

A MODEL AND ALGORITHM FOR MULTIDIMENSIONAL SCALING WITH EXTERNAL CONSTRAINTS ON THE DISTANCES

INGWER BORG

RHEINISCH-WESTFÄLISCHE TECHNISCHE HOCHSCHULE

JAMES C. LINGOES

THE UNIVERSITY OF MICHIGAN

A method for externally constraining certain distances in multidimensional scaling configurations is introduced and illustrated. The approach defines an objective function which is a linear composite of the loss function of the point configuration X relative to the proximity data P and the loss of X relative to a pseudo-data matrix R . The matrix R is set up such that the side constraints to be imposed on X 's distances are expressed by the relations among R 's numerical elements. One then uses a double-phase procedure with relative penalties on the loss components to generate a constrained solution X . Various possibilities for constructing actual MDS algorithms are conceivable: the major classes are defined by the specification of metric or nonmetric loss for data and/or constraints, and by the various possibilities for partitioning the matrices P and R . Further generalizations are introduced by substituting R by a set of R matrices, R_i , $i = 1, \dots, r$, which opens the way for formulating overlapping constraints as, e.g., in patterns that are both row- and column-conditional at the same time.

Key words: constrained multidimensional scaling, hypothesis testing, geometric models, nonlinear optimization.

1. Introduction

Particularly in the context of nonmetric multidimensional scaling, it has been repeatedly pointed out [cf., e.g., Guttman, 1968] that minimizing a general loss function like the one defined by Kruskal's stress measure or Guttman's coefficient of alienation does not always impose enough constraints upon the solution. There are many cases where one knows or hypothesizes more about the point configuration X and/or its functional relation to the data than to simply say that the data can be mapped into distances on X by some monotonic function. If such structural expectations are not incorporated into the general loss function as additional constraints upon the solution X , it is hard to say if there exists a configuration which satisfies the hypotheses.

Consider the following two cases: (a) The graph of the optimization function is very flat in the neighborhood of the (local) minimum solution point; (b) the optimization function has many local minima with almost the same value but quite different coordinates. In both instances, the loss function has, so-to-speak, a "low degree of uniqueness" which can imply the existence of different X 's which are qualitatively distinct but practically equal in alienation. If one wishes to test a structural hypothesis about one's data by scaling them via an MDS procedure, it is necessary to know what loss would be incurred if the solution X were forced to be feasible in terms of constraints that go beyond the simple general restrictions contained in the alienation coefficient, for example. If, under these additional constraints, the loss measure is not appreciably higher than the one for the general, un-

Requests for reprints should be sent to James C. Lingoes, The University of Michigan, 1005 North University Building, Ann Arbor, Michigan 48109.

restricted model, the hypotheses are apparently shown to be compatible with the data, provided, of course, the representation is reasonably good in terms of absolute alienation. On the other hand, in a more exploratory context, a configuration generated via a general MDS procedure may turn out to have properties which are rather unattractive. The researcher may then decide that he does not want to interpret the solution before these psychological dissonances were eliminated. Thus, after this preliminary general MDS analysis, a rescaling of the data would become necessary where certain characteristics of the solution—for example, certain order relations among the inter-point distances—are defined to be inadmissible.

A third and more technical instance in which further constraints on X are absolutely essential is related to cases where one would obtain degenerate or quasi-degenerate solutions. Such degeneracies can be (a) a consequence of the data as, for example, if SSA-I is used on a data matrix which can be partitioned such that all proximities within partitions are greater than those between partitions, or (b) they can occur, independently of whatever the data are, as a consequence of the properties of the MDS procedure as in the case of an unconditional unfolding analysis [Carroll, 1979].

Because of these and other less general reasons, a number of different approaches for restricting the general MDS loss function have been proposed recently. Shepard & Crawford [Note 6] introduce a procedure for constraining the shape of the monotone mapping function to convexity, concavity, local linearity, etc. [cf. also, Shepard, 1974]. Another approach is taken by Bentler & Weeks [1978]: they allow one to reduce the cardinality of the parameter space for X by specifying functional relationships among the point coordinates or by setting some of these values equal to an externally provided constant value. A similar line of thought underlies Bloxom's [1978] and Carroll et al.'s [1980] methods for constraining MDS solutions in N spaces, i.e., in the context of individual differences scaling. A generalization of such restricted metric MDS approaches which allows one to specify linear and nonlinear constraints has recently been described by deLeeuw & Heiser [in press]. Noma & Johnson [Note 4, 5], on the other hand, developed a procedure in which one can constrain the order of point projections on the dimensions of the solution space. This approach is generalized by Skarabis [Note 7, 1978] who discusses various procedures designed to satisfy order constraints on dimensional projections as well as on inter-point distances. These latter constraints cannot be made by specifying projection orders, except in special cases like a conditional ordering on the distances. Yet, they seem to be particularly pertinent to nonmetric multidimensional scaling and many questions asked in this context. A completely different approach is taken by Guttman [Note 1] who sketches an algorithm that restricts the MDS solution to satisfy certain regional hypotheses derived from facet theory.

In this paper, we will consider a very general alternative approach to constrained metric and nonmetric multidimensional scaling. Our procedure is designed such that the metric and/or order constraints on the distances of X can be formulated in a particularly simple and yet flexible way by using a pseudo-data matrix which describes, by the relations among its numerical elements, the external restrictions to be imposed on X . The method is first described in general; then, a specification of the general formulation is outlined; finally, some applications help illustrate some of the issues involved in such MDS problems.

2. General Definitions and Equations

In the following section, we state some definitions and equations that are basic to all MDS procedures that will be discussed in this paper.

