



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).

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.

This chapter is an adapted version of 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, 71–98.

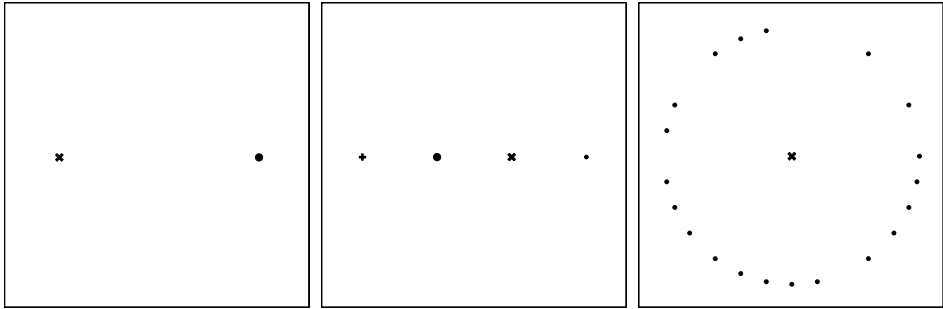


Figure 4.1 Typical degenerate solutions. The big dot represents multiple ideal points, the little dot represents a single ideal point, the cross represents multiple stimulus points, and the plus sign represents a single stimulus point.

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-Vogeleang (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 $i = 1, \dots, n$ and $k = 1, \dots, p$, 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} ($j = 1, \dots, m$ and $k = 1, \dots, p$) and rows \mathbf{y}_j . The data of the problem are either a set of dissimilarities δ_{ij} , or a set of similarities ρ_{ij} , where the indices have the same range as before. In both cases, we need to find a new set of quantities γ_{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

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

and for the total group of individuals we obtain

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

The notation $\sigma_{\text{row}}^2(\boldsymbol{\gamma}_i, \mathbf{x}_i, \mathbf{Y})$ and $\sigma_{\text{r}}^2(\boldsymbol{\Gamma}, \mathbf{X}, \mathbf{Y})$ 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 \hat{d}_{ij} , called *d-hats*, and the resulting function $\sigma_{\text{r}}^2(\hat{\mathbf{D}}, \mathbf{X}, \mathbf{Y})$ is now only a function of \mathbf{X} and \mathbf{Y} , since $\hat{\mathbf{D}}$ 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 Γ or \mathbf{X} and \mathbf{Y} , or both. For example, the square of Kruskal's STRESS-1 is given by

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

where $\eta_1^2(\mathbf{X}, \mathbf{Y}) = w_{++}^{-1} \sum_{i=1}^n \sum_{j=1}^m w_{ij} d^2(x_i, y_j)$. 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 $d(x_i, y_j) = d$ 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 Ω is always a cone, that is, it always satisfies the condition that $\Gamma \in \Omega$ implies $\beta\Gamma \in \Omega$, for any non-negative scalar β . Finally, presence of an intercept in the transformation means that Ω is closed under nondecreasing linear transformations, that is, it satisfies the condition that $\Gamma \in \Omega$ implies $(\beta\Gamma + \alpha\mathbf{E}) \in \Omega$, for non-negative α and β , 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 Ω is closed under nondecreasing linear transformations. Then it is valid to select from the cone the pseudo-distances $\alpha \mathbf{E} \in \Omega$ for some positive α , so that R-STRESS becomes

$$\sigma_r^2(\alpha \mathbf{E}, \mathbf{X}, \mathbf{Y}) = \frac{1}{w_{++}} \sum_{i=1}^n \sum_{j=1}^m w_{ij} (\alpha - d(\mathbf{x}_i, \mathbf{y}_j))^2.$$

Taking $\mathbf{y}_j = \mathbf{0}$ and $\mathbf{x}_i = \alpha \|\mathbf{a}_i\|^{-1} \mathbf{a}_i$, where \mathbf{a}_i is some arbitrary p -vector, yields $d(\mathbf{x}_i, \mathbf{y}_j) = \alpha$, 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 $d(\mathbf{x}_i, \mathbf{y}_j) = \alpha$. 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 $\alpha \mathbf{E} \in \Omega$. \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 $\eta_1^2(\mathbf{X}, \mathbf{Y}) = \alpha^2 > 0$.

