Sonderdruck aus: „Archiv für Psychologie", *128* (1976) 240–266

Aus dem Psychologischen Institut der Universität Heidelberg

Nonmetric Multidimensional Scaling without Disparities and Derivatives: A Rankcorrelation-Orientated Approach Through L_1-Approximation

By

Claus Möbus

With 3 figures

(Received May 5, 1975)

AKADEMISCHE VERLAGSGESELLSCHAFT
WIESBADEN
1976

Résumés

A new nonmetric multidimensional scaling technique is presented. The method is transformation-free (no monotonic data transformations) and operates directly on distances and data-rankorders. Numeric examples demonstrate that this method is not inferior to the classical nonmetric techniques. An analysis of LINGOES & ROSKAM's (1973) order-4 matrices produced results with a still improved rank-order-isomorphy.

Eine neue nonmetrische multidimensionale Skalierungstechnik wird vorgestellt. Die Methode verzichtet auf monotone Datentransformationen und benutzt nur Distanzen und Rangordnungen innerhalb der Daten. Numerische Beispiele demonstrieren, daß die Methode den klassischen nonmetrischen Skalierverfahren nicht unterlegen ist. Eine Analyse von LINGOES und ROSKAM (1973) Ordnungs-4 Matrizen brachte Ergebnisse mit einer noch besseren Rangordnungsisomorphie.

Background

Nonmetric multidimensional scaling as proposed by KRUSKAL (1964a, 1964b) and GUTTMAN (1968) involves minimization of a badness-of-fit measure. Minimization is done iteratively. Each iteration consists of two stages:

$$\text{Stage 1: } k = k+1$$
$$_1L_k\,(d_k|\hat{d}_{k-1}) = \min! \tag{1}$$
$$\text{Stage 2:}$$
$$_2L_k\,(\hat{d}_k|d_k) = \min! \tag{2}$$

where: k = number of iteration
$L(d|\tilde{d})$ means minimization of L with fixed $\tilde{d}$ and variable d
(L = statistical loss-function)
d = set of distances (derived from the model)
$\tilde{d}$ = set of disparities (derived from the data by means of monotone regression or rankimage transformation).
The disparities are "close" to the d but constrained to some criterion of monotonicity with the data.

Convergence of this process is obtained if ${}_1L_k = {}_2L_k$ or $d_k = d_{k+1}$ or $\tilde{d}_k = \tilde{d}_{k+1}$. We realize that the set $\tilde{d}$ is *absolutely necessary* for this kind of nonmetric "two-stage"-scaling. In our "one-stage" algorithm we can omit the disparities without any loss of information.

There remains still one point of criticism: implicit weighting of "error" by quadratic "error" terms. L is in most programs a quadratic "error" function (least-squares principle): each "error" $(d_{ij} - \tilde{d}_{ij})$ is weighted by a factor $(d_{ij} - \tilde{d}_{ij})$. Larger departures from monotonicity are weighted more than really necessary.

If we look at L as an euclidean distance function, which measures the distance of the configuration from the ideal configuration, we can use the same arguments which were written about the problem of implicit weighting in MINKOWSKY-distance-functions (CROSS, 1965; WENDER, 1969; COOMBS, DAWES & TVERSKY, 1970; AHRENS, 1972, 1973). An L with greater sensivity and without implicit weighting of departures from monotonicity should look like the city-block-distance (L_1-approximation). Look at (3) and (4).

Implicit weighting of discrepancies by MINKOWSKY-*distance-formula:*

$$d_r(i,j) = \left[\sum_k |x_{ik} - x_{jk}|^r\right]^{1/r} = \sum_k \left[\frac{|x_{ik} - x_{jk}|}{d_r(i,j)}\right]^{r-1} \cdot |x_{ik} - x_{jk}| \tag{3}$$

r = metric-parameter
$(1 \leq r \leq \infty)$

$$= \sum \left\{ \begin{matrix} \text{weight of} \\ \text{specific-distance} \end{matrix} \right\} \left\{ \begin{matrix} \text{dimension-} \\ \text{specific-distance} \end{matrix} \right\}$$

Implicit weighting of departures from monotonicity by two-stage lossfunction L*:*

$$L_r(\tilde{d}, d) = \sum_{ij} \left[\frac{|\tilde{d}_{ij} - d_{ij}|}{N L_r(\tilde{d}, d)}\right]^{r-1} \cdot |\tilde{d}_{ij} - d_{ij}| \tag{4}$$

$$= \sum \left\{ \text{weight} \right\} \cdot \left\{ \begin{matrix} \text{departure from} \\ \text{monotonicity} \end{matrix} \right\}$$

if r = 2, then least-squares-method (L_2-approximation)
N = normalization factor which is specific for the chosen algorithm
L_r = L_r-approximation

Only in the case of L_1-approximation (SPÄTH, 1973, 1974) are the errors $|\tilde{d}_{ij}-d_{ij}|$ weighted equally with a factor 1. Because of the lack of a statistical error theory the desirable statistical properties of L_2-approximation (least squares estimators) do not hold in nonmetric scaling contexts. Therefore, the only reason for using L_2-approximation lies in the smoothness of the loss function, so that one can use standard gradient techniques for minimization.

A further point to improve is the *nonmetric aggregation of individual data* which is treated rather cumbersomely in the classical approaches. We shall show below (20) how this can be done very easily and efficiently, even for large data samples.

The degree of approximation is evaluated in most "two-stage" programs by means of three measures of fit: KRUSKAL's stress 1 and stress 2 and GUTTMAN's Φ.

KRUSKAL's *stress 1:*

$$L = S_1^2 = \frac{\sum_{i>j} (\tilde{d}_{ij} - d_{ij})^2}{\sum_{i>j} d_{ij}^2} = \frac{\sum_{i>j} |\tilde{d}_{ij} - d_{ij}||\tilde{d}_{ij} - d_{ij}|}{\sum_{j>i} d_{ij}^2} \tag{5}$$

KRUSKAL's *stress 2:*

$$L = S_2^2 = \frac{\sum_{i>j} (\tilde{d}_{ij} - d_{ij})^2}{\sum_{i>j} (d_{ij} - \bar{d}\ )^2} = \frac{\sum_{i>j} |\tilde{d}_{ij} - d_{ij}||\tilde{d}_{ij} - d_{ij}|}{\sum_{i>j} (d_{ij} - \bar{d})^2} \tag{6}$$

where: $\tilde{d}_{ij}$ = disparities as a result of KRUSKAL's monotone regression
$\bar{d}$ = mean of all distances
d_{ij} = dissimilarity

GUTTMAN's *normalized phi:*

$$L = \Phi = \frac{\sum_{i>j} (\tilde{d}_{ij} - d_{ij})^2}{2 \sum_{i>j} d_{ij}^2} = \frac{\sum_{i>j} |\tilde{d}_{ij} - d_{ij}||\tilde{d}_{ij} - d_{ij}|}{2 \sum_{i>j} d_{ij}^2} \tag{7}$$

where: $\tilde{d}_{ij}$ = disparities as a result of GUTTMAN's rank-image transformation.

The distinction between S_1 and S_2 is only relevant in the case of "off-diagonal" data matrices.

A remarkable step in a new "one-stage" was taken by JOHNSON[1] (1973) by introducing a type of stress (loss), which operates only on distances and rank-orders derived from the data. Monotone data-transforms are not used any longer. Minimization reduces to an one-stage process with no iterations:

$$\text{Stage 1: } k = 1$$
$$L_k\,(d_k) = \min!$$

JOHNSON's measure of stress is free of disparities and depends solely on distances and data-rankorders. His algorithm is restricted to the euclidean metric and symmetric proximity matrices.

JOHNSON's *stress:*

$$L = S^2 = \frac{\sum\limits_{\substack{i>j\\k>l\\IJ<KL}} \delta_{ij,\,kl}(d_{ij}^2 - d_{ik}^2)^2}{\sum\limits_{\substack{i>j\\k>l\\IJ<KL}} (d_{ij}^2 - d_{kl}^2)^2} = \min! \tag{8}$$

where:

$$\delta_{ij,\,kl} = \begin{cases} 1, \text{ if sign } (d_{ij} - d_{kl}) \\ \quad \neq \text{ sign } (\hat{d}_{ij} - \hat{d}_{kl}) \\ 0, \text{ otherwise and if } d_{ij} = d_{kl} \end{cases}$$
$$IJ = (i-2)(i-1)/2 + j$$
$$KL = (k-2)(k-1)/2 + l$$
$$\hat{d}_{ij} = \text{dissimilarity between } i \text{ and } j$$
$$d_{ij} = \text{distance between } i \text{ and } j$$

The restriction $IJ < KL$ was not introduced by JOHNSON, but it is helpful because it halves computer time for evaluation of L. If the distances are in the wrong order (= inversion of distances), the numerator is increased by a "well-defined" amount. Are on the other hand the distances in the right order (pro-version of distances) the numerator is not increased. Ties are treated according to two options: KRUSKAL's PRIMARY (distances may differ even in the case of tied proximities) and SECONDARY (distances may not differ in the case of tied proximities) approach. In the first case the term $(\hat{d}_{ij}^2 - \hat{d}_{kl}^2)^2$ should be added to neither numerator nor denominator of L. In the second case $(\hat{d}_{ij}^2 - \hat{d}_{kl}^2)^2$ should be added to both numerator and denominator. If $\hat{d}_{ij}$ is unknown all comparisons involving d_{ij} are completely left out (missing data option). In JOHNSON's stress (8) we notice an even greater weighting of error than in (5)–(7).