(2.1) $P = (p_{ij})$: an $n \times n$ symmetric matrix of proximity measures with

- n being the number of objects or stimuli. Some elements of P may be undefined (missing data).
- (2.2) $X = (x_{ia})$: an $n \times m$ matrix of coordinates with m being the dimensionality of the configuration that X describes.
- (2.3) $D = (d_{ij})$: an $n \times n$ matrix of distances on X .
- (2.4) $R = (r_{ij})$: an $n \times n$ matrix of pseudo-data which expresses, by the relations among its numerical elements, the restrictions to be imposed on the corresponding elements in P . The matrix R may only be partially defined, or, expressed differently, some r_{ij} 's may be "missing data". Moreover, R may be partitioned by defining any disjoint subsets of its elements.
- (2.5) $\Delta_P = (\delta_{ij(P)}) = T(P)$: an $n \times n$ matrix generated from P by the transformation T . T is a set of functions with elements defined for the respective partitions of P . T is "admissible" in the sense that each of its elements preserves those properties of its domain variables which are considered meaningful. In addition, T optimizes the correspondence between P and D in some sense, i.e., T is either a least-squares fitting transformation of P with respect to D , or a rank-image permutation of D relative to P .
- (2.6) $\Delta_R = (\delta_{ij(R)}) = M(R)$: an $n \times n$ matrix which is a function of R . The transformation M is defined for specified values of R only. Analogous to T in (2.5), M optimizes a relationship between D and R which is either a least-squares fit of R to D or a rank-image permutation of D relative to R .

Properties of M and T

2.7.1 Monotone functions. For nonmetric definitions of M or T , the following monotonicity cases are relevant [Lingoes & Roskam, 1973]: (a) $p_{ij} > p_{kl}$ (P being a set of similarity data.) $\Rightarrow \delta_{ij} < \delta_{kl}$, $i \neq j$, $k \neq l$. This defines semi-strong monotonicity if some values in P are tied, since $p_{ij} = p_{kl}$ apparently implies nothing about the order relation of the images of p_{ij} and p_{kl} . Thus, ties in P can be "broken" in Δ , i.e., Δ is less restricted and can be constructed to fit D better. If no ties exist in P , we have strong monotonicity. (b) $p_{ij} > p_{kl} \Rightarrow \delta_{ij} < \delta_{kl}$ & $p_{ij} = p_{kl} \Rightarrow \delta_{ij} = \delta_{kl}$. This is the more general case of strong monotonicity where ties in P are considered meaningful and have to be preserved in the solution X . (c) $p_{ij} > p_{kl} \Rightarrow \delta_{ij} \leq \delta_{kl}$. This defines weak monotonicity if there are ties in P . Otherwise, we have semi-weak monotonicity with the more general case being analogous to (b).

2.7.2 Rank images. In a nonmetric context, M and T may generate so-called rank images of D relative to P or R , respectively, denoted by D_P^* and D_R^* . Letting Y stand for either P or R , this means that the operators induce a permutation in D such that the ranks of the elements in the resulting $\Delta = D^*$ correspond to the ranks of the elements in the reference sequence Y . The nature of the correspondence is determined by the respective definition of monotonicity and, of course, by the existence of ties in Y and/or D . If ties exist in D , the permutation is not determined within the respective sets of tied values. If ties exist in Y , the values in D^* corresponding to those in the respective tied sets of Y are permuted such that a better fit between Y and D^* is achieved, provided the monotonicity criterion allows such untying (primary approach to ties). Otherwise, the values in these subsets of D^* are averaged (secondary approach to ties) [Guttman, 1968; Roskam, 1977].

2.7.3 Monotone regression. Another choice for M and/or T in a nonmetric procedure is Kruskal's monotone regression [Kruskal, 1964]. The resulting sequence of fitting values, $\Delta = \hat{D}_v$, is required to be monotone in some sense with Y , where Y is equal to P or R . Under the stronger model, which is related to choosing the secondary approach to ties, the values in D corresponding to each respective set of tied values in Y are averaged first. The

resulting sequence of numbers is then iteratively partitioned into larger and larger subsets until the averages of the elements of these subsets of "blocks" are weakly monotone with Y . Once this is achieved, all values belonging to one block are set equal to this block's average. The weaker model drops some constraints as a consequence of requiring only the primary mapping of ties which permits different distances for tied data. Here, the tied elements of Y are permuted first within their respective tie-blocks such that the corresponding distances all form monotonically increasing sequences (for dissimilarity data). One then proceeds as before via blocking and averaging.

2.7.4 Metric mappings. The simplest metric MDS procedure is certainly the linear one which defines Δ as follows:

$$\Delta = a + bY,$$

where Y is equal to P or R , and

$$(2.7.4') \quad b = \frac{[\sum_{ij}(d_{ij} - \bar{d})(y_{ij} - \bar{y})]}{\sum_{ij}(y_{ij} - \bar{y})^2},$$

$$a = \bar{d} - b\bar{y},$$

where \bar{d} and \bar{y} are the means of those elements of D and Y , respectively, to which the summation sign in (2.7.4') refers. This simple regression problem is generalized in an obvious way if partitions are defined on Y : the summations then extend over the respective subsets only, i.e., a separate regression equation is set up for each partition. A substantial number of nonlinear models can also be handled by the linear approach if the regression function is linearizable. (Some of the more common linearizing transformations are described by Daniel & Wood [1971]). In such a situation, one has to substitute Y in (2.7.4') by Y^* , where Y^* is the linearized Y . For example, for the model $y_{ij} = a \exp(bd_{ij})$, one sets $y_{ij}^* = \log(y_{ij})$. Other, strictly nonlinear models are principally also conceivable in this context, but they certainly require special study. We shall not pursue this issue here.

Loss Functions

The multidimensional scaling problem can now be formulated as

$$(2.8.1) \quad L = \alpha L_P + \beta L_R = \min,$$

where $\alpha, \beta \geq 0$, $\alpha + \beta = 1$, and

$$(2.8.2) \quad L_P = \sum_{ij}(d_{ij} - \delta_{ij(P)})^2,$$

$$(2.8.3) \quad L_R = \sum_{ij}(d_{ij} - \delta_{ij(R)})^2.$$

The trivial solution where all points in X are identical is excluded by a norming constraint on X .

