PSYCHOMETRIKA—VOL. 45, NO. 1.
MARCH, 1980

COMPONENT MODELS FOR THREE-WAY DATA: AN ALTERNATING LEAST SQUARES ALGORITHM WITH OPTIMAL SCALING FEATURES

RICHARD SANDS AND FORREST W. YOUNG

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

A review of the existing techniques for the analysis of three-way data revealed that none were appropriate to the wide variety of data usually encountered in psychological research, and few were capable of both isolating common information and systematically describing individual differences. An alternating least squares algorithm was proposed to fit both an individual difference model and a replications component model to three-way data which may be defined at the nominal, ordinal, interval, ratio, or mixed measurement level; which may be discrete or continuous; and which may be unconditional, matrix conditional, or row conditional. This algorithm was evaluated by a Monte Carlo study. Recovery of the original information was excellent when the correct measurement characteristics were assumed. Furthermore, the algorithm was robust to the presence of random error. In addition, the algorithm was used to fit the individual difference model to a real, binary, subject conditional data set. The findings from this application were consistent with previous research in the area of implicit personality theory and uncovered interesting systematic individual differences in the perception of political figures and roles.

Key words: individual differences, measurement level.

The discovery of meaningful relations underlying complex bodies of data is important in many areas of psychology. One of the most commonly used procedures for this purpose is component analysis. The classical component-analysis procedure is applicable to two-way data, such as the scores of a group of subjects on a battery of tests.

Often, however, psychological data are three-way (or multi-way). Common examples of such data are multi-trait multi-method matrices [Campbell & Fiske, 1959], the semantic differential [Osgood, Suci, & Tannenbaum, 1957], and personality trait ratings [Rosenberg & Sedlak, 1972]. Because these data are three-way, they pose a problem for the classical component-analysis procedures.

Early solutions to this problem [Abelson, 1960; Cattell, 1952; Guttman, 1958] involved reducing the data from three- to two-way. For example, semantic differential data is frequently reduced to a concept by scale matrix by averaging across subjects. Or, if individual differences are of interest, the data can be averaged across concepts or scales. However, there are a number of problems with such solutions. First, these solutions cannot be employed when differences in all three ways are of interest, because they do not allow the simultaneous analysis of all three ways. Second, if there are substantial differences between subjects, then a solution based on the average may represent only a small minority of the subjects or, even worse, none at all. Third, the use of any reduction method involving arithmetic operations such as averaging is only appropriate for interval or ratio data.

For these reasons, investigators have recently proposed a number of direct solutions

This paper is part of a Thesis performed by Richard Sands under the direction of Forrest Young at the L. L. Thurstone Psychometric Laboratory, University of North Carolina at Chapel Hill. Thanks are extended to Drs. Charles Schmidt and Andrea Sedlak for the use of their political role data set.

Reprints and programs may be obtained from Forrest Young, Psychometric Laboratory, Davie Hall 013-A, University of North Carolina, Chapel Hill, NC 27514

0033-3123/80/0300-2808$00.75/0

for three-way data [Carroll & Chang, 1970; Israelsson, 1969; Tucker, 1963, 1964, 1966; Harshman, Note 1; Jenrich, Note 2; Kroonenberg & de Leeuw, 1980]. These procedures are extensions of classical two-way component analysis. But because all of these new procedures place stringent requirements on the measurement characteristics of the data, none of them are applicable to most of the data types usually encountered in psychological research.

Three measurement characteristics can be used to classify the types of data encountered in psychological research [de Leeuw, Young, & Takane, 1976; Young, de Leeuw, & Takane, 1976; Takane, Young, & de Leeuw, 1977; Young et al., Note 3]. These are measurement process, measurement level, and measurement conditionality. These three measurement characteristics describe a wide variety of data, including data in which the variables' measurement levels and/or measurement processes are different. Recently, Takane, Young, and de Leeuw [1978] have developed a two-way component-analysis procedure appropriate to such data. However, there are no three-way procedures appropriate to such a wide variety of data. Both the reduction and direct three-way procedures, mentioned previously, are applicable to only unconditional interval or ratio data.

