METRIC MULTIDIMENSIONAL UNFOLDING

WILLEM HEISER
JAN DE LEEUW

Department of Datatheory
University of Leiden
Wassenaarseweg 80
Leiden
The Netherlands

ABSTRACT

A solution of the metric multidimensional unfolding
problem is stated as a special case of the general
multidimensional scaling method of de Leeuw and
Heiser (1977), which guarantees convergence to a local
minimum of stress. Because the number of local minima
is usually very large, considerable attention is paid
to the algebraic solution of the unfolding problem
which we use as an initial configuration. Here we dis-
tinguish three approaches: no centering, single cen-
tering and double centering, which differ in the way
they treat the nonlinearity of the problem. The con-
tributions of Ross and Cliff (1964) and Schönemann
(1970) are discussed within this framework. The various
approaches are evaluated in terms of the stress they
produce in our iterative program, but the current state
of affairs does not permit any definite conclusions.

1.0 Introduction

In multidimensional unfolding problems we consider a nonnegative datamatrix Δ of order $n \times m$, whose elements are interpreted as measures of dissimilarity between the n (row)objects $\underline{R}=\{r_1, r_2, \ldots, r_n\}$ and m (column)objects $\underline{C}=\{c_1, c_2, \ldots, c_m\}$. Thus δ_{ij} is the dissimilarity between objects r_i and c_j. In a psychological context the row objects are often called subjects, the column objects stimuli, and the dissimilarities are derived from preference judgments.

Multidimensional unfolding techniques represent both row and column objects as points $\underline{X}=\{x_1, x_2, \ldots, x_n\}$ and $\underline{Y}=\{y_1, y_2, \ldots, y_m\}$ in a metric space $<\Omega, d>$ in such a way that the distances $d(x_i, y_j)$ are approximately equal to the dissimilarities δ_{ij}. We sometimes write d_{ij} or $d_{ij}(X,Y)$ for $d(x_i, y_j)$. In this paper we study representations of $r_1, r_2, \ldots, r_n$ and $c_1, c_2, \ldots, c_m$ in the space of all p-tuples of real numbers, in which the metric is defined by the euclidean norm. Thus a representation of $r_1, r_2, \ldots r_n$ is the $n \times p$ configuration matrix X, with elements x_{ia}, a representation of $c_1, c_2, \ldots, c_m$ is the $m \times p$ configuration matrix Y, with elements y_{ja} and the $d_{ij}(X,Y)$ are euclidean distances, defined on the rows of X and Y by

$$d_{ij}(X,Y) = \left[\sum_{a=1}^{p} (x_{ia} - y_{ja})^2 \right]^{\frac{1}{2}}. \tag{1}$$

It is convenient to rephrase the multidimensional unfolding problem as a special kind of multidimensional scaling problem, where in general we represent N objects $\underline{O}=\{o_1, o_2, \ldots, o_N\}$ as N points $\underline{Z}=\{z_1, z_2, \ldots, z_N\}$ in metric space such that the interpoint distances $d_{ik}(Z)$ are approximately equal to the interobject dissimilarities, collected in the matrix $\Gamma=\{\gamma_{ik}\}$. To evaluate the badness-of-fit of a particular configuration Z, we use the loss function

$$\sigma_0(Z) = \sum_{i=1}^{N} \sum_{k=1}^{N} w_{ik}(\gamma_{ik} - d_{ik}(Z))^2, \tag{2}$$

where $W=\{w_{ik}\}$ is a nonnegative square matrix of given weights and the summation is over $i<k$. We can base our unfolding technique on a general multidimensional scaling algorithm that minimizes (2) by considering:

$$\underline{O} = \{ \underline{R} \cup \underline{C} \}, \tag{3a}$$

$$\underline{Z} = \{ \underline{X} \cup \underline{Y} \}, \tag{3b}$$

$$\{ \underline{R} \cap \underline{C} \} = \phi , \tag{3c}$$

$$\{ \underline{X} \cap \underline{Y} \} = \phi , \tag{3d}$$

and partitioning the matrices Z, W and Γ as follows:

$$Z = \begin{bmatrix} X \\ \cdots \\ Y \end{bmatrix} , \tag{4a}$$

$$W = \begin{bmatrix} \phi & : & U \\ \cdots & : & \cdots \\ U' & : & \phi \end{bmatrix} , \tag{4b}$$

$$\Gamma = \begin{bmatrix} \phi & : & \Delta \\ \cdots & : & \cdots \\ \Delta' & : & \phi \end{bmatrix} . \tag{4c}$$

Here the partitioning of W parallels that of Γ and N=n+m. Now the loss function (2) transforms into

$$\sigma_0(Z) = \sum_{i=1}^{n} \sum_{k=n+1}^{n+m} w_{ik}(\gamma_{ik} - d_{ik}(z))^2 , \tag{5}$$

and our unfolding technique will minimize

$$\sigma_1(X,Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ij}(\delta_{ij} - d_{ij}(X,Y))^2 . \tag{6}$$

In principle, the generality of an unfolding technique is determined by the generality of the multidimensional scaling technique on which it is based. So, given an appropriate multidimensional scaling algorithm, we can do noneuclidian unfolding by generalizing definition (1) to general Minkovski metrics. Also, if only the rank order of the dissimilarities is given, we may use a nonmetric multidimensional scaling algorithm to do nonmetric unfolding. In some applications, where we cannot assume comparability of intersubjective utilities, we need a row-conditional (non)metric algorithm to do the job. Apart from these generalizations, we may handle missing data by setting $u_{ij}=0$ for all pairs i,j for which no observation is available.

In this paper we only treat the simplest case: metric euclidean unweighted unfolding. There are at least two reasons for this. In the first place, the impressive succes of the nonmetric approach in multidimensional scaling problems depends critically on the great number of ordinal restrictions that are imposed

on the distances; for each of the $\frac{1}{2}N(N-1)$ dissimilarities, we have in principle $\frac{1}{2}N(N-1)$ order restrictions of the form $\gamma_{ik} > \gamma_{lm}$. In an unfolding problem with the same number of parameters (say $(n+m)p$), the number of restrictions imposed by the data $(\frac{1}{2}(n \times m)(n \times m-1))$ decreases rapidly and moreover depends on the ratio n/m; if only row-conditional comparisons are allowed, this again reduces the number of restrictions to $n \times \frac{1}{2}m(m-1)$. This effect is illustrated in table 1.