For the conditional case, we call a solution (\mathbf{X}, \mathbf{Y}) absolutely degenerate if and only if R-STRESS (4.2) is zero and $d(\mathbf{x}_i, \mathbf{y}_j) = d_i$ 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 Ω_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 $\mathbf{x}_i = \alpha_i \|\mathbf{a}_i\|^{-1} \mathbf{a}_i$ for any positive α_i , together with $\mathbf{y}_i = \mathbf{0}$. Thus, \mathbf{x}_i is in fact completely arbitrary, and solutions of this type have the property that the matrix of pseudo-distances satisfies $\Gamma = \boldsymbol{\alpha} \mathbf{e}_m$, with \mathbf{e}_m an m -vector of ones and $\boldsymbol{\alpha}$ 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 (\mathbf{X}, \mathbf{Y}) is called *partially degenerate* if and only if R-STRESS (4.1) is zero and $d(\mathbf{x}_i, \mathbf{y}_j) = d_i$ for some i ($1 \leq i < n$), 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 Ω or Ω_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 $\sigma_{\text{row}}^2(\hat{\mathbf{d}}_i, \mathbf{x}_i, \mathbf{Y})$ 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 $\eta_2^2(\mathbf{x}_i, \mathbf{Y}) = w_{i+}^{-1} \sum_{j=1}^m w_{ij} (d(\mathbf{x}_i, \mathbf{y}_j) - \bar{d}_i)^2$, where $\bar{d}_i = w_{i+}^{-1} \sum_{j=1}^m w_{ij} d(\mathbf{x}_i, \mathbf{y}_j)$, the square of Kruskal’s STRESS-2 generalized to the weighted row-conditional or split-by-rows case is defined as

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

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 $\eta_2^2(\mathbf{x}_i, \mathbf{Y}) = 0$, 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 $\sigma_2^2(\mathbf{X}, \mathbf{Y})$ 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

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

with $\eta_3^2(\gamma_i) = w_{i+}^{-1} \sum_{j=1}^m w_{ij} (\gamma_{ij} - \bar{\gamma}_i)^2$, where $\bar{\gamma}_i = w_{i+}^{-1} \sum_{j=1}^m w_{ij} \gamma_{ij}$. 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 \mathbf{X} and \mathbf{Y} for any given Γ 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 $v(\mathbf{a})$ can be written as

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

It is clear from (4.6) that $v(\mathbf{a})$ achieves its minimal value when all a_t are equal, that is, $v(\mathbf{a}) = 0$ iff $a_t = \bar{a}$ 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).

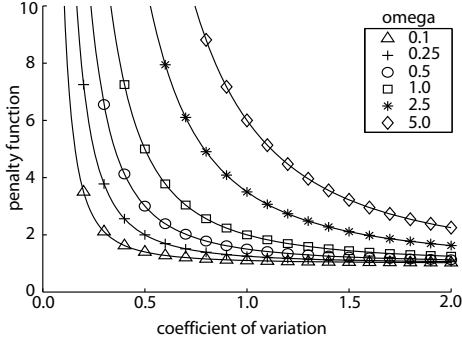


Figure 4.2 Penalty function $\mu(\Gamma)$ versus variation coefficient $v(\Gamma)$ for different values of ω (see legend).

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

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

where λ is a lack-of-penalty parameter ($0 < \lambda \leq 1$) that controls the balance between the R-STRESS $\sigma_r^2(\Gamma, \mathbf{X}, \mathbf{Y})$ and the penalty function $\mu(\Gamma)$. When λ 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 Γ , so that for any given value of Γ minimizing $\sigma_p^2(\Gamma, \mathbf{X}, \mathbf{Y})$ amounts to minimizing $\sigma_r^2(\Gamma, \mathbf{X}, \mathbf{Y})$. The penalty function is chosen to be a multiplicative factor, and such that it approaches 1 when $v(\Gamma)$ becomes large, and grows to infinity when $v(\Gamma)$ goes to zero. A simple set of functions that has these properties is defined as

