Behaviormetrika
1977, No. 4, 29-43

ON THE RELATIONS AMONG FOUR METHODS OF MULTIDIMENSIONAL SCALING

Yoshio Takane*

Some intriguing relationships among four methods of metric and nonmetric multidimensional scaling (MDS) are explicated. It is shown that all four methods of MDS considered here amount to solving, explicitly or implicitly, the stationary point of a matrix which can be generally represented as $A'HA$, where A is a difference matrix (to be defined), and where H depends on a particular criterion being optimized. H may be a matrix of fixed constants or of functions of unknown parameters (stimulus coordinates) of the representation model. A conceptual distinction is made as to the scale level of measurement in reference to MDS methods and solutions.

1. Introduction

Okamoto and Toda (1973) compared two methods of multidimensional scaling (MDS) in terms of their performance. The two methods discussed are Hayashi's quantification method IV (Hayashi, 1952) and Torgerson-Gower type (Torgerson, 1952; Gower, 1966) principal coordinates method of MDS, which may be more appropriately ascribed to Young and Householder (1938). (For simplicity and for the purpose of avoiding the unnecessary confusion, however, the former will be referred to as Q4 and the latter as the T-G method.) In this paper we explicate some interesting relationships among four methods of multidimensional scaling. It is deemed to shed a further light on the characterizations of the two methods, Q4 and the T-G method, in a more general perspective by clarifying certain aspects of formal structures of solutions. It is interesting to note that the two methods have been used as initialization procedures (to obtain a good initial start) in two representative methods of nonmetric MDS, Q4 in SSA (Smallest Space Analysis: Guttman, 1968) and the T-G method in TORSCA (Young, 1968) and in KYST (Kruskal, Young & Seery, 1973), which is a merger of TORSCA and MDSCAL (Kruskal, 1964a, 1964b).

2. Representations of solutions

The problem we face with MDS in general can be stated as follows. Given a set of empirical dissimilarities (or similarities) between pairs of stimuli assumed to be measured at a certain scale level, we are to find a configuration of points in a space

Key Words and phrases; multidimensional scaling (MDS), metric and nonmetric MDS, quantification method IV (Q4), Torgerson-Gower type principal coordinates method (the T-G method), the Young-Housholder transformation, scalar product optimization, distance optimization, squared distance optimization, optimal scaling, alternating least squares (ALS), monotone convergence, scale levels ot measurement, admissible transformation

* University of Tokyo Department of Psychology, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113. The author is currently at McGill University, Department of Psychology, 1205 McGregor Avenue, Montreal, PQ, Canada, H3A 1B1.

whose mutual distances, in some sense, best represent the observed dissimilarities.[1] Three conditions must be prescribed; one is the assumed scale level of measurement of observed dissimilarities, another is the specific form of a distance function to be employed, and the third is the exact statement of the optimality of representations.

The first point relates to what empirical relations we may legitimately use in deriving the spatial representation, and to how we concepualize the nature of errors. The point will be more fully discussed in Section 3 of this paper, since it plays a crucial role in the metric and nonmetric distinction of MDS methods. Throughout Section 2 we assume that the observed dissimilarities are fixed constants, an assumption which is usually made in the metric MDS.

We will not be concerned with the form of distance functions in any general perspective. We simply assume the Euclidean distance model,

$$d_{ij}^2 = \sum_{a=1}^{t} (x_{ia} - x_{ja})^2 \tag{1}$$

where d_{ij} is the distance between stimuli i and j, x_{ia} is the coordinate of stimulus i on dimension a, and t is the dimensionality of the space. The Euclidean distance is the only metric to which all of the methods to be considered in this paper can apply. We further assume that t is given.

In the remaining of this section we will elaborate the third point in more detail, which amounts to how we measure the goodness of representations, and how we attain the optimality. Specifically we attempt to characterize four methods of MDS in terms of 1) optimization criteria and 2) formal structures of solutions. The methods to be discussed are Hayashi's Q4, the T-G method for scalar product optimization, Guttman's C-matrix method for distance optimization, and de Leeuw's successive refactoring method for squared distance optimization. We note that a distance optimization method has been originated by Kruskal's NDSCAL, and that a squared distance optimization method has been implemented by ALSCAL (Takane, Young & de Leeuw, 1977).

2.1 Symbolism

The following symbols will be used in the following discussion.

e or e_{ij}: observed similarity (between stimuli i and j)
o or o_{ij}: observed dissimilarity (between stimuli i and j)
d or d_{ij}: distance (between stimuli i and j)
n: the number of stimuli
E: an $n \times n$ matrix whose (i, j) element is e_{ij}
O: an $n \times n$ matrix whose (i, j) element is o_{ij}
D: an $n \times n$ matrix whose (i, j) element is d_{ij}
$O^{(2)}$: an $n \times n$ matrix whose (i, j) element is o_{ij}^2

1) The MDS, in its broader sense, only presupposes multiplicity of representations. However, we use the term in a narrower sense; i.e., a method which involves distance models of some form for the representation of dissimilarity data.