The purpose of this work is to describe and evaluate a new procedure appropriate for a wide variety of three-way data. This procedure fits several component models to three-way data which may be defined at nominal, ordinal, interval, ratio, or mixed levels; which may be discrete or continuous; and which may be unconditional, matrix conditional, or row conditional.

Models for Three-Way Data

There are a number of component-analysis models for three-way data. (Jöreskog, 1971, proposed three-way factor-analysis models. However, this discussion is limited to component models, those models where uniquenesses are not estimated.) These models are extensions of the familiar model for two-way data. The most general of these models is Tucker's Three Mode Factor Analysis model [Tucker, 1963, 1964, 1966]. This model, which is a component model, not a factor analysis (since there is no provision for uniquenesses) is expressed as

$$z_{ijk} = \sum_s \sum_t \sum_u f_{is} x_{jt} y_{ku} c_{stu} ;$$

where z_{ijk} is the i, j, k^{th} element of the three-way data matrix Z; the coefficients f_{is}, x_{jt}, and y_{ku} are entries in the component matrices F, X, and Y; and the coefficients c_{stu} are entries in a three-way matrix, C, which Tucker calls the "core matrix."

The matrices F, X, and Y give the scores of the variables, concepts, and individuals on the "idealized" component variables. It is not necessary that there be the same number of components for each way. In order to obtain a unique solution, F, X, and Y are assumed to be columnwise orthonormal. (Columnwise orthonormal will be used to designate a matrix A where $A'A = I$. Columnwise orthogonal will designate a matrix A, where $A'A = \text{DIAG}$.) C is an s by t by u matrix which gives the interrelations that connect the three sets of components. According to Levin [1965], the inner core matrix gives the scores of each "idealized" object on each "idealized" variable for each "idealized" individual. Application of the TMFA model has been limited, possibly because the model is hard to understand, making interpretation of results difficult.

If the TMFA model is restricted by assuming $Y = I$, then the model becomes

$$z_{ijk} = \sum_s \sum_t f_{is} x_{jt} c_{stk} ,$$

where c_{stk} is an element of the k^{th} individual's two-way core matrix, and the other terms are as previously defined. Israelsson [1969], Jennrich [Note 2], and Kroonenberg and de Leeuw [1980] proposed this model.

This model is more conveniently expressed as

$$Z_k = FC_kX' \tag{1}$$

where Z_k is the two-way data matrix and where C_k is the two-way core matrix for the k^{th} individual. The core matrices can be interpreted as before. However, it is more appealing to think of C_k as defining the k^{th} individual's metric or rotation of the common components, F or X. (C_k can operate on either F or X. For the purposes of interpretation it will be assumed that C_k operates on X.) For this reason, it seems appropriate to call this a "generalized subjective metrics model." The cross product model derived from (1),

$$B_k = Z_k{}'Z_k = XC_k'F'FC_kX' = XC_k^*X',$$

is identical to Schonemann's [1972] "generalized subjective metrics model" for MDS, where $C_k^* = C_k'F'FC_k$.

There is a straightforward geometrical interpretation for this model. Each C_k describes a unique oblique rotation (linear combinations) of the common component scores X. So, the k^{th} individual's component scores are formed by the linear combinations, C_k, of the set of common component scores X. And the same linear combinations, as specified by F, of these individualized component variables describe each person's original data. To obtain a unique solution, F and X are assumed to be columnwise orthonormal.

There are several additional simplifying assumptions which can be made. First, we can assume that there are the same number of components for each way. Then we can impose further constraints on C_k in (1). While a variety of constraints could be investigated, we look into the situation where $C_k = W_k$, giving what we call the "weighted model":

$$Z_k = FW_kX', \tag{2}$$

where W_k is diagonal for all k. Harshman [Note 1] proposed this model. It is only necessary to restrict the size of F and X to obtain a unique solution. Thus, the particular orientation of F and X are determined by the model itself.

In this model different individual weighting of the common component variables gives the individualized component scores. There is no rotation of the common component scores. Individuals differ only with respect to the salience of each component. This model is analogous to the INDSCAL cross product model [Carroll & Chang, 1970].

