PSYCHOMETRIKA—VOL. 46, NO. 4.
DECEMBER, 1981

NONMETRIC MAXIMUM LIKELIHOOD MULTIDIMENSIONAL SCALING FROM DIRECTIONAL RANKINGS OF SIMILARITIES

YOSHIO TAKANE

MCGILL UNIVERSITY

and

J. DOUGLAS CARROLL

BELL LABORATORIES

A maximum likelihood estimation procedure is developed for multidimensional scaling when (dis)similarity measures are taken by ranking procedures such as the method of conditional rank orders or the method of triadic combinations. The central feature of these procedures may be termed directionality of ranking processes. That is, rank orderings are performed in a prescribed order by successive first choices. Those data have conventionally been analyzed by Shepard-Kruskal type of nonmetric multidimensional scaling procedures. We propose, as a more appropriate alternative, a maximum likelihood method specifically designed for this type of data. A broader perspective on the present approach is given, which encompasses a wide variety of experimental methods for collecting dissimilarity data including pair comparison methods (such as the method of tetrads) and the pick-M method of similarities. An example is given to illustrate various advantages of nonmetric maximum likelihood multidimensional scaling as a statistical method. At the moment the approach is limited to the case of one-mode two-way proximity data, but could be extended in a relatively straightforward way to two-mode two-way, two-mode three-way or even three-mode three-way data, under the assumption of such models as INDSCAL or the two or three-way unfolding models.

Key words: rank orders, model selection, AIC, face perception.

Introduction

In ranking procedures such as the method of triadic combinations [Richardson, 1938] or the method of conditional rank orders [Young, 1975] the rank order judgments may be characterized as *directional.* For example, in the method of triadic combinations stimuli are presented in triads. The subject is asked to decide which two of the three are most alike, and then which two are most different. Furthermore, a stimulus pair whose members are neither most alike nor most different is deduced from the previous two judgments to obtain a single ranking of three dissimilarities defined on a triad of stimuli. Thus, the ranking is conditional on the triad of stimuli, and it is directional since it is performed in a specific order.

In the method of conditional rank orders one of n stimuli is designated as a standard stimulus. The subject is asked to choose the most similar stimulus to the standard among $n - 1$ stimuli, and then after this stimulus is eliminated, to choose the one which is now

The first author's work was supported partly by the Natural Sciences and Engineering Research Council of Canada, grant number A6394. Portions of this research were done while the first author was at Bell Laboratories. MAXSCAL-4.1, a program to perform the computations described in this paper can be obtained by writing to: Computing Information Service, Attention: Ms. Carole Scheiderman, Bell Laboratories, 600 Mountain Ave., Murray Hill, N.J. 07974. Thanks are due to Yukio Inukai, who generously let us use his stimuli in our experiment, and to Jim Ramsay for his helpful comments on an earlier draft of this paper. Confidence regions in Figures 2 and 3 were drawn by the program written by Jim Ramsay. We are also indebted to anonymous reviewers for their suggestions.

Requests for reprints should be sent to Yoshio Takane, Department of Psychology, McGill University, 1205 Docteur Penfield Ave., Montreal, Quebec H3A, 1B1, Canada.

0033-3123/81/1200-3044$00.75/0

most similar to the standard among the $n-2$ remaining stimuli, proceeding in this way until all $n-1$ stimuli are completely rank ordered relative to the standard. The whole process is repeated with a different stimulus used as a standard stimulus in turn until all n stimuli serve as a standard. Again the ranking process is directional (being performed from the smallest to the largest dissimilarity) and conditional upon the standard stimulus.

Dissimilarity data arising from ranking procedures have been analyzed either by Torgerson's [1952] classical multidimensional scaling (MDS) or by the type of nonmetric MDS originally introduced by Shepard and Kruskal [Shepard, 1962; Kruskal, 1964a, b; McGee, 1966; Guttman, 1968; Roskam, 1970; Young, 1975]. In this paper we propose, as a more appropriate alternative, a maximum likelihood estimation (MLE) method specifically designed for such directional rank order data. The principal advantage of the maximum likelihood method over other fitting procedures is that it permits various (asymptotically valid) statistical inferences, provided that the distributional assumptions are correct. For maximum likelihood MDS procedures for other types of dissimilarity data, see Ramsay [1977, 1978, 1980b] and Takane [1978a, b, 1981].

The Method

Our general approach in this paper may be called a "parametric approach" to nonmetric scaling. In this approach nonmetric data are viewed as incomplete data [Dempster, Laird & Rubin, 1977] conveying only ordinal information about distances. An unobserved metric process conveying complete information about distances is assumed to underlie the nonmetric data generation process, and a specific information reduction mechanism is postulated between the two kinds of processes. The likelihood function is specified for observed nonmetric data, which are related to distances based on some parametric assumptions about the underlying metric process.

The Likelihood Function

Let us assume that a set of n stimuli have an A-dimensional representation in the euclidean space. That is, d_{ij}, the distance between stimuli i and j, is given by

$$d_{ij} = \left\{ \sum_{a=1}^{A} (x_{ia} - x_{ja})^2 \right\}^{1/2}, \tag{1}$$

where x_{ia} is the coordinate of stimulus i on dimension a. It is straightforward to extend the current approach to the INDSCAL model [Carroll & Chang, 1970]. However, in this paper we deliberately restrict our attention to the simple euclidean model defined above. Let us further assume that d_{ij} as defined above is error-perturbed, and generates an error-perturbed metric process, $\lambda_{ijk}^{(t)}$, for replication k and at occasion t; i.e.,

$$\lambda_{ijk}^{(t)} = \varepsilon(d_{ij}, e_{ijk}^{(t)}), \tag{2}$$

where $e_{ijk}^{(t)}$ is an error random variable and where ε is a joint function (to be made more explicit) of the distance and the error component. We consider the following two error models in this paper:

$$\lambda_{ijk}^{(t)} = d_{ij} + e_{ijk}^{(t)} \qquad \text{(Additive error model)} \tag{3}$$

where $e_{ijk}^{(t)} \sim N(0, \sigma_k^2/2)$, and

$$\lambda_{ijk}^{(t)} = d_{ij}\, e_{ijk}^{(t)} \qquad \text{(Multiplicative error model)} \tag{4}$$

where $\ln e_{ijk}^{(t)} \sim N(0, \sigma_k 2/2)$. The replication may be taken over a sample of individuals (i.e., replication k is individual k). The σ_k^2 allows for possible individual differences in dispersion.

We may optionally set $\sigma_k^2 = \sigma^2$ for all k. These are basically the principal error models considered by Ramsay [1977] in his maximum likelihood approach to metric MDS. We refer the reader to Ramsay's paper for a discussion of the psychological and empirical rationale for these particular models.

Our problem is to specify the likelihood of an observed set of ranked dissimilarities given a set of distances which are error-perturbed in a specific way. A model of psychological processes by which complete metric information is "collapsed" into incomplete ordinal information must be an intrinsic part of the likelihood function. The problem of parameter estimation is then to find a set of distances with the structure defined in (1) such that they maximize the likelihood of the particular observed data.

For explanatory purposes we only discuss the specification of the likelihood function for the method of conditional rank orders in some detail. (The extension to the method of triadic combinations is straightforward.) As noted earlier, the central feature of this method is the directionality of the ranking processes. Given that the ranking process is directional, a rank order may be regarded as resulting from successive first choices. The likelihood of a ranking is then defined by the product of the likelihoods of successive first choices, assuming that each successive first choice is made statistically independently of the others. (This assumption may seem unrealistically strong at first glance. However, as will be shown later, it follows from a much weaker assumption.) The joint likelihood of several rankings arising from different conditionalities may, in turn, be defined by the product of the likelihoods of those rankings. It is thus sufficient to specify the likelihood of a first choice.