---

[1] I am very grateful to IAN SPENCE and FORREST W. YOUNG for reading an earlier draft of the manuscript. They also informed me that GUTTMAN (at the XIXth Congress for Psychology in 1969 in London) and DE LEEUW (unpublished manuscripts) were the very first to introduce the concept of transformation-free scaling.

We think, that only the discrepancy $|d_{ij} - d_{kl}|$ should be of any interest, if d_{ij} and d_{kl} show an inversion. This discrepancy is hidden in JOHNSON's stress. It is weighted by a factor, which deteriorates the *orderisomorphy* between the dissimilarity data and the distances. In a later section we will see that this is true at least for order-4 matrices which were used by LINGOES & ROSKAM (1973) comparing various scaling algorithms. If we look at the numerator we can factor out:

$$(d_{ij}^2 - d_{kl}^2)^2 = \underbrace{|d_{ij} - d_{kl}|}_{\substack{\text{discrepancy} \\ = \text{amount of} \\ \text{inversion}}} \cdot \underbrace{|(d_{ij} - d_{kl})(d_{ij} + d_{kl})^2|}_{\text{weighting factor}} \tag{9}$$

Discrepancies are weighted much more in the case of great distances. The algorithm tends to reduce the weighting factor first. Reducing the amount of discrepancies is not of primary interest. This will sometimes happen at the cost of:

$$(L) = S_{-p} = \frac{S_-}{S_- + S_+} = \frac{\sum \delta_{ij,\,kl}}{\sum \delta_{ij,\,kl} + \sum \varepsilon_{ij,\,kl}} = \delta_p$$

where: S_+ = Sum of proversions
S_- = Sum of inversions
δ = 1, if an inversion occurs; 0, otherwise
ε = 1, if an proversion occurs; 0, otherwise

δ_p possesses a strong connection to KENDALL's τ, to KENDALL-SILLITTO's τ and GOODMAN-KRUSKAL's γ (KENDALL, 1948; SILLITTO, 1947; 1959) as will be shown in the "Method"-section. JOHNSON's algorithm pulls and pushes points with great distances into a "best" position even at the cost of an increasing number of inversions. Thus *orderisomorphy* between data and model (one of the main goals of nonmetric MDS) is deteriorated. Although S_{-p} reflects the number of violations of rankisomorphy it is not appropriate for a loss function. Being a step-function it is awkward to minimize: one needs special algorithms for integer programming. Further, S_{-p} only takes the *occurrence* but not the severity of each inversion into account. But, because we are provided with metric information on the model side, we can extract more information than S_{-p} does. Thus we can improve S_{-p}, which leads us to (10).

Method

On the basis of SHEPARD's verbal definition of nonmetric MDS we want to propose a badness-of-fit measure for a "one-stage" procedure without any implicit weighting of "error" or discrepancy (MÖBUS, 1974).

SHEPARD (1972) gave the clear and distinctive definition of nonmetric scaling:

"We seek, simply, that configuration of n points in the (Euclidean) space of smallest possible dimension such that, to an acceptable degree of approximation, the resulting interpoint distances d_{ij} are monotonically related to the given proximity data in the sense that

$$d_{ij} < d_{kl} \text{ whenever } \hat{d}_{ij} < \hat{d}_{kl} \text{ ."}$$

Our stress can be looked at an nonmetric analogue of S_{-p}, which plays a prominent role in rank-correlation methodology (at least in the KENDALL-tradition). Thus minimizing our stress is equivalent to maximizing a nonmetric analogue to KENDALL's (purely ordinal) rank-correlation coefficient τ. Thus being a "pairwise" method at the first glance it is now obvious that this kind of stress uses information from the full rank-orders. Our algorithm can be used for all meaningful distance-formulas. We extend the stress-measure to the case of square nonsymmetric un/conditional matrices and show how to aggregate individual data the nonmetric way.

Our type of stress: "Proportion of weighted inversions S_{w-p}:

$$L = S_{w-p} = \frac{\sum_{\substack{i>j \\ k>l \\ IJ<KL}} \delta_{ij,\,kl} |d_{ij} - d_{kl}|}{\left[\sum_{\substack{i>j \\ k>l \\ IJ<KL}} \delta_{ij,\,kl} |d_{ij} - d_{kl}| \right] + \left[\sum_{\substack{i>j \\ k>l \\ IJ<KL}} \varepsilon_{ij,\,kl} |d_{ij} - d_{kl}| \right]} \tag{10}$$

where:

δ	ε	if $(\hat{d}_{ij} - \hat{d}_{kl})$	and	$(d_{ij} - d_{kl})$
0	1	> 0		> 0
0	0	> 0		$= 0$
1	0	> 0		< 0
PRIMARY (SECONDARY)				
0 (1)	0 (0)	$= 0$		> 0
0	0	$= 0$		$= 0$
0 (1)	0 (0)	$= 0$		< 0
1	0	< 0		> 0
0	0	< 0		$= 0$
0	1	< 0		< 0

In our method we compare pairs of dissimilarities with pairs of distances. (10) is simple to read. The inequalities $i > j$ and $k > l$ restrict us to the lower half of the symmetric dissimilarity matrix. The inequality $IJ < KL$ saves us from unnecessary comparisons. If the distances are not in the same order as the dissimilarities, we notice an inversion and $\delta = 1$. The severity of the inversion is equal to $|d_{ij} - d_{kl}|$. This amount is added to the numerator. In the case of orderisomorphy, (distances and dissimilarities show the correct order), we notice a proversion and $\varepsilon = 1$. The actual errors are added in the numerator and the maximum possible errors are added in the denominator. Thus S_{w-p} is normalized and lies between 0 and 1.

Transforming S_{w-p} we get a coefficient ψ which shows a strong resemblance to KENDALL's τ and which can be looked at its nonmetric analogue:

$$-1 \leq \psi = 1 - 2S_{w-p} \leq +1 .$$

This can be seen very easy if we transform S_{-p} (or δ_p) into KENDALL's τ (no ties) or into GOODMAN & KRUSKAL's γ (ties and PRIMARY approach) just the same way:

$$-1 \leq \tau = 1 - 2S_{-p} \leq +1 .$$

We can summarize the comparison between S_{w-p}, S_{-p} and JOHNSON's stress as follows:

	desirable weighting of inversion δ	undesirable weighting of error $\lvert d_{ij}-d_{kl}\rvert$	transformation to
S_{-p}	no	no	KENDALL's τ (without ties) GOODMAN & KRUSKAL's γ (with ties and PRIMARY approach)
S_{w-p} (10)	yes: $\lvert d_{ij}-d_{kl}\rvert$	no	nonmetric analogue ψ to KENDALL's τ
JOHNSON's stress (8)	yes: $\lvert d_{ij}-d_{kl}\rvert$	but: $\lvert (d_{ij}-d_{kl}) / (d_{ij}+d_{kl})^2 \rvert$	

We hope that statistical inferences in nonmetric multidimensional scaling will be facilitated in the future by ψ. Minimization has to be done under the side-condition

$$\sum x^2 = 1 \quad \begin{matrix}\text{Normalisation of} \\ \text{configuration}\end{matrix} \tag{11}$$

which leaves stress invariant.

There is still another interpretation of stress. (4), (5)–(7) define the euclidean distance of the configuration from the ideal configuration in a stimulus-pair space and (8) shows similarity to an euclidean distance of an empirically derived configuration to the ideal configuration in a "complete comparison space". Our stress can be interpreted as a city-block distance in that space.

Despite the conceptual advantages of (10) it cannot be minimized with standard gradient methods. This is due to the fact that L cannot be differentiated at each point. In some areas where no further orderviolations have been found, L is constant and the gradient is zero. In some other areas where $\delta_{ij,\,kl}$ switches over from zero to one the gradient is undetermined. These are reasons for using a gradient-free hill-climbing method.

The original version of the gradient-free hill-climbing method was published by NELDER & MEAD (1965). An improved and corrected version was published by TIEDE (1973). We used the latter version but with some modifications in the starting simplex.

Starting configuration: To save computer-time we choose the approach which was recommended by LINGOES & ROSKAM (1973, p. 18). We transform the dissimilarity matrix in a matrix of ranks. Computing scalar products according to the formula:

$$b_{ij} = \bar{\varrho}_i + \bar{\varrho}_j - \left\{ \frac{(n-1)\,(1+r)}{2n} + \varrho_{ij} \right\} \qquad \begin{array}{l} r = n(n-1)/2 \\ \varrho_{ij} = \text{rank of } \hat{d}_{ij} \\ \bar{\varrho}_i = \text{mean of } i\text{-th row} \end{array}$$

and solving for the roots and vectors of B offers us the initial configuration.

Starting simplex: The function we want to minimize is: $F = F(X_1, X_2, \ldots, X_i, \ldots, X_n)$. The side condition is: $\sum x_i^2 = 1$. The argument vector is a vector in configuration space (KRUSKAL, 1964b). We choose a starting point $P_0 = (x_1, x_2, \ldots, x_i, \ldots, x_n)$ in configuration space under the side condition, which means that P_0 lies in minimization space (= surface of a hypersphere in configuration space with radius 1). P_0 consists of the components of the normalized eigenvectors of the matrix B. Later in the algorithm P_0 can be a result of a step in optimizing the starting simplex. We are now moving parallel to the first coordinate axis with a step-width $d_1 \neq 0$ in order to get the point $P'_{11} = (x_1 + d_1, x_2, \ldots, x_i, \ldots, x_n)$ and the corresponding values $F_0 = F(P_0)$ and $F'_{11} = F(P'_{11})$. If we move backwards we get $P'_{12} = (x_1 - d_1, x_2, \ldots, x_i, \ldots, x_n)$ and $F'_{12} = F(P'_{12})$. Now we choose the smaller value $F'_1 = \min(F'_{11}, F'_{12})$ and the corresponding point for P'_1. If $F'_1 > F(P_0)$ we turn over to the next dimension and move along the second coordinate axis with a P_0 being replaced by the best point on the previous path along the axis. In this case $P_0 = P_{0(\text{old})}$. The movement along the second axis produces $P'_{21} = (x_1, x_2 + d_2, \ldots, x_i, \ldots, x_n)$ etc.