	MDS	UNF 1.00	UNF 1.50	UNF 2.33	UNF 4.00	ROW 1.00	ROW 1.50	ROW 2.33	ROW 4.00
10	990	300	276	210	120	50	36	21	8
20	17955	4950	4560	3486	2016	450	336	210	96
N 30	94395	25200	23220	17766	10296	1575	1188	756	360
40	307810	79800	73536	56280	32640	3800	2880	1848	896
50	749700	195000	179700	137550	79800	7500	5700	3675	1800

Table 1. Number of restrictions imposed by ordinal data. For the unfolding cases (UNF) and the row-conditional unfolding cases (ROW), N=n+m and 4 different ratio's n/m are tabulated.

So, for the not quite uncommon row-conditional 40×10 unfolding problem, the number of restrictions (1800) is more than 400 times less than that number for the equal parameter multidimensional scaling case (749700). If ties in the data are treated by the so called primary approach (if $\delta_{ij} = \delta_{kl}$, then d_{ij} need not to be equal to d_{kl}), or if missing data are present, the situation deteriorates even more. As a result of all this, the configuration of points is less well "tightened" by the data compared with the complete multidimensional scaling case and problems arise in the form of "degenerate" solutions, "spurious" dimensions and local minima.

The second reason for restricting ourselves to the simplest case is that, even in the metric approach, the local minimum problem is very serious; i.e., if we do not start an iterative technique in the neighbourhood of the global minimum, we are almost sure that we get caught in a local one. Therefore, we will concentrate in section 2 on the algebraic analysis of the unfolding problem, based on the properties of the squared euclidean distances. We then use in section 3 the configurations $\hat{X}, \hat{Y}$ which are solutions of the algebraic problem as a start for an iterative technique which minimizes (6).

1.1 Other work

The one-dimensional unfolding model was proposed by Coombs (1950,1964) as a consequence of his theory of preferential choice; in his terminology, the one-dimensional configuration Y is called the J-scale and each row of Δ an I-scale, which may be thought of as the J-scale folded at the ideal point x_i with only the rank order of the stimuli given in order of increasing distance from the ideal point. The extension to multidimensional unfolding was made possible by the work of Bennett and Hays (1960,1961; also see Coombs 1964). These older approaches are non-metric, not only in the sense that rankorders within rows of Δ are used, but they also end up with a representation consisting of p partial orders of projections on the axes of euclidean p-space.

The trick of specializing an iterative multidimensional scaling program such that it becomes an unfolding technique was developed by several people in the late sixties (cf. Kruskal and Carroll, 1969), but the problem of finding a good initial configuration for this special purpose seems to be a bit neglected. The algebraic approach goes back to Coombs and Kao (1960), who conjectured a connection with principal component analysis, Ross and Cliff (1964), who proved some theorems about this conjecture and Schönemann (1970), who proposed a method for recovering X and Y from the double centered squared distances (cf. Gold 1973).

2.0 The algebraic approach

In this section we study the algebraic properties of the unfolding problem. We will distinguish three cases: no centering, single centering and double centering. These cases differ in the way they use information from the data. We will also present three algorithms, the last of which is very closely related to the one proposed by Schönemann (1970).

2.1 No centering

We suppose that the n×m matrix $D^{(2)} = \{d_{ij}^2(X,Y)\}$ is given. According to the euclidean assumption,

$$d_{ij}^2(X,Y) = \sum_{a=1}^{p} (x_{ia} - y_{ja})^2$$

$$= \sum_{a=1}^{p} x_{ia}^2 + \sum_{a=1}^{p} y_{ja}^2 - 2 \sum_{a=1}^{p} x_{ia} y_{ja}. \qquad (8)$$

The problem is, how to reconstruct X and Y from $D^{(2)}$. Ross and Cliff (1964) al-

ready observed, that (8) can be rewritten in matrix notation as

$$D^{(2)} = \begin{bmatrix} & \vdots & & \vdots & \\ e & \vdots & \alpha & \vdots & X \\ & \vdots & & \vdots & \end{bmatrix} \begin{bmatrix} \cdots & \beta' & \cdots \\ \cdots & e' & \cdots \\ \cdots & Z' & \cdots \end{bmatrix} \tag{9}$$

with e a vector with all elements equal to unity (the number of elements of e follows from the context), with $Z=-2Y$, with a vector α defined by

$$\alpha_i = \sum_{a=1}^{p} x_{ia}^2 , \tag{10a}$$

and a vector β defined by

$$\beta_j = \sum_{a=1}^{p} y_{ja}^2 . \tag{10b}$$

An immediate consequence is the following theorem (also due to Ross and Cliff, and conjectured in a somewhat imprecise form by Coombs and Kao 1960):

Theorem 1. A necessary condition for the solvability of (8) is $\text{rank}(D^{(2)}) \leqslant p + 2$.

The necessary condition in theorem 1 is interesting, but hopelessly inadequate for most practical purposes. It does not use any of the special properties of (9), only the number of rows and columns in the decomposition. The condition is neces-sary and sufficient for $D^{(2)}$ to be representable as a vector or inner product model in p+2 dimensions. Thus if the unfolding model is true, applying principal component analysis is uneconomical, in the sense that it gives two extra "arti-ficial" dimensions.

We now suppose $n \geqslant m \geqslant p+2$ and $\text{rank}(D^{(2)})=p+2$. Suppose

$$D^{(2)} = GH' \tag{11}$$

is a full-rank decomposition of $D^{(2)}$. The whole problem then amounts to finding the direction of the two "artificial" dimensions in the column space of G. It is well known that decomposition (11) is unique up to a nonsingular transformation; i.e., we can always find

$$G^* = GU , \tag{12a}$$
$$H^* = HU^{-1} , \tag{12b}$$
$$D^{(2)} = G^* H^{*'} , \tag{12c}$$

where U is an arbitrary nonsingular $(p+2) \times (p+2)$ matrix. It follows that what we are looking for is a nonsingular T such that

$$GT = \left[\begin{array}{c:c:c} e & \alpha & X \end{array} \right] \tag{13a}$$

and

$$HS = \left[\begin{array}{c:c:c} \beta & e & Z \end{array} \right] \tag{13b}$$

with

$$S = (T^{-1})' \tag{13c}$$

We proceed the analysis asymmetrically, because we will try to reconstruct T from (13a) and then define a solution for S to be the inverse of the solution for T. We partition T as follows:

$$T = \left[\begin{array}{c:c:c} t_1 & t_2 & T_3 \end{array} \right] \tag{14}$$

with t_1 and t_2 vectors of length p+2 and T_3 $(p+2) \times p$. Rewriting (13a), we get

$$Gt_1 = e \, , \tag{15a}$$
$$Gt_2 = \alpha \, , \tag{15b}$$
$$GT_3 = X \, , \tag{15c}$$

and from definition (10a) it follows that

$$\alpha = diag(XX') \, , \tag{16}$$

where diag(A) denotes the vector of diagonal elements of A. Combining (15b) and (15c) we must have

$$Gt_2 = diag(GT_3 T_3' G') . \tag{17}$$

We can summarize the development as follows:

<u>Theorem 2</u>. Suppose $rank(D^{(2)})=p+2$. Then (8) is solvable for X and Y if and only

if (15a) and (17) are solvable for t_1, t_2, T_3.