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

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

The shape of (4.8) as a function of v is illustrated in Figure 4.2 for different values of ω . 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 γ_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 $v(\gamma_i)$, we use the harmonic mean to aggregate them, yielding the penalty function

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

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

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

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 λ and ω 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 λ and ω , 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 λ and ω , based on the results of several pilot studies. Values for λ ranged from 0.1 to 1 with increments of 0.1 and values for ω were chosen as 0.1, 0.25, 0.5, 1, 2.5, and 5. A strong penalty is marked by small values for λ in combination with large values for ω , 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%	λ	10	0.1, 0.2, 0.3, ..., 0.9, 1.0
Data Level	2	Interval, Ordinal	ω	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 $\sigma_r^2(\Gamma, \mathbf{X}, \mathbf{Y})$, the (conditional) variation coefficient of the pseudo-distances, that is, $v(\Gamma)$ for the unconditional case and $v_c(\Gamma) = [n^{-1} \sum_{i=1}^n v^{-1}(\gamma_i)]^{-1}$ for the row-conditional case, and the variation coefficient of the distances $v(\mathbf{D})$.

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 configuration. 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 parameters 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 remaining 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 reasons 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 $n m(m-1)/2$ inequalities to be satisfied, compared to a maximum $n m(n m - 1)/2$ 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 often 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 λ in the neighborhood of 0.5.

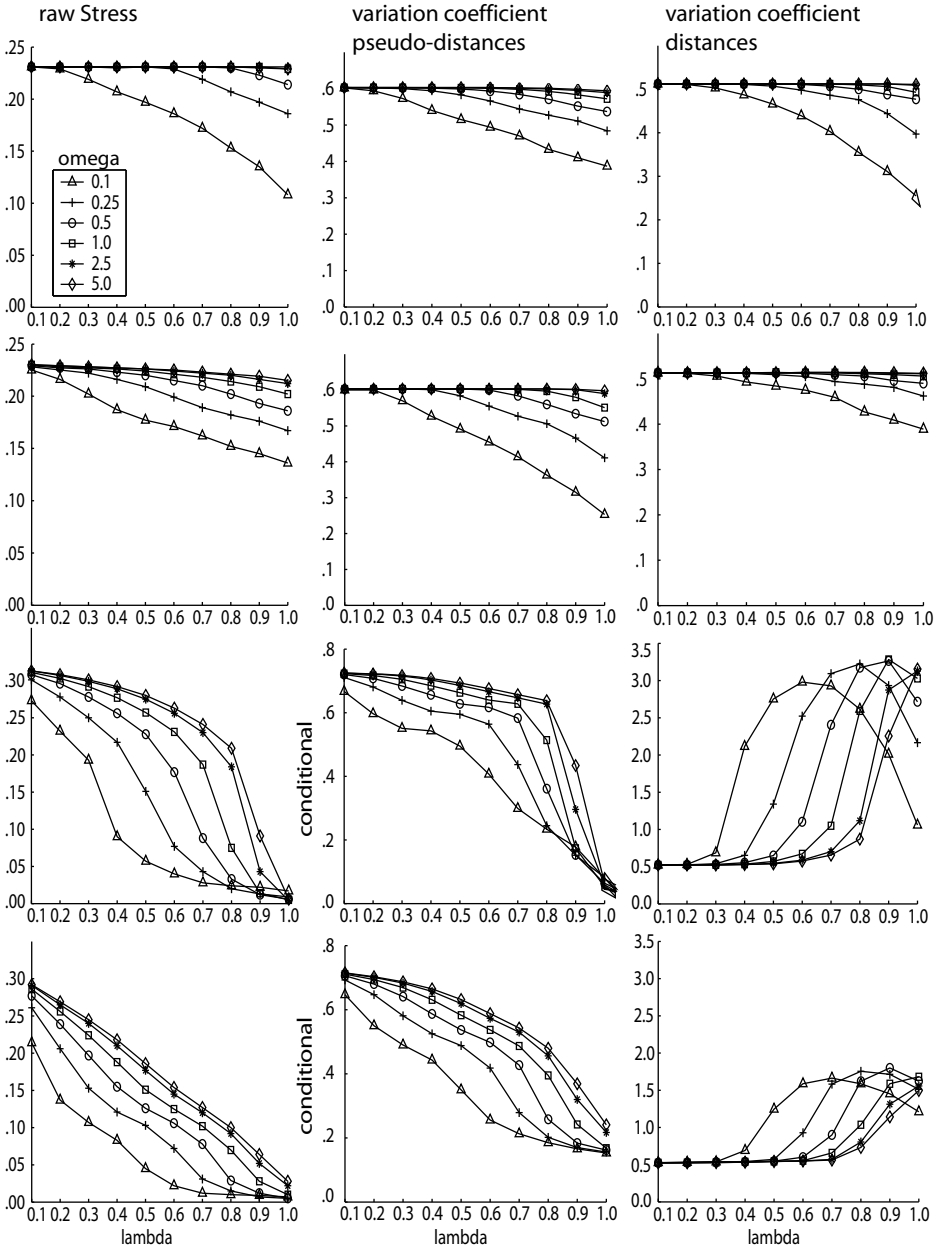
For the error-perturbed condition, the results are presented in Figure 4.3. Note that the average R-STRESS for the error-perturbed condition imposed during the simulation process is 0.245.