It is interesting to note that this and the TMFA model are the only symmetric models. That is, no matter which way one fits the model to the data there will be a set of components for each way. This is most obvious when one expresses the model (2) as

$$z_{ijk} = \sum_t f_{it}x_{jt}w_{kt},$$

where w_{kt} are the diagonal elements of W_k. This model seems to be the most direct extension of the component-analysis model for two-way data, which is expressed as

$$z_{ij} = \sum_t f_{it}x_{jt}.$$

If we assume that $W_k = I$ for all k, we obtain what we call the "replicated model":

$$Z_k = FX', \tag{3}$$

where F must be columnwise orthogonal and X columnwise orthonormal to obtain a unique solution. The interpretation of this model is that the same component scores represent all individuals. That is, there are no individual differences. The model assumes that individuals' data are replications of each other.

In this paper we develop an algorithm for fitting the weighted and replicated models to data with the variety of measurement levels, processes, and conditionality commonly encountered in laboratory and field research. It performs both internal and external [Carroll, 1972] analyses. That is, for external analyses, F and/or X can be constrained to a priori sets of parameter values. Also, the algorithm will fit the replications model, since it is a special case of the weighted model. Finally, the algorithm has been formulated such that it can be expanded to fit the "generalized subjective metrics model" discussed in this section.

The ALSCOMP3 Algorithm

This section presents the details of an *a*lternating *l*east *s*quares optimal scaling algorithm for *comp*onent analysis of *3*-way data.

The ALSCOMP3 algorithm involves two major phases and two minor phases. The first major phase involves obtaining the least squares estimates of the optimally scaled data Z_k^* under the assumption that F, X, and W_k are constants. Specifically, this phase solves the conditional least squares problem which minimizes the loss function

$$\phi^2 = \sum_i \sum_j \sum_k (z_{ijk}^* - \hat{z}_{ijk})^2 = \sum_k tr\,(Z_k^* - FW_kX')'(Z_k^* - FW_kX'), \tag{4}$$

where z_{ijk}^* and $\hat{z}_{ijk}$ are, respectively, the optimally scaled data and the model estimates obtained from FW_kX', and Z_k^* is a scale by object matrix of optimally scaled data for the k^{th} subject. Using the notation of Takane et al. [1977] this problem is denoted by $\text{MIN}_{Z^*}[\phi^2(Z^*/F,X,W)]$, where Z^* is the collection of Z_k^* matrices, and W is the collection of W_k matrices. The second major phase involves obtaining the conditional least squares estimates for the three model parameter subsets. The first subphase solves the problem $\text{MIN}_F[\phi^2(F/X,W,Z^*)]$, the second subphase solves the problem $\text{MIN}_X[\phi^2(X/F,W,Z^*)]$, and the third the problem $\text{MIN}_W[\phi^2(W/F,X,Z^*)]$. The two minor phases are initialization and termination. The algorithm begins with the initialization phase, then alternates between the two major phases and the termination phase. The details of these phases are discussed in the next few sections.

Initialization Phase

The initialization procedure solves for the initial values of F, X, and W_k using the observed data Z_k. This initialization procedure is analogous to the algebraic solution for subjective metrics proposed by Schönemann [1972]. However, it differs from his solution because Z_k is rectangular, as opposed to a square symmetric matrix. (This approach was suggested by Yoshio Takane.) This solution is exact for the error-free, ratio-level, unconditional measurement case. Recall the weighted model is

$$Z_k = FW_kX'. \tag{5}$$

This model does not have a unique solution, because

$$Z_k = (FD_1)\,(D_1^{-1}W_kD_2)\,(D_2^{-1}X') = F^*W_k^*X^{*\prime},$$

where D_1 and D_2 are diagonal matrices. Since the parameters of the model are only defined up to joint dilations and reflections, we must restrict the size of F, X, or W_k. We choose to resolve this indeterminacy by arbitrarily restricting the size of the W_k matrices such that