Let $d_{i(mk)}$ represent the distance between standard stimulus i and whatever the stimulus judged m^{th} most similar to the standard in replication k. (If that stimulus is j, then $d_{i(mk)} = d_{ij}$.) Let $\lambda^{(t)}_{i(mk)}$ denote the random metric process corresponding to $d_{i(mk)}$ generated at occasion t. The occasion in this case refers to each successive first choice. We assume that the probability, $p^{(1)}_{ik}$, that the dissimilarity corresponding to $d_{i(1k)}$ is judged smallest among $n-1$ dissimilarities is equal to the probability that the random metric process $\lambda^{(1)}_{i(1k)}$ corresponding to $d_{i(1k)}$ turns out to be the smallest of all $n-1$ random processes at occasion 1. That is,

$$p^{(1)}_{ik} = \Pr(\lambda^{(1)}_{i(1k)} < \lambda^{(1)}_{i(2k)}, \ldots, \lambda^{(1)}_{i(1k)} < \lambda^{(1)}_{i(n-1,k)}). \tag{5}$$

Define

$$\mathbf{B}^{(1)} = \begin{bmatrix} 1 & \overbrace{-1 \quad 0 \quad \cdots}^{n-1} & 0 \\ 1 & 0 \quad -1 \quad \cdots & 0 \\ \vdots & \vdots \quad \vdots & \vdots \\ 1 & 0 \quad 0 \quad \cdots & -1 \end{bmatrix} \Big\} n-2$$

and

$$\boldsymbol{\lambda}^{(1)}_{ik} = \begin{bmatrix} \lambda^{(1)}_{i(1k)} \\ \vdots \\ \lambda^{(1)}_{i(n-1,k)} \end{bmatrix} \quad \text{and} \quad \mathbf{d}^{(1)}_{ik} = \begin{bmatrix} d_{i(1k)} \\ \vdots \\ d_{i(n-1,k)} \end{bmatrix}.$$

Then under the additive error assumption $p^{(1)}_{ik}$ can be more explicitly written as

$$p^{(1)}_{ik} = \int_{R_{ik}^{(1)}} f(\mathbf{z}^{(1)}_{ik})\, d\mathbf{z}^{(1)}_{ik}, \tag{6}$$

where

$$\underline{\mathbf{z}_{ik}^{(1)}} = \mathbf{B}^{(1)}\boldsymbol{\lambda}_{ik}^{(1)} \sim N(\mathbf{B}^{(1)}\mathbf{d}_{ik}^{(1)}, \mathbf{B}^{(1)}\boldsymbol{\Sigma}_{ik}^{(1)}\mathbf{B}^{(1)\prime}), \tag{7}$$

and where $R_{ik}^{(1)}$ is a multidimensional region such that $\mathbf{B}^{(1)}\boldsymbol{\lambda}_{ik}^{(1)} \leq \mathbf{0}$. The $\boldsymbol{\Sigma}_{ik}^{(1)}$ in (7) is the covariance of $\boldsymbol{\lambda}_{ik}^{(1)}$. In case of the multiplicative error model (4), the elements of $\boldsymbol{\lambda}_{ik}^{(1)}$ and $\mathbf{d}_{ik}^{(1)}$ should be redefined as $\ln \lambda_{i(mk)}^{(1)}$ and $\ln d_{i(mk)}$, respectively.

If the elements of $\boldsymbol{\lambda}_{ik}^{(1)}$ are independently normally distributed with variance $\sigma_k^2/2$; i.e.,

$$\boldsymbol{\Sigma}_{ik}^{(1)} = \left(\frac{\sigma_k^2}{2}\right)\mathbf{I}, \tag{8}$$

then the covariance of $\mathbf{z}_{ik}^{(1)}$ is simplified into

$$\mathbf{B}^{(1)}\boldsymbol{\Sigma}_{ik}^{(1)}\mathbf{B}^{(1)\prime} = \sigma_k^2 \begin{bmatrix} 1 & & \frac{1}{2} \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ \frac{1}{2} & & 1 \end{bmatrix}. \tag{9}$$

The important consequence of (9) is that $p_{ik}^{(1)}$ in (6) can be well approximated by the multivariate logistic distribution [Bock, 1975; pp. 521–522],

$$p_{ik}^{(1)} \cong \left[1 + \sum_{q=1}^{n-2} \exp(c_k \mathbf{b}_q^{(1)\prime}\mathbf{d}_{ik}^{(1)})\right]^{-1}, \tag{10}$$

where $\mathbf{b}_q^{(1)\prime}$ is the q^{th} row vector of $\mathbf{B}^{(1)}$ (the prime indicates a row vector), and where c_k is a dispersion parameter which is approximately $\pi/(3^{1/2}\sigma_k)$. This distribution is much easier to evaluate than the multivariate normal integral required in (6). The $p_{ik}^{(1)}$ in (10) is also interesting in its own sake, since in the case of the multiplicative error model it is equivalent to Luce's model for the first choice [Luce, 1959]. (See Appendix.)

The independence assumption in (8), however, may be restrictive in practical situations. Fortunately, we can relax (8) into

$$\underline{\boldsymbol{\Sigma}_{ik}^{(1)}} = \left(\frac{\sigma_k^2}{2}\right)\mathbf{I} + \mathbf{h}_{ik}^{(1)}\mathbf{1}' + \mathbf{1}\mathbf{h}_{ik}^{(1)\prime}, \tag{11}$$

where $\mathbf{h}_{ik}^{(1)}$ is an arbitrary vector and $\mathbf{1}$ is a vector of ones [Takane, 1978a]. The same covariance matrix given in (9) can be derived from this weaker assumption as well. This can be easily seen by pointing out that $\mathbf{B}^{(1)}\mathbf{1} = \mathbf{0}$ so that the second and the third terms on the right hand side of (11) vanish when $\boldsymbol{\Sigma}_{ik}^{(1)}$ is pre- and postmultiplied by $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(1)\prime}$, respectively. The fact that (11) is sufficient to obtain (9) has considerable import, since the former is much weaker than the latter, and is consequently more realistic in practical situations. Note that (8) follows as a special case of (11) by setting $\mathbf{h}_{ik}^{(1)} = \mathbf{0}$. Note also that an equal covariance case also follows from (11) by setting $\mathbf{h}_{ik}^{(1)}$ to be a constant vector. Essentially the same condition as (11) has been postulated by Huynh and Feldt [1970] as the minimal condition to be satisfied in a univariate repeated measures analysis of variance. Indeed, all mutually orthogonal contrasts are statistically independent under this assumption.

In order to obtain $p_{ik}^{(2)}$ (the probability that the stimulus corresponding to $d_{i(2k)}$ is judged the next most similar stimulus to i), we define $\mathbf{B}^{(2)}$ by eliminating the last row and the last column of $\mathbf{B}^{(1)}$, $\mathbf{d}_{ik}^{(2)}$ by eliminating the first element of $\mathbf{d}_{ik}^{(1)}$ and $\boldsymbol{\lambda}_{ik}^{(2)\prime} = (\lambda_{i(2k)}^{(2)} \ldots, \lambda_{i(n-1,k)}^{(2)})$. We may then express the $p_{ik}^{(2)}$ in a form analogous to (6) and (7). Note that it is

assumed that an entirely new set of random processes $\lambda_{i(mk)}^{(2)}$ $(m = 2, \ldots, n-1)$ are generated for this judgment (as indicated by the new occasion index). In general, we have

$$p_{ik}^{(m)} \cong \left[1 + \sum_{q=1}^{n-m-1} \exp(c_k \mathbf{b}_q^{(m)\prime} \mathbf{d}_{ik}^{(m)})\right]^{-1} \tag{12}$$

for the m^{th} first choice after the $m-1$ most similar stimuli are eliminated from the comparison set. Then $\mathbf{b}_q^{(m)\prime}$ is the q^{th} row vector of $\mathbf{B}^{(m)}$, which is obtained by deleting the last $m-1$ rows and columns of $\mathbf{B}^{(1)}$. For $m = n-1$ we have $p_{ik}^{(n-1)} = 1$.