Numerical Solution

Problem (2.8.1) represents a standard MDS loss function which can be solved by an iterative method such as the gradient minimization employed in most MDS procedures. Thus, we rewrite (2.8.1) as

$$(2.9.1) \quad L^{(s)} = \alpha^{(s)} L_P^{(s)} + \beta^{(s)} L_R^{(s)} = \min,$$

where s is an iteration index. Setting $\alpha^{(s)} = k$ and $\beta^{(s)} = c$ leads to problems previously considered in KYST [Kruskal et al., Note 2] and ALSCAL [Takane et al., 1977], for example. (The usual psychological context is scaling of replicated data with P being the data col-

lected at time t_1 and R being the replication.) For the present purpose of constraining MDS solutions, we define

$$(2.9.2) \quad \lim_{s \rightarrow \infty} \beta^{(s)} = 1.$$

The relative weight of the loss component related to the external constraints R becomes thus greater and greater as the iterative minimization process proceeds, which eventually forces X to satisfy the constraints as nearly as possible. Solving (2.9.1) under condition (2.9.2) differs from minimizing (2.8.3) only if D is not completely determined by (2.8.3). In other words, there must be some leeway for the data to influence X . There are various possibilities for such a situation, e.g.: R contains many “missing data”, while P is complete; R is partitioned such that M represents a set of functions as, for example, in case of a row-conditional matrix, while P is unconditional; R contains a large number of ties, and M and T are monotone mappings using the primary approach to ties, while P contains values that are all distinct; etc.

An obvious generalization (not previously introduced to simplify notation) is the following:

$$(2.9.3) \quad L^{(s)} = \alpha^{(s)} L_P^{(s)} + \beta^{(s)} [L_{R_1}^{(s)} + \dots + L_{R_r}^{(s)}].$$

We will show later that the case where $r > 1$ is indeed needed for relatively simple constraint patterns. Of course, (2.9.3) could also be further generalized by allowing for more than just one P matrix, for example.

3. A Specification of the General Approach: the CMDA Procedure

An algorithm which represents a specification of the general formulation outlined above has been developed and programmed [Lingoes & Borg, 1978]. The procedure is a “nonmetric” MDS method which allows the user to impose certain order constraints on the distances of X . In the following, we will denote this method as CMDA, an acronym for “constrained/confirmatory monotone distance analysis”.

CMDA makes extensive use of previous work by Guttman [1968] and Lingoes & Roskam [1973]. It suffices, therefore, to give but a brief outline of the algorithm [see also: Borg & Lingoes, 1979].

Assuming the existence of an initial configuration X , we seek to improve X relative to L via a movement in the direction of steepest descent in Re^m . The minimization of L is conditional to fixed targets Δ_P and Δ_R , respectively. These targets are computed first in cycle s . The required gradient is then found via partial differentiation of (2.9.1) with respect to X . One obtains the correction formula

$$(3.1) \quad x_{ia}^{(t+1,s)} = x_{ia}^{(t,s)} - \left(\frac{1}{2n}\right) [\alpha^{(s)} \nabla L_P + \beta^{(s)} \nabla L_R],$$

where

$$(3.2) \quad \nabla L_P = 2 \sum_j \left(\frac{1 - \delta_{ij(P)}}{d_{ij}} \right) (x_{ia} - x_{ja}),$$

and ∇L_R is analogous. For undefined elements in P or R , one sets $\delta_{ij(P)}/d_{ij} = \delta_{ij(R)}/d_{ij} = 1$. Correction (3.1) is carried out for all coordinates iteratively until the t -process converges reasonably well. The obtained X then gives rise to new targets Δ_P and Δ_R in $s + 1$. As to the penalty weights $\alpha^{(s)}$ and $\beta^{(s)}$, we set $\alpha^{(s=0)} = \beta^{(s=0)} = 0.5$ and $\alpha^{(s+1)} = k/1.1^s$, for $k < 1.1$, and $\beta^{(s+1)} = 1 - \alpha^{(s+1)}$.

The target matrices Δ_P and Δ_R are constructed via rank-image transformations for a number of initial s -iterations and later—when the minimization process seems to have converged essentially or starts to diverge—via the monotone regression approach. This tactic has the advantage of speed as well as preserving distinctness of the points in the beginning and later exploits the least-squares minimization property of monotone regression to ensure convergence [de Leeuw & Heiser, 1977].

The generalized loss function (2.9.3) is easily implemented by setting $\nabla L_R = \nabla L_{R_1} + \dots + \nabla L_{R_r}$ in (3.1). CMDA presently allows for two R matrices.

Partitions on P or R and the chosen approach to treatment of ties are incorporated into CMDA by constructing the target matrices Δ_P and Δ_R accordingly. Partitions simply imply that the respective subsets of Δ_P and Δ_R are generated independently of each other. For rank images, this means that D is not permuted into the unconditional rank order of P or R , but rather that each respective partition of D is permuted into the ranks of the corresponding partition of P or R . In case of monotone regression, the disparities are constructed for each partition in turn. In the primary approach to ties, ties in the data are untied by ranking tied data according to the ranking of the corresponding elements in P and R , respectively. Using the secondary approach to ties and rank images entails averaging all fitting values in Δ for each set of ties in the data. For monotone regression targets, the secondary approach to ties requires a prepartitioning of D [see (2.7.3)].

As to a “rational” initial configuration, a number of possibilities is conceivable. None of them is very satisfactory since R can be constructed in so many different ways. However, since CMDA’s purpose is closely related to the existence and testing of structural hypotheses, it seems most appropriate to use a starting configuration that is either directly generated from a substantive theory or is the solution of an unconstrained MDS analysis modified such that it reflects the substantive hypotheses.

4. *Some Illustrative Applications of CMDA*

In the following, we will illustrate the CMDA procedure with the help of some real data applications. No attempt will be made to go into much substantive detail: the purpose of the examples is rather to demonstrate how the procedure can be used.