$$W_{\cdot} = \frac{1}{N}\sum_k W_k = I\,, \tag{6}$$

where N is the number of subjects. Combining (6) and (5) and averaging gives

$$Z_{\cdot} = \frac{1}{N}\sum_k Z_k = FX'\,. \tag{7}$$

We can use any method to find a factoring of $Z_{\cdot}$. Once $\hat{F}$ and $\hat{X}$ are obtained by an appropriate method of factoring (ALSCOMP3 uses an Eckart-Young, 1936, decomposition), we have

$$Z_{\cdot} = \hat{F}\hat{X}' = (\hat{F}A)\,(A^{-1}\hat{X}')\,, \tag{8}$$

where A is any nonsingular rotation matrix. Combining (7) and (8) gives

$$F = \hat{F}A\,, \tag{9}$$

and

$$X = \hat{X}A^{-1\prime}\,. \tag{10}$$

So to find F and X we need to find A, since $\hat{F}$ and $\hat{X}$ are known from the factoring (8).

To solve for A substitute (9) and (10) into (5) which gives

$$Z_k = \hat{F}AW_kA^{-1}\hat{X}' = \hat{F}C_k\hat{X}'\,, \tag{11}$$

where $C_k = AW_kA^{-1}$. Since we know $\hat{F}$, $\hat{X}$, and Z_k we can find C_k from (11), where

$$C_k = (\hat{F}'\hat{F})^{-1}\hat{F}'Z_k\hat{X}(\hat{X}'\hat{X})^{-1}\,. \tag{12}$$

Using (12) to obtain C_k we can find A. Since $C_k = AW_kA^{-1}$ and W_k is diagonal for all k, A must be the right eigenvectors of C_k. (Because C_k is not symmetric it has different right and left eigenvectors. That is, $A^{-1} \neq A'$. Throughout this derivation, A is used to designate the right eigenvectors from the right eigenvector equation $C_kA = AW_k$.)

This suggests that the model only fits if all C_k from (12) have the same eigenvectors A. However, this will not be the case in most situations with fallible data. Although we could use the eigenvectors of any C_k as an estimate of A, it is more practical to obtain an estimate of A based on some sort of group average [Schönemann, Carter, & James, Note 6]. The problem is that the group average has multiple roots since

$$C_{\cdot} = \frac{1}{N}\sum_k C_k = A\left(\frac{1}{N}\sum_k W_k\right)A^{-1} = AIA^{-1} = I\,.$$

However, to avoid this difficulty we can use the average of some integral power of the C_k. This gives

$$C_{\cdot}^{r} = \frac{1}{N}\sum_k C_k^r = A\left(\frac{1}{N}\sum_k W_k^r\right)A^{-1} = ADA^{-1}\,, \tag{13}$$

where D is a diagonal matrix and the C_k are estimated by (12). Thus, we can find A from an eigenvalue decomposition of $C_{\cdot}^{r}$. (Following Schönemann et al., Note 6, ALSCOMP3 squares the C_k prior to averaging. Note that because the C_k are asymmetric, $C_{\cdot}^{r}$ is asymmetric. Because this is not the standard eigenvalue problem, a separate routine is necessary to obtain the eigenvectors of $C_{\cdot}^{r}$. Therefore, $C_{\cdot}^{r}$ was decomposed by routines from EISPACK, Note 7, an eigensystem subroutine package for real general matrices. The decomposition of an asymmetric matrix such as $C_{\cdot}^{r}$ may result in complex eigenvectors. However, complex eigenvectors were only encountered in the three-dimensional analysis

of the real data to be discussed.) Given this estimate of A we find F and X from $\hat{F}$ and $\hat{X}$ by (9) and (10). And from F, X, and Z_k we can find the W_k by regression. This will be explained in detail in the discussion of the model estimation phase.

The initialization procedure is almost identical for an external analysis which allows for an arbitrary rotation of the known X. This model is

$$Z_k = FW_kA^{-1}X', \tag{14}$$

where X is known and A^{-1} is some unknown nonsingular rotation. Assuming the same size restriction expressed in (6), (14) becomes

$$Z. = FA^{-1}X'. \tag{15}$$

Since we know X, we can find $\hat{F}$ by regression, where

$$\hat{F} = FA^{-1},$$

and from (15)

$$\hat{F} = Z.\,X(X'X)^{-1}. \tag{16}$$

And from (14) and (16)

$$Z_k = \hat{F}AW_kA^{-1}X',$$

which is identical to (11), except that we know X. The solution for A, F, and W_k proceeds exactly as before except X is used instead of $\hat{X}$.