It is clear that (15a) determines t_1 uniquely:

$$t_1 = G^+ e , \tag{18}$$

where G^+ is the left inverse of G. We now analyse (17) in detail. We can rewrite it as

$$G t_2 = \text{diag}(GMG') \tag{19}$$

$$M = T_3 T_3' \tag{20}$$

Thus (20) requires that M is symmetric, positive semidefinite, with $\text{rank}(M) \leqslant p$. Equation (19) is a linear homogeneous system in M and t_2. We investigate its solution space. First define the $n \times \frac{1}{2}(p+2)(p+3)$ "columnwise direct product" (G×G) by:

$$(G \times G)_{i(kl)} = 2 g_{ik} g_{il} \qquad (k>l) \tag{21a}$$

$$(G \times G)_{i(kl)} = g_{ik}^2 \qquad (k=l) \tag{21b}$$

Here we write (kl) for $1 + \frac{1}{2}k(k-1)$ with $k \geqslant l$, which makes (kl) run from 1 to $\frac{1}{2}(p+2)(p+3)$. We also define the $\frac{1}{2}(p+2)(p+3)$-element vector m by

$$m_{(kl)} = (M)_{kl} . \tag{21c}$$

So m is M, strung out as a vector. In definitions (21) we have made use of the symmetry of M to avoid unnecessary linear dependencies in the columns of (G×G). We may now rewrite (19) as

$$G t_2 = (G \times G) m . \tag{22}$$

So one of our "artificial" dimensions is contained in the column space of (G×G). Moreover, it turns out that all columns of G are contained in (G×G). To show this, we define the $\frac{1}{2}(p+2)(p+3) \times (p+2)$ matrix Ψ by

$$\Psi_{(kl)u} = \frac{1}{2}(\tau_k \delta^{lu} + \tau_l \delta^{ku}) , \tag{23}$$

where τ_k is the k'th element of the solution of (18) and δ^{lu} is Kronecker's δ.

Then

$$\{(G\times G)\Psi\}_{iu} = \sum_{k>1}\sum_{l} g_{ik}g_{il}(\tau_k\delta^{lu} + \tau_l\delta^{ku}) + \sum_{k=1}\sum_{l} g_{ik}g_{il}\cdot\tfrac{1}{2}(\tau_k\delta^{lu} + \tau_l\delta^{ku})$$

$$= \tfrac{1}{2}\sum_{k}\sum_{l} g_{ik}g_{il}\tau_k\delta^{lu} + \tfrac{1}{2}\sum_{k}\sum_{l} g_{ik}g_{il}\tau_l\delta^{ku}$$

$$= \tfrac{1}{2}\sum_{l} g_{il}\delta^{lu} + \tfrac{1}{2}\sum_{k} g_{ik}\delta^{ku} = \tfrac{1}{2}g_{iu} + \tfrac{1}{2}g_{iu} = g_{iu} \tag{24}$$

where we have used (15a). Thus

$$(G\times G)\Psi = G \tag{25}$$

and (22) transforms into

$$(G\times G)\Psi t_2 = (G\times G)m \ . \tag{26}$$

Now suppose E contains a basis for the null space of $(G\times G)$, i.e. the subspace $\{x\,|\,(G\times G)x=0\}$. That the rank of $(G\times G)$ is at least one less than its number of columns can be easily demonstrated. Suppose p=1; we write x for X and x^2 for α. Then if a transformation exist such that (13a) is true, there must be an accompanying transformation of the columns of $(G\times G)$ into

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ e & 2x^2 & x^4 & 2x & 2x^3 & x^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{27}$$

where the powers are understood to be columnwise direct products again. Clearly (27) contains one linear dependency. Unfortunately however, we have not been able to derive any general results on the dimensionality of E. We therefore take it to be, say, $q<\tfrac{1}{2}(p+2)(p+3)$ and all solutions of (22) are given by

$$m = \Psi t_2 + E\theta \tag{28}$$

with both t_2 and θ (a vector of length q) completely arbitrary. From m of the form (28) we can recover t_2 by

$$t_2 = G^+(G\times G)m \ . \tag{29}$$

We now need some identification constraints. Suppose we have chosen our full

rank decomposition (11) such that $G'G=I$. We may require without loss of generality that the row points are centered which is true i.f.f.

$$e'GT_3 = 0 \tag{30a}$$

i.f.f.

$$t_1'T_3 = 0 \tag{30b}$$

i.f.f

$$t_1'Mt_1' = 0 \tag{30c}$$

Here we have used (15c),(18),(20) and the positive semi-definiteness of M. Now define the $(p+2)\times(p+2)$ matrix P as

$$P = (I - \frac{t_1 t_1'}{t_1' t_1}) \tag{31}$$

Then the row points are centered i.f.f.

$$PMP = M . \tag{32}$$

To use this condition, we rewrite (28) in matrix form:

$$M = t_1 t_2' + t_2 t_1' + \sum_v^q \theta_v \varepsilon_v , \tag{33}$$

where the ε_v correspond with the vectors in the null space basis. Substituting (32) in (33) we get:

$$M = \sum_v^q \theta_v P \varepsilon_v P . \tag{34}$$

We summarize the developments in a new theorem:

Theorem 3. Suppose $\text{rank}(D^{(2)})=p+2$, and suppose $D^{(2)}=GH'$ is a full rank decomposition of $D^{(2)}$ such that $G'G=I$. Then (17) is solvable for t_2 and T_3 if and only if we can find an M in the subspace defined by (34) which is positive semi-definite of rank p.

