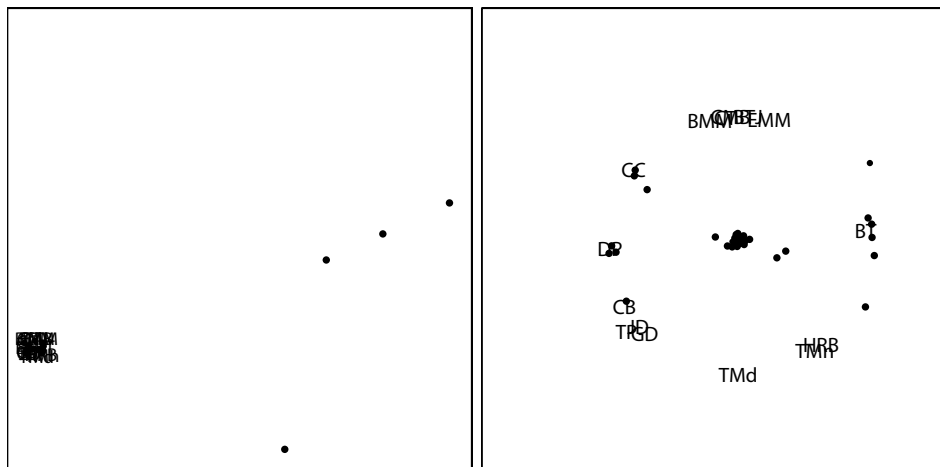


Figure 2.3 Configuration for the breakfast data when minimizing Stress-2 with a convergent algorithm. The right-hand panel depicts a detail of the left-hand panel.

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).

2.9 DESARBO AND RAO, 1984

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 δ_{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 an empirical issue, depending upon the data.

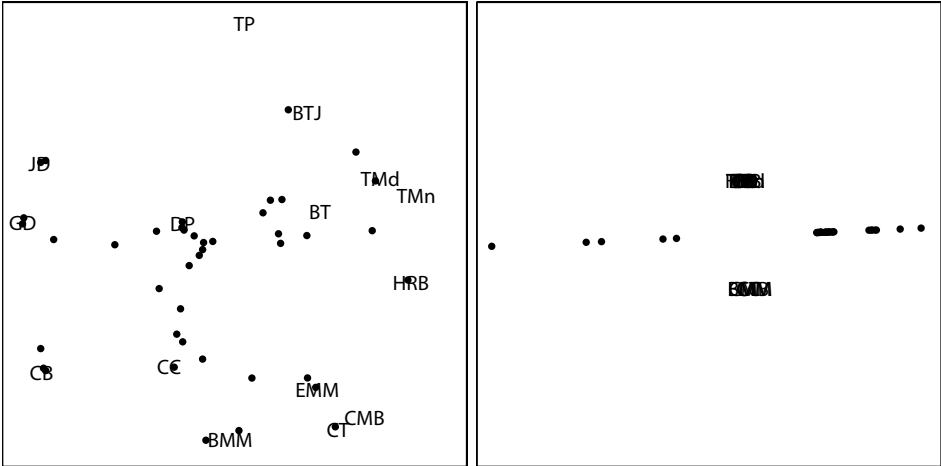


Figure 2.4 GENFOLD solution with $p=2$ for the breakfast data minimizing Stress-2 (left-hand panel) and Stress-1 (right-hand panel).

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 led 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., $\gamma_l - \gamma_{l-1} \geq 0$, smoothness is defined as a condition on the second order differences, as $|\theta_l - \theta_{l-1}| \leq \bar{\theta}$, where $\theta_l = \gamma_l - \gamma_{l-1}$ and $\bar{\theta}$ 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

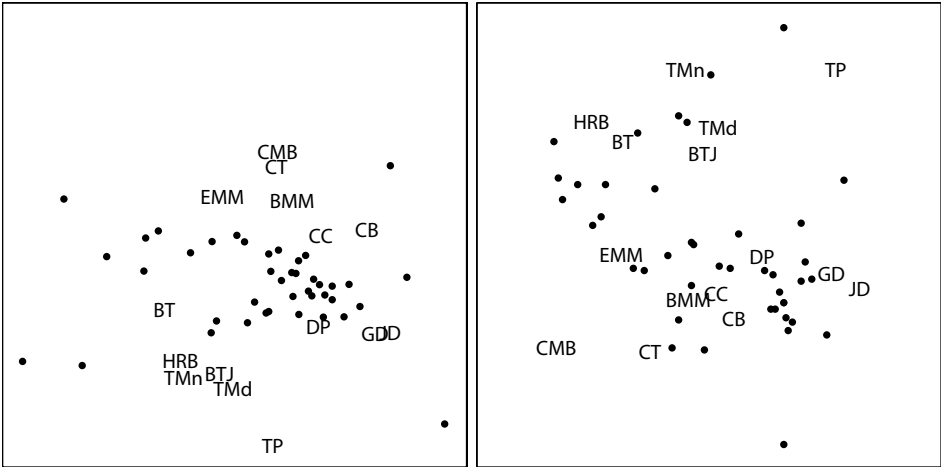


Figure 2.5 SMACOF-3b solution for the breakfast data (left-hand panel) and NEWFOLD solution for the breakfast data (right-hand panel).

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).

2.11 KIM, RANGASWAMY, AND DESARBO, 1999

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 Δ , 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.