$D^{(2)}$: an $n \times n$ matrix whose (i, j) element is d_{ij}^2

n^*: $n^*=n(n-1)/2$

D_s: a diagonal matrix of order n^* whose diagonal elements are, say, s_{ij}'s arranged in a prescribed order (e.g., $s_{12}, s_{13}, \cdots, s_{1n}, s_{23}, \cdots, s_{2n}, \cdots, s_{(n-1)n}$). s_{ij} may be $-e_{ij}$, o_{ij}, o_{ij}^2, d_{ij}^2 or o_{ij}/d_{ij}. We assume that observations are symmetric; i.e., $e_{ij}=e_{ji}$ and $o_{ij}=o_{ji}$.

A: an $n^* \times n$ difference matrix which takes pairwise differences of stimulus coordinates. For $n=4$, A looks like:

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

J: a centering operator matrix of order n (Takeuchi & Yanai, 1972). Let p_{ij} be the (i, j) element of J, then $p_{ij}=\delta_{ij}-1/n$ where δ_{ij} is a Kronecker delta. Note that

$$A1_n = 0_{n*} \tag{2}$$

where 1_n is an n-component vector whose elements are all unity and 0_{n*} an n^*-component zero vector, and that

$$A'A = nJ . \tag{3}$$

X: an $n \times t$ matrix of stimulus coordinates whose (i, a) element is x_{ia}. We assume, without loss of generality, that

$$JX = X . \tag{4}$$

which implies that X is columnwise centered.

diag(S): a diagonal matrix whose diagonal elements are those of S for any square matrix S.

off(S): off(S)=S–diag(S).

2.2 Quantification method IV (Q4)

The Q4 maximizes

$$\theta_1 = \frac{1}{2} tr(-ED^{(2)}) = tr(X'A'D_{-e}AX) \tag{5}$$

under the ortho-normalization restriction that $X'X=I$.[2] It follows immediately that the above optimality condition leads to a stationary equation which amounts to

2) The normalization restriction is not included in Hayashi's original formulation (Hayashi, 1952). The XQ for any positive diagonal matrix Q is legitimate as a solution. However, the solution derived under the restriction in the text is the only case in which distances can be defined in the usual way (i.e., by (1)). It is conventional to use $Q=\Delta^{1/2}$ where Δ is a diagonal matrix of t dominant eigenvalues of matrix G in (6).

solving (normalized) eigenvectors[3] corresponding to t dominant eigenvalues of a certain matrix. This matrix can be written, using the above notation, as

$$G = A'D_{-e}A\,. \tag{6}$$

It can be readily seen from this representation of G that it is positive semidefinite if all $-e_{ij}$ are nonnegative[4], and that the solution X remains invariant over positive affine transformations of $-e_{ij}$[5]. For the later comparison it is convenient to represent G yet in another form:

$$G = A'\left[\mathrm{diag}\left(-\frac{1}{2}A(-E)A'\right)\right]A\,. \tag{7}$$

2.3 Torgerson-Gower type scalar product optimization method (the T-G method)

Define

$$\theta_2 = tr(P-XX')^2 \tag{8}$$

where

$$P = -\frac{1}{2}JO^{(2)}J\,. \tag{9}$$

The matematical operation expressed by (9) is often called the Young-Householder transformation.[6] It is well known that the minimum of (8) is attained by

$$X = U_t\Delta_t^{1/2} \tag{10}$$

where Δ_t is the diagonal matrix of t dominant eigenvalues of P, and U_t is the matrix of corresponding (normalized) eigenvectors. To make X defined in (10) the unique solution the restriction that $X'X=\Delta_t$ has been imposed. It is also tacitly assumed that at least t dominant eigenvalues are nonnegative[7], and that they are all distinct (and t'th eigenvalue is distinct from $t+1$'st). Obviously the solution remains

3) Still some indeterminacy exists in the solution. Any orthogonal transforms of the matrix of eigenvalues are also solutions.

4) This property has been noted and proved by Okamoto and Toda (1973). Non-negative definiteness of G can be easily verified by noting that $G=A'D_{-e}^{1/2}D_{-e}^{1/2}A=B'B$ where $B=D_{-e}^{1/2}A$. The decomposition $D_{-e}=D_{-e}^{1/2}D_{-e}^{1/2}$ is justified by the non-negativity of all $-e_{ij}$. Positive definiteness follows immediately from (3).

5) A formal proof of this property has been given in Saito, Ogawa and Nojima (1972). In reference to the representation of G matrix in (6), however, it can be more easily shown that

$$G^* = A'D_{b(-e)+c}A$$

for some constants b and c has an identical set of eigenvectors to those of G (Eigenvalues may well be different). The G^* can be expanded as

$$G^* = bA'D_{-e}A+cA'A = bG+ncJ\,.$$

Now the J has $n-1$ unit eigenvalues and a single zero eigenvalues. Eigenvectors corresponding to the $n-1$ unit eigenvalues cannot be uniquely determined, but an arbitrary set of $n-1$ orthogonal eigenvectors can be constructed in the subspace of dimensionality $n-1$ (orthogonal to the subspace of dimensionality 1 spanned by the n-component constant vector which is an eigenvector of G^* corresponding to the zero eigenvalue), and they can be made identical to the $n-1$ nonconstant eigenvectors of G corresponding to the $n-1$ nonzero eigenvalues. Since G also has a constant eigenvector corresponding to the zero eigenvalue, all of the n eigenvectors of G^* and G should agree with each other, establishing the desired result.

essentially invariant (i.e., invariant except for the change in the overall size of the configuration) over positive similarity transformations of observed dissimilarities. However, the solution is generally not affine invariant. It is noted that a monotone invariant procedure (called INDISCAL: de Leeuw, Takane & Young, in preparation) has been developed within the scalar product optimization framework.