So according to theorem 3, we end up with a nonlinear problem and its conditions for solvability and uniqueness are complicated (if we merely assume that M must be positive semi-definite, and forget about its rank, then useful results can be obtained).

We propose an algorithm that minimizes the loss function

$$\zeta = \frac{tr(M - T_3 T_3')^2}{trM^2} \tag{35}$$

over the $(p+2) \times p$ matrices T_3 and over all M of the form (34). The algorithm is an alternating least squares method also used in INDISCAL (de Leeuw, Takane and Young 1977). We alternate computing a new optimal θ for fixed T_3 (a simple linear regression problem), and computing a new optimal T_3 for fixed M (a truncated eigenvalue problem), until convergence has been reached. Of course, when we choose q=1 only one iteration is needed.

Now suppose $\hat{t}_1$ is the solution of (18), $\hat{t}_2$ is a solution of (29) and $\hat{T}_3$ a solution of (35). Then the general solution for T is of the form

$$\hat{T} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \xi & 0 \\ 0 & 0 & \xi^{\frac{1}{2}}K \end{bmatrix}, \tag{36}$$

with $\xi > 0$ an arbitrary constant and K an arbitrary $p \times p$ rotation matrix. Define $\hat{S} = (\hat{T}^{-1})'$ and partition $\hat{S}$ as

$$\hat{S} = \begin{bmatrix} \hat{s}_1 & \vdots & \hat{s}_2 & \vdots & \hat{S}_3 \end{bmatrix}. \tag{37}$$

We now must have

$$\hat{Gt}_1 = e \tag{38a}$$
$$\xi G \hat{t}_2 = \alpha \tag{38b}$$
$$\xi^{\frac{1}{2}} G \hat{T}_3 K = X \tag{38c}$$

as well as

$$H \hat{s}_1 = \beta \tag{39a}$$
$$\xi^{-1} H \hat{s}_2 = e \tag{39b}$$
$$\xi^{-\frac{1}{2}} H \hat{S}_3 K = Z \tag{39c}$$

As can be seen, we need not to bother about the rotation, as it transforms X

and Y in the same way, and we can use (39b) to determine ξ. We finally find the solution for X and Y:

$$\hat{X} = \hat{\xi}^{\frac{1}{2}}G\hat{T}_3 \qquad (40a)$$

$$\hat{Y} = -\frac{1}{2}\hat{\xi}^{-\frac{1}{2}}H\hat{S}_3 \qquad (40b)$$

If (8) is not exactly solvable, we use least squares procedures for all sub-problems. There is one additional problem, however. Equation (22) will no longer have perfect solutions in general, because (25) is not necessarily true anymore. Consequently we may need a least squares technique (some canonical correlation method) to find a "best" solution, but if a good fit is possible there will be a number of canonical correlations very close to unity. It seems better to work with (28) and to define E as an approximate null space basis of (G×G). The problem then becomes when a small eigenvalue is considered to be equal to zero. Again the fact that we do not have information on the rank of (G×G) in the perfect case proves to be a nuisance.

2.2 Single centering

In this section we will freely use some of the same symbols as in the preceeding one, but with a different meaning. Again we suppose without loss of generality that the row points are centered. We now remove the column means of $D^{(2)}$:

$$\dot{d}_{ij}^{(2)} = d_{ij}^{(2)} - d_{.j}^{(2)}$$

$$= \sum_{a=1}^{p} x_{ia}^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{p} x_{ia}^2 - 2\sum_{a=1}^{p} x_{ia}y_{ja} \qquad (41a)$$

$$d_{.j}^{(2)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{p} x_{ia}^2 + \sum_{a=1}^{p} y_{ja}^2 \qquad (41b)$$

and in matrix notation (41a) becomes

$$\dot{D}^{(2)} = \left[\alpha \ \vdots \ X \right] \left[\begin{array}{c} e' \\ \cdots\cdots \\ Z' \end{array} \right] \qquad (42)$$

where α is now defined as

$$\alpha_i = \sum_{a=1}^{p} x_{ia}^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{p} x_{ia}^2 . \qquad (43)$$

This gives us another result proved by Ross and Cliff (1964).

Theorem 4. A necessary condition for the solvability of (41a), and thus of (8), is $\mathrm{rank}(\dot{D}^{(2)}) \leqslant p+1$.

In the rest of this section we ignore (41b), and we try to find X and Y from (42). Again we suppose that $\mathrm{rank}(\dot{D}^{(2)}) = p+1$, and again

$$\dot{D}^{(2)} = GH' \tag{44}$$

is a full rank decomposition of $\dot{D}^{(2)}$. Now we get

$$T = \left[t_1 \; \vdots \; T_2 \right] \tag{45}$$

with t_1 of length $p+1$ and T_2 $(p+1) \times p$. Furthermore

$$Gt_1 = (G \times G)m \; , \tag{46}$$

where the new $(G \times G)$ is a columncentered version of the old one:

$$(G \times G)_{i(kl)} = 2g_{ik}g_{il} - \frac{1}{n} \sum_{i=1}^{n} 2g_{ik}g_{il} \qquad (k > 1) \tag{47a}$$

$$(G \times G)_{i(kl)} = g_{ik}^2 - \frac{1}{n} \sum_{i=1}^{n} g_{ik}^2 \qquad (k = 1) \tag{47b}$$

and m is $M = T_2 T_2'$, strung out as a vector. This time, however, it is not possible to find an explicit representation like (28) for the solutions of (46). We may derive from (46), supposing $G'G = I$ again,

$$Gt_1 = GG'(G \times G)m \tag{48}$$

and consequently

$$(I - GG')(G \times G)m = 0 \; . \tag{49}$$

So if E contains a basis for the null space of $(I - GG')(G \times G)$, then

$$m = E\theta \tag{50}$$

or, in matrix form

$$M = \sum_{v}^{q} \theta_v \varepsilon_v \ . \tag{51}$$

The nonlinear problem now is to find θ such that M given by (51) is positive semi-definite of rank p. Again there is no information about the dimensionality of the null space E; in situations with non-exact fit we have to compute an approximate null space.

It is clear that single centering makes the unfolding problem more simple, in the sense that less arbitrary decisions have to be made, it is much clearer how to proceed. On the other hand, we use far less information from the data (by ignoring (41b)). We do not give an explicit algorithm, because after M, T_2 and t_1 are computed it is fairly obvious what to do. It is also clear that another single centering analysis can be done by removing row means instead of column means. But this amounts to the same thing, due to the symmetry of (8) in X and Y.