If arbitrary rotations of X are not allowed the initialization procedure is simpler. In this case the model is the same as that expressed in (5) except that X is known. One must still restrict the size of W_k, which results in (7). However, since X is known, the least squares estimate for F is the regression equation

$$F = Z.X(X'X)^{-1}.$$

Again, the W_k are found by regression.

Optimal Scaling Phase

This phase of the algorithm obtains the optimally scaled data Z_k^* from the observations Z_k and the model estimates $\hat{Z}_k$. There are two constraints on the Z_k^*; (1) the relationship between Z_k^* and Z_k must conform to the specified measurement restrictions, and (2) the Z_k^* must be least squares estimates of the $\hat{Z}_k$. By assuming that F, W, and X are known constants, this phase solves the conditional least squares problem $\text{MIN}_{Z^*}[\phi^2(Z^*|F,W,X)]$.

Compute model estimates. The first step is to compute $\hat{Z}_k$ from the current values of F, X, and W_k by (5).

Optimal scaling. The second step is to actually perform the optimal scaling. For most types of data these procedures are familiar. The optimally scaled data for the nominal discrete case are the means of the model estimates for each category of the observations. The ordinal continuous and discrete measurement restrictions correspond to Kruskal's [1964] primary and secondary monotonic transformations. Interval and ratio discrete transformations are regressions of $\hat{Z}_k$ on Z_k with or without an intercept. The nominal continuous transformation is the two-step process described by Young et al. [1976]. They give more detailed formulations for these procedures.

Normalization. The final step in the optimal scaling phase is to normalize the solution. There are two distinct normalization problems. The first concerns the normalization of the model parameters so as to define a unique solution. The second concerns the nor-

malization of the loss function. The normalization of the loss function must be performed after each iteration in order to avoid certain degenerate solutions. However, the normalization of model parameters need only be done prior to the printing of the final solution.

The need for normalization of the model parameters is a result of the indeterminacy of the model expressed in (5). Harshman [Note 1] recognized this indeterminancy, noting that two of the three parameter subsets must be constrained. However, since optimal scaling is included in ALSCOMP3 it is necessary to impose constraints on all three model parameter subsets. Therefore, for unconditional data the following set of normalization constraints are imposed on the final solution:

$$f_t'f_t = N_i \qquad \text{for all } t,$$

$$x_t'x_t = N_j \qquad \text{for all } t,$$

and

$$\sum_t \sum_k w_{kt} = N_k R;$$

where f_t and x_t are, respectively, the t^{th} column vectors of F and X; and N_i, N_j, N_k, and R are the number of scales, objects, subjects, and dimensions, respectively. For subject conditional data the constraint on W is

$$\sum_t w_{kt} = R \qquad \text{for all } k.$$

And for row conditional data it is necessary to constrain F such that

$$\sum_t f_{it} = R \qquad \text{for all } i.$$

Finally, if the data are row conditional and the assumed measurement level is not ratio, each column of X must be centered. That is, under these conditions it is necessary to impose the constraint

$$X_t'1 = 0 \qquad \text{for all } t,$$

where 1 is a column vector of ones.

The second normalization consideration is normalizing the loss function. In order to avoid certain degenerate solutions, it is more desirable to minimize a normalized loss function. However, the normalized loss function is the ratio of two quadratic forms. This produces a quite complicated minimization problem. Conveniently, the normalized loss function can be minimized by adjusting the optimally scaled data obtained by solving the unnormalized problem. For this reason, ALSCOMP3 finds the parameters which solve the unnormalized loss function, and then adjusts them appropriately. The appropriate adjustment depends on the exact form of the desired normalized loss function.

There are three relevant questions in determining the exact form of the normalized loss function. (1) Should normalization be within partitions? (2) Should normalization be with respect to model estimates or to the optimally scaled data? (3) Should normalization be by sums of squares or by the variance?