Otherwise, in the case of $F'_1 \leq F(P_0)$, it seems to be promising to take a further step parallel to the first coordinate axis in a favorable direction. We seek the point $P''_1 = (x'_1 + d''_1, x_2, \ldots, x_i, \ldots, x_n)$. $d''_1 = \mathrm{SIGN}^1(d_1 + d_1, x'_1 - x_1)$. If $F''_1 \leq F'_1$. P_0 is replaced by P''_1, otherwise, P_0 is replaced by P'_1. Now we possess a better P_0. We can take this P_0 and go along axis 2. All described steps are computed for the higher dimensions ceteris paribus, with $P_{0(\mathrm{new})}$ = best point we found by moving along the previous axis. Each path along the i-th of the n coordinate axes produces a "best" point P_{0i}. At the end of our search through the axes we get a set of $n+1$ points $\{P_0, P_{01}, \ldots, P_{0i}, \ldots, P_{0n}\}$ in configuration space, which we call the *starting simplex*.

Evaluation of function: After collecting and comparing all values of the starting simplex we look for the highest (F_h), the second highest (F_s) and the lowest (F_1) values: $F_h = F(P_h)$, $F_s = F(P_s)$, $F_1 = F(P_1)$. Because P_h is the worst point of the starting simplex it would be wise to substitute it by a better one.

Reflexion: For this purpose it is necessary to calculate a weighted centroid of the reduced starting simplex, which does not contain P_h:

$$\frac{\sum_{j,\, j \neq h} \frac{G_j}{\sum_j G_j} P_j}{n} = \bar{P} \ (= \text{centroid of simplex}) \tag{12}$$

$$G_j = (F_h - F_j) + 0.7 \cdot (F_h - F_1). \tag{13}$$

Numerical experimentation has shown, that (12) is mobile, so that the danger of being trapped in local minima is reduced. The "jumpy" character and the great number of function evaluations are good aid in this respect. Points with smaller function values attract the centroid with a "greater force" than points with greater function values will do.

Now we have to substitute P_h by that point we get by a "reflexion" operation:

$$\vec{P} = (1 + \alpha) \cdot \bar{P} - \alpha P_h \quad \text{where } \alpha > 0 \ (= \text{coefficient of reflection: in this program} = 1). \tag{14}$$

If the result is $F_1 \leq \vec{F} < F_s$ we replace P_h by $\vec{P}$, but if $\vec{F} < F_1$ we are encouraged to continue our search in this direction and perform an "expansion" operation on the simplex.

---

[1] FORTRAN-IV function

Expansion: By an expansion we get the point:

$$\vec{P}\,\vec{\ } = (1+\gamma)\cdot\vec{P} - \gamma\vec{P} \quad \text{where } \gamma > 1 \text{ (= coefficient of expansion: in this program = 2).} \tag{15}$$

Then we compute $\vec{F}\,\vec{\ }$ and compare this value with the previously obtained F_h, F_s, and F_1. If $F_1 \leq \vec{F}\,\vec{\ }$ we substitute P_h by $\vec{P}$. If on the other hand $\vec{F}\,\vec{\ } < F_1$, P_h will be replaced by $\vec{P}\,\vec{\ }$ if $F(\vec{P}\,\vec{\ }) < F(\vec{P})$ and by $\vec{P}$ if $F(\vec{P}) \leq F(\vec{P}\,\vec{\ })$. In cases of $F_s \leqslant \vec{F}$ our reflexion was not very successful. We now proceed to make a contraction:

Contraction: We distinguish two cases. If $F_s < F_h < \vec{F}$ we advance directly to a contraction, otherwise, if $F_s < \vec{F} \leqslant F_h$ we substitute (before the contraction) P_h by $\vec{P}$. The contraction formula is:

$$\vec{P}\,\vec{\ } = (1-\beta)\cdot\bar{P} + \beta P_h \quad \text{where } 1 > \beta > 0 \text{ (= coefficient of contraction: in this program = 0.5)} \tag{16}$$

$\vec{P}\,\vec{\ }$ as the result of the contraction is always pushed away from the point with the highest value (either $\vec{P}$ or P_h). This result seems to be very desirable. We now compare $\vec{F}\,\vec{\ }$ with F_h. If $\vec{F}\,\vec{\ } \leqq F_h$ we replace P_h by $\vec{P}\,\vec{\ }$ and start with a new iteration. If, however, $F_h < \vec{F}\,\vec{\ }$ reflection and contraction showed no improvement and we have to shrink the whole simplex.

Shrinkage of the whole simplex: The center of shrinkage is P_1. All other points are moved toward this "best" point:

$$P_j = (P_j + P_1)/2. \tag{17}$$

Then a new iteration is started.

Program

The most important thing we have to consider in a computer program for our version of nonmetric MDS is the speed of a FUNCTION, which computes the distances and S. For this purpose all arrays are stored one-dimensionally. Index computation is made only by additions and subtractions. "Slow" operations such as multiplications and divisions or $\sqrt{\ }$ are used only in absolutely necessary instances.

Each time we calculate F (this is done by calling FUNCTION), we normalize the configuration. Thus minimization is done on a surface of a p-dimensional hypersphere (p = number of stimuli · number of stimulus-dimensions) with radius 1 in configuration space. For more details look into the flow chart (Fig. 1).

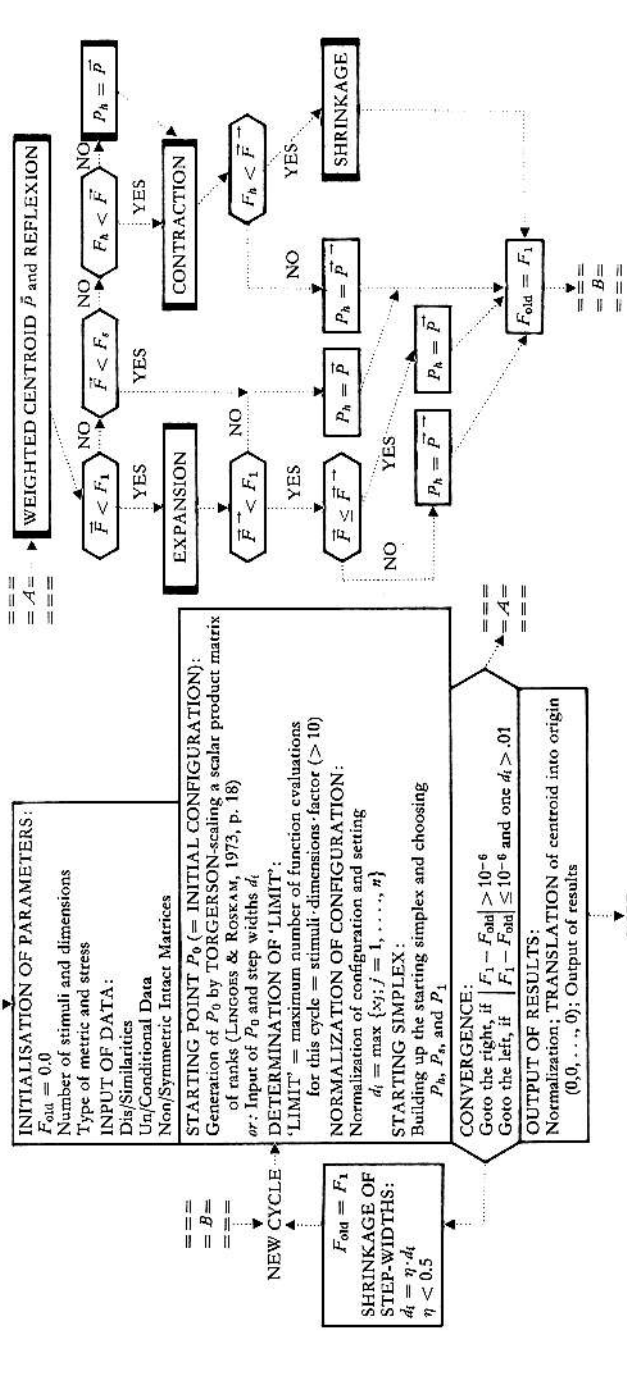


Fig. 1. Flow chart of our scaling method

Some people will hesitate in using a "search" technique. However, one ought to keep in mind that these techniques as well as gradient-methods belong to the hill-climbing algorithms. All hill-climbing methods are in fact search-techniques, some of them using gradients. It is interesting to know that the more advanced gradient techniques (e.g. the conjugate gradient method) include gradient-free *linear* searches to improve results. Nearly all numerical studies have shown that without a good initial configuration one often gets trapped in a local minimum when using a gradient-method. This is not true for the *revised* simplex. But in order to get even more safety we estimate the initial configuration (= P_0) by TORGERSON-scaling the matrix of rankscalar-products. The numerical precision of the method can be seen in Table 1 where coordinates of points coincide at least for two decimals. This is true for points 1, 3, 7, 9 (dimension 1) for 1, 3, 7, 9 (dimension 2), for 2, 8 (dimension 1), for 4, 6 (dimension 2), for 4, 6 (dimension 1), for 2, 8 (dimension 2) and for 5 (dimension 1 and 2) in the two-dimensional solution.