The first application reanalyzes Ekman’s [1954] well-known data on the similarity of 14 colors. These data were previously investigated by various researchers using different MDS techniques [e.g., Coombs, 1964; Shepard, 1962, 1978]. It was uniformly found that a distance representation is possible in a two-dimensional space. In addition, the expected circular order of the color-points was also shown to result from the scaling, thus confirming the color circle as a principle of color perception. Using SSA-I, for example, one obtains the configuration represented by the open circles in Figure 1. The solution is in excellent agreement with the data, the alienation being $K = 0.03$. One notices, however, that the points do not fall on a perfect circle, although very nearly so. We will now use CMDA to “project” them on a circle such that the projection points are as compatible with Ekman’s data—in the sense of the loss function L_P —as possible. The specifications made for the procedure are as follows: (a) As an initial configuration, the SSA-I solution, augmented by a fifteenth point somewhere in the center of the circular point arrangement, is used. (b) The P matrix corresponds to Ekman’s data matrix, but is augmented by a dummy variable which represents the “center point” used in (a); the relations of this dummy variable to the other objects are defined to be missing data. (c) The constraint matrix R is also 15×15 , containing the same value (e.g., all 1’s or 2’s) in row/column 15 and “missing data” otherwise. (d) The secondary approach to ties on R is chosen which forces all points representing real data to be equidistant from the center point. The thus constrained MDS solution is found in but a few iterations. It corresponds to the solid points

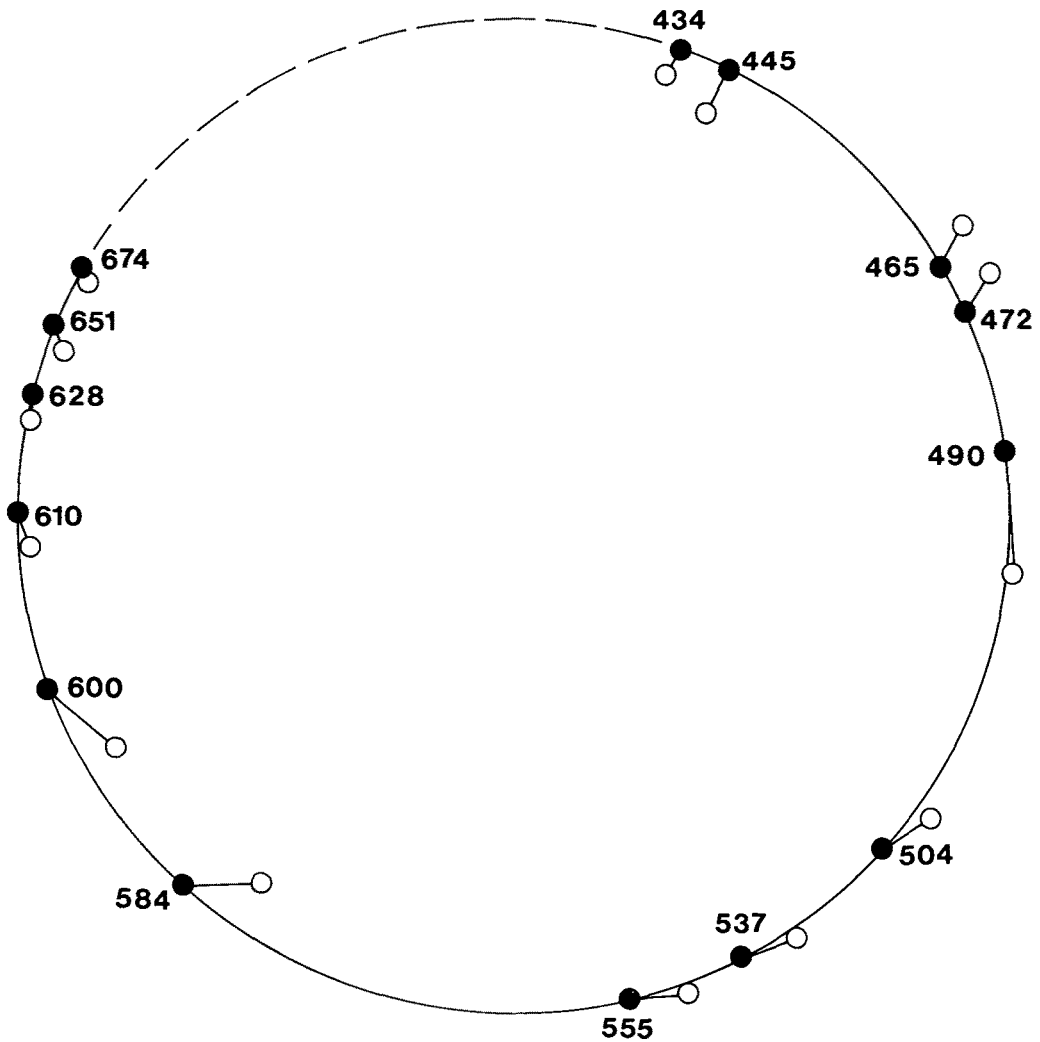


FIGURE 1

SSA-I representation for Ekman's color data (open circles); CMDA solution for same data (solid circles); Numbers indicate wavelength of stimuli.

in Figure 1. Not surprisingly, its alienation is only very slightly higher ($K = 0.04$) than the one for the SSA solution ($K = 0.03$) which shows that the loss function is pretty flat in the neighborhood of the SSA solution. From this, one may derive the substantive conclusion that the elliptic deformation of the color circle in the unconstrained solution is probably psychologically meaningless.

We now consider some data on the perception of simple geometrical objects. A problem that has been of considerable interest for some time is how the judgments that an individual gives with respect to the similarity of simple geometrical stimuli can be explained. These stimuli are commonly constructed by systematically varying some of their attributes like "length" and "width" for rectangles [Krantz & Tversky, 1975; Schönemann, 1977], "area" and "eccentricity" for ellipses [Noma & Johnson, Note 4], "length" and "inclination" of straight lines [Eisler & Roskam, 1977], etc. Our illustration uses data from Noma & Johnson [Note 4] on ellipse perception. They asked subjects to evaluate the

pairwise similarities of 16 ellipses constructed according to the two-dimensional design depicted in Figure 2. It was hypothesized that the individuals base their judgements on the design dimensions but may use monotone psychophysical functions to distort them.