Kruskal and Carroll [1969], Takane et al. [1977], and Roskam [Note 8] addressed the first question for a variety of MDS models. They argued that normalization must be within data partitions to avoid certain degenerate solutions. These arguments and conclusions apply to the models fit by ALSCOMP3. Therefore, it is necessary to normalize the ALSCOMP3 loss function within partitions.

Considering the second question, whether to normalize by the model estimates or by the optimally scaled data, both Kruskal and Carroll [1969] and Young [1972] argued that the choice is arbitrary. However, Takane et al. [1977] argued that, for the row conditional case of the INDSCAL model, normalization must be by the optimally scaled data. They noted that the distances in the INDSCAL model are jointly determined to a single multiplicative constant. For this reason, they argued that adjusting each row of the distance matrix (normalization by distances) by different multiplicative constants violates the model. However, recent arguments by Young et al. [Note 4] present an alternative view suggesting that one may normalize by either the model estimates or the optimally scaled data without the necessity of adjusting model estimates. Although the decision now appears arbitrary, the ALSCOMP3 loss function is normalized by the optimally scaled data.

Kruskal and Carroll [1969] considered the third question: whether to normalize by the variances or sums of squares. They emphasized the importance in MDS models of normalizing by the variances when the data is row conditional. This is equally important, even with unpartitioned data, for multiplicative models like those in ALSCOMP3. Consider a simple case of unpartitioned data from one individual. If the loss function is normalized by the sums of squares of Z_k^*, then a perfect fit can always be obtained by representing all scales and objects by the same point. Since F and X have constant column vectors, $\hat{Z}_k^*$ will be a matrix of constants. And for any transformation the least squares estimate of Z_k^* will be the matrix of constants, $\hat{Z}_k$. This means that the numerator of the loss function would be zero, and the denominator some constant. The value of the loss function would be zero, indicating a perfect fit. However, if one uses the variances of the Z_k^*, then as this solution is approached the denominator will also approach zero, producing an ill-behaved function whose value may be quite different from zero, thereby avoiding some degenerate solutions. Kruskal and Carroll state that, for MDS, the loss function becomes inflated as the denominator approaches zero. However, Sands [Note 9, Appendix B] shows that in cases where the numerator is also approaching zero the function may be inflated or deflated.

Considering these three points, ALSCOMP3 uses a loss function which is normalized within partitions by the variance of the optimally scaled data. This loss function is

$$\phi_n^2 = \frac{1}{N_p} \sum_p \frac{(Z_p^* - \hat{Z}_p)'(Z_p^* - \hat{Z}_p)}{(Z_p^* - \bar{Z}_p^*)'(Z_p^* - \bar{Z}_p^*)}, \tag{17}$$

where N_p is the number of partitions, Z_p^* is a column vector of the optimally scaled data in the p^{th} partition, $\hat{Z}_p$ is a column vector of the model estimates in the p^{th} partition, and $\bar{Z}_p^*$ is the average of the optimally scaled data in the p^{th} partition.

Recall that we are solving the unnormalized problem expressed in (4) and desire to minimize the normalized loss function (17). To accomplish this it is necessary to adjust the optimally scaled data obtained from solving the unnormalized problem. The appropriate adjustment [Appendix A] is to multiply the optimally scaled data for each partition by the constant

$$a_p \frac{(\hat{Z}_p - \bar{\hat{Z}}_p)'(\hat{Z}_p - \bar{\hat{Z}}_p)}{(Z_p^* - \bar{Z}_p^*)'(Z_p^* - \bar{Z}_p^*)}; \tag{18}$$

and add the constant

$$b_p = (1 - a_p)\bar{\hat{Z}}_p,$$

where a_p is from (18), $\bar{\hat{Z}}_p$ is the average of the model estimates in the p^{th} partition, and the other terms are as previously defined.