We may then state p_{ik}, the probability of a complete ranking of $n-1$ stimuli with respect to stimulus i, as

$$p_{ik} = \prod_{m=1}^{n-1} p_{ik}^{(m)}. \tag{13}$$

This assumes the statistical independence of $p_{ik}^{(m)}$ and $p_{ik}^{(m')}$ $(m \neq m')$. The statistical independence, of course, can be obtained, if all elements in $\boldsymbol{\lambda}_{ik}^{(m)}$ and those in $\boldsymbol{\lambda}_{ik}^{(m')}$ are statistically independent. However, it can be deduced from a much weaker assumption, namely

$$\boldsymbol{\Sigma}_{ik}^{(mm')} = \mathbf{1}\mathbf{u}' + \mathbf{v}\mathbf{1}', \tag{14}$$

where $\boldsymbol{\Sigma}_{ik}^{(mm')}$ is the covariance matrix between $\boldsymbol{\lambda}_{ik}^{(m)}$ and $\boldsymbol{\lambda}_{ik}^{(m')}$, $\mathbf{u}$ and $\mathbf{v}$ are arbitrary vectors, but with appropriate numbers of components. The $\mathbf{1}$ is a vector of unities, but the two $\mathbf{1}$'s in (14) are different in dimensionality. Under the assumption of (14), $\mathbf{B}^{(m)}\boldsymbol{\Sigma}_{ik}^{(mm')}\mathbf{B}^{(m')\prime}$ vanishes, so that $\mathbf{z}_{ik}^{(m)} = \mathbf{B}^{(m)}\boldsymbol{\lambda}_{ik}^{(m)}$ and $\mathbf{z}_{ik}^{(m')} = \mathbf{B}^{(m')}\boldsymbol{\lambda}_{ik}^{(m')}$ are independent of each other, and consequently so are the $p_{ik}^{(m)}$ and $p_{ik}^{(m')}$.

The joint likelihood of several conditional rankings is now stated as

$$L_k = \prod_{i=1}^{n} p_{ik}, \tag{15}$$

and the joint likelihood for the entire set of observations as

$$L = \prod_{k} L_k. \tag{16}$$

Again the statistical independence between $p_{ik}^{(m)}$ and $p_{i'k'}^{(m')}$ $(i \neq i')$ can be deduced, if the covariance matrix between $\boldsymbol{\lambda}_{ik}^{(m)}$ and $\boldsymbol{\lambda}_{i'k}^{(m')}$ has a similar structure to (14). On the other hand, $\boldsymbol{\lambda}_{ik}^{(m)}$ and $\boldsymbol{\lambda}_{i'k'}^{(m')}$ $(k \neq k')$ may be assumed independent, if replications are taken over different individuals.

The (log of) L can be maximized by various numerical techniques. MAXSCAL-4.1, the computer program which performs the necessary computations, uses Fisher's scoring algorithm for maximization (see Rao, 1952, for example). A detailed description of this algorithm as applied to similar situations, as well as of its relation to the Gauss-Newton method for weighted least squares problems, can be found in Takane [1978a].

The necessary derivatives for this optimization are given as follows: For the additive error model we may write

$$p_{ik}^{(m)} = \frac{\exp(s_k d_{i(mk)})}{u_{ik}^{(m)}}, \tag{17}$$

where $s_k = -c_k$, and $u_{ik}^{(m)} = \sum_{j=m}^{n-1} \exp(s_k d_{i(jk)})$. Then

$$\frac{\partial \ln p_{ik}^{(m)}}{\partial x_{qa}} = s_k \frac{\partial d_{i(mk)}}{\partial x_{qa}} - \frac{1}{u_{ik}^{(m)}} \sum_{j=m}^{n-1} \frac{\partial \exp(s_k d_{i(jk)})}{\partial x_{qa}}. \tag{18}$$

and

$$\frac{\partial \ln p_{ik}^{(m)}}{\partial s_k} = d_{i(mk)} - \frac{1}{u_{ik}^{(m)}} \sum_{j=m}^{n-1} d_{i(jk)} \exp(s_k d_{i(jk)}), \tag{19}$$

where

$$\frac{\partial \exp(s_k d_{i(jk)})}{\partial x_{qa}} = s_k \exp(s_k d_{i(jk)}) \frac{\partial d_{i(jk)}}{\partial x_{qa}} \tag{20}$$

and

$$\frac{\partial d_{i(jk)}}{\partial x_{qa}} = \frac{(\delta_{iq} - \delta_{(jk)q})(x_{ia} - x_{(jk)a})}{d_{i(jk)}}. \tag{21}$$

In the above formula δ .. is the Kronecker delta (i.e., $\delta_{iq} = 1$ when $i = q$, and $\delta_{iq} = 0$ when $i \neq q$. Parenthesized subscript (jk) in $\delta_{(jk)q}$ should be replaced by a single stimulus index. If it is i, then $\delta_{(jk)q} = \delta_{iq}$, and δ_{iq} is as defined above.) For the multiplicative error model we have

$$p_{ik}^{(m)} = \frac{(d_{i(mk)})^{s_k}}{v_{ik}^{(m)}}, \tag{22}$$

where $v_{ik}^{(m)} = \sum_{j=m}^{n-1} (d_{i(jk)})^{s_k}$. We obtain

$$\frac{\partial \ln p_{ik}^{(m)}}{\partial x_{qa}} = \frac{s_k}{d_{i(mk)}} \frac{\partial d_{i(mk)}}{\partial x_{qa}} - \frac{1}{v_{ik}^{(m)}} \sum_{j=m}^{n-1} \left[\frac{\partial (d_{i(jk)})^{s_k}}{\partial x_{qa}} \right] \tag{23}$$

and

$$\frac{\partial \ln p_{ik}^{(m)}}{\partial s_k} = \ln d_{i(mk)} - \frac{1}{v_{ik}^{(m)}} \sum_{j=m}^{n-1} (\ln d_{i(jk)})(d_{i(jk)})^{s_k} \tag{24}$$

where

$$\frac{\partial (d_{i(jk)})^{s_k}}{\partial x_{qa}} = s_k (d_{i(jk)})^{(s_k - 1)} \frac{\partial d_{i(jk)}}{\partial x_{qa}}. \tag{25}$$

The expression for $\partial d_{i(jk)}/\partial x_{qa}$ is given in (21).

Little modification is necessary to extend the above formulation to the method of triadic combinations. In this method conditional rank orders are taken on dissimilarities defined on triads of stimuli rather than on dissimilarities with a common standard stimulus. Also, each conditional ranking is performed in the order of the smallest, then the largest (and then the intermediate) rather than from the smallest to the largest dissimilarity (as in the method of conditional rank orders). However, these differences are minor, and can be easily accomodated within the present framework.

Missing Data

Each conditional ranking does not have to be complete in the method of conditional rank orders, and there are two possible cases. The major distinction lies in whether or not an experimenter has experimental control over missing data elements.

In one case the experimenter presents all $n - 1$ comparison stimuli, but obtains only n^* first choices. In this case we may simply take the product of the likelihoods of n^* successive first choices, while the likelihood of each successive first choice remains the same

as before. That is,

$$p_{ik} = \prod_{m=1}^{n^*} p_{ik}^{(m)} \tag{26}$$

where $p_{ik}^{(m)}$ is given in (12).