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	λ	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): $\sigma_r(\Gamma, X, Y)$, $v_c(\Gamma)$, and $v(\mathbf{D})$, with lines showing the averages over 50 replications for different values of ω (the legend for all panels is displayed in the top-left-hand panel).



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 λ 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 λ to 0.5 specifies the function as a usual square root, whereas ω 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.

4.5 APPLICATIONS

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 ALSICAL (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 γ_{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 Bergermaier (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 (ρ ; RHO), and Kendall's tau- b coefficient (τ_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 (v). 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.

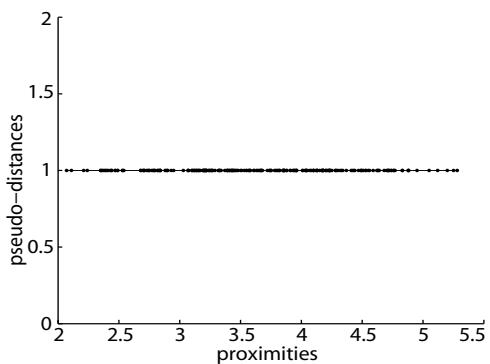
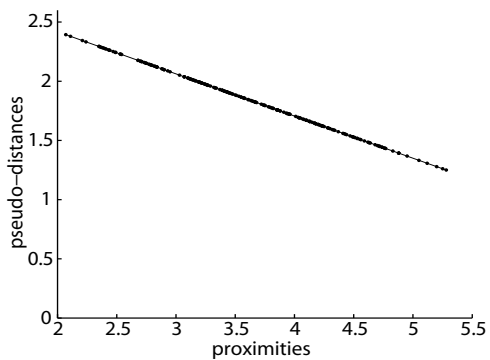
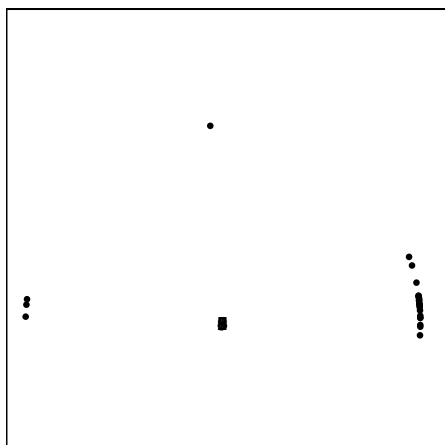
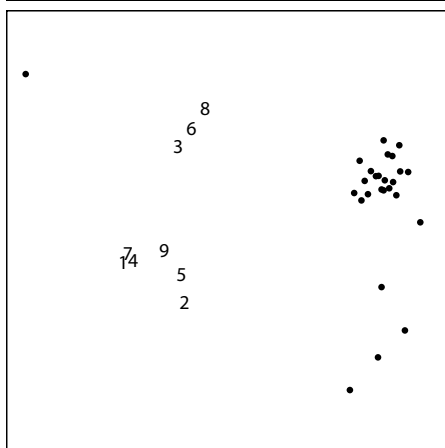
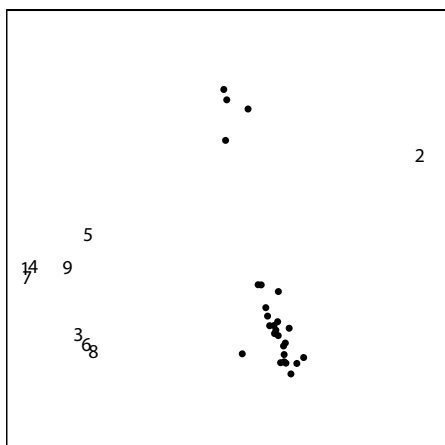
Finally, we provide a non-degeneracy index (D-INDEX), suggested by Shepard (1974), and an intermixedness index (I-INDEX), suggested by DeSarbo, Young, and Rangaswamy (1997) (for formulas, see Technical Appendix G, pages 220–221). Shepard described “a rough index of the *nondegeneracy* of a solution” 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, $I_1 = \ln(\bar{d}_y/\bar{d}_{xy})$, $I_2 = \ln(\bar{d}_x/\bar{d}_{xy})$, and $I_3 = \ln(\bar{d}_y/\bar{d}_x)$, where \bar{d}_x and \bar{d}_y are the average within-set-distances and \bar{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 ALSICAL (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 $p \rightarrow \infty$.