The normalized loss function (19) can be interpreted as the percent of the variance of

the optimally scaled data which is not accounted for by the model estimates obtained from (5). This implies that the normalized loss function should be bounded by 0 and 1. It is quite obvious that when $Z_p^* = \hat{Z}_p$, for all p, $\phi_n^2 = 0$. However, it appears that when Z_p^* is a constant vector, which occurs when the observations, Z_p, and the model estimates, $\hat{Z}_p$, are orthogonal, that ϕ_n^2 approaches infinity. In fact, this is not the case, as Z_p^* approaches a vector of constants, ϕ_n^2 approaches unity [Sands, Note 9, Appendix B].

Termination Phase

The termination phase is extremely simple since an iteration of an ALS procedure never worsens the value of the loss function [de Leeuw et al., 1976; de Leeuw, Note 10]. This phase estimates the improvement in fit by comparing the value of the loss function for the current iteration to the value for the previous iteration. If this improvement is less than the specified criterion the iterations are terminated. The model parameters are normalized and the final solution printed. If the improvement is greater than the criterion, the next phase is executed.

Model Estimation Phase

This phase successively estimates the three model parameter subsets F, W, and X. The estimation of each parameter subset is a subphase which solves a conditional least squares problem. This procedure is similar to canonical decomposition [Carroll & Chang, 1970] and Jennrich's quick algorithm [Harshman, Note 1].

The solution to the conditional least squares problem in each subphase is most complicated for row conditional data. For this type of data it is necessary to estimate each row of F and W separately. Although this solution is the most general, and appropriate to both matrix and unconditional data, it is computationally inefficient. Therefore, the least squares equations for each subphase are presented for each of the three conditionality patterns.

Estimation of F. The first subphase solves the conditional least squares problem $\text{MIN}_F[\phi^2(F|X, W, Z^*)]$, where ϕ^2 is defined by (17). For the row conditional case (17) can be expressed by

$$\phi^2 = \frac{1}{N_p} \sum_i \sum_k d_{ik}^{-1}(Z_{ik}^* - F_i W_k X')(Z_{ik}^* - F_i W_k X')', \tag{19}$$

where Z_{ik}^* is the i^{th} row vector (i^{th} scale) of the Z_k^* matrix, F_i is the i^{th} row vector of F, and $d_{ik} = (Z_{ik}^* - \bar{Z}_{ik}^*)(Z_{ik}^* - \bar{Z}_{ik}^*)'$. Because the partitioning of the data involves the elements of both F and W it is necessary to solve for each row of F (i.e., F_i) separately. That is, one must minimize each

$$\phi_i^2 = \frac{1}{N_p} \sum_k d_{ik}^{-1}(Z_{ik}^* - F_i W_k X')(Z_{ik}^* - F_i W_k X')' \tag{20}$$

with respect to F_i. Expanding (20), taking the partial derivatives of ϕ_i^2 with respect to F_i, and setting these partial derivatives to zero implies

$$\sum_k d_{ik}^{-1} Z_{ik}^* X W_k = F_i \sum_k d_{ik}^{-1} W_k X' X W_k.$$

Solving for F_i gives

$$F_i = \left(\sum_k d_{ik}^{-1} Z_{ik}^* X W_k \right) \left(\sum_k d_{ik}^{-1} W_k X' X W_k \right)^{-1}.$$

When the data is matrix conditional or unconditional, d_{ik}^{-1} is constant for all i. Under these conditions the loss function (17) can be expressed as

$$\phi^2 = \frac{1}{N_p} \sum_k \text{tr}[d_k^{-1}(Z_k^* - FW_kX')(Z_k^* - FW_kX')'], \tag{21}$$

where d_k is the variance of the optimally scaled data in the k^{th} partition. Under these conditions it is not necessary to solve separately for each row of F. The solution for F is

$$F = \left(\sum_k d_k^{-1} Z_k^* X W_k \right) \left(\sum_k d_k^{-1} W_k X' X W_k \right)^{-1}. \tag{22}$$