On the other hand, the experimenter may restrict the comparison stimuli to a subset of $n-1$ stimuli, and obtain a rank ordering among only the restricted set of comparison stimuli. Let n^* be the number of comparison stimuli. We define the probability of the m^{th} first choice ($m < n^*$) in a manner analogous to (12), but excluding those terms not related to n^* comparison stimuli. That is,

$$p_{ik}^{(m)} = \left[1 + \sum_{q=1}^{n^*-m} \exp(c_k \mathbf{b}_q^{(m)\prime} \mathbf{d}_{ik}^{(m)}) \right]^{-1}. \tag{27}$$

The likelihood of a ranking is defined as in (26). Note that this is equivalent to redefining the conditionality of a ranking (the set of dissimilarities on which a ranking is obtained).

Tied Observations

Since a continuous distribution is assumed on distances, tied ranks should not occur theoretically, and it may be possible, depending on the experimental procedure we use, to completely avoid tied ranks. Nevertheless it is desirable for a nonmetric multidimensional scaling procedure to have some provisions for treatment of tied observations, since they may occur frequently in practice.

There are two cases to be distinguished, one we call a weak tie and the other we call a strong tie. Two dissimilarities are said to be in a weak tie when they are not explicitly compared and consequently not rank ordered. For example, an experimental procedure may require the subject to choose the n^* most similar stimuli to a standard, but not necessarily to obtain a rank ordering among the n^* picked stimuli. Then the n^* dissimilarities (corresponding to the n^* picked stimuli) are "tied" in this sense. A strong tie, on the other hand, is characterized by empirical indistinguishability. That is, two dissimilarities are explicitly compared, but the subject is unable to rank order them.

The two kinds of ties should be treated differently, analogous to Kruskal's [1964b] primary and secondary approaches to ties. In the primary approach no constraints are imposed among distances corresponding to tied observations except that they should satisfy, as much as possible, the order restrictions imposed by the larger and the smaller dissimilarities than themselves. This approach is suitable for the treatment of weak ties. In the secondary approach, on the other hand, distances corresponding to tied observations are required to be as close to each other as possible. This treatment of ties is suitable for the treatment of strong ties, since dissimilarities corresponding to strong ties are deemed empirically indistinguishable.

Suppose dissimilarities corresponding to $d^{(m)}$ and $d^{(m+1)}$ are tied. Then the primary approach defines $p^{(m)}$ (the probability of the m^{th} first choice in a certain ranking) excluding the term related to $d^{(m+1)}$, and $p^{(m+1)}$ excluding the term related to $d^{(m)}$. That is, $d^{(m+1)}$ does not have any effects on $p^{(m)}$, and $d^{(m)}$, in turn, does not have any effects on $p^{(m+1)}$. The secondary approach, on the other hand, defines both $p^{(m)}$ and $p^{(m+1)}$ including both terms related to $d^{(m)}$ and $d^{(m+1)}$. In this case $d^{(m)}$ and $d^{(m+1)}$ are bound to be close to each other, if $p^{(m)} \times p^{(m+1)}$ is to be maximized.

In general, conditionalities of rankings can be of any form. The method of conditional rank orders and the method of triadic combinations are but two special cases of such

instances. Dissimilarities in different conditionalities are not explicitly compared, and thus considered to be in weak tie relations. The primary approach to weak ties described above is consistent with our previous treatment of conditionalities.

Generalizations

We may generalize our basic results on directional ranking procedures to other experimental procedures, including such procedures as the method of tetrads and the pick-M method of similarities. In the method of tetrads four stimuli are presented in two pairs, and the subject is asked to judge which pair of stimuli is more similar (or dissimilar). In the pick-M method of similarities, the subject is required to pick out the M most similar stimuli to a standard stimulus from a well defined set of alternative stimuli. In what follows we show that these procedures are in fact two special cases of directional ranking procedures. We have kept our exposition as succinct as possible, so that nonmathematical readers may skip this entire section without loss of continuity.

Let S be the set of dissimilarities defined on a set of pairs of stimuli. We define three kinds of relations on $S \times S$, order relation, strong tie relation and weak tie relation. We assume that for each pair of dissimilarities one of the three relations hold; i.e., for $a, b \in S$,

i. $a > b$ or $a < b$ (order relation)
ii. $a \sim b$ (strong tie), or
iii. $a \leftrightarrow b$ (weak tie).

Let S_i and S_j be subsets of S. We write $S_i > S_j$ ($S_i < S_j$), $S_i \sim S_j$ or $S_i \leftrightarrow S_j$, when $\forall a \in S_i$ and $\forall b \in S_j$, $a > b$ ($a < b$), $a \sim b$ or $a \leftrightarrow b$, respectively. The S_i and S_j may not be disjoint.

Define the set of separate conditionalities Q whose i^{th} member is denoted by $S_i \subset S$. $\forall S_i, \forall S_j \in Q$ $(i \neq j)$, $S_i \leftrightarrow S_j$. Let the number of elements in S_i be L. The procedure which requires the subject to rank order M elements ($M \leq L$) in S_i partitions S_i into subsets $S_i^{(M)}$ and $S_i^{(L-M)}$, where $S_i^{(L-M)}$ is possibly null (when $M = L$). The $S_i^{(M)}$ is the set of the M smallest dissimilarities in S_i, while $S_i^{(L-M)}$ is the set of $L - M$ largest dissimilarities in S_i. Obviously $S_i^{(M)} < S_i^{(L-M)}$. Either the order or strong tie relations are defined among the elements of $S_i^{(M)}$, whereas the weak tie relations are defined among the elements of $S_i^{(L-M)}$.

With appropriate choice of the member of S_i we have the method of tetrads (including the method of triads as a special case) when $L = M = 2$, the method of triadic combinations when $L = M = 3$, and the method of conditional rank orders when $L = M = n - 1$ (where n is the number of stimuli). The two incomplete (missing data) cases of the method of conditional rank orders can be obtained by setting $L < M = n - 1$ and $L = M < n - 1$. (Of course, we may set $L < M < n - 1$ to generate yet another variation of the method of conditional rank orders.)

With minor modification the above formulation can be extended to the pick-M method of similarities. This method also requires the subject to partition the S_i into two disjoint subsets, $S_i^{(M)}$ and $S_i^{(L-M)}$. However, this time $S_i^{(L-M)}$ cannot be null (i.e., $M < L$). We have $S_i^{(M)} < S_i^{(L-M)}$ as before. However, the weak tie relation is defined among the elements of $S_i^{(M)}$ as well as among the elements of $S_i^{(L-M)}$.

A whole variety of experimental procedures, other than those already mentioned, can be generated by specifying different Q's. MAXSCAL-4.1 is capable of performing MDS for any definitions of Q, provided that each S_i is at most matrix conditional (i.e., dissimilarities are not compared across different replications). We call this general case where Q can be arbitrary the method of general directional rank orders.

Note, however, that the above discussion is exclusively based on the algebraic structures of data obtained by the various data collection methods. Little statistical consider-

ations are taken into account, which may cause some problem in justifying statistical assumptions underlying the methods. For example, we must assume statistical independence of weak ties, which may or may not be valid empirically.

Results and Discussion

As has been alluded to earlier, maximum likelihood estimation offers various advantages over other fitting procedures. The asymptotic chi square goodness of fit statistic can be readily derived from the general principle of likelihood ratio [Wilks, 1962]. In addition the AIC statistic [Akaike, 1974] can be used for identifying a best fitting model, where the asymptotic chi square test is not feasible. Furthermore, the knowledge of the asymptotic behavior of maximum likelihood estimators allows us to draw asymptotic confidence regions (of a prescribed size) surrounding estimated stimulus points to indicate the degree of precision of the estimation. In this section we demonstrate these features of maximum likelihood estimation in the specific context of multidimensional scaling.