Computer time for stress evaluations increases very fast with a growing number of comparisons. At the present time we recommend this method only up to 20 points. It is however possible to compare the proximity matrix with a chess-board. Each "white" matrix entry is completely left out in stress evaluation thus simulating the missing-data approach. Further research is needed in evaluating the relation between "precision of scaling solution" and "number of orderrelations".

At the same time this apparent drawback of one-stage scaling is compensated by the rapid aggregation of individual data matrices. This was not possible with the same efficiency in classical nonmetric scaling.

Handling ties, missing data, nonsymmetric conditional and unconditional 'diagonal' data

Missing data and ties are handled according to JOHNSON's proposals (1973, p. 14). If $\hat{d}_{ij}$ is unknown, all comparisons which involve d_{ij} are left out. In the case of tied data the values of δ and ε are taken from the table on page 245.

In order to analyse a nonsymmetric 'diagonal' matrix in which all elements are comparable (=*intact, unconditional asymmetric* proximity matrix, GREEN & CARMONE, 1970, p. 31) we minimize (18):

$$L = S_{w-p} = \frac{\sum_{IJ<KL} \delta_{ij,\,kl} |d_{ij} - d_{kl}|}{\sum_{IJ<KL} \delta |d_{ij} - d_{kl}| + \sum_{IJ<KL} \varepsilon |d_{ij} - d_{kl}|} = \min! \qquad (18)$$

where: $IJ = (i-1)\cdot n + j$
$KL = (k-1)\cdot n + l$
n = number of points.

When we have a nonsymmetric 'diagonal' matrix whose elements are comparable only within a row (= *intact, conditional* matrix) we use (19):

$$L = S_{w-p} = \frac{\sum_{j<l} \delta_{ij,\,il} |d_{ij} - d_{il}|}{\sum_{j<l} \delta |d_{ij} - d_{il}| + \sum_{j<l} \varepsilon |d_{ij} - d_{il}|} = \min! \qquad (19)$$

Nonmetric aggregation of individual data

If there is *one* dis/similarity matrix for *each* person, it is not correct to aggregate individual information by averaging, when the data are less than intervalscaled. Nonmetric aggregation can be handled very easy (under the assumption of homogenity of individuals) by redefinition of $\delta_{ij,\,kl}$ and $\varepsilon_{ij,\,kl}$:

$$\delta_{ij,\,kl} = \frac{\sum_{n=1}^{N} \delta_{ij,\,kl,\,n}}{N} \qquad \varepsilon_{ij,\,kl} = \frac{\sum_{n=1}^{N} \varepsilon_{ij,\,kl,\,n}}{N} \qquad (20)$$

= proportion of order-isomorphy violations for comparison (ij, kl) in the whole sample

= proportion of proversions for comparison (ij, kl) in the whole sample

Local minima and 'Clustered' solutions

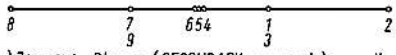
Each numeric minimization technique has to face local minimum problems. However, the danger of getting trapped into a local minimum is reduced by various features of the algorithm: choosing a quasimetric initial configuration by TORGERSON-scaling the matrix of rank-scalar products (= initial P_0 in configuration space); fast changing directions of search; switching between contractions and expansions.

In some cases we get a clustered solution. If all dis/similarities of one stimulus to the remaining stimuli are greater/smaller than the remaining *inter*-stimuli dis/similarities we get a "clustered" configuration. The corresponding point is moved far away and the rest of the stimuli is collapsed into one point.

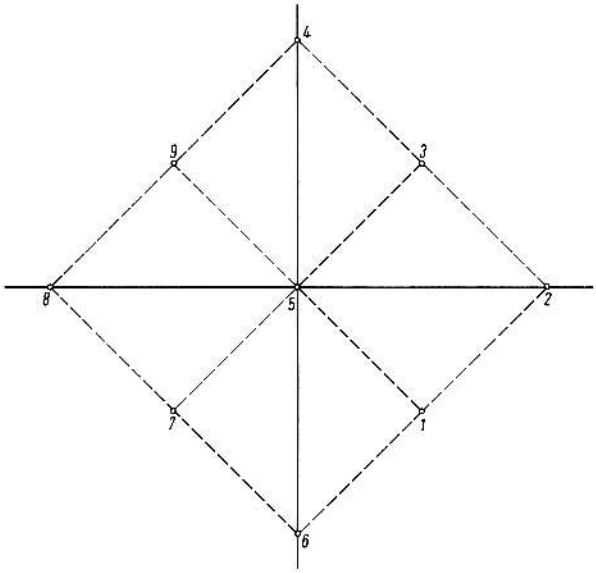
Numerical examples

We scaled MESSICK & ABELSON's (1956) famous example with (8) and (10) to demonstrate the numeric precision of the search procedure (SECONDARY approach). The "data" consisted of the interpoint distances of a unit square.

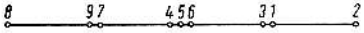
Results can be seen in Table 1 and Figure 2.



a) JOHNSON's Stress (SECONDARY approach): one dim.



b, d) two-dimensional solutions



c) Our Stress (SECONDARY approach): one dim.

Fig. 2. Scaling solutions of MESSICK's and ABELSON's "Square" Data (1956) on the basis of an intact unconditional symmetric proximity matrix

a: JOHNSON's Stress (SECONDARY approach) MINKOWSKY $R = 2$ $S = .36222$ one-dimensional solution

b: JOHNSON's Stress (SECONDARY approach) MINKOWSKY $R = 2$ $S = .00085$ two-dimensional solution

c: Our Stress (SECONDARY approach) MINKOWSKY $R = 2$ $S = .30286$ one-dimensional solution

d: Our Stress (SECONDARY approach) MINKOWSKY $R = 2$ $S = .00001$ two-dimensional solution

The analysis of nonsymmetric square proximity matrices is demonstrated with an example from COOMBS, DAWES & TVERSKY (1970, p. 73ff.). See Table 2 and Figure 3. "Clustered" solutions are shown in Table 3.

Table 1. *Scaling solutions for* MESSICK & ABELSON*'s data in euclidean space with (8) and (10)*

	JOHNSON's Stress (SECONDARY approach)						Our Stress (SECONDARY approach)					
	I	I	II	I	II	III	I	I	II	I	II	III
1	.23028	.20322	−.20409	.20672	−.19834	.01246	.30553	.20637	−.20141	.20658	−.21411	−.04522
2	.62553	.40772	−.00139	.40509	.00961	.00376	.57380	.40778	.00495	.39423	−.00282	.05840
3	.23577	.20427	.20319	.19701	.20129	−.05451	.27111	.20141	.20636	.20958	.21105	−.04534
4	.01021	.00096	.40786	−.00821	.40240	−.04372	−.03728	−.00496	.40778	.00200	.38471	.08635
5	−.00047	.00052	.00005	−.00115	−.00421	−.02560	−.00008	−.00000	−.00000	.00081	−.00009	−.09464
6	−.01179	−.00145	−.40817	.01012	−.39631	.08275	.03734	.00495	−.40778	−.00353	−.38474	.08643
7	−.23308	−.20417	−.20306	−.19843	−.20684	.01989	−.27106	−.20141	−.20636	−.20888	−.21107	−.04871
8	−.62629	−.40777	.00121	−.40351	−.00380	.03443	−.57381	−.40778	−.00495	−.39510	.00299	.05178
9	−.23016	−.20328	.20439	−.20763	.19620	−.02945	−.30555	−.20636	.20141	−.20571	.21408	−.04904
S	.36222	.00085			.00158		.30286	.00001			.00431	Stress
ϱ	.3748	.9678			.9680		.4221	.9680			.9681	rho
SS	1.0000	.50028	.49972	.49114	.48027	.02860	1.00000	.50013	.49987	.48411	.47681	.03908

ϱ = SPEARMAN's rank-correlation coefficient between interpointdistances and dissimilarities (= data)
SS = Sum of squares of coordinates on this dimension (can be used as a measure of weight for this dimension)

Table 2. *Example for an analysis of an intact, nonsymmetric matrix:* Number of references in row journal to column journal, 1964, in: COOMBS, DAWES & TVERSKY (1970, p. 73ff.)

	Number of references in row journal to column journal								Matrix of residuals (raw data corrected by subtracting out row and column means)							
		1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8
1	AJP	119	8	4	21	—	1	85	2	AJP	48.	−107.	− 42.	− 88.	− 70.	− 39.	− 68.	− 40.
2	JASP	32	510	16	11	73	9	119	4	JASP	−105.	327.	− 97.	−165.	− 63.	− 97.	−100.	−105.
3	JAP	2	8	84	1	7	8	16	10	JAP	− 55.	− 94.	50.	− 95.	− 50.	− 19.	−124.	− 19.
4	JCPP	35	8	—	533	—	1	126	1	JCPP	− 93.	−165.	−104.	365.	−128.	− 97.	− 85.	− 99.
5	JCP	6	116	11	1	255	7	12	7	JCP	− 82.	− 17.	− 53.	−226.	136.	− 51.	−159.	− 53.
6	JEdP	4	9	7	—	3	52	27	5	JEdP	− 50.	− 89.	− 23.	− 93.	− 50.	28.	−109.	− 21.
7	JExP	125	19	6	70	—	—	586	15	JExP	− 18.	−169.	−113.	−112.	−142.	−112.	360.	−100.
8	Pka	2	5	5	—	13	2	13	58	Pka	− 50.	− 92.	− 23.	− 91.	− 39.	− 20.	−122.	33.