The solutions of ALSICAL 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 ALSICAL 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 ALSICAL, since the program erroneously provides distances instead of pseudo-distances in its output.

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

	ALSICAL	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
TAU _b	0.824	0.835	0.000	0.495	0.844
Variability					
var(D)	0.020	0.022	0.000	0.158	0.143
v(D)	0.144	0.150	0.000	0.433	0.404
var(Γ)	N/A	0.020	0.000	0.165	0.165
v(Γ)	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.

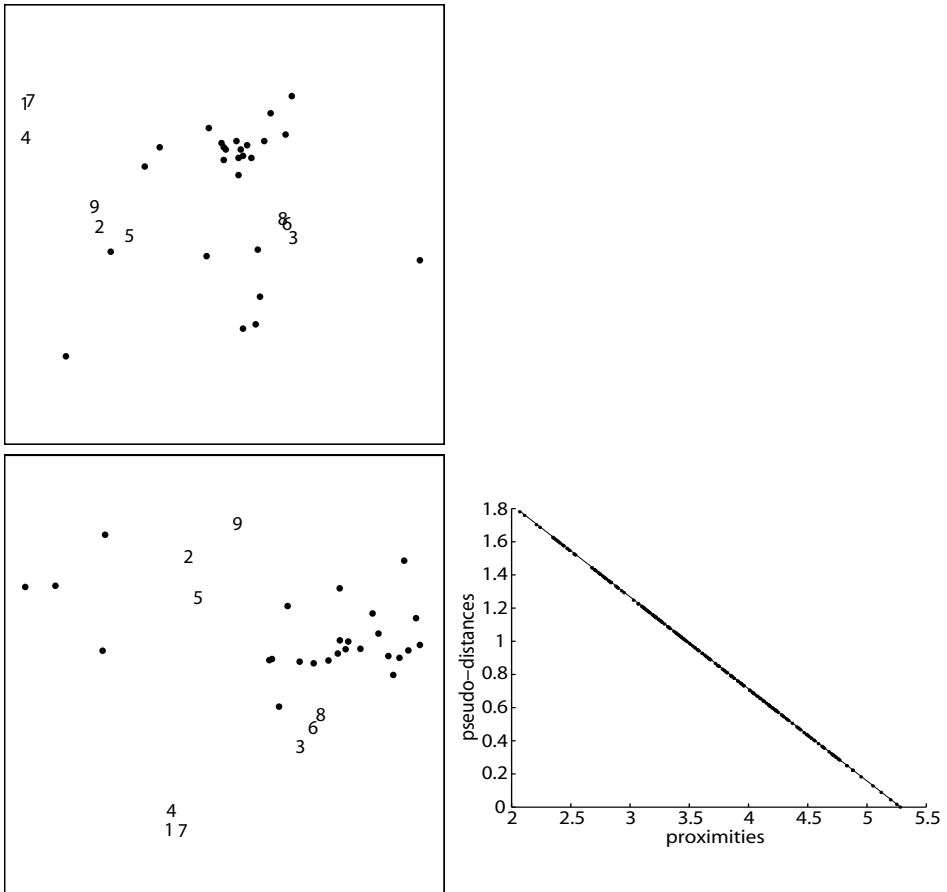


Figure 4.5 Configurations (left-side panels) and transformation plot (right-side panel) for the brewery data set with an unconditional interval transformation using NEWFOLD (top panel) and PREFSCAL with $\lambda=0.5$ and $\omega=0.5$ (bottom panels).

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 ALSICAL 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 ALSICAL and KYST. However, the fit measures for NEWFOLD are not good; these measures are even worse than those for ALSICAL 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.