Scaling the individuals' similarity data via SSA-I, one finds one subject whose data lead to a representation (Figure 3) that is grossly angle-distorting or non-conformal with $K = 0.15$. Testing different initial configurations further suggests that this solution is indeed the global minimum for the unconstrained MDS problem. Apparently, the result is not in agreement with the substantive theory and might lead to a rejection of the perceptual model. In order to evaluate the significance of the finding, we now use CMDA to enforce conformality of the representation as follows: (a) As an initial configuration, the design lattice in Figure 2 is used. (b) With $d(i, j)$ denoting the distance between points i and j , nine constraint sets are enforced on X : $d(1, 2) = d(5, 6) = d(9, 10) = d(13, 14)$, $d(2, 3) = d(6, 7) = d(10, 11) = d(14, 15)$, $d(3, 4) = d(7, 8) = d(11, 12) = d(15, 16)$; $d(1, 5) = d(2, 6) = d(3, 7) = d(4, 8)$, $d(5, 9) = d(6, 10) = d(7, 11) = d(8, 12)$, $d(9, 13) = d(10, 14) = d(11, 15) = d(12, 16)$; and $d(1, 6) = d(2, 5)$, $d(6, 11) = d(7, 10)$, $d(11, 16) = d(12, 15)$. (b) The secondary approach to ties is used on R and all elements within a constraint set are set equal to some constant number k . Under these conditions, one obtains the representation in Figure 4 which is, of course, in perfect agreement with the hypothesis but only slightly higher in alienation ($K = 0.16$) than the one in Figure 2. (Although we have succeeded here in imposing restrictions on the coordinates by constraining the distances, it seems generally easier to handle such cases with one of the procedures designed specifically for that purpose [e.g., Bentler & Weeks, 1978; Noma & Johnson, Note 4, 5; de Leeuw & Heiser, in press; etc.]). The results illustrate one of the points made in the introductory paragraph of this paper: the loss criterion employed in general MDS procedures like SSA-I may be too general for the problem at hand. Tailoring the optimization problem to the substantive research problem can make the MDS solution much more meaningful.

As a final example, we will now investigate a data set imposing order constraints only. Guttman [1965] reports a matrix of intercorrelations for eight intelligence tests (Table 1). A theoretical study of the content of these tests leads Guttman to the formulation of a two-facet definition with the facets $\{Ac = \text{achievement}, An = \text{analytical}\}$ and $\{N = \text{numerical}, P = \text{pictorial}\}$. The respective structuples for the tests are given in Table 1. Note that there are just four different test types: the digit following the structuple denotes a replication. If the content definitions are actually useful for organizing empirical observations, one should expect to find the following structure of items if Table 1 is represented via a nonmetric MDS procedure (Figure 5).

TABLE 1

Intercorrelations for Eight Intelligence Test Items

Structuple*	NAc1	NAc2	NAn1	PAn1	PAn2	PAc1	PAc2	PAc3
1:NAc1	*	.67	.40	.19	.12	.25	.26	.39
2:NAc2	.67	*	.50	.26	.20	.28	.26	.38
3:NAn1	.40	.50	*	.52	.39	.31	.18	.24
4:PAn1	.19	.26	.52	*	.55	.49	.25	.22
5:PAn2	.12	.20	.39	.55	*	.46	.29	.14
6:PAc1	.25	.28	.31	.49	.46	*	.42	.38
7:PAc2	.26	.26	.18	.25	.29	.42	*	.40
8:PAc3	.39	.38	.24	.22	.14	.38	.40	*

* See text for meaning of structuples.

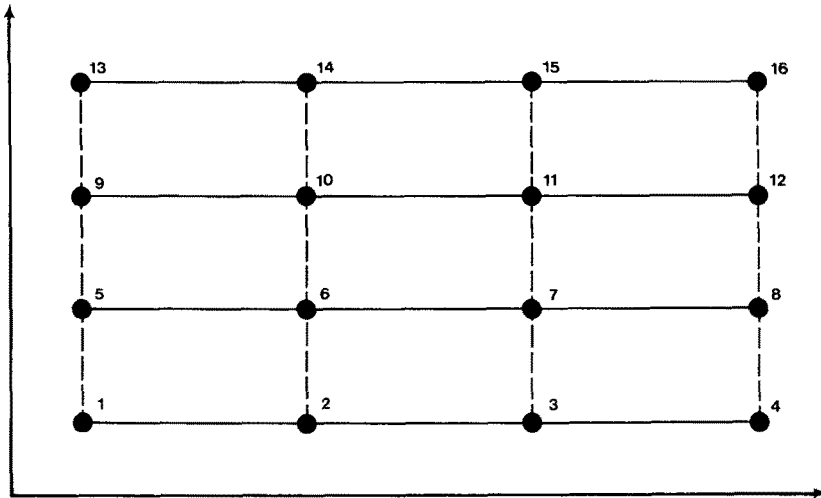


FIGURE 2
Design configuration for Noma & Johnson's ellipse data.

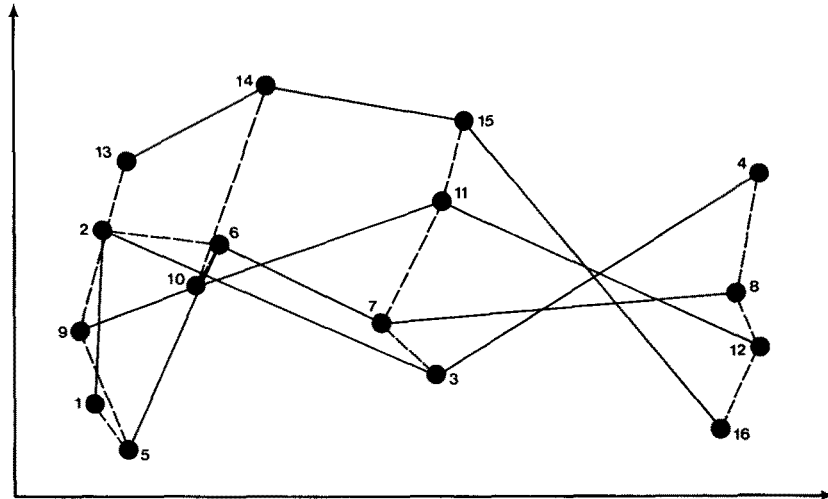


FIGURE 3
SSA-I solution for ellipse similarity data of subject from Noma & Johnson's experiment.