By noting (3) P can be restated as

$$P = A'\Big[-\frac{1}{2}AO^{(2)}A'\Big]A/n^2$$

$$= A'\Big[D_{o^2}+\mathrm{off}\Big(-\frac{1}{2}AO^{(2)}A'\Big)\Big]A/n^2\,, \tag{11}$$

which may render more direct comparisons with other methods.

2.4 C-matrix method for distance optimization

Curiously a distance optimization method has been originated as a nonmetric procedure (Kruskal, 1964a, 1964b). It is because the T-G method is vastly superior when the observed dissimilarities are measured up to a ratio scale. Nonetheless Kruskal type iterative procedure can be applied, at least theoretically, to metric data as well.

Let

$$\theta_3 = tr(O-D)^2\,. \tag{12}$$

The soft-squeeze[8] phase of Guttman's SSA (Guttman, 1968) also optimizes θ_3 under an appropriate normalization restriction on X. This leads to a stationary equation of the form

$$X = \frac{1}{\lambda}C(X)X/n \tag{13}$$

---

6) It has been proved by Young and Householder (1938) that $D^{(2)}$ is transformed into XX' by (9) in an exact sense. This can be readily seen by noting that $D^{(2)}$ can be expanded as

$$D^{(2)} = x^{(2)}1_n' - 2XX' + 1_n x^{(2)\prime}$$

where $x^{(2)}$ is an n-component vector whose i'th element is the sum of squares of coordinates of stimulus i across dimensions, namely $\sum_{a=1}^{t} x_{ia}^2$. We apply $J1_n = 0_n$ and Eq. (4) to obtain the desired result.

7) In case not all t dominant eigenvalues are non-negative, the LS estimate of X in the sense of (8) is given by replacing negative eigenvalues in Δ_t by zero (de Leeuw, 1974).

8) The fact that the soft-squeeze phase is suffifficient to optimize (12) has been overlooked by Guttman (1968) and his collaborators (Lingoes & Roskam, 1973). This is partly because the gradient (of θ_3 with respect to X) does not vanish at the stationary point of X in (15), since in nonmetric procedures observations are not fixed constants but may be monotonically transformed, so that θ_3 can always be made smaller by shrinking the size of X while X is normalized in each iteration. A further complication is introduced by the use of non-optimal monotonic transformation of data called the rank-image, which often causes fluctuation of θ_3 in the process of iterations. This observation has tempted Guttman to propose a supplementary process (called the hard-squeeze phase) for more rigorous optimization. However, it has turned out that the hard-squeeze procedure still suffers from the use of the rank-image transformation; either the gradient is only approximate (one-step procedure) or the monotonic transformation is non-optimal (two-step procedure).

where λ is the normalization factor, and where

$$C(X) = A'D_{o/d}A\,, \tag{14}$$

assuming that $d_{ij} \neq 0$ for $i \neq j$.[9] Equation (13) is, of course, not a completely explicit form of the solution for X. Rather, it is used as an updating equation for successive approximations of X. When this fact is taken into account, it is written as

$$X^{(k+1)} = \frac{1}{\lambda^{(k)}} C(X^{(k)})X^{(k)}/n \tag{15}$$

where the parenthesized superscript indicates the iteration number. If the normalization restriction on X is such that

$$tr(X^{(k+1)\prime}X^{(k+1)}) = 1\,,$$

then $\lambda^{(k)}$ is given by

$$\lambda^{(k)} = tr(X^{(k)\prime}C(X^{(k)})^2X^{(k)})^{1/2}/n\,.$$

The iterative scheme (15) is called Guttman's C-matrix method, which has been shown to be a special case of Kruskal's steepest descent algorithm (Gleason, 1967). Note that $C(X)$ is a function of unknown parameter X, and consequently should be updated for the new estimate of X. Otherwise Eq. (15) constitutes a power-method-like iterative procedure. If X^* is a fixed point of (15), then obviously its column vectors, when appropriately rotated, should correspond to *some* (denormalized) eigenvectors of $(1/\lambda^*)C(X^*)/n$. However, those eigenvectors do not necessarily correspond to the t dominant eigenvalues, nor are the dimensionwise variances of X^* (diagonal elements of $X^{*\prime}X^*$ assuming that X^* is already rotated to principal axes orientation; i.e., $X^{*\prime}X^*$ itself is diagonal) proportional to the magnitudes of eigenvalues associated with the eigenvectors which correspond to the solution, which should all be equal since we have $a \cdot C(X) = X$ for some constant a. (Some numerical results regarding this fact will be presented in the discussion section.)

The convergence characteristic of (15) has always been observed (Lingoes & Roskam, 1973). The formal proof of the monotone convergence property of the matrix method has been established by de Leeuw (1976) based on the convexity property of distance functions.