	KRUSKAL'S 'conditional' approach		GUTTMAN-LINGOES' (split opt.)		Solutions by search procedure: JOHNSON'S				*Our* stress			
1	AJP	.125	.910	232	−215	.1570	.0406	−.0184	−.1100	.1569	−.0142	.1411	.0038
2	JASP	−1.382	−.644	−935	−613	−.0354	−.4472	−.1290	.5025	−.0864	−.4811	−.0633	−.4902
3	JAP	− .261	−.237	−139	411	−.1980	.0411	.2460	.0032	−.1290	.0436	−.1321	.0399
4	JCPP	1.302	.366	1125	298	.1343	.4581	−.1074	−.4975	.1183	.5084	.1010	.5216
5	JCP	− .924	−.002	−731	127	−.2240	−.2379	.1539	.2661	−.2331	−.2289	−.1885	−.2369
6	JEdP	− .180	.324	−109	355	−.1588	.0459	.1421	−.0084	−.1290	.0436	−.1318	.0399
7	JExP	.904	−.924	592	−791	.4666	.0418	−.4423	−.1031	.4313	.0850	.4053	.0817
8	Pka	.416	.207	− 33	427	−.1417	.0577	.1545	−.0530	−.1290	.0436	−.1318	.0401
	Stress² 1 = .108					'unconditional' $S^2 = 0.0699$		'conditional' $S^2 = 0.0601$		'unconditional' $S = 0.1081$		'conditional' $S = 0.0962$	

1 AJP = American Journal of Psychology
2 JASP = Journal of Abnormal & Social Psychology
3 JAP = Journal of Applied Psychology
4 JCPP = Journal of Comparative & Physiol. Psych.
5 JCP = Journal of Consulting Psychology
6 JEdP = Journal of Educational Psychology
7 JExP = Journal of Experimental Psychology
8 Pka = Psychometrika

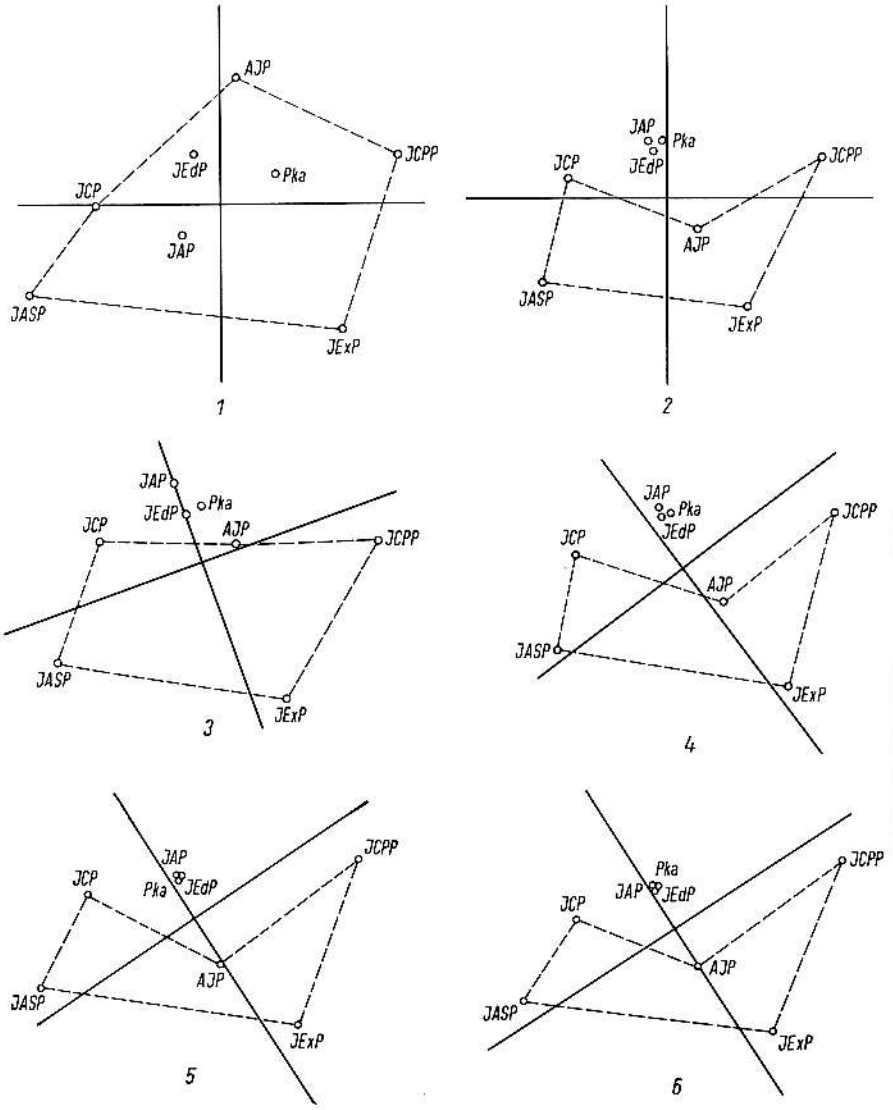


Fig. 3. Scaling solutions of COOMBS, DAWES & TVERSKY's (1970, p. 73 ff.) data (matrix of residuals, Table 2)

1: KRUSKAL's configuration (COOMBS *et al.*, 1970, p. 75) based on the 'conditional' approach

2: GUTTMAN-LINGOES' configuration (op. cit.) based on the 'conditional' approach

3: Our configuration based on JOHNSON-type stress and on the 'conditional' approach

4: Our configuration based on JOHNSON-type stress and on the 'unconditional' approach

5: Our configuration based on our (18) *stress* and on the '*unconditional*' approach

6: Our configuration based on the our (19) *stress* and on the '*conditional*' approach

Distinct order-4 matrices

In Tables 4–7 we compare our scaling results with results obtained by LINGOES & ROSKAM (1973, p. 56) who evaluated the KRUSKAL- and GUTTMAN-LINGOES-scaling approaches. LINGOES & ROSKAM explain their motivation for analysing distinct order-4 matrices (LINGOES & ROSKAM, 1973, p. 55):

"We shall be concerned ... with a fairly detailed analysis of the behavior of various algorithms on all possible order-4 distinct matrices (without ties). While admittedly of slight substantive import, such analyses could shed some further light on such issues as robustness, convergence problems, the role of the initial configuration in obtaining minimal solutions, the relationships among converged solutions based on different constraints, etc. One of the chief advantages of working with such smal matrices is that we can investigate rather exhaustively a large number of possibilities at a low cost and provide a criterion set of results that might be used by others for testing different algorithms. While some of our results would seem to be a function of small n, the main conclusions are consonant with those obtained from analyses of much larger matrices (since, without the latter, there would be little ground for generalization). It is our opinion that, if anything, these analyses of order-4 matrices highlight some of the problems that are only hinted at in the analyses of larger matrices."

LINGOES & ROSKAM divided the 30 order-4 matrices into three groups: group A with matrices which have a perfect strongly monotonic fit in one dimension and group B with 14 matrices which produce zero stress solutions (the number of different distances (n_d) is less than six which occurred by either collapsing stimuli into one point or by tying distances) and group C with matrices which possess a non-zero one dimensional fit (Table 4). We use the same partition to facilitate comparisons. Our original scaling solutions are given in Table 5a and the rounded and normed solutions in Table 5b. For the matrices in section A and B we obtained zero stress solutions which fairly agree with those from LINGOES & ROSKAM. Main differences occur only in section C. We analysed matrices 20–30 (LINGOES & ROSKAM's rounded and normed solutions, our original and our rounded and normed solutions) with JOHNSON's (8) and our stress (10) (= J-Stress and O-Stress), percentage of order-violations or inversions (DELTAP), SPEARMAN's rank correlation coefficient rho, KENDALL's tau (in the case of untied distances) or GOODMAN & KRUSKAL's γ (in the case of tied distances). The results can be seen in Table 6a–c.

K and S solutions have been already compared by LINGOES & ROSKAM. Taking S_{-p} (= DELTAP/100) and τ as *accompanying* measures of goodness of orderisomorphy a certain superiority of our algorithm (at least with these data sets) cannot be denied. It seems to be better to tie distances than leaving them inverted and facing a deteriorated orderisomorphy. Let us take matrix 30, where the greatest discrepancy between our stress solution and the S-solution can be observed, as an example. For the S-solution τ is .20, for our original solution .60 and for the rounded and normed solution .78. What

Table 3. *Examples for clustered solutions*

Completely degenerate solution[1]

Dissimilarity data

2	1.07					
3	1.42	0.34				
4	0.72	0.36	0.70			
5	1.17	0.20	0.25	0.45		
6	1.11	0.67	0.50	0.55	0.47	
7	0.75	0.33	0.67	0.03	0.42	0.56

Configuration in two-dimensional space (City – block – metric)

1	0.33158	0.79041	
2	−0.05526	−0.13174	
3	−0.05526	−0.13174	
4	−0.05526	−0.13174	$\frac{\sum\delta\lvert d_{ij}-d_{kl}\rvert}{\sum\lvert d_{ij}-d_{kl}\rvert}=.0$
5	−0.05526	−0.13174	
6	−0.05526	−0.13174	
7	−0.05526	−0.13174	

[1] The dissimilarities between stimulus 1 and the set {2–7} are greater than the interset dissimilarities: all stimuli are collapsed into two points.