Data

A small experiment was conducted to collect relevant data. Seventeen schematic faces (Figure 1) were used in this experiment. Those stimuli were originally used by Inukai, Nakamura and Shinohara [Note 1] in their study to identity determinants of face perception. They were constructed by combining two of the most important determinants of facial expressions, namely the curvature of lips and the curvature of eyes. The stimuli, 4 cm in height and 3 cm in width, were each pasted on a 7.5 cm × 12.5 cm index card. Dissimilarity judgments were obtained from five subjects (two male and three female university students) by the method of conditional rank orders. The order in which each stimulus served as a standard stimulus was randomized over subjects. An average subject took 70 minutes to complete the task.

Identifying the Best fitting Model

In order to choose the best fitting model the data were analyzed under all combinations of two error assumptions (the additive and the multiplicative error models), two conditions about error variance (constant or variable over subjects) and two dimensionalities (two and three dimensions). In certain cases four and five dimensional solutions were also obtained. The results are summarized in Table 1. Each cell of the table contains three numbers; the top one is the log likelihood of the model, and the one in the middle is the value of the AIC statistic [Akaike, 1974], which is defined by

$$\text{AIC} = -2 \ln L + 2 \text{ d.f.},$$

where the d.f. is the effective number of parameters in the model given at the bottom of the cell (enclosed in parentheses). Note that the d.f. used here are for the degrees of freedom of the model, *not* for the error degrees of freedom. An important feature of AIC is that it explicitly takes into account the number of model parameters in evaluating the goodness of fit of a model used to describe the data. The AIC is a badness-of-fit index, so that the smaller value indicates a better fit. A relatively nontechnical discussion on this statistic may be found in Takane [1981].

The AIC statistic may be used when models to be compared are not necessarily hierarchically structured. This feature is very convenient when we compare, for example, the goodness of fit of the two error models since the usual asymptotic chi square test is not feasible for such comparisons. The d.f. of the model is calculated by $NA - A(A + 1)/2 + r$, where r is the number of estimated dispersion parameters. This r is zero when the dispersions are assumed constant over subjects, and is equal to $N - 1$ when they are assumed

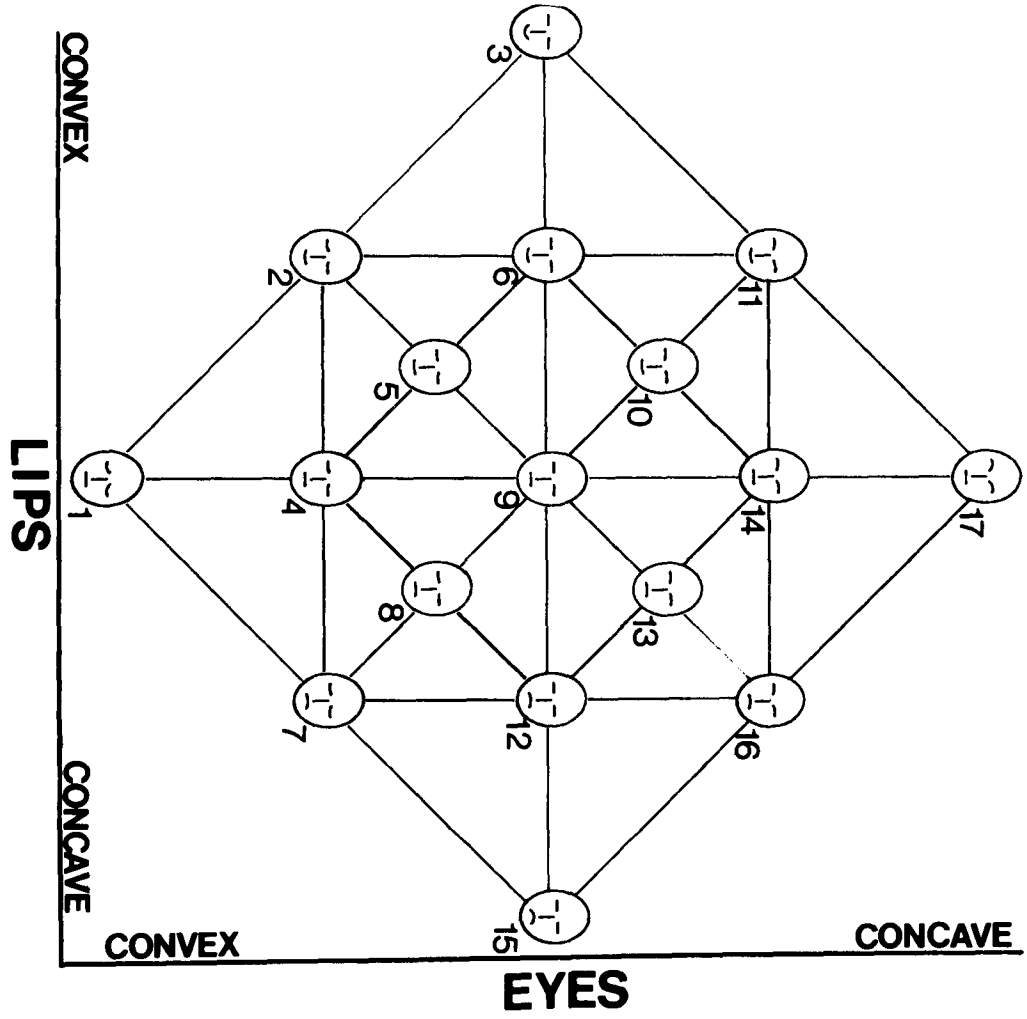

FIGURE 1
The stimulus configuration in terms of the curvatures of lips and eyes (From Inukai, et al., Note 1).

to vary over subjects. This is so, because there is a trade off between the size of stimulus configuration and the dispersion parameter, and consequently we can either fix the size of configuration or fix one of the dispersion parameters.

The AIC is unfortunately based on asymptotic properties of maximum likelihood estimators (as is the asymptotic chi square). This means that, strictly speaking, it is only applicable to large-sample data. Although in all cases we shall deal with in this paper we have at least ten times as many observations as there are parameters to be estimated, we are not entirely sure whether it is sufficient. Furthermore, with models in which the number of parameters increases with increasing number of replications (e.g., the model which allows for individual differences in dispersion), data are never sufficiently large to warrant the asymptotic properties of MLE. Consequently, we have to rely on some heuristic rule in the following discussion. The work [Takane & Carroll, Note 2] is in process, however, to develop a precise modification rule to the basic formula of AIC when the asymptotic properties of MLE may not be exploited.

Choosing a model with a minimum AIC value, when models are in fact hierarchically

Table 1

Summary of the results of MAXSCAL-4.1

analyses of the face data

		Unconstrained solutions dimensionality				Constrained solutions dimensionality	
		5	4	3	2	4	2
Additive error							
Constant	ln(L)			-1737.9	-1810.6		
variance	AIC			3565.8	3683.1		
	d.f.			(45)	(31)		
Nonconstant	ln(L)	-1668.6	-1680.9	-1711.5	-1784.8	-1755.5	-1847.5
variance	AIC	3485.2	3485.8	3521.1	3639.6	3554.9	3727.0
	d.f.	(74)	(62)	(49)	(35)	(22)	(16)
Multiplicative error							
Constant	ln(L)			-1748.9	-1817.6		
variance	AIC			3587.9	3697.2		
	d.f.			(45)	(31)		
Nonconstant	ln(L)			-1720.1	-1797.6		
variance	AIC			3538.2	3665.3		
	d.f.			(49)	(35)		

structured, is equivalent to choosing a more restricted model, whenever the asymptotic chi square is smaller than 2 × d.f. (where the d.f. is now the difference in the numbers of parameters between two models being compared), and to choosing a less restricted model otherwise. Ramsay [1980a] has shown that the asymptotic chi square tends to be inflated for small samples, but that there is a remarkable regularity in the way it is inflated. Noting this fact he has developed a simple correction formula for the asymptotic chi square, which is obtained by just multiplying some constant to the chi square criterion value. This constant should represent how the asymptotic chi square tends to be inflated under specific circumstances. Interestingly, this constant never seems to exceed 3. Furthermore, it reduces to approximately 1.5, when there are 15 stimuli and five complete replications. We may exploit this fact to construct a tentative rule of thumb. That is, we favor a less restricted

model whenever the asymptotic chi square exceeds 3 × d.f. (3 = 1.5 × 2). Otherwise we choose a more restricted model. This decision rule is equivalent to adding 3 × d.f. (rather than 2 × d.f.) to $-2 \ln L$ in the formula of AIC. (Note, however, that the values of AIC in Table 1 are evaluated according to the unmodified formula.)