2.3. Double centering

We now describe an approach which is evidently inspired by the Torgerson-Gower approach to metric multidimensional scaling. We first define the centering operator

$$J = (I - \frac{ee'}{e'e}) \tag{52}$$

with e a vector with all elements equal to one (the number of elements of e follows from the context). Furthermore, we form the n×m matrix C by double centering $-\frac{1}{2}D^{(2)}$:

$$C = -\frac{1}{2}JD^{(2)}J \ . \tag{53}$$

According to (9), the matrix expression for $D^{(2)}$ is

$$D^{(2)} = e\beta' + \alpha e' - 2XY' \ . \tag{54}$$

Substituting (54) in (53) we get

$$C = (X - e\mu')(Y - e\nu')' \ , \tag{55}$$

where $\mu = \frac{1}{n}e'X$ and $\nu = \frac{1}{m}e'Y$ are the centroids of X and Y, respectively. From (55) we obtain another Ross-Cliff theorem.

Theorem 5. A necessary condition for the solvability of (55), and thus of (8), is rank(C)<p.

Here we have something quite different from the previous sections, because there is no partitioning, and there are no internal constraints. As a consequence we have to use the removed row- and column means of $D^{(2)}$ later on to determine X and Y and cannot simply ignore them, as we ignored (41b). Suppose again that rank(C)=p and suppose C=GH' is a full rank decomposition. It follows that

$$GT = X - e\mu' \tag{56a}$$
$$HS = Y - e\nu' \tag{56b}$$
$$TS' = I \tag{56c}$$

and obviously

$$X = GT + e\mu' \tag{57a}$$
$$Y = HS + e\nu' \tag{57b}$$

We now derive some intermediate expressions.

$$XX' = GMG' + (\mu'\mu)ee' + GT\mu e' + e\mu'T'G' \tag{58a}$$
$$YY' = HM^{-1}H' + (\nu'\nu)ee' + HS\nu e' + e\nu'S'H' \tag{58b}$$
$$XY' = GH' + (\mu'\nu)ee' + GT\nu e' + e\mu'S'H' \tag{58c}$$

where M=TT'. Furthermore

$$\alpha e' = (G\times G)me' + (\mu'\mu)ee' + 2GT\mu e' \tag{58d}$$
$$e\beta' = en'(H\times H)' + (\nu'\nu)ee' + 2e\nu'S'H' \tag{58e}$$

where (G×G) is defined as before (cf.21) and (H×H) is defined analogously on the columns of H. Substituting (58c),(58d) and (58e) in (54) we get

$$D^{(2)} = (G\times G)me' + en'(H\times H)' + 2GT(\mu - \nu)e' - 2e(\mu - \nu)'S'H' +$$
$$+ (\mu - \nu)'(\mu - \nu)ee' - 2GH' . \tag{59}$$

This is a matrix formulation of a result from Schönemann (1970). Following him, we now define a matrix F as

$$F = D^{(2)} + 2GH' \tag{60}$$

F must be of the form

$$F = \phi e' + e\psi' + \chi ee' \tag{61}$$

with

$$\phi = (G \times G)m + 2GT(\mu - \nu) \tag{62a}$$
$$\psi = (H \times H)n - 2HS(\mu - \nu) \tag{62b}$$
$$\chi = (\mu - \nu)'(\mu - \nu) \tag{62c}$$

At this point there are probably several ways to proceed. Schönemann (1970) subtracts the last row of F from the other ones to get rid of the terms with ψ and χ. We follow this asymmetric approach (solving for m and $(\mu-\nu)$ and defining a solution for S as the inverse of a solution for T), but prefer the somewhat less arbitrary procedure of centering F:

$$JF = J(G \times G)me' + 2 \, JGT(\mu - \nu)e' \, . \tag{63}$$

To simplify things a bit, we define the matrix K as

$$K = \left[J(G \times G) \vdots JG \right] \tag{64}$$

of order $n \times \frac{1}{2}p(p+3)$ and the vector of parameters ξ as

$$\xi = \begin{bmatrix} m \\ \dotsb \\ 2T(\mu-\nu) \end{bmatrix} \tag{65}$$

so that (63) transforms into

$$JF = K\xi e' \tag{66}$$

Consequently, in the perfect case we can use any column of JF, say f_j, to solve the nonhomogeneous system of equations $f_j = K\xi$. In the fallible case, we may simply average the f_j and solve the resulting linear regression problem. Having obtained a solution $\hat{\xi}$, we may transform it back to $\hat{M}$, $\hat{T}$, $\hat{S}$ and $(\hat{\mu}-\hat{\nu})$, and again assuming without loss of generality that one of the configuration matrices is centered, we may solve for X and Y from (57).

3.0 The iterative approach

In this section we briefly review the SMACOF-3 program, designed to minimize (6). It is based on the general multidimensional scaling technique discussed in de Leeuw (1977) and de Leeuw and Heiser (1977), for which convergence to a local minimum of (2) is assured. We will try out all three cases from section 2 as a starting configuration for the iterative technique, first on a set of dissimilarities of which we are sure the model holds and then in a real data situation.

3.1 SMACOF-3

The general iterative algorithm is as follows. Suppose Z_μ is our current best solution (we write μ for iterations). The basic iteration is

$$Z_{\mu+1} = V^+ B(Z_\mu) Z_\mu . \tag{67}$$

Here the matrix $B(Z_\mu)$ is defined by

$$b_{ik}(Z_\mu) = \frac{-w_{ik}\gamma_{ik}}{d_{ik}(Z_\mu)} \qquad \text{if } i \neq k \tag{68a}$$

$$b_{ii}(Z_\mu) = \sum_{l \neq i}^{N} \frac{w_{il}\gamma_{il}}{d_{il}(Z_\mu)} \tag{68b}$$

$$b_{ik}(Z_\mu) = 0 \qquad \text{if } d_{ik}(Z_\mu) = 0 \tag{68c}$$

Furthermore, V^+ is the Moore-Penrose inverse of V, which is defined by

$$v_{ik} = -w_{ik} \qquad \text{if } i \neq k \tag{69a}$$

$$v_{ii} = \sum_{l \neq i}^{N} w_{il} \tag{69b}$$

In our unfolding case we apply the partitionings defined in (4) and get

$$Z_\mu = \begin{bmatrix} X_\mu \\ \cdots \\ Y_\mu \end{bmatrix} \tag{70}$$

$$W = \begin{bmatrix} \phi & \vdots & ee' \\ \cdots & \cdots & \cdots \\ ee' & \vdots & \phi \end{bmatrix} \tag{71}$$