Clustered solution[2]

Dissimilarity data

2	0.63				
3	0.43	0.81			
4	0.03	0.61	0.42		
5	0.55	0.10	0.92	0.52	
6	0.16	0.47	0.37	0.13	0.54

Configuration in two-dimensional space (City – block – metric)

1	−0.19221	−0.09605	
2	0.38554	−0.09608	
3	−0.19441	0.48035	
4	−0.19221	−0.09605	$\frac{\sum\delta\lvert d_{ij}-d_{kl}\rvert}{\sum\lvert d_{ij}-d_{kl}\rvert}=.01545$
5	0.38552	−0.09609	
6	−0.19223	−0.09608	

[2] The dissimilarities within the sets {1, 4, 6} and {2, 5} are smaller than the other between-set-dissimilarities: all stimuli are collapsed into three points.

Table 4. *All distinct order-4 matrices and their minimum configurations obtained by* LINGOES & ROSKAM

Group A ($K = S = 0$; $n_d = 6$)

No.	Matrix	Configurations			
1	125, 34, 6	− 42.9,	− 14.3,	−100.0,	100.0
2	125, 43, 6	− 33.3,	11.1,	−100.0,	100.0
3	125, 36, 4	− 71.4,	−100.0,	− 14.3,	100.0
4	125, 46, 2	− 55.5,	−100.0,	11.1,	100.0
5	135, 46, 2	− 66.7,	−100.0,	33.3,	100.0

Group B ($K = S = 0$; $n_d < 6$)

No.	Matrix	Configurations			
6	136, 45, 2	−100.0,	−100.0,	54.4,	100.0
7	126, 35, 4	−100.0,	−100.0,	− 55.3,	100.0
8	124, 35, 6	− 75.7,	− 75.7,	−100.0,	100.0
9	126, 45, 3	−100.0,	−100.0,	0.0,	100.0
10	123, 54, 6	0.0,	0.0,	−100.0,	100.0
11	123, 45, 6	0.0,	0.0,	−100.0,	100.0
12	134, 65, 2	−100.0,	−100.0,	100.0,	100.0
13	135, 64, 2	−100.0,	−100.0,	100.0,	100.0
14	136, 54, 2	−100.0,	−100.0,	100.0,	100.0
15	124, 36, 5	−100.0,	−100.0,	−100.0,	100.0
16	126, 34, 5	−100.0,	−100.0,	−100.0,	100.0
17	124, 53, 6	− 24.7,	24.7,	−100.0,	100.0
18	134, 56, 2	− 70.1,	−100.0,	70.1,	100.0
19	124, 56, 3	− 33.3,	−100.0,	33.3,	100.0

Group C ($K, S > 0$)

No.	Matrix	*K*-solution				*S*-solution			
20	125, 63, 4	− 42.9,	42.9,	−100.0,	100.0	−41.4,	41.4,	−100.0,	100.0
21	126, 53, 4	− 41.4,	41.4,	−100.0,	100,0	−41.4,	41.4,	−100.0,	100.0
22	126, 43, 5	− 62.8,	−4.9,	−100.0,	100.0	−62.7,	−4.9,	−100.0,	100.0
23	124, 63, 5	− 23.1,	53.8,	−100.0,	100.0	−25.6,	51.0,	−100.0,	100.0
24	124, 65, 3	− 23.8,	−100.0,	55.1,	100.0	−25.6,	−100.0,	51.0,	100.0
25	125, 64, 3	− 41.5,	−100.0,	45.1,	100.0	−54.1,	−100.0,	38.1,	100.0
26	126, 54, 3	− 52.2,	−100.0,	35.6,	100.0	−54.1,	−100.0,	38.1,	100.0
27	123, 46, 5	− 40.4,	−100.0,	−0.7,	100.0	−35.9,	−100.0,	5.9,	100.0
28	123, 56, 4	− 38.0,	−100.0,	6.3,	100.0	−35.9,	−100.0,	6.0,	100,0
29	123, 65, 4	− 25.7,	−100.0,	100.0,	22.4	−11.2,	−100.0,	100.0,	46.2
30	123, 64, 5	− 30.2,	−100.0,	100.0,	28.2	−11.2,	100.0,	−100.0,	46.3

This matrix is adapted from LINGOES & ROSKAM (1973, p. 56). Matrix entries for ranks are listed for upper-half of matrix, i.e.: $p_{12} = 1$, $p_{13} = 2$, $p_{14} = 5$, $p_{23} = 3$, $p_{24} = 4$, $p_{34} = 6$ for matrix No. 1. The scale values of the configurations are computed and rounded by: $y_{ia} = 200\ (x_{ia} - x_{\min,\ a})/d_{\max} - 100$ where: $d_{\max} = \max(x_{ia}) - \min(x_{ia})$ on dimension a. K is the coefficient of alienation in GUTTMAN & LINGOES' SSA-I-program and S is stress 1 in KRUSKAL's M-D-SCAL program (LINGOES & ROSKAM, 1973, p. vii).

Table 5a. *All distinct order-4 matrices and their minimum configurations as a result of "Nonmetric scaling without disparities and derivatives"* (original scale values: Sum of squares = 1)

No.	Original configurations							
	JOHNSON's Stress				Our Stress			
1	−.11005	−.04381	−.61685	.77071	−.21062	.05015	−.60102	.76149
2	−.10683	.09077	−.69170	.70776	−.21980	.13389	−.63851	.72442
3	−.34132	−.47635	.00741	.81027	−.34132	−.47635	.00741	.81027
4	−.26773	−.63169	.20527	.69416	−.28561	−.61367	.19035	.70893
5	−.41387	−.55390	.32015	.64762	−.41387	−.55390	.32015	.64762
6	−.48146	−.48108	.34629	.61625	−.34899	−.34897	.19404	.50392
7	−.31357	−.31331	−.22929	.85617	−.28867	−.28867	−.28867	.86602
8	−.28175	−.28224	−.28243	.84643	−.04406	−.04404	−.65834	.74644
9	−.45203	−.45116	.15029	.75289	−.45231	−.45295	.15222	.75304
10	.00042	−.00082	−.70683	.70723	.00002	−.00001	−.70703	.70702
11	−.00004	.00054	−.70711	.70661	.00011	−.00062	−.70673	.70724
12	−.49437	−.50181	.49819	.49800	−.49970	−.49991	.49949	.50012
13	−.49987	−.50010	.50019	.49978	−.49992	−.49992	.49986	.49998
14	−.49837	−.49825	.49858	.49804	−.49982	−.49993	.49990	.49985
15	−.28867	−.28867	−.28867	.86602	−.28867	−.28867	−.28867	.86602
16	−.26102	−.25929	−.26102	.78133	−.28867	−.28867	−.28867	.86602
17	−.15101	.15107	−.69082	.69076	−.16435	.16457	−.68783	.68761
18	−.43528	−.55076	.43330	.55274	−.34020	−.61984	.34018	.61987
19	−.22317	−.67081	.22331	.67066	−.22348	−.66982	.22337	.66993
20	−.00119	−.70677	.70727	.00069	−.22361	.22359	−.67081	.67083
21	−.70703	−.00047	.00031	.70718	−.24098	.14858	−.63040	.72280
22	−.42013	.10596	−.46053	.77470	−.25274	.08423	−.58972	.75823
23	−.15002	.42456	−.74821	.47367	−.22357	.22357	−.67074	.67074
24	−.15589	−.74473	.42637	.47424	−.22361	−.67081	.22361	.67081
25	−.39845	−.55848	.38701	.56993	−.49992	−.49996	.49979	.50009
26	−.44830	−.53077	−.35169	−.62739	−.45004	−.45003	.15004	.75003
27	−.07940	−.67580	.03045	.72475	−.25320	−.59178	.08435	.76063
28	−.08524	−.68311	.04482	.72353	−.22358	−.67074	.22358	.67075
29	−.12956	−.76264	.44668	.44552	−.22351	−.66830	.66913	.22267
30	−.11612	.43020	−.75827	.44419	−.00019	.00020	−.70582	.70581

Stress-values in Table 6b are based on these original configurations.

Table 5b. *All distinct order-4 matrices and their minimum configurations as a result of "Nonmetric scaling without disparities and derivatives"*