An interesting fact exploited by de Leeuw in the convergence proof of the C-matrix method is that the minimization of θ_3 over all (centered) X is equivalent to the maximization of

$$\theta_3' = \frac{1}{2}tr(OD) = tr(X'C(X)X) \tag{16}$$

over all (centered) X such that $(1/2)trD^2 = n$ $tr(X'X) = n$ (or any arbitrary constant, for that matter). The apparent similarity of θ_3' to θ_1 defined in (5) may been seen if we

9) All gradient type methods within the distance optimization framework (e.g., MDSCAL, SSA) tacitly assume this. Otherwise the gradient is not properly defined (The d_{ij} appears in the denominator of the gradient). The point relates to the identifiability property of the MDS methods discussed by Okamoto and Toda (1973). Thus, algorithmically the gradient type methods do not possess the identifiability property. However, it has been shown by de Leeuw (1976) that defining $o_{ij}/d_{jj} = 0$ for which $d_{ij} = 0$ $(i \neq j)$ is the rational treatment consistent with the monotone convergence property of the C-matrix method.

put $O=-E$ and replace D by $D^{(2)}$ in (16). However, for the maximization of θ_1 to be equivalent to the minimization $tr(-E-D^{(2)})^2$, a least squares (LS) criterion obtained by replacing O by $-E$ and D by $D^{(2)}$ in θ_3, the associated normalization restriction must be $tr\ D^{(2)2}=n$, which is clearly different from the normalization restriction adopted in Q4. Unfortunately the optimization of θ_1 with this new restriction is unwieldy. The Q4 does not seem to bear any simple relationship to any LS methods considered in this paper. (However, the optimization method for $tr(O^{(2)}-D^{(2)})^2$ will be discussed in the next section, which is equivalent to $tr(-E-D^{(2)})^2$ if we put $O^{(2)}=-E$).

Finally, it is noted that a convergent algorithm can be readily constructed within the distance optimization framework (SMACOF: de Leeuw, 1976) by combining the C-matrix method and Kruskal's LS monotonic transformation of observed dissimilarities.

2.5 Successive refactoring method for squared distance optimization (S-R method)

A squared distance optimization principle has been found particularly useful in the nonmetric extention (ALSCAL: Takane, Young & de Leeuw, 1976) to the individual differences model (Carroll & Chang, 1970) in MDS. ALSCAL can of course be applied to metric data and to the simple (unweighted) Euclidean distance model as well. However, for the present purpose (of comparing the formal structures of various solutions of MDS methods) a particularly interesting procedure is de Leeuw's (1975a) ELEGANT algorithm, which is also based on the alternating least squares (ALS) methodology (Young, de Leeuw & Takane, 1976) as is ALSCAL, and which optimizes the same LS criterion (defined on the squared distances) that ALSCAL optimizes.

Let

$$\theta_4 = tr(O^{(2)}-D^{(2)})^2 . \tag{17}$$

We are to minimize θ_4 with respect to X under an identification restriction that $X'X=\Delta$ (a diagonal matrix), which amounts to solving a stationary equation of the form:

$$A'\,[\{D_{o\bullet}-\mathrm{diag}\,(AXX'A')\}/n^2]\,AX = 0 . \tag{18}$$

Equation (18) may be solved iteratively. Let

$$B(X) = A'\,[D_{o\bullet}+\mathrm{off}\,(AXX'A')]\,A/n^2 , \tag{19}$$

then

$$B(X)X = X\Delta . \tag{20}$$

If the X in $B(X)$ is held constant (which can be legitimately done in the ALS framework), Eq. (20) states an eigen equation with Δ being a diagonal matrix of t dominant eigenvalues of $B(X)$. We may use the estimate of X from the previous iteration, say k'th, which is denoted as $X^{(k)}$, to define $B(X^{(k)})$[10], then $X^{(k+1)}$ can be obtained by the eigenvalue-vector decomposition of $B(X^{(k)})$. θ_4 is ultimately optimized by successive refactorings of $B(X)$ which is redfined in each iteration for the new X.

10) $B(X^{(k)})$ should be kept constant until $X^{(k+1)}$ converges completely (if Clint and Jennings type simultaneous power method is used), which ensures the LS estimate of X for fixed B. The stationary equation is $B(X^{(k)})X^{(k+1)}=X^{(k+1)}\Delta^{(k+1)}$ when the iteration number is explicitly indicated, and not $B(X^{(k)})X^{(k)}=X^{(k+1)}$ as in the C-matrix method.

Let

$$H^{(k)} = D_{o^2} + \text{off}\,(AX^{(k)}X^{(k)\prime}A')\,,$$

then

$$\frac{1}{2}\theta_4^{(k+1)} = tr(H^{(k)} - AX^{(k+1)}X^{(k+1)\prime}A')^2$$

$$= tr(H^{(k)2}) - n^2 tr\Delta^{(k+1)2}\,.$$

We see that the minimum of $\theta_4^{(k+1)}$ is attained by the set of eigenvectors corresponding to the t dominant eigenvalues of $B(X^{(k)})$. We have

$$X^{(k+1)\prime}X^{(k+1)} = \Delta^{(k+1)}\,.$$

In computation it is more convenient to use

$$B(X) = A'(D_{o^2} - D_{d^2})A/n^2 + XX'$$

where $D_{d^2} = \text{diag}(AXX'A')$.

Again, the extention to nonmetric data is straightforward using Kruskal's LS monotonic transformation.