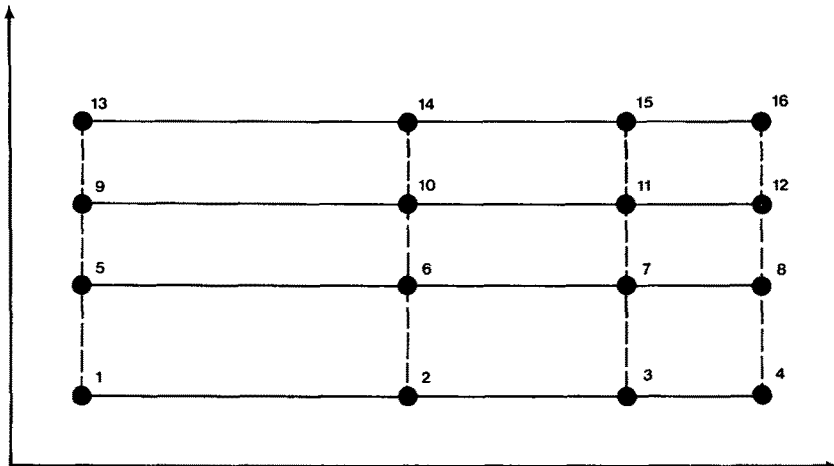


FIGURE 4
CMDA solution for subject from Figure 3.

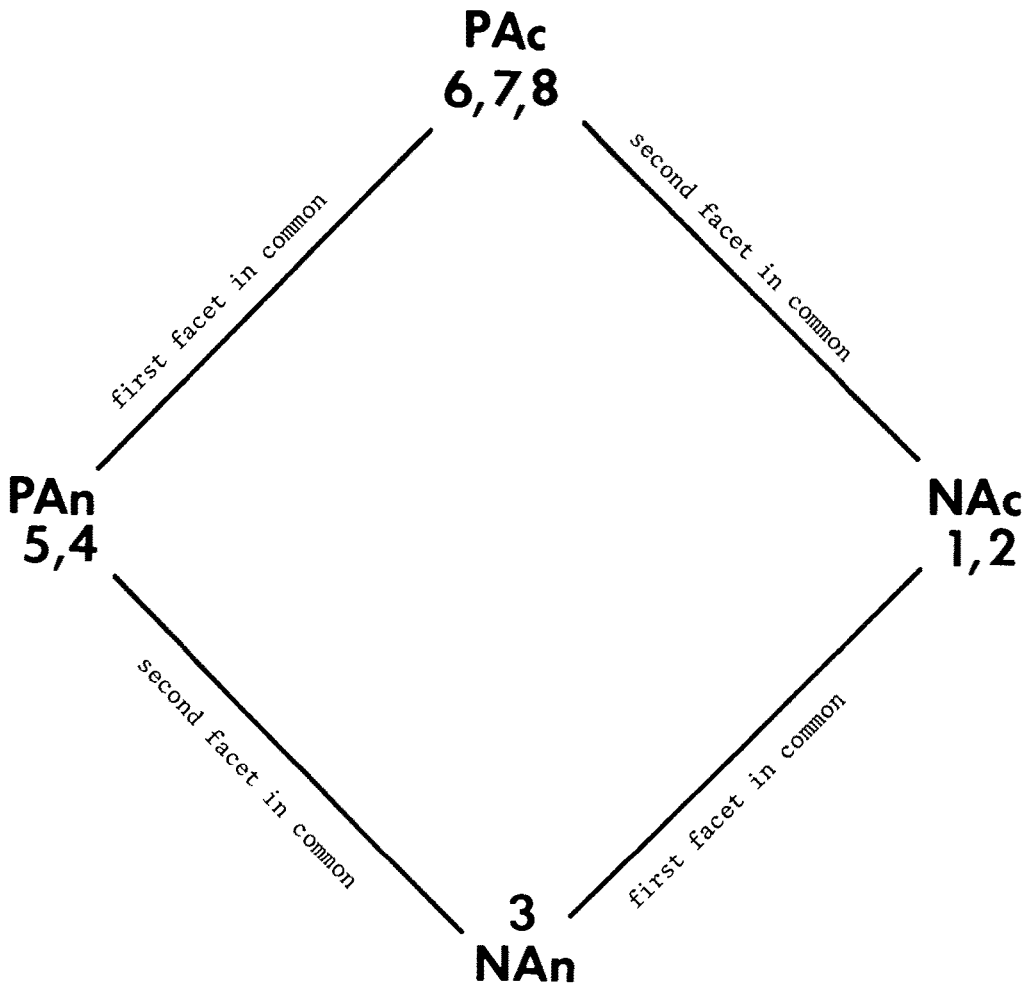


FIGURE 5
Expected MDS representation for data in Table 1.

Scaling the data via SSA-I leads indeed to a configuration which comes close to fulfilling the expectations depicted in Figure 5. However, the configuration is too compressed in the vertical direction which brings points 3 and 6, for example, so close together that $d(3, 6) < d(1, 6)$, contrary to the hypothesis. The question that arises now is just how much such an elliptical compression of Figure 5 is due to data constraints. An answer is given by enforcing the various conditional orders implied in the facet definitions as side constraints for the MDS analysis. For that purpose, we use the matrix in Table 2 as R and employ the primary approach to ties.

Note that by using the primary approach to ties, R imposes actually substantially fewer constraints on X than P , since P contains much fewer ties. To be more precise, P specifies $378 - 11 = 367$ order conditions, but R only $378 - 151 = 227$, where a complete order of eight elements implies 378 restrictions and the subtraction factors are due to the existence of ties. (It should be noted, however, that the number of constraints or even the ratio of the number of constraints in P and R is limited informationally. The constraint pattern, as well as which constraints are being imposed vis-à-vis the data, are better in-

TABLE 2

Constraint Pattern for
Intelligence Test Items*

ID	1	2	3	4	5	6	7	8
1	--	5	4	3	3	4	4	4
2	5	--	4	3	3	4	4	4
3	4	4	--	4	4	3	3	3
4	3	3	4	--	5	4	4	4
5	3	3	4	5	--	4	4	4
6	4	4	3	4	4	--	5	5
7	4	4	3	4	4	5	--	5
8	4	4	3	4	4	5	5	--

*See Table 1 for identification.

dicators of how much leeway there is for P to influence the solution [cf., Lingo, Note 3]). Using CMDA, one obtains the configuration in Figure 6 with $K = 0.15$. The alienation increment of the CMDA solution relative to the SSA-I result (0.15 vs. 0.03) is quite substantial considering that only eight variables are being represented. It may lead the researcher to think about additional facets of item content in order to explain the compression of the expected circular point order (cf. also Borg & Lingo, 1978, for some further analyses).