Notice, in the unconditional case the d_k are equal for all k. This allows the solution for F in (22) to be simplified to

$$F = \left(\sum_k Z_k^* X W_k \right) \left(\sum_k W_k X' X W_k \right)^{-1}.$$

Estimation of X. The second subphase solves the conditional least squares problem $\text{MIN}_X[\phi^2(X|F, W, Z^*)]$. To solve this problem for all types of conditional data it is most convenient to express the loss function (17) as

$$\phi^2 = \frac{1}{N_p} \sum_k \text{tr}(Z_k^* - FW_kX')'D_k^{-1}(Z_k^* - FW_kX'), \tag{23}$$

where D_k is a diagonal matrix. The form of the D_k matrix depends on the type of conditionality. When the data is row conditional, the i^{th} diagonal element of the D_k matrix is the variance of the optimally scaled data for the i^{th} row of the k^{th} subject. When the data is matrix conditional, D_k is a constant diagonal matrix having entries equal to the variance of the optimally scaled data for the k^{th} subject. Finally, when the data is unconditional, D_k is a constant diagonal matrix equal for all k. In this case the diagonal elements of all the D_k matrices are equal to the variance of the unpartitioned optimally scaled data.

Expanding (23), taking the partial derivatives of ϕ^2, with respect to X, setting these to zero and solving for X gives

$$X = \left(\sum_k Z_k^{*\prime} D_k^{-1} F W_k \right) \left(\sum_k W_k F' D_k^{-1} F W_k \right)^{-1}.$$

Recall, when the data is unconditional, the D_k matrices are equal constant diagonal matrices for all k. This allows the simplification of the solution for X, as derived above, to

$$X = \left(\sum_k Z_k^{*\prime} F W_k \right) \left(\sum_k W_k F' F W_k \right)^{-1}.$$

Estimation of W. The third subphase solves the conditional least squares problem $\text{MIN}_W[\phi^2(W|F,X,Z^*)]$. To solve this problem it is easiest to express the model as

$$Z_j = WF_jX';$$

where Z_j is a subject by object matrix for the j^{th} scale, F_j is a diagonal matrix formed from the j^{th} row vector of F, and W is a subject by dimensions matrix whose k^{th} row vector is equal to the diagonal of W_k. Using this notation the loss function (17) can be expressed as

$$\phi^2 = \frac{1}{N_p} \sum_k \sum_j d_{kj}^{-1}(Z_{kj}^* - W_kF_jX')(Z_{kj}^* - W_kF_jX')';$$

where Z_{kj}^* is the k^{th} row vector (k^{th} subject) of the Z_j^* matrix, W_k is the k^{th} row vector of W, and $d_{kj} = (Z_{kj}^* - \bar{Z}_{kj}^*)(Z_{kj}^* - \bar{Z}_{kj}^*)'$. Because the partitioning of the data involves elements of

both W and F, when the data is row conditional it is necessary to solve separately for each row of W (i.e., W_i). That is, one must minimize each

$$\phi_k^2 = \frac{1}{N_p} \sum_j d_{kj}^{-1}(Z_{kj}^* - W_k F_j X')(Z_{kj}^* - W_k F_j X')' \tag{24}$$

with respect to W_k.

Expanding (24), taking the partial derivatives with respect to W_k, setting these to zero, and solving for W_k gives

$$W_k = \left(\sum_j d_{kj}^{-1} Z_{kj}^* X F_j \right) \left(\sum_j d_{kj}^{-1} F_j X' X F_j \right)^{-1}. \tag{25}$$

When the data is matrix conditional or unconditional d_{kj} is equal for all j. Under these conditions the loss function (17) can be expressed as

$$\phi^2 = \frac{1}{N_p} \sum_j \operatorname{tr}(Z_j^* - W F_j X')' D_j^{-1} (Z_j^* - W F_j X'), \tag{26}$$

where D_j is a diagonal matrix equal for all j. For these types of conditionality it is not necessary to estimate each row of W separately. Expanding (26), taking partial derivatives with respect to W, setting these to zero, and solving for W gives

$$W = \left(\sum_j Z_j^* X F_j \right) \left(\sum_j F_j X' X F_j \right)^{-1}. \tag{27}$$

External analysis. If the analysis is external the subphases corresponding to the fixed parameter sets are skipped. If the external analysis allows for arbitrary rotation of the known X, it is necessary to solve for this rotation matrix. The model underlying this analysis is given in (14) and the solution for the rotation matrix is

$$