2.6 Discussion

It has been shown that in all cases we have discussed solutions are given as stationary points of matrices which can be generally represented as $A'HA$. The specific forms of H depend on the particular optimization criteria being optimized, and are summarized in Table 1. Elements of H are fixed constants in Q4 and in the T-G method for scalar product optimization, while they are functions of unknown parameters (hence written as $H(X)$ in the table) in the C-matrix method for distance optimization and in S-R method for squared distance optimization. This difference in the formation of H matrices makes the first major distinction of the solutions. If H is a matrix of fixed constants, the solution is analytic (at least in principle) and is exact, otherwise it is iterative and only approximate.

Table 1

Comparisons of H matrices in the four methods of MDS

Name	Criterion	Representation of matrix H
Q4	MAX $tr(OD^{(2)})$	$H = D_{-e} = \text{diag}\left(-\frac{1}{2}A(-E)A'\right) = D_{-e}$
T-G method	MIN $tr(P-XX')^2$	$H = -\frac{1}{2}AO^{(2)}A'/n^2 = \left[D_{o^2} + \text{off}\left(-\frac{1}{2}AO^{(2)}A'\right)\right]/n^2$
C-matrix method	MIN $tr(O-D)^2$	$H(X) = D_{o/d}$
S-R method	MIN $tr(O^{(2)}-D^{(2)})^2$	$H(X) = [D_{o^2} + \text{off}(AXX'A')]/n^2$

In Table 1 we find an interesting status of Q4 in reference to the other three methods based on some LS criteria; Q4 can be regarded as an approximate method in two different ways.

First, let $o^2 = -e$, and ignore all off-diagonal elements of H in the T-G and S-R methods, then they both reduce to a form essentially equivalent to Q4 except for a

multiplicative constant. It is difficult, however, if not totally impossible, to predict what extraneous effects this reduction (of zeroing all off-diagonals) may have on the solution of MDS.[11]

Second, let $o=-e$ and assume that all d_{ij}'s are constant in the C-matrix method, then $H(X)$ for the C-matrix method reduces to a matrix of fixed constants, and

$$C(X) = C \propto AD'_{-e}A$$

which is equivalent to G in Q4 (Eq. (6)). Guttman (1968) proposed to use this approximate C-matrix method for obtaining an initial estimate of stimulus configuration in SSA. His intuitive rationale for the procedure is as follows. Suppose n stimuli are arranged in such a way that they are equi-distant from each other. This would be the case if no information whatsoever is available as to their mutual distances. How should one then modify the equi-distant configuration (which in itself is not informative at all) in light of the information regarding dissimilarities between stimuli? One defines the C-matrix with available information, observed dissimilarities and equi-distances, and obtains the approximate C-matrix method. The $-e_{ij}$ may be converted into an initial distance rank number as in MINISSA (Lingoes & Roskam, 1973). Guttman refers this initialization procedure to his paper published in the middle of 1940's (Guttman, 1946). However, it does not seem to be too obvious, at least to the present author, where in the paper the method is specifically referred to.

Another remarkable property of Q4 can be revealed by looking at it as an approximate C-matrix method. Okamoto and Toda (1973) compared the speed of decrease in the magnitudes of eigenvalues of G and P matrices. They consistently observed that the decrease is more rapid with P than with G, which implies that for fixed dimensionality the solution X obtained by the spectral decomposition of P contains more information pertaining to P than the solution X obtained from G contains information pertaining to G. Indeed, in case of the T-G method a rapid decrease in the size of eigenvalues of P is a desirable property of solutions since it is explicitly intended to be so; the optimization criterion for T-G method can be rewritten as

$$\theta_2 = tr\Delta^2_{n-t}$$

where Δ_{n-t} is a diagonal matrix of $n-t$ non-dominant eigenvalues of P, and hence the minimization of θ_2 is equivalent to putting $n-t$ non-dominant eigenvalues as close to zero as possible. However, in case of Q4 regarded as an approximate C-matrix method, just the opposite is true. Suppose $o_{ij}=d_{ij}$ for all i and j (the case of perfect fit), then the C-matrix reduces to

$$C(X) = nJ.$$

It is well known that J has $n-1$ eigenvalues of size unity and one zero eigenvalue. Thus eigenvalues of the C-matrix in this limiting situation are completely levelled (except

11) On the contrary, the difference between solutions of the T-G and S-R methods are very subtle. In the former off-diagonal elements of H are derived from observations, whereas in latter they are functions of distances (which presuppose that the stimulus coordinates are already known). It can be shown, however, that when $O^{(2)}=D^{(2)}$, H matrices in the two methods become identical and that $\theta_2=\theta_4$.

of course for one zero eignevalue). Therefore we conjecture that the more homogeneous are the diagonal elements of $D_{o/d}$ (which is what happens when the fit is improved in the sense of θ_3), the more levelled are the eigenvalues of the C-matrix.