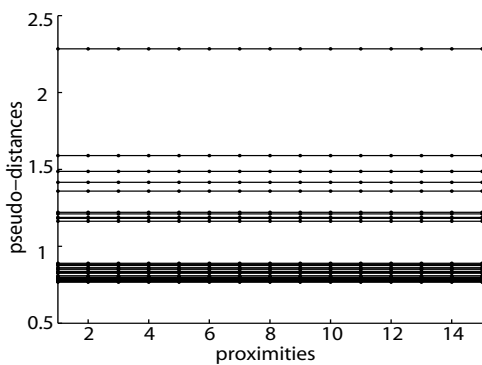
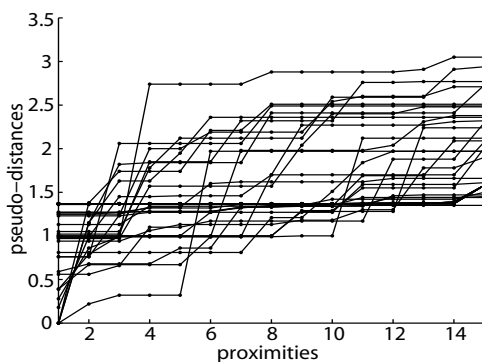
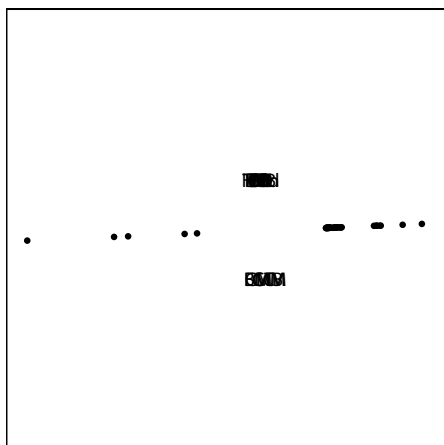
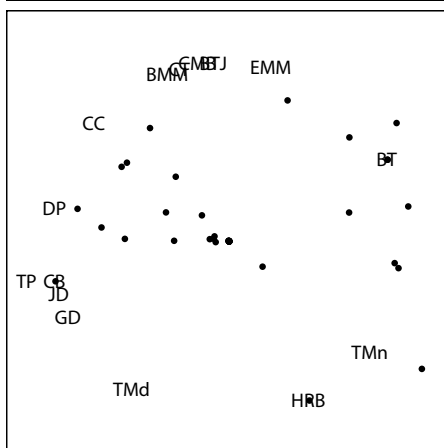
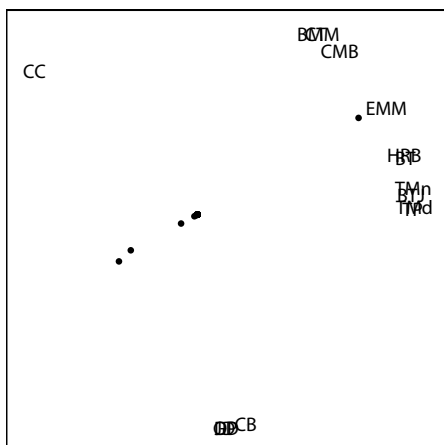
The PREFSCAL solution in Figure 4.5 (bottom panels) provides the best fit statistics, non-degeneracy index D-INDEX, and intermixedness index I-INDEX, while maintaining sufficient variability in both pseudo-distances and distances. Yet the PREFSCAL configuration is in close agreement with the metric solution 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 ALSICAL, 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 ALSICAL and GENFOLD show degenerate solutions. For ALSICAL, 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 o/o. Actually, the value becomes $\epsilon/\epsilon^2 = \epsilon^{-1}$, where ϵ is a very small number: The numerator indicates that the weighted sum-of-squares loss function is close to zero (ϵ),