$$\Gamma = \begin{bmatrix} \phi & \vdots & \Delta \\ \cdots & \cdots & \cdots \\ \Delta' & \vdots & \phi \end{bmatrix} \tag{72}$$

$$V = \begin{bmatrix} mI & \vdots & -ee' \\ \cdots & \cdots & \cdots \\ -ee' & \vdots & nI \end{bmatrix} \tag{73}$$

$$V^+ = \begin{bmatrix} \frac{1}{m}(I - \frac{ee'}{n+m}) & \vdots & \phi \\ \cdots & \cdots & \cdots \\ -\frac{1}{m} \cdot \frac{ee'}{n+m} & \vdots & \frac{1}{n}(I - \frac{ee'}{m}) \end{bmatrix} \tag{74}$$

$$B = \begin{bmatrix} P & \vdots & A \\ \cdots & \cdots \\ A' & \vdots & Q \end{bmatrix} \tag{75}$$

$$a_{ij} = \frac{-\delta_{ij}}{d_{ij}(X_\mu, Y_\mu)} \tag{76a}$$

$$a_{ij} = 0 \qquad\qquad \text{if } d_{ij}(X_\mu, Y_\mu) = 0 \tag{76b}$$

$$p_{ii} = \sum_{l=1}^{m} \frac{\delta_{il}}{d_{il}(X_\mu, Y_\mu)} \tag{77a}$$

$$p_{ij} = 0 \qquad\qquad \text{if } i \neq j \tag{77b}$$

$$q_{jj} = \sum_{l=1}^{n} \frac{\delta_{lj}}{d_{lj}(X_\mu, Y_\mu)} \tag{77c}$$

$$q_{ij} = 0 \qquad\qquad \text{if } i \neq j \tag{77d}$$

With these specifications, our updates are (omitting V^+):

$$x_{ia}^{(\mu+1)} = \left(\sum_{l=1}^{m} \frac{\delta_{il}}{d_{il}}\right) x_{ia}^{(\mu)} - \sum_{j=1}^{m} \frac{\delta_{ij}}{d_{ij}} y_{ja}^{(\mu)} \tag{78a}$$

$$y_{ja}^{(\mu+1)} = \left(\sum_{l=1}^{n} \frac{\delta_{lj}}{d_{lj}}\right) y_{ja}^{(\mu)} - \sum_{i=1}^{n} \frac{\delta_{ij}}{d_{ij}} x_{ia}^{(\mu)} \tag{78b}$$

The effect of premultiplying by V^+ is, in words, to divide the updates (78a) by m, then center the updates (78b) and divide by n and finally readjust both of them such that they are jointly centered. We go on with these iterations until the improvement of stress (6) is no longer appreciable.

3.2 Comparisons if the model holds

We will study the behaviour of the three algorithms of section 2 in terms of the initial and final stress values they produce in the SMACOF-3 program. For this purpose we take the distances between 20 row- and 7 columnpoints, more or less randomly distributed in 2-space, and create three different levels of precision in the data: the distances themselves (PERFECT), the distances rounded to one significant digit (ROUNDED) and the distances dichotomized below and above their mean (BINARY). The results are tabulated an table 2.

	No centering		Single centering		Double centering	
PERFECT	0.000007	0.000004	0.000010	0.000004	0.000011	0.000006
ROUNDED	0.004023	0.001618	0.003610	0.001623	0.003118	0.001610
BINARY	0.461769	0.186396	0.422152	0.208834	-	-

Table 2. Initial and final values of stress under three levels of precision in the data.

In the perfect case, there is nothing for SMACOF to improve. In the rounded case, the initial values differ somewhat, but essentially they all converge to the same stress (and, at least to the naked eye, to the same (intended) configuration). In the binary case, the double centered algorithm breaks down because the second eigenvalue of M in (58) becomes negative. The other two reproduce the right configuration (with fairly large local distortions), but SMACOF can improve stress a lot (only small local distortions remain).

3.3 Comparisons in a real data situation

We now analyse a set of dissimilarities taken from Gold (1958). The data are reproduced in table 3. The 17 row objects represent several properties which are possibly related to social power in a group of middle class american children. Eight groups of children (labeled A - H) from about five to twelve years old judged all properties according to their importance in social relations.

|	A	B	C	D	E	F	G	H
1. Smart at school	13.5	13	17	15	16	17	17	16
2. Good ideas how to have fun	1	17	13	6	6	10	9	4
3. Good at making things	13.5	6.5	12	15	13	12.5	15	14
4. Good at games with running and throwing	16.5	3	14	17	17	11	13	13
5. Knows how to fight	12	4	11	15	14	15	16	12
6. Strong	9.5	13	15	13	12	16	14	15
7. Acts friendly	2	15.5	3	3	3	5	2	2
8. A good person to do things with	9.5	1	4	11	9	6	6	9
9. Asks you to do things in a nice way	5.5	5	1	4	4	2	1	5
10. Doesn't start fights and doesn't tease	5.5	11	7.5	1	7	1	4	7
11. Knows how to act so people will like him	15	13	5	2	2	8	5	3
12. Plays with you a lot	3	8.5	9	10	8	9	11	6
13. Likes to do the same things you like to do	5.5	6.5	10	5	1	7	8	1
14. Nice looking	11	10	7.5	12	15	14	10	17
15. Has things you'd like to have	16.5	15.5	16	7	10	12.5	12	10
16. Gives you things	8	8.5	6	9	11	3	7	11
17. Does things for you	5.5	2	2	8	5	4	3	8

Table 3. Ranks of items by per cent of times they were rated "very important"; low value - most important. Taken from Gold (1958).

The details of data collection and group composition do not bother us here, we will just present six solutions (three initial and three final ones, see figures 1 to 6), and discuss some striking features.

figure 1.

figure 2.