In Table 2 we present lists of eigenvalues of several matrices. The first three columns have been extracted from Okamoto and Toda (1973), who analyzed three sets of similarity measures derived from the same set of similarity data concerning 22 food textures by Q4. The three sets of measures are monotonically related with each other. Eigenvalues listed in the fourth column are obtained by Q4 using rank numbers of similarities (obtained by ranking similarities, which is still another monotonic transformation of the original similarities) as input data. They correspond to the eigenvalues of the initial matrix in MINISSA. In columns 5 and 6 two sets of eigenvalues of the C-matrices at the convergence points of SMACOF (a nonmetric procedure combining features of the C-matrix method and of Kruskal's LS monotonic transformations of observed dissimilarities to optimize a slightly generalized version of θ_3: de Leeuw, 1976) are listed for two and four dimensional solutions, respectively. The $-e_{ij}$ is used as the observed dissimilarity. We see that the degree of levelings of eigenvalues is roughly in the order of Q4(2)<MINISSA (initial)<Q4(1)<Q4(3)<SMACOF (dimension 2)<SMACOF(dimension 4). At the bottom of each column a statistic indicating the goodness of nonmetric fit is presented. It is found that the better fit

Table 2 Distributions of eigenvalues

	Q4			MINISSA (initial)	SMACOF†	
	(1)	(2)	(3)		dim. 2	dim. 4
1	6.71	7.78	5.88	7.54	5.60	4.94
2	5.64	5.87	5.60	5.67	5.26	4.92
3	5.05	5.73	5.13	5.58	5.17	4.86
4	5.04	5.70	5.08	5.57	5.07	4.84
5	5.02	5.44	5.07	5.52	4.97	4.814
6	5.01	5.25	5.02	5.31	4.84	4.806
7	4.90	4.80	4.92	4.75	4.83	4.789
8	4.89	4.63	4.90	4.69	4.82	4.779
9	4.82	4.62	4.70	4.64	4.74	4.76
10	4.81	4.49	4.65	4.59	4.65	4.747
11	4.77	4.43	4.62	4.57	4.643	4.7438(3)
12	4.71	4.41	4.56	4.54	4.6405(2)	4.7436(4)
13	4.57	4.39	4.53	4.47	4.6400(1)	4.741 (2)
14	4.53	4.35	4.52	4.37	4.62	4.739 (1)
15	4.42	4.26	4.51	4.29	4.59	4.731
16	4.34	4.11	4.50	4.18	4.58	4.714
17	4.25	4.06	4.49	4.05	4.53	4.711
18	4.23	4.03	4.41	3.98	4.47	4.703
19	4.14	4.01	4.37	3.95	4.46	4.696
20	4.11	3.85	4.21	3.93	4.45	4.68
21	4.04	3.79	4.21	3.86	4.42	4.55
22	0	0	0	0	0	0
θ_3^*	.177	.436	.181	.435	.057	.026

† SMACOF results are those obtained with data set 1. Results with other data sets are essentially equivalent to the presented results.

Eigenvalues corresponding to solution vectors should be all identical. They are slightly different here because of insufficient number of iterations.

roughly corresponds to the more leveled distribution of eigenvalues.[12] The leveling of eigenvalues is a desirable characteristic in case of Q4 in terms of the distance optimization. It does not seem to be meaningful to decide the correct dimensionality of a solution by merely looking at the distribution of eigenvalues of G matrix in Q4.

The parenthesized numbers in columns 5 and 6 indicating dimensions of stimulus coordinates in the order of magnitude are put to eigenvalues corresponding to the eigenvectors of solutions. It can be observed that they do not necessarily correspond to the first t dominant eigenvalues, as has been alluded to earlier. Dimensionwise dispersions of stimulus coordinates are 5.786 and 3.243 for dimensions 1 and 2, respectively, in two dimensional solution, and 7.611, 4.715, 2.358 and 1.508 in four dimensional solution. Evidently the sizes of eigenvalues corresponding to the eigenvectors of solution, which are theoretically equal at a stationary point of the C-matrix method, convey no information as to the relative importance of dimensions.

3. Measurement levels and invariance properties of the MDS methods

Up to this point observed dissimilarities (or similarities) are assumed to be fixed constants. That is, the observed data are metric. However, as it has been suggested earlier, the three methods, T-G, C-matrix and S-R methods, all based on LS criteria, can be, and in fact have been extended to nonmetric procedures. The extention has been carried out by allowing observed data to be monotonically transformed so that they in the LS sense best accord with the model predictions (which therefore is called optimal scaling of observed data). Algorithmically this amounts to alternate applications of model estimation procedures and some forms of quadratic programming procedures. However, a more rigorous distinction of the concepts, metric and nonmetric, seems to be in order at this point. Roughly speaking this is the problem, which has been known as the scale levels of measurement in psychometric literature.

We define the scale level of data as the type of functional form through which model predictions are related to observations, or vice versa. In the presence of various measurement errors a qualification "within a reasonable amount of errors" may be necessary to the above statement. Note that observations are the empirical dissimilarity measures and model predictions are the distances derived from the stimulus configuration in the cortext of multidimensional scaling. As you may have noticed, this definition of the scale level is somewhat different from that of Stevens' (1951) who was the first to define the scale level of measurement in terms of invariance characteristic of certain relations between data elements (more precisely, between numbers assigned to data elements), and who inadvertently assumed that those invariant relations are intrinsic properties of measurement, the presumption to which we do not quite agree. On the contrary, we define the measurement level by the relation which exists between

12) Some complication arises by the use of the nonmetric fit measure in all cases. In the first four cases, at least, the θ_3 may reflect the levelings of eigenvalues more accurately. The θ_3^* is defined by

$$\theta_3^* = tr(O^*-D)^2/tr(D^2)$$

where O^* is the matrix of optimally transformed dissimilarities under the monotonic restriction.

two distinct entities, observations and predictions. The same set of numbers may well represent the different scale properties depending on which models the particular set of data are related to.