With the above qualification in mind let us start with the comparison between the additive error model and the multiplicative error model. We see that in all four corresponding solutions (two conditions on the variance structure times two dimensionalities) obtained under these two error models, the AIC is consistently smaller in the additive error model. For these data the additive error model seems to fit better than the multiplicative error model. Our experience with other data sets also indicates that the additive error model is generally superior to the multiplicative error model, when the type of judgment involved is direct comparison of two or more dissimilarities [Takane, 1978b], though the converse is true for rating judgment [Takane, 1981].

Assuming that the additive error model is more appropriate, we may now compare the constant variance and the nonconstant variance assumptions. For both two and three dimensional solutions the values of AIC are found smaller in the nonconstant variance assumption ($\chi^2 = 51.6$ with 4 d.f. for the two dimensional solution, and $\chi^2 = 52.8$ with 4 d.f. for the three dimensional solution. Since the values of the asymptotic chi square are both more than ten times larger than the corresponding d.f., it may be safely concluded that there are substantial individual differences in dispersions.

We are now in the position to determine appropriate dimensionality of the representation space. The comparison between two and three dimensional solutions in the nonconstant variance additive error model reveals that the third dimension is clearly significant ($\chi^2 = 146.6$ with 14 d.f.), implying that the appropriate dimensionality is at least three. It is possible that the dimensionality is even higher, so that four and five dimensional solutions were also obtained under the equivalent conditions. The fourth dimension seems to be significant ($\chi^2 = 61.3$ with 13 d.f.), while the fifth dimension is not. The value of the asymptotic chi square (=24.6) representing the difference between the four and five dimensional solutions is only slightly larger than twice the corresponding d.f.(=12). Note that the five dimensional solution still has the minimum AIC value according to the original formula, but the four dimensional solution would have been the minimum AIC solution, if we had taken into account the correction factor for small samples (i.e., if we had added 3 × d.f. to $-2 \ln L$ instead of 2 × d.f.).

The four dimensional solution is depicted in Figures 2 and 3 for selected pairs of dimensions along with 95% asymptotic confidence regions for the estimated points. The first two dimensions roughly correspond with the two defining properties of the stimuli; dimension 1 represents the lips dimension, while dimension 2 represents the eyes dimension. Although the stimuli are physically two dimensional, the third and the fourth dimensions are clearly interpretable. While it is possible to interpret the unrotated third and fourth dimensions (designated as dimension 3 and dimension 4, respectively in Figure 3), the interpretation would be much more straightforward if they are rotated for about 60° counter clockwise, as indicated in Figure 3. In the figure the new dimensions are designated as dimension 3′ and dimension 4′. (The dimensions are ordered in terms of the amount of stimulus variability they can account for.) Dimension 3′ represents "skew-symmetry" about the eyes dimension. It can be seen in Figure 3 that the stimuli, which occupy symmetric locations with respect to the neutral eye curvature level (i.e., flat eyes) in Figure 1, take approximately equal coordinate values on dimension 3′. For example, stimuli 1 and 17 are very close on this dimension, since they have exactly opposite curvatures of eyes. Dimension 4′ represents essentially the same thing for the lips dimension. (i.e., it represents "skew-symmetry" about the lips dimensions.) Stimuli 3 and 15, for example, take approximately equal coordinate values on this dimension.

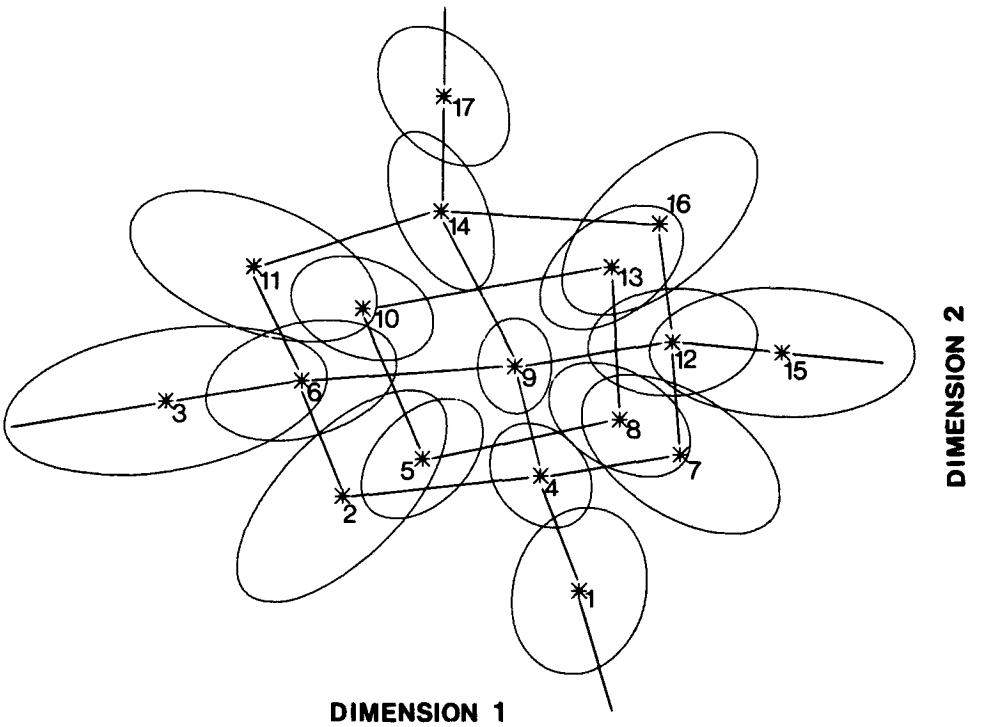

FIGURE 2.
The two-dimensional plot of the best fitting four dimensional solution from MAXSCAL-4.1: Dimension 1 versus Dimension 2.