Group A (Stress = 0; $n_d = 6$)		Rounded and normed configurations							
No.	Matrix	Johnson's Stress				Our Stress			
1	125, 34, 6	− 27.0,	− 17.4,	−100.0,	100.0	− 42.7,	− 4.4,	−100.0,	100.0
2	125, 43, 6	− 16.4,	11.8,	−100.0,	100.0	− 38.6,	13.3,	−100.0,	100.0
3	125, 36, 4	− 79.0,	−100.0,	− 24.8,	100.0	− 79.0,	−100.0,	− 24.8,	100.0
4	125, 46, 3	− 45.1,	−100.0,	26.3,	100.0	− 50.4,	−100.0,	21.6,	100.0
5	135, 46, 2	− 76.7,	−100.0,	45.6,	100.0	− 76.7,	−100.0,	45.5,	100.0
Group B (Stress = 0; $n_d < 6$)									
No.	Matrix								
6	136, 45, 2	−100.0,	−100.0,	50.8,	100.0	−100.0,	−100.0,	27.3,	100.0
7	126, 35, 4	−100.0,	−100.0,	− 85.6,	100.0	−100.0,	−100.0,	−100.0,	100.0
8	124, 35, 6	−100.0,	−100.0,	−100.0,	100.0	− 12.5,	− 12.5,	−100.0,	100.0
9	126, 45, 3	−100.0,	−100.0,	− 0.0,	100.0	−100.0,	−100.0,	0.3,	100.0
10	123, 54, 6	0.0,	− 0.1,	−100.0,	100.0	0.0,	0.0,	−100.0,	100.0
11	123, 45, 6	0.0,	0.0,	−100.0,	100.0	0.0,	0.0,	−100.0,	100.0
12	134, 65, 2	−100.0,	−100.0,	100.0,	100.0	−100.0,	−100.0,	100.0,	100.0
13	135, 64, 2	−100.0,	−100.0,	100.0,	100.0	−100.0,	−100.0,	100.0,	100.0
14	136, 54, 2	−100.0,	−100.0,	100.0,	100.0	−100.0,	−100.0,	100.0,	100.0
15	124, 36, 5	−100.0,	−100.0,	−100.0,	100.0	−100.0,	−100.0,	−100.0,	100.0
16	126, 34, 5	−100.0,	−100.0,	−100.0,	100.0	−100.0,	−100.0,	−100.0,	100.0
17	124, 53, 6	− 21.9,	21.9,	−100.0,	100.0	− 23.9,	23.9,	−100.0,	100.0
18	134, 56, 2	− 79.1,	−100.0,	78.4,	100.0	− 54.9,	−100.0,	54.9,	100.0
19	124, 56, 3	− 33.3,	−100.0,	33.3,	100.0	− 33.3,	−100.0,	33.3,	100.0
Group C (Stress > 0)									
No.	Matrix								
20	125, 63, 4	− 0.2,	−100.0,	100.0,	0.1	− 33.3,	33.3,	−100.0,	100.0
21	126, 53, 4	−100.0,	− 0.1,	0.1,	100.0	− 42.4,	15.1,	−100.0,	100.0
22	126, 43, 5	− 93.5,	− 8.3,	−100.0,	100.0	− 50.0,	0.0,	−100.0,	100.0
23	124, 63, 5	− 2.1,	92.0,	−100.0,	100.0	− 33.3,	33.3,	−100.0,	100.0
24	124, 65, 3	− 3.4,	−100.0,	92.2,	100.0	− 33.3,	−100.0,	33.3,	100.0
25	125, 64, 3	− 71.6,	−100.0,	67.6,	100.0	−100.0,	−100.0,	100.0,	100.0
26	126, 54, 3	− 85.7,	−100.0,	52.4,	100.0	−100.0,	−100.0,	0.0,	100.0
27	123, 46, 5	− 14.8,	−100.0,	0.9,	100.0	− 49.9,	−100.0,	0.0,	100.0
28	123, 56, 4	− 15.0,	−100.0,	3.5,	100.0	− 33.3,	−100.0,	33.3,	100.0
29	123, 65, 4	4.7,	−100.0,	100.0,	100.0	− 33.5,	−100.0,	100.0,	33.2
30	123, 64, 5	6.8,	97.7,	−100.0,	100.0	− 0.0,	0.0,	−100.0,	100.0

n_d = number of different distances; there are 30 unique matrices of untied values, when we have 4 stimuli/points: Theorem (Lingoes & Roskam, 1973, p. 58): There are exactly $\{{n \atop 2}\}!/n!$ *inequivalent* (distinct) symmetric matrices with constant diagonal elements and one each of the integers (ranking numbers) 1, 2, 3, ... $n(n-1)/2$ in the off-diagonal cells; where A is not equivalent to B iff there is no permutation matrix P such that $PAP' = B$ and $n \geq 3$ for $a_{ij} \neq a_{ii}$ $(i \neq j)$. The original scale values where normed the way Lingoes & Roskam did (look at the bottom of Table 4).

Table 6a. LINGOES' *and* ROSKAM's *normed*

No.	Matrix	*K*-solution				OS	JS	DELTAP	Rho τ, γ	
20	125634	−42.9	42.9	−100.0	100.0	.17	.36	30.77	.62	**.38**
21	126534	−41.4	41.4	−100.0	100.0	.16	.37	30.77	.62	**.38**
22	126435	−62.8	−4.9	−100.0	100.0	.06	.17	20.00	.83	**.60**
23	124635	−23.1	53.8	−100.0	100.0	.10	.22	20.00	.77	**.60**
24	124653	−23.8	−100.0	55.1	100.0	.10	.21	20.00	.77	**.60**
25	125643	−41.5	−100.0	45.1	100.0	.14	.35	26.67	.66	**.47**
26	126543	−52.2	−100.0	35.6	100.0	.13	.35	20.00	.77	**.60**
27	123465	−40.4	−100.0	−0.7	100.0	.10	.18	20.00	.77	**.60**
28	123564	−38.0	−100.0	6.3	100.0	.09	.17	20.00	.77	**.60**
29	123654	−25.7	−100.0	100.0	22.4	.16	.22	26.67	.54	**.47**
30	123645	−30.2	−100.0	100.0	28.2	.21	.28	33.33	.49	**.33**

OS = Our Stress
JS = JOHNSON's Stress

Table 6b. *Configurations as a*

No.	Matrix	*J*-solution				OS	JS	DELTAP	Rho τ, γ	
20	125634	−0.0012	−0.7068	0.7073	0.0007	.40	.27	46.67	.20	.07
21	126534	−0.7070	−0.0005	0.0003	0.7072	.40	.27	40.00	.26	.20
22	126435	−0.4201	0.1060	−0.4605	0.7747	.08	.09	20.00	.83	.60
23	124635	−0.1500	0.4246	−0.7482	0.4737	.15	.15	20.00	.77	.60
24	124653	−0.6559	−0.7447	0.4264	0.4742	.12	.31	20.00	.77	.60
25	125643	−0.3984	−0.5585	0.3870	0.5699	.13	.28	26.67	.71	.47
26	126543	−0.4483	−0.5308	0.3517	0.6274	.11	.27	20.00	.77	.60
27	123465	−0.0794	−0.6758	0.0304	0.7247	.10	.11	26.67	.71	.47
28	123564	−0.0852	−0.6831	0.0448	0.7235	.10	.11	20.00	.77	.60
29	123654	−0.1296	−0.7626	0.4467	0.4455	.24	.18	40.00	.43	.20
30	123645	−0.1161	0.4302	−0.7583	0.4442	.23	.18	33.33	.54	.33

J-solution = Solution on the Basis of JOHNSON's Stress (8)
O-solution = Solution on the Basis of Our Stress (10)

Table 6c. *Normed and rounded configurations*

No.	Matrix	*J*-solution				OS	JS	DELTAP	Rho τ, γ	
20	125634	−0.2	−100.0	100.0	0.1	.40	.27	46.67	.20	**.07**
21	126534	−100.0	−0.1	0.1	100.0	.40	.27	46.15	.15	**.08**
22	126435	−93.5	−8.3	−100.0	100.0	.08	.09	20.00	.83	**.60**
23	124635	−2.1	92.0	−100.0	100.0	.15	.15	20.00	.77	**.60**
24	124653	−3.4	−100.0	92.2	100.0	.15	.15	26.67	.71	**.47**
25	125643	−71.6	−100.0	67.6	100.0	.13	.28	26.67	.71	**.47**
26	126543	−85.7	−100.0	52.4	100.0	.11	.27	20.00	.77	**.60**
27	123465	−14.8	−100.0	0.9	100.0	.10	.11	26.67	.71	**.47**
28	123564	−15.0	−100.0	3.5	100.0	.10	.11	20.00	.77	**.60**
29	123654	4.7	−100.0	100.0	100.0	.24	.18	38.46	.44	**.23**
30	123645	6.8	97.7	−100.0	100.0	.23	.18	33.33	.54	**.33**

and rounded configurations

S-solution				OS	JS	DELTAP	Rho	τ, γ
−41.4	41.4	−100.0	100.0	.16	.37	30.77	.62	**.38**
−41.4	41.4	−100.0	100.0	.16	.37	30.77	.62	**.38**
−62.7	−4.9	−100.0	100.0	.06	.17	20.00	.83	**.60**
−25.6	51.0	−100.0	100.0	.10	.23	26.67	.71	**.47**
−25.6	−100.0	51.0	100.0	.10	.23	20.00	.77	**.60**
−54.1	−100.0	38.1	100.0	.14	.35	26.67	.71	**.47**
−54.1	−100.0	38.1	100.0	.13	.34	20.00	.77	**.60**
−35.9	−100.0	5.9	100.0	.10	.16	26.67	.71	**.47**
−35.9	−100.0	6.0	100.0	.09	.16	20.00	.77	**.60**
−11.2	−100.0	100.0	46.2	.18	.19	33.33	.49	**.33**
−11.2	100.0	−100.0	46.2	.20	.20	40.00	.43	**.20**

result of our algorithm

Our-solution				OS	JS	DELTAP	Rho	τ, γ
−0.2236	0.2236	−0.6708	0.6708	.14	.41	20.00	.77	.60
−0.2410	0.1486	−0.6304	0.7228	.14	.43	20.00	.77	.60
−0.2527	0.0842	−0.5897	0.7582	.05	.22	13.33	.89	.73
−0.2236	0.2236	−0.6707	0.6707	.07	.29	7.69	.88	.85
−0.2021	−0.6064	0.2021	0.6064	.07	.29	15.38	.79	.69
−0.4999	−0.5000	0.4998	0.5001	.13	.35	26.67	.71	.47
−0.4500	−0.4500	0.1500	0.7500	.08	.32	15.38	.77	.69
−0.2532	−0.5918	0.0843	0.7606	.10	.22	20.00	.77	.60
−0.2236	−0.6707	0.2236	0.6707	.07	.17	20.00	.77	.60
0.2235	0.6683	−0.6691	−0.2227	.14	.25	20.00	.66	.60
−0.0002	0.0002	−0.7058	0.7058	.10	.40	20.00	.77	.60

as a result of our algorithm

Our-solution				OS	JS	DELTAP	Rho	τ, γ
− 33.3	33.3	−100.0	100.0	**.14**	.41	15.38	.79	**.69**
− 42.4	15.1	−100.0	100.0	**.14**	.43	13.33	.83	**.73**
− 50.0	0.0	−100.0	100.0	**.05**	.22	7.69	.91	**.85**
− 33.3	33.3	−100.0	100.0	**.07**	.29	7.69	.88	**.85**
− 33.3	−100.0	33.3	100.0	**.07**	.29	15.38	.79	**.69**
−100.0	−100.0	100.0	100.0	**.13**	.35	12.50	.62	**.75**
−100.0	−100.0	0.0	100.0	**.08**	.32	9.09	.77	**.82**
− 49.9	−100.0	0.0	100.0	**.10**	.22	21.43	.75	**.57**
− 33.3	−100.0	33.3	100.0	**.07**	.17	15.38	.77	**.69**
− 33.5	−100.0	100.0	33.2	**.14**	.25	20.00	.66	**.60**
0.0	0.0	−100.0	100.0	**.10**	.40	11.11	.68	**.78**