The close relationship of the present definition and that of Stevens' can be readily seen by noting that the type of functional forms between observations and predictions is also invariant over a certain class of transformations of data, and that the transformations must have the same type of functional forms as functions whose forms are to be preserved through the transformations. The class of transformations over which the type of functional forms is invariant is called the admissible transformation of the data associated with the specific level of measurement. Thus, it can be said that our definition also rests on the invariance characteristic, but the invariance is in the relations between observations and predictions and not in the intrinsic relations among observations themselves.[13]

We presume that any methods of analysis (deemed to handle data measured at a certain scale level) must give esssentially invariant results (i.e., estimates of model parameters) over admissible transformations of data. The required property of invariance, however, is not in an absolute sense, since model parameters may be subject to change if the transformation preserves their scale properties invariant.

This presumed invariance property is demonstrated with SMACOF using Okamoto and Toda's data under the ordinal measurement assumption. Table 3 presents three stimulus configurations obtained from three different sets of data, but are related monotonically with each other. Agreements are quite impressive.[14] (Results with data sets 1 and 2 are identical up to three decimal points. Only a result with data set 3 is slightly different from the other two.)

Obviously neither Q4 nor the T-G method (as it were) enjoy the invariance property over monotonic transformations of observed dissimilarities, as has been abundantly demonstrated by Okamoto and Toda (1973) and by Hayashi (1973). Derived stimulus configurations may suffer from haphazard effects of monotonic distortions of observations, however robust the methods may seem against certain degrees of systematic perturbations of data. In terms of the above invariance criterion the T-G method (as well as others LS methods) is a ratio scale procedure, while Q4 is an interval scale procedure (showing the invariance over affine transformations of

13) Two closely related, but distinct, concepts often confused with each other and with the scale level of data are the scale level of model predictions and the scale level of model parameters. The former is defined by the invariance characteristic of defining properties of a specific model over a certain class of transformations of model predictions. For example, the defining properties of distances, the three metric axioms, remain valid over similarity transformations of distances, but not necessarily over other more general types of transformations. Hence, distance (as a model) is said to constitute a ratio scale. The scale level of model parameters, on the other hand, is defined in reference to model predictions. It is the type of transformations of model parameters over which model predictions remain "essentially" unchanged.

The stimulus coordinates in the distance model constitute dimensionwise interval scales with a common multiple across dimensions, since the intrinsic properties of distances are invariant over positive affine transformations of X.

14) Stimuli 18 and 19 should have an identical location. Results indicate a slight displacement in all cases perhaps due to insufficient numbers of iterations.

Table 3

SMACOF solutions with three sets of data

Dimension	Data 1			Data 2			Data 3		
		1	2		1	2		1	2
Stimulus	1	1.101	.101	1	1.101	.101	1	1.097	−.104
	2	.884	.265	2	.884	.265	2	.894	−.281
	3	1.232	−.638	3	1.232	−.638	3	1.214	.652
	4	1.159	−.838	4	1.159	−.838	4	1.167	.845
	5	1.327	−.612	5	1.327	−.612	5	1.297	.638
	6	1.196	−.101	6	1.196	−.101	6	1.221	.072
	7	1.352	−.168	7	1.352	−.168	7	1.311	.209
	8	1.036	.251	8	1.036	.251	8	1.041	−.253
	9	.545	−.493	9	.545	−.493	9	.546	.471
	10	1.118	.000	10	1.118	.000	10	1.111	.001
	11	1.246	−.333	11	1.246	−.333	11	1.253	.341
	12	.244	1.253	12	.244	1.253	12	.251	−1.266
	13	−.325	1.357	13	−.325	1.357	13	−.323	−1.370
	14	−.294	1.357	14	−.294	1.357	14	−.294	−1.371
	15	−.620	1.244	15	−.620	1.244	15	−.625	−1.259
	16	−1.278	.353	16	−1.278	.353	16	−1.284	−.366
	17	−1.724	−.357	17	−1.724	−.357	17	−1.733	.342
	18	−1.823	−.493	18	−1.823	−.493	18	−1.806	.512
	19	−1.818	−.478	19	−1.818	−.478	19	−1.804	.498
	20	−1.833	−.464	20	−1.833	−.464	20	−1.841	.469
	21	−1.863	−.477	21	−1.863	−.477	21	−1.831	.494
	22	−.860	−.730	22	−.860	−.730	22	−.863	.726
Dispersion		5.786	3.243		5.786	3.243		5.763	3.284
Stress		.057			.057			.057	
Iterations		50			50			50	

The over-all size of configurations is such that $tr(X'X)=nt$.

data). However, it should be noted that in case of Q4 this invariance property does not conform to the type-of-functional-form definition of the scale level of observed data. The Q4 reveals the invariance property over any affine transformations of data irrespective of the "true" functional relationship between observations and predictions.[15] This is due to its somewhat arbitrary construction of the optimization criterion in Q4.