At this point, it seems necessary to add a few more comments on multiple R 's. What we have attempted in Table 2 was, in fact, to specify a constraint pattern for a circumplex. However, upon close investigation, the R matrix turns out to be slightly over-restrictive, i.e., it imposes more side constraints than are actually needed in this context. If one defines an ordinal circumplex by the property that the distances from each respective point to all other points form a function with only one bending point, then this implies an R that is row- and column-conditional at the same time. Such a constraint is actually easily formulated by using the same R matrix twice, i.e., column-conditionally as R_1 and row-conditionally as R_2 . The fact, however, that the facet design for the data in Table 1 implies a weak order only, requires slightly more exotic splittings: conditionalities are here to be set up over the item "blocks" $\{1, 2\}$, $\{3\}$, $\{4, 5\}$, and $\{6, 7, 8\}$. This could be achieved by using a separate R matrix for each super-row and super-column defined by the item blocks [Kruskal, Note 8], but CMDA thus far handles only two R matrices.

5. Discussion

The proposed approach for externally constraining the distances in an MDS configuration uses the so-called exterior penalty function method. There are, of course, a great variety of other procedures in nonlinear programming which could also be considered. One of the more popular alternatives is the gradient projection method, previously used by Skarabis [Note 7] in this context. The projection method, however, assumes that the projection point of ∇_p on the tangent plane to the feasible region is "close" to a point on the surface of the feasible region. If this is not so, the procedure becomes very complicated indeed [Luenberger, 1973]. Since one should not expect that the feasible set (defined by R) is always convex in MDS, the projection method does not seem to offer any advantages over the chosen penalty function approach. A more important reason for adopting the latter method is, however, that a feasible region might not even exist in some

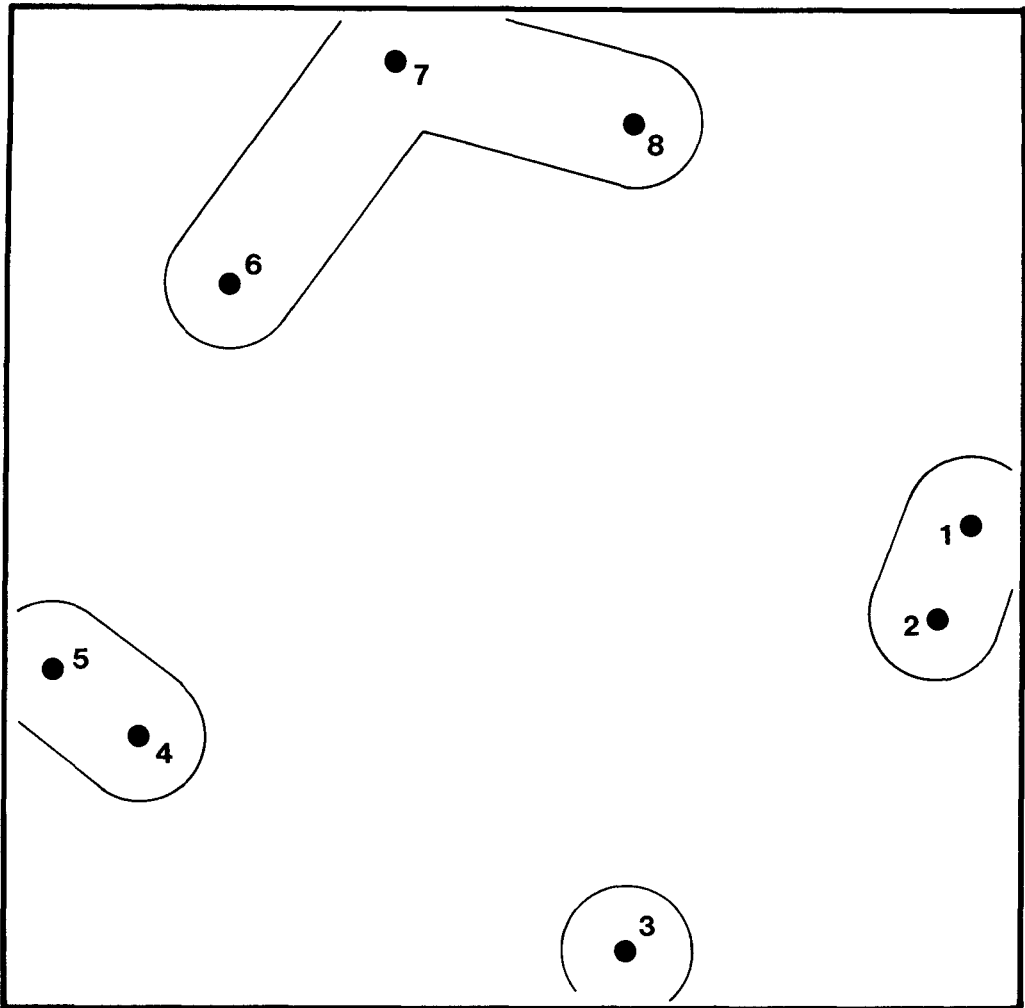


FIGURE 6
CMDA solution for data in Table 1.

applications, i.e., one may define R in such a way that the constraints cannot be perfectly satisfied in a given dimensionality, but the approximate imposition of structure seems sufficient for the purpose at hand or leads to a solution that is visually more accessible to interpretation. In such a case, it is completely unclear how a projection method could be constructed at all.

Although the penalty function method seems to have principal advantages in the considered MDS context, there are some problems associated with it that deserve further study. The most important and obvious question is related to the choice of the weights α and β . The starting values $\alpha^{(0)} = \beta^{(0)} = 0.5$ have been found not to be optimal in all cases. Rather, a higher β -weight right from the start is often better. As to the shifting of the relative weight to β in the s -process, it is apparent that a fast incrementing of β causes the trajectory of the gradient search to curve relatively fast toward the feasible region such that the final X satisfies all constraints but is not optimal in terms of L_p . One can test for such a solution, however, by restarting the procedure with new weights, say, $\alpha^{(0)} = \beta^{(0)} = 0.5$, etc., thus moving in "leaps" toward the optimal X , or, if the optimal X has been reached already, returning to this solution due to the effect of the increasing β .