2-DIMENSIONAL INITIAL SOLUTION. DIMENSION 1 (X-AXIS) VERSUS DIMENSION 2 (Y-AXIS)
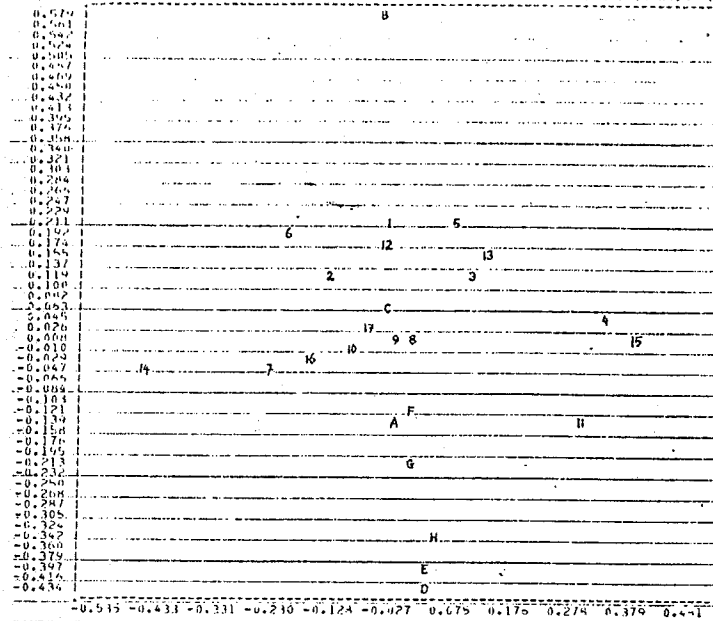
Start
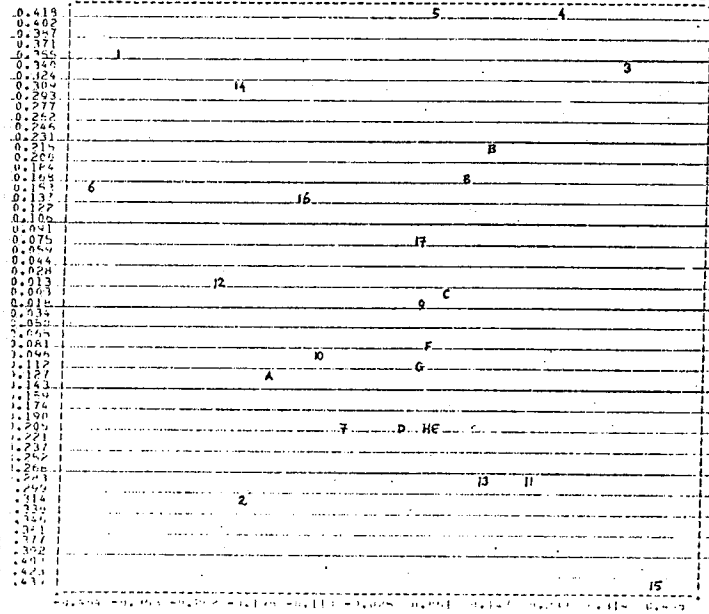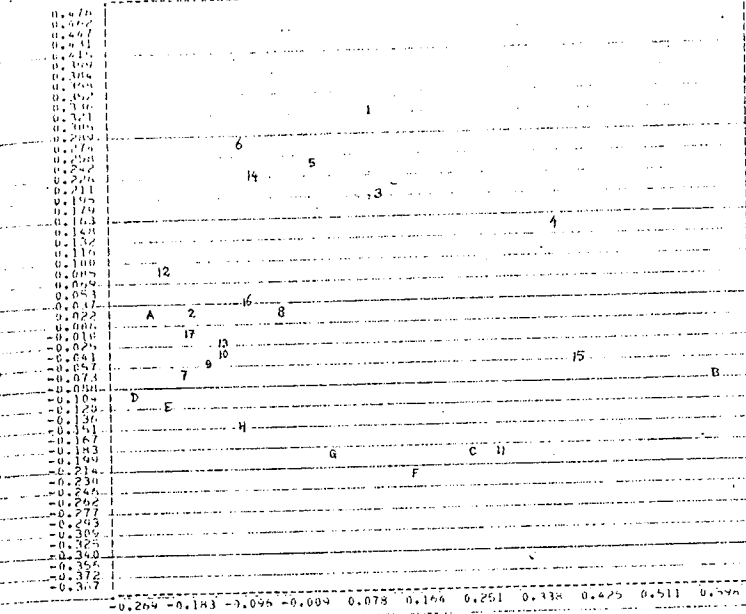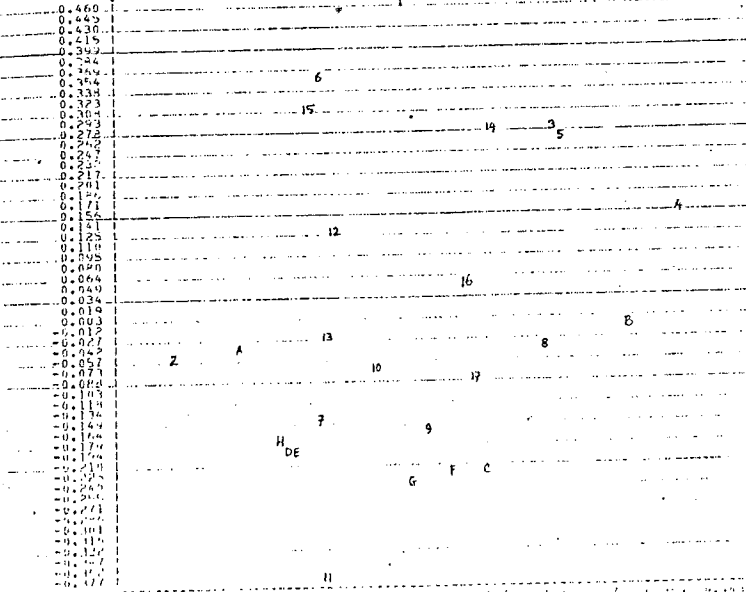*Single centering*
Stress = 0.231916

figure 3.