We maintain that the invariance property which is not conformable to the type of true functional relationships between observations and predictions is not very meaningful, since, say, monotone-invariant procedures can otherwise be easily constructed by simply rank-ordering observed dissimilarities. Another more powerful concept should be established which can effectively exclude the possibility of the "trivial" invariance property as possessed by Q4, and which also generalizes the concept of "faithfulness of representation" (Okamoto & Toda, 1973). The "recovery of stronger scale properties" seems to serve this purpose. To explicate, a procedure

15) The Q4 is approximate no matter what conditions are satisfied, while the T-G method is exact when the data satify certain conditions. As approximate methods (such as initialization procedurds) both work reasonably well. Discussion as to the superiority of one over the other in this respect is somewhat elusive, since it is not very difficult to construct examples on which either one works better.

ought to be able to recover the true functional relationship even when a weaker scale level assumption is made, provided that the data convey enough information to constrain model parameters uniquely under the weaker assumption. All nonmetric versions of the LS methods discussed in this paper, at least in principle, are supposed to have this desirable feature (despite the existence of the well-known numerical difficulty called a local minimum problem).

4. Summary

In this paper a unified way of representing solutions of four methods of MDS, Hayashi's quantification method IV (Q4), Torgerson-Gower type principal coordinates method (the T-G method), Guttman's C-matrix method and de Leeuw's successive refactoring method, is discussed. The discussion, while primarily not concerned with the development of a new procedure, is deemed to have some informative values to MDS methodologists.

5. Acknowledgement

The author is grateful to Drs. Shiba and Yanai for their helpful comments on the earlier draft of this paper.

References

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

Gleason, T.C. (1967). A general model for nonmetric multidimensional scaling. Michigan Mathematical Psychology Program, MMPP 67–3, Ann Arbor, Michigan.

Gower, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, **53**, 325–338.

Guttman, L. (1946). An approach to quantifying paired comparisons and rank order. *Annals of Mathematical Statistics*, **17**, 144–163.

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

Hayashi, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics*, **3**, 69–98.

Hayashi, C. (1973). e_{ij} gata suryoka to Torgerson-Gower no hoho o megutte (On e_{ij}-type quantification method and Torgerson-Gower method). *Nihon Tokeigakkaishi (Journal of the Japanese Statistical Society)*, **3**, 41–53.

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

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

Kruskal, J.B., Young, F.W. & Seery, J. (1973). How to use KYST; Unpublished user's manual, Bell Laboratories, Murray Hill, N.J.

de Leeuw, J. (1974). Finding a positive semidefinite matrix of prescribed rank r in a nonlinear differentiable manifold; Technical Report, Bell Laboratories, Murray Hill, N.J.

de Leeuw, J. (1975). An alternating least squares approach to squared distance scaling; Unpublished manuscript, The University of Leiden, Leiden.

de Leeuw, J. (1976). Applications of convex analysis to multidimensional scaling; *Proceedings of thd European Meeting of Statistics*, Grenoble.

Lingoes, J.C. & Roskam, E.E. (1973). A mathematical and empirical analysis of two multidimensional scaling algorithms. *Psychometrika Monograph Supplement* **19**.

Okamoto, M. & Toda, J. (1973). Suryokariron 4-rui to shuzahyo bunsekiho (Fourth quantification method and principal coordinate analysis). *Nihon Tokeigakkaishi* (*Journal of the Japanese Statistical Society*), **3**, 41–53.

Saito, T., Ogawa, S. & Nojima, E. (1972). Deta kaseki (1): suryoka ni kansuru sogo hokoku (Data analysis I: a summary review article on quantification theory). *Soken Kiyo*, **2**, 23–114.

Stevens, S.S. (1951). Mathematics, measurement and psychophysics; In S.S. Stevens (ed.). *Handbook of experimental psychology*. New York: Wiley.

Takane, Y., Young, F.W. & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometirka*, **42**, 7–67.

Takeuchi, K. & Yanai, H. (1972). *Tahenryo kaiseki no kiso* (*Foundations of multivariate analysis*); Tokyo: Toyo Keizai Shinpo.

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

Young, F.W. (1968). TORSCA-9: a FORTRAN IV program for nonmetric multidimensional scaling. *Behavioral Science*, **13**, 343–344.

Young, F.W., de Leeuw, J. & Takane, Y. (1976). Quantifying qualitative data; Psychometric Laboratory Report #149, The University of North Carolina, Chapel Hill, N.C.

Young, G. & Householder, A.S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, **3**, 19–22.

(Received February, 1977)

Takane, Y. On the relations among four methods of Multidimensional scaling. Behaviormetrika, 1977, 4, 29-43.

ERRATA

page	line	wrong	correct
29	3 (footnote)	Housholder	Householder
30	25	NDSCAL	MDSCAL
31	8 - 14	$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$
32	6 (footnote)	Positive definiteness	Positive semidefiniteness
34	23	aC(X) = X	aC(X)X = X
34	26	the matrix	the C-matrix
39	31	cortext	context
43	last	February, 1977	December, 1976
40	31	others	other
37	6	$C(X) = C \propto A\ D'_{-e} A$	$C(X) = C \propto A'\ D_{-e} A$
42	3 (below)	thd	the
35	26	= 0	= XΔ