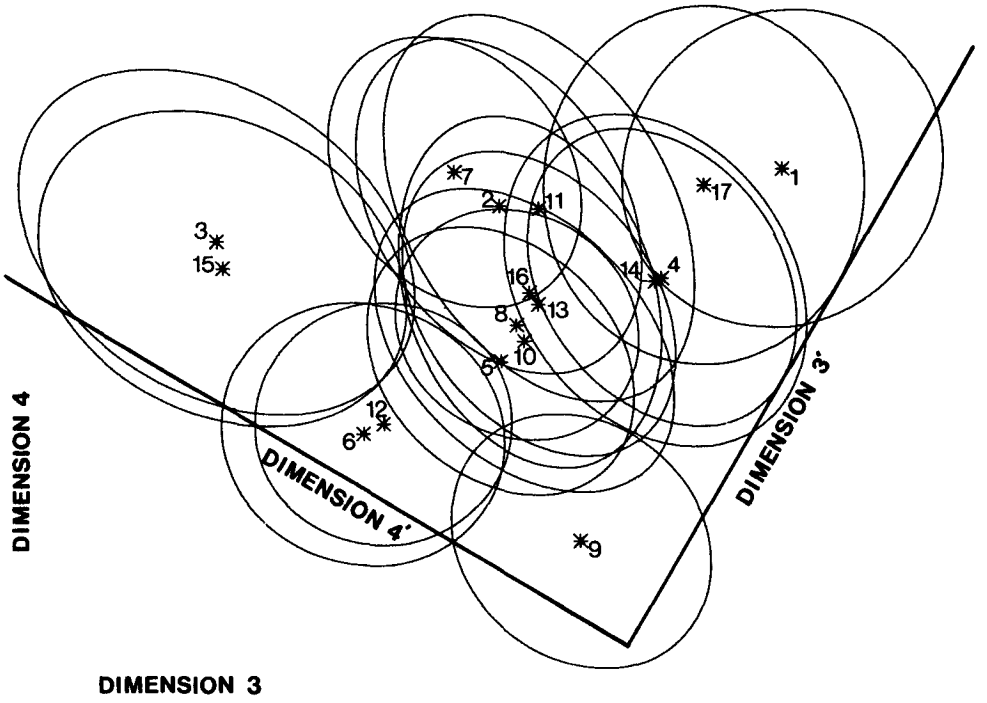

FIGURE 3
The two-dimensional plot of the best fitting four dimensional solution from MAXSCAL-4.1: Dimension 3 versus Dimension 4.

Curvatures equal in absolute value but opposite in sign are perceived to be more similar than would be predicted from linear dimensions of curvature. The linear dimensions are folded (or bent) at neutral curvature levels in order to accomodate the perceived similarities between "skew-symmetric" curvatures. Figure 4 depicts how the two defining properties of the stimuli are folded (or bent) at the neutral curvature levels. (The top figure is the plot of dimension 1 against dimension 4′; the bottom one is the plot of dimension 2 against dimension 3′. In making these plots dimensions 1 and 2 were rotated slightly (about 10° clockwise in the plane spanned by these dimensions) so that they more closely match with the two defining physical dimensions of the stimuli.) It may be observed in Figure 4 that there is a subtle difference in the ways in which the two dimensions are folded (or bent). For the lips dimension the bent is gradual, while for the eyes dimension it is rather sharp. At the moment we are not sure whether this difference has any psychological significance.

A "folded" dimension may be interpreted psychologically as representing "intensity" analogous in some way to the social utility dimension discussed by Coombs and Kao [1960]. On the other hand, it may be just an artifact of the noneuclidean nature of the perceptual space. For example, as pointed out by Chang and Carroll [1980] in the context of individual differences MDS analysis of color normal and color deficient subjects, possible idiosyncratic monotonic transformations of a common stimulus dimension by different subjects may well give rise to a "folded" dimension like the ones we have observed in our study. At present, however, there is no definite evidence to favor one over the other. The

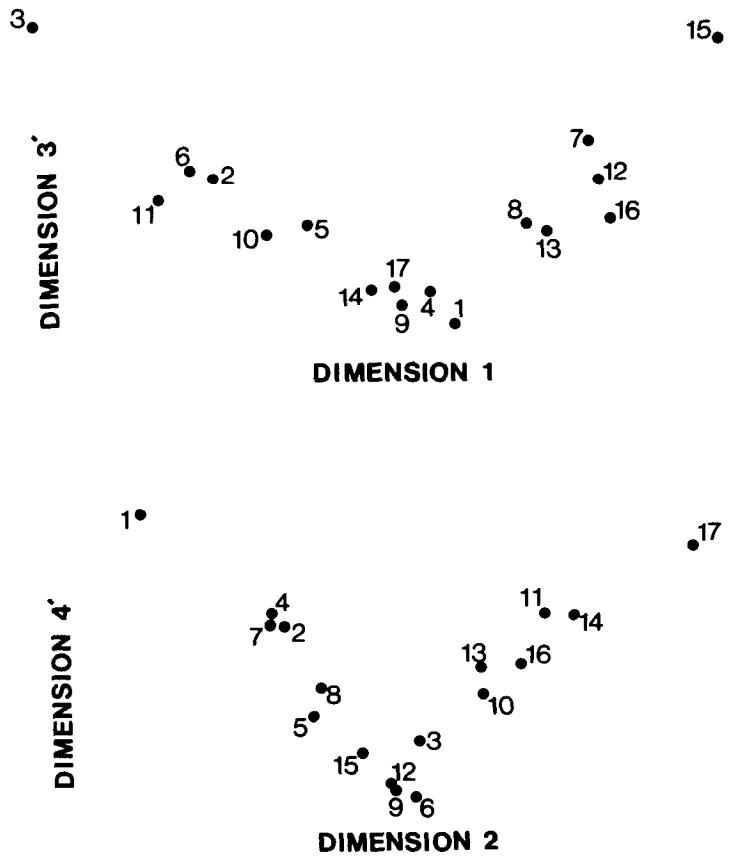

FIGURE 4
The plots of the "folded" dimensions: Dimension 1 versus dimension 4′ (top) and dimension 2 versus dimension 3′ (bottom).

important point is that a "folded" dimension is a rule rather than an exception [Chang & Carroll, 1980; Takane, 1981].

Since the stimuli were constructed by incomplete factorial combinations of two physical properties, we may formally test whether those physical properties can account for the subjective dissimilarities among the stimuli. This hypothesis is geometrically very similar to the configuration depicted in Figure 1, except that equal intervals are not necessarily assumed between adjacent curvature levels. We may add the "skew-symmetric" hypothesis to obtain a four dimensional hypothesis. Dimensions 1 and 2 represent the curvature of lips and the curvature of eyes, respectively. The third dimension represents the "skew-symmetry" about the eye dimension. On this dimension the coordinates are assumed equal for stimuli 3, 6, 9, 12 and 15, stimuli 5, 8, 10 and 13, stimuli 2, 4, 7, 11, 14 and 16, and stimuli 1 and 17, respectively. Similarly, the fourth dimension represents the "skew-symmetry" about the lips dimension. The coordinates for stimuli 1, 4, 9, 14 and 17, stimuli 5, 8, 10 and 13, stimuli 2, 6, 7, 11, 12 and 16, and stimuli 3 and 15 are assumed equal on this dimension. Coordinates which are assumed equal were treated as if they were single parameters in the estimation procedure.

These hypotheses were fitted under the nonconstant variance additive error assumption. Among the two constrained solutions the four dimensional solution has the minimum AIC value of 3554.9 ($\chi^2 = 184.0$ with 6 d.f.). However, this value is still substantially larger than that of the corresponding unconstrained solution. The asymptotic chi square representing the difference between the two solutions is 149.2, which is more than three times its corresponding d.f. ($=40$).

Asymptotic Confidence Regions

The inverse of the information matrix evaluated at the maximum likelihood estimates of parameters is known to give asymptotic variance and covariance estimates of estimated parameters. It has been shown by Ramsay [1978] that the regular inverse may be replaced by the Moore-Penrose inverse when the information matrix is singular due to nonuniqueness of parameters in multidimensional scaling. Exploiting this fact and the asymptotic normality of the maximum likelihood estimators, we may draw asymptotic confidence regions for any subsets of parameters. In MDS we are typically interested in confidence regions drawn separately for each estimated point. In cases of a four dimensional configuration, each confidence region forms a four dimensional hyper ellipsoid, which can be orthogonally projected onto two dimensional subspaces each formed by two of the four dimensions. The projection of a hyper ellipsoid forms an ellipse like those depicted in Figures 2 and 3. In drawing those ellipses a small-sample correction factor, developed by Ramsay for this MULTISCALE [Ramsay, 1980a], was applied. It generally has an effect of making them somewhat larger. Unfortunately, it is impossible to visualize four dimensional hyper ellipsoid based on these ellipses.