Table 7. *Comparison of S-solution and our stress solution for matrix 30*

S-solution					*Our stress solution* (rounded and normed)		
Pair of points	Rank of dissimilarity	Disparity	d_{ij}	Rank of d_{ij}	d_{ij}	Rank of d_{ij}	Pair of points
3–2	6	200.0	200.0	6	100.0	3.5	3–2
4–3	5	146.3	146.3	5	200.0	6	4–3
4–2	4	77.8	53.7	1	100.0	3.5	4–2
4–1	3	77.8	57.5	2	100.0	3.5	4–1
3–1	2	77.8	88.8	3	100.0	3.5	3–1
2–1	1	77.8	111.2	4	0.0	1	2–1

d_{ij}	*in S-solution* shows inversions with d_{kl}	proversions with d_{kl}	d_{ij}	*in our stress solution* shows inversions with d_{kl}	proversions with d_{kl}
d_{21}	d_{31}	d_{43}	d_{21}		d_{31}
	d_{41}	d_{32}			d_{41}
	d_{42}				d_{42}
					d_{43}
					d_{32}
d_{31}	d_{41}	d_{43}	d_{31}		d_{43}
	d_{42}	d_{32}			
d_{41}	d_{42}	d_{43}	d_{41}		d_{43}
		d_{32}			
d_{42}		d_{43}	d_{42}		d_{43}
		d_{32}			
d_{43}		d_{32}	d_{43}	d_{32}	

$$S_{-p} = \frac{6}{9+6} = 0.40 = \text{DELTAP/100} \qquad S_{-p} = \frac{1}{8+1} = 0.11 = \text{DELTAP/100}$$

$$\gamma = 0.20 \qquad \gamma = 0.78$$

happened? This can be seen in Table 7. After selecting matrix 30 we compared the *S*-solution (last line in Table 6a) with our solution (last line in Table 6c). The results in the lower half in Table 7 show that the number of inversions in our solution is lower (11%) than in the *S*-solution (40%). This is a result of tying distances. If the rankorder of the dissimilarities is so "complicated" that one can not find a rankpreserving isomorphic mapping, it seems to be better to tie distances than to leave them wrongly inverted. This means that the algorithm tends to reduce the number of "group C" solutions and to increase the number of "group B" solutions. The postulation that a good nonmetric

algorithm should avoid tied distances has to be modified: A good algorithm should not tie distances as long as it does not create new inversions. It seems not to make sense in multidimensional scaling to demand different distances at the cost of lower order isomorphy.

Relations between one-stage and two-stage stress

It is possible to transform our stress (10) into a classical two-stage measure (1, 2) similar to KRUSKAL's stress (5).

For a two-stage stress we need a set of disparities, which can be interpreted as *ideal distances* with the same rankorder as the dissimilarities. It is interesting to see which parts of our stress can be looked at "disparities" or ideal orderpreserving distances. If we reorder the numerator of (10), we get a "two-stage" stress with disparites, as follows:

$$S_{w-p} = \frac{\sum_{ij} \{d_{ij} - d_{ij} + \sum_{kl} \delta_{ij,kl} \, |d_{ij} - d_{kl}|}{\text{denominator } D} \qquad (21) = (10)$$

$$= \frac{\sum_{ij} \{d_{ij} - \tilde{d}_{ij}\}}{D}$$

where: $\tilde{d}_{ij} = d_{ij} - \sum_{kl} \delta_{ij,kl} \, |d_{ij} - d_{kl}|$ is the "disparity".

The disparity (ideal distance) is equal to the distance minus that amount ($\sum \delta |d_{ij} - d_{kl}|$) which destroys the rankisomorphy with the data. Because the disparity is never greater than the distance there is no need for squaring error terms in the numerator to get positive values (L_1-approximation).

References

AHRENS, H. J., Zur Verwendung des Metrikparameters multidimensionaler Skalierungen bei der Analyse von Wahrnehmungsstrukturen. Z. exp. angew. Psychol. 1972, *19*, 173–195.

AHRENS, H. J., Zur Bedeutung der Metrik in multidimensionalen Ähnlichkeitsskalierungen. Bericht über den 27. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1970. Göttingen: Hogrefe, 1973, 221–228.

BRENT, R. P., Algorithms for minimization without derivatives. Englewood Cliffs, N. J.: Prentice-Hall, Inc. 1973.

COOMBS, C. H., DAWES, R. M. & TVERSKY, A., Mathematical Psychology: An Elementary Introduction. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1970.

CROSS, D. V., Metric Properties of Multidimensional Stimulus Generalization, in: MOSTOFSKY, D. I. (ed), Stimulus Generalization. Stanford: University Press, 1965, 72–93.

DAVIDON, W. C., Variance Algorithm for Minimization. Comp. J. 1968, *10*, 406–410.

FLETCHER, R. & POWELL, M. J. D., A Rapidly Convergent descent method for minimization. Comp. J. 1963, *6*, 163–168.

GOODMAN, L. A. & KRUSKAL, W. H., Measures of Association for Cross-classifications. J. Am. Stat. Ass., 1954, *49*, 732–764.

GOODMAN, L. A. & KRUSKAL, W. H., Measures of Association for Cross-classification: II Further Discussion and References. J. Am. Stat. Ass. 1959, *54*, 123–163.

GREEN, P. E. & CARMONE, F. J., Multidimensional Scaling and Related Techniques in marketing analysis. Boston, Mass: Allyn & Bacon, 1970.

GUTTMAN, L., A General Nonmetric Technique for Finding the Smallest coordinate space for a Configuration of Points. Psychometrika 1968, *33*, 469–506.

JOHNSON, R. M., Pairwise Nonmetric Multidimensional Scaling. Psychometrika 1973, *38*, 11–18.

KENDALL, M. G. & STUART, A., The Advanced Theory of Statistics, Inference and Relationship (vol. II), London: Griffin & Co. Ltd., 1967².

KRUSKAL, J. B., Nonmetric Multidimensional Scaling: A Numeric Method. Psychometrika 1964a, *29*, 115–129.

KRUSKAL, J. B., Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. Psychometrika 1964b, *29*, 1–27.

LINGOES, J. C. & ROSKAM, E. E., A Mathematical and Empirical Analysis of Two Multidimensional Scaling Algorithms. Psychometrika Monograph Supplement 19, 1973.

MESSICK, S. J. & ABELSON, R. P., The additive constant problem in Multidimensional Scaling. Psychometrika 1956, *21*, 1–17.

MÖBUS, C., Ein Beitrag zur metrischen und nonmetrischen Analyse von nichtsymmetrischen Proximitätsmatrizen mit multidimensionalen Skalierungsmodellen, Heidelberg: Wirtschafts- und Sozialwissenschaftliche Fakultät der Universität (Inaug. Diss.), 1974.

NELDER, J. A. & MEAD, R., A Simplex Method for Function Minimization. Comp. J. 1965, *7*, 308–313.

ROSKAM, E. A., A Comparison of Principles for Algorithm Construction in Nonmetric Scaling. Michigan Mathematical Psychology Program Technical Report 1969, *2*, 1–58.

SHEPARD, R. N., ROMNEY, A. K. & NERLOVE, S. B. (eds.), Multidimensional Scaling: Theory and Applications in the Behavioral Sciences. Vol. I. New York: Seminar Press 1972.

SPÄTH, H., Algorithmen für elementare Ausgleichsmodelle. München: Oldenbourg, 1973.

SPÄTH, H., Algorithmen für multivariable Ausgleichsmodelle. München: Oldenbourg, 1974.

SILITTO, G. P., The Distribution of KENDALL's τ coefficient in Rankings Containing Ties. Biometrika 1947, *34*, 36–44.

TIEDE, M., Theorie und Praxis eines neuen Suchverfahrens, Statistische Hefte 1973, *14*, 39–58.

WENDER, K., Die psychologische Interpretation nichteuklidischer Metriken in der multidimensionalen Skalierung. Darmstadt: Philosophische Fakultät der Technischen Hochschule 1969 (Dissertation).

YOUNG, F. W., Polynomial Conjoint Analysis Algorithms, in: SHEPARD, R. N., ROMNEY, A. K. & NERLOVE, S. B., Multidimensional Scaling, New York: Seminar Press, Vol. I, 1972, 83–103.

Anschrift des Verfassers:

Dr. C. MÖBUS, Psychologisches Institut der Universität Heidelberg
Hauptstraße 47–51, 6900 Heidelberg