2-DIMENSIONAL FINAL SOLUTION. DIMENSION 1 (X-AXIS) VERSUS DIMENSION 2 (Y-AXIS)

SMACOF-3
based on
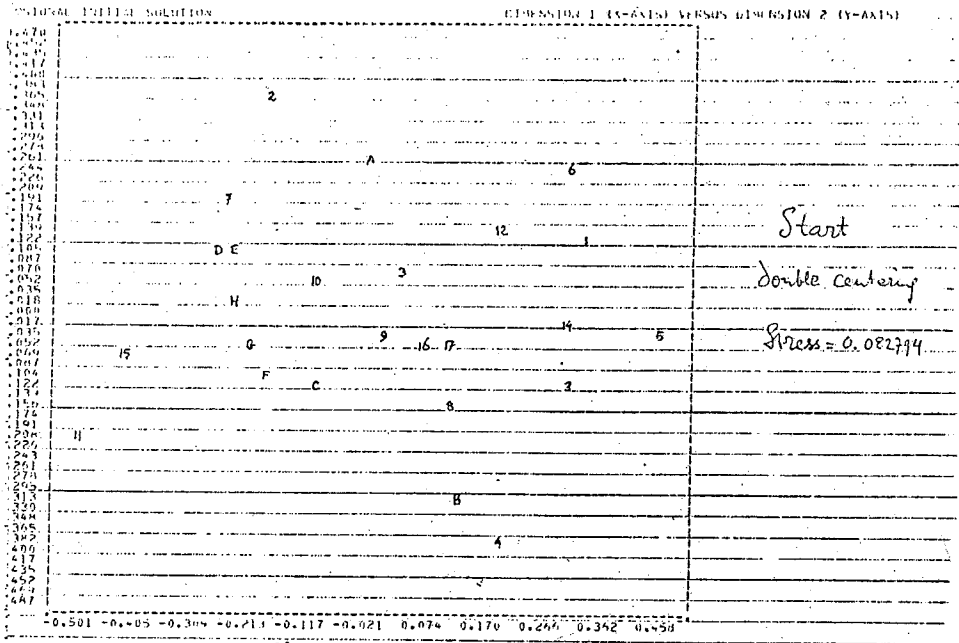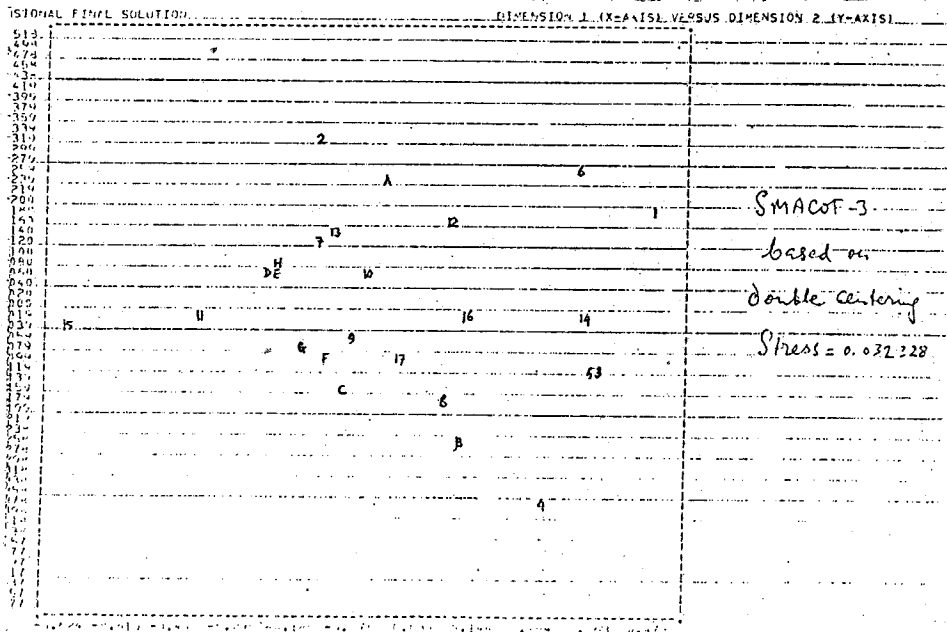*single centering*
Stress = 0.038786

figure 4.

figure 5.

figure 6.

In the first place, neither the three initial nor the three final solutions
resemble each other very much (there is more similarity between each initial
solution and its own final one). Furthermore, the final stresses roughly are
the same, which illustrates again the local minimum problem. Of course, the
six solutions are not completely different. The column points nearly always
lie along a curved line in the order A - (D E H) - (G F C) - B except in
the "no centering start" configuration, which is severely elongated with A
out of its common position. Some of the row points are always close together
(cf. 16,17 and 8 with 8 always in the direction of B), others are always far
apart (cf. 2 anf 4, the first one being typical for roup A and the second one
for B). Also, some points move a lot (cf.15), others hardly never (point 9 is
always near the centroid). These observations are rather typical for real data
situations.

Although in this example the double centering approach does produce the lowest
stress values, we think it is to early to give any definite recommendations
regarding the quality of the different approaches to metric multidimensional
unfolding.

References

Bennett,J.F, and W.L.Hays, 1960. Multidimensional unfolding: determining the
    dimensionality of ranked preference data, PM, 25, 27 - 43.
Coombs, C.H., 1950. Psychological scaling without a unit of measurement, Psych.
    Review, 57, 148 - 158.
Coombs, C.H., 1964. A theory of data. Wiley, New York.
Coombs, C.H., and R.C.Kao, 1960. On a connection between factor analysis and
    multidimensional unfolding, PM, 25, 219 - 231.
Gold, E.M., 1973. Metric unfolding: Data requirement for unique solution and
    classification of Schönemann's algorithm, PM, 38, 555 - 569.
Gold, M., 1958. Power in the classroom, Sociometry, 21, 50 - 60.
Hays, W.L. and J.F.Bennett, 1961. Multidimensional unfolding: determining
    configuration from complete rank order preference data, PM, 26, 221 -238.
Kruskal, J.B. and J.D.Carroll, 1969. Geometrical models and badness-of-fit
    functions. In: P.R.Krishnaiah, ed. Multivariate Analysis II, 639 - 671.
    Academic Press, New York.

de Leeuw, J., 1977. Applications of convex analysis to multidimensional csa
    In: J.R. Barra e.a., eds. Recent Developments in Statistics. North Holla⸱
    Publishing Company, Amsterdam.

de Leeuw, J. and W.J. Heiser, 1977. Convergence of correction matrix algori⸱
    for multidimensional scaling. In: J. Lingoes, ed. Geometric Representati
    of relational data. Mathesis Press, Michigan.

de Leeuw, J., Y. Takane and F.W. Young. An indicator method of multidimensi
    scaling (in preparation).

Ross, J. and N. Cliff, 1964. A generalization of the interpoint distance mo⸱
    PM, 35, 349 - 366.

Schönemann, P.H., 1970. On metric multidimensional unfolding, PM, 35, 349 -