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 ALSICAL (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 (ϵ^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 ALSICAL 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 ALSICAL, KYST, GENFOLD, NEWFOLD, and PREFSCAL for the breakfast data set.

	ALSICAL	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					
var(\mathbf{D})	0.013	0.122	0.091	0.197	0.192
$v(\mathbf{D})$	0.114	0.374	0.317	0.495	0.483
var $_c(\Gamma)$	N/A	0.003	0.000	0.369	0.233
$v_c(\Gamma)$	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).

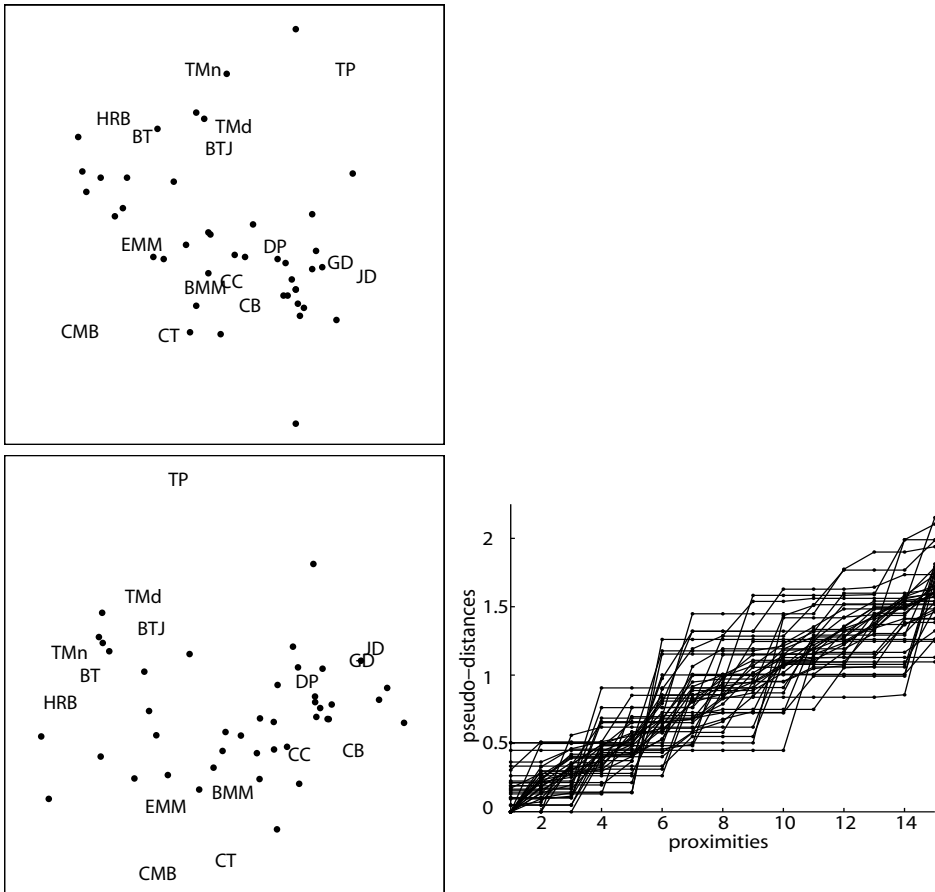


Figure 4.7 Configurations (left-side panels) and transformation plot (right-side panel) for the breakfast data set with row-conditional ordinal transformations using NEWFOLD (top panel) and PREFSCAL with $\lambda=0.5$ and $\omega=0.3$ (bottom panels). For the breakfast plotting codes, see Table 2.1.

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 *nondegeneracy* 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 intersets 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 o/o. 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 lose 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.