Leaving these technical problems behind, one can nevertheless see from the illustrative applications of the CMDA method that one point made at the beginning of this paper is actually valid: data structures may often be represented by a variety of MDS configurations that are not "much" different in alienation. This allows one to pick a more meaningful solution rather than having to deal with some unconstrained and blindly descriptive representation which may suggest that certain substantive expectations are not compatible with the data. On the other hand, the constrained scaling method is not particularly powerful in discriminating among competing theories. In a number of empirical applications, we have found that the difference in alienation for various constraint conditions related to different hypotheses is often very "small". This, of course, raises the question of how to evaluate such differences, and, more generally, whether global fit coefficients are very useful in this context at all. We cannot offer any simple answers to these questions at this time, but it should be pointed out that studying point-alienation coefficients [Borg, 1978] or the matrix of "residuals" [Kruskal & Wish, 1978] can provide more detailed information as to the sources of alienation.

REFERENCE NOTES

1. Guttman, L. *Faceted SSA-I*. Unpublished Manuscript, Jerusalem: Israel Institute of Applied Social Research, 1976.
2. Kruskal, J. B., Young, F. W., & Seery, J. B. *How to use KYST-2, a very flexible program to do multidimensional scaling and unfolding* (Technical Report). Murray Hill, N.J.: Bell Laboratories, ca. 1978.
3. Lingo, J. C. *Testing regional hypotheses in multidimensional scaling*. Paper presented at the Second International Symposium on Data Analysis and Informatics. Versailles, France, Oct. 17-19, 1979.
4. Noma, E. & Johnson, J. *Constraining nonmetric multidimensional scaling configurations* (Technical Report No. 60). Ann Arbor: Human Performance Center, The University of Michigan, August 1977.
5. Noma, E. & Johnson, J. *Confirmatory multidimensional scaling*. Unpublished manuscript, Michigan Mathematical Psychology Department, 1977.
6. Shepard, R. N. & Crawford, G. *Multidimensional scaling based on fitting of constrained difference functions*. Paper presented at the U.S.-Japan Seminar: Theory, Methods, and Applications of Multidimensional Scaling and Related Techniques. La Jolla, Cal., 1975.
7. Skarabis, H. *Multidimensional scaling with prescribed structures I*. Unpublished manuscript, Freie Universität, Berlin, Germany, ca. 1976.
8. Kruskal, J. B. Personal communication, 1978.

REFERENCES

- Bentler, P. M. & Weeks, D. G. Restricted multidimensional scaling models. *Journal of Mathematical Psychology*, 1978, 17, 138-151.
- Bloxom, B. Constrained multidimensional scaling in N spaces. *Psychometrika*, 1978, 43, 397-408.
- Borg, I. PAL: point-wise alienation coefficients in multidimensional scaling. *Journal of Marketing Research*, 1978, 15, 478-479.
- Borg, I. & Lingo, J. C. Multidimensional scaling with side constraints on the distances. In J. C. Lingo, E. E. Roskam, & I. Borg (Eds.) *Geometric representations of relational data* (2nd ed.) Ann Arbor, Mich.: Mathesis Press, 1979.
- Carroll, J. D. Models and methods for multidimensional analysis of preferential choice (or other dominance) data. In E.-D. Lantermann & H. Feger (Eds.) *Similarity and choice*. Bern, Switzerland: Huber, 1979.
- Carroll, J. D., Pruzansky, S., & Kruskal, J. B. CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, 1980, 45, 3-24.
- Coombs, C. H. *A theory of data*. New York: Wiley, 1964 (reprinted by Mathesis Press: Ann Arbor, Mich., 1976).
- Daniel, C. & Wood, F. S. *Fitting equations to data*. New York: Wiley, 1971.
- de Leeuw, J. & Heiser, W. Convergence of correction-matrix algorithms for multidimensional scaling. In J. C. Lingo (Ed.) *Geometric representations of relational data*. Ann Arbor, Mich.: Mathesis Press, 1977.
- de Leeuw, J. & Heiser, W. Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (Ed.) *Multivariate analysis-V*. Amsterdam, The Netherlands: North Holland (in press).
- Eisler, H. & Roskam, E. E. Multidimensional similarity: An experimental and theoretical comparison between vector, distance, and set theoretical models. *Acta Psychologica*, 1977, 41, 1-46, 355-363.
- Ekman, G. Dimensions of color vision. *Journal of Psychology*, 1954, 38, 467-474.
- Guttman, L. A faceted definition of intelligence. *Scripta Hierosolymitana*, 1965, 14, 166-181.

- Guttman, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 1968, 33, 469-506.
- Krantz, D. & Tversky, A. Similarity of rectangles: An analysis of subjective dimensions. *Journal of Mathematical Psychology*, 1975, 12, 4-34.
- Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964, 29, 1-27.
- Kruskal, J. B. & Wish, M. *Multidimensional scaling*. Beverly Hills, Cal.: Sage Publications, 1978.
- Lingoes, J. C. & Borg, I. CMDA-U: Confirmatory monotone distance analysis-unconditional. *Journal of Marketing Research*, 1978, 15, 610-611.
- Lingoes, J. C. & Roskam, E. E. A mathematical and empirical study of two multidimensional scaling algorithms. *Psychometrika*, 1973, 38, Monograph Supplement.
- Luenberger, D. G. *Introduction to linear and nonlinear programming*. Reading, Mass.: Addison-Wesley, 1973.
- Roskam, E. E. A general system for nonmetric data analysis. In J. C. Lingoes (Ed.) *Geometric representations of relational data*. Ann Arbor, Mich.: Mathesis Press, 1977.
- Schönemann, P. H. Similarity of rectangles. *Journal of Mathematical Psychology*, 1977, 16, 161-165.
- Shepard, R. N. Representation of structure in similarity data: Problems and prospects. *Psychometrika*, 1974, 39, 373-421.
- Shepard, R. N. The circumplex and related topological manifolds in the study of perception. In S. Shye (Ed.) *Theory construction and data analysis in the behavioral sciences*. San Francisco: Jossey-Bass, 1978.
- Skarabis, H. Multidimensional scaling with prescribed structures. In H. Skarabis & P. Sint (Eds.) *Compstat lectures 1*. Wien, Austria: Physika Verlag, 1978.

Manuscript received 8/28/78

First revision received 4/2/79

Final version received 8/9/79