Discussion

We have seen an example of analysis which can be performed by MAXSCAL-4.1. Although this example pertains to conditional rank order data, essentially the same type of analyses can be performed for other types of directional rank order data. As should be clear from the definition of the likelihood function [(12), (14) & (15)], the directionality of ranking processes is one of the most crucial ingredients of the current procedure. It may be that this procedure proves to be very robust against violation of this basic assumption, so that it may be safely used for nondirectional rank order data as well. At present, however, we have no hard evidence which indicates that this is the case, though a study is being undertaken which is aimed at systematically examining the robustness of the current procedure against nondirectional ranking processes [Takane & Carroll, Note 2].

Aside from the robustness question there is a good reason to favor directional ranking procedures over nondirectional ones; the former generally provides dissimilarity rank orders which are more consistent with true orderings of underlying distances. To see this a number of dissimilarity rankings were generated from assumed stimulus configurations by both the directional and the nondirectional ranking process. Correlations (Kendall's tau) were calculated between true rankings of distances and rankings obtained by the two processes. For two levels of the error examined [$\sigma = .2$ and .4 for $\text{tr}(X'X) = n$, which is set to 10] the correlations were uniformly higher for the directional ranking processes. This implies that directional ranking procedures can provide more reliable rankings of dissimilarities than their nondirectional counterparts. Note that in the latter error processes are generated once for all per ranking, whereas in the former they are assumed to be generated for each successive first choice.

The empirical results indicating superior validity of directional rankings can be rationalized in terms of the following argument. For sake of simplicity let us consider the case of ranking of only three objects (although the arguments for $n > 3$ should follow straightforwardly). For $n = 3$, the whole difference between directional and nondirectional rankings arises from the probability that the object whose "true" rank is 2 is in fact given an actual rank of 2. (The process for choosing the first ranked object is identical for both processes, while the third ranked item is completely specified once the first two ranks are specified.) Denoting this probability as $\Pr(r = 2 \mid R = 2)$ where R denotes the "true" rank and r the observed rank, we can see that the conditional probability is just the joint probability that object 2 is not maximal in the first place and *is* maximal in the second. In the nondirectional case the first and second events must be negatively correlated, since a small value associated with object 2 will tend to make that object nonmaximal on both the first and second first choices. Thus the fact that object 2 was not maximal in the first ranking tends to imply that the value associated with it was in fact relatively small, and thus decreases the probability that that object is (correctly) maximal in the second first choice. In the directional case, on the other hand, two events are independent, so that the nonmaximality of object 2 in the first place has no implied effect on its position in the second first choice. Consequently, $\Pr(r = 2 \mid R = 2)$ for directional rankings must be greater than $\Pr(r = 2 \mid R = 2)$ for nondirectional rankings. Since in this particular case this conditional probability completely determines the relative validity of the two processes, it follows that directional rankings (at least for $n = 3$) are more valid than nondirectional ones. A similar line of argument may be applied to the case of $n > 3$.

REFERENCES NOTES

1. Inukai, Y., Nakamura, K., Shinohara, M. A multidimensional scaling analysis of judged dissimilarities among schematic facial expressions. *Bulletin of Industrial Product Research Institute*, 1977, *81*, 21–30 (in Japanese).
2. Takane, Y., and Carroll, J. D. On the robustness of AIC in the context of maximum likelihood multidimensional scaling. Manuscript in preparation.

REFERENCES

Akaike, H. A new look at the statistical model identification. *IEEE Transactions on automatic control*, 1974, *19*, 716–723.

Bock, R. D. *Multivariate statistical methods in behavioral research.* New York: McGraw Hill, 1975.

Carroll, J. D., & Chang, J. J. Analysis of individual differences in multidimensional scaling via an *N*-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 1970, *35*, 283–319.

Chang, J. J., & Carroll, J. D. Three are not enough: An INDSCAL analysis suggesting that color space has seven (± 1) dimensions. *Color research and applications*, 1980, *5*, 193–206.

Coombs, C. H., & Kao, R. C. On a connection between factor analysis and multidimensional unfolding. *Psychometrika*, 1960, *25*, 219–231.

Dempster, A. P., Laird, N. M., & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *The Journal of the Royal Statistical Society*, Series B. 1977, *39*, 1–38.

Guttman, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 1968, *33*, 469–506.

Huynh, H. & Feldt, L. S. Conditions under which mean square ratios in repeated measurements designs have exact *F*-distributions. *Journal of the American Statistical Association*, 1970, *65*, 1582–1589.

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964 (a), *29*, 1–27.

Kruskal, J. B. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 1964 (b), *29*, 115–129.

Luce, R. D. *Individual choice behavior : A theoretical analysis.* New York: Wiley, 1959.

McGee, V. The multidimensional scaling of "elastic" distances. *The British Journal of Mathematical and Statistical Psychology*, 1966, *19*, 181–196.

Ramsay, J. O. Maximum likelihood estimation in multidimensional scaling. *Psychometrika*, 1977, *42*, 241–266.

Ramsay, J. O. Confidence regions for multidimensional scaling analysis. *Psychometrika*, 1978, *43*, 145–160.

Ramsay, J. O. Some small sample results for maximum likelihood estimation in multidimensional scaling. *Psychometrika*, 1980 (a), *45*, 139–144.

Ramsay, J. O. Joint analysis of direct ratings, pairwise preferences and dissimilarities. *Psychometrika*, 1980 (b), *45*, 149–165.

Rao, C. R. *Advanced statistical methods in biometric research.* New York: Wiley, 1952.

Richardson, M. W. Multidimensional psychophysics. *Psychological Bulletin*, 1938, *35*, 659–660.

Roskam, E. E. The methods of triads for multidimensional scaling. *Nederlands Tijdschrift Voor de Psychologie en haar grensgebieden*, 1970, *25*, 404–417.

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function, I & II. *Psychometrika*, 1962, *27*, 125–140 & 219–246.

Takane, Y. A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent—theory. *Japanese Psychological Research*, 1978 (a), *20*, 7–17.

Takane, Y. A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent—evaluations. *Japanese Psychological Research*, 1978 (b), *20*, 105–114.

Takane, Y. Multidimensional successive categories scaling: A maximum likelihood method. *Psychometrika*, 1981, *46*, 9–28.

Torgerson, W. S. Multidimensional scaling: I. Theory and method. *Psychometrika*, 1952, *17*, 401–419.

Wilks, S. S. *Mathematical statistics.* New York: Wiley, 1962.

Young, F. W. Scaling replicated conditional rank-order data. In Heise, D. R. (ed.), *Sociological Methodology*. San Francisco: Jossey-Bass Publishers, 1975.

Manuscript received 11/24/80
Final version received 6/18/81

APPENDIX

We show the equivalence of (10) to Luce's model for the first choice [Luce, 1959]. Assuming the multiplicative error model we may write (10) as

$$p_{ik}^{(1)} \cong \left[1 + \sum_{q=1}^{n-2} \exp\{c_k(\ln d_{i(1k)} - \ln d_{i(q+1,k)})\}\right]^{-1}$$

$$= \left[1 + \sum_{q=1}^{n-2} \exp\{\ln(d_{i(q+1,k)}/d_{i(1k)})^{-c_k}\}\right]^{-1}$$

$$= \frac{1}{\left[1 + \sum_{q=1}^{n-2} (d_{i(q+1,k)}/d_{i(1k)})^{-c_k}\right]}$$

$$= \frac{(d_{i(1k)})^{s_k}}{\sum_{q=0}^{n-2} (d_{i(q+1,k)})^{s_k}},$$

which is Luce's model. The last equation can be derived by multiplying both the numerator and the denominator by $(d_{i(1k)})^{-c_k}$ and by replacing $-c_k$ by $s_k(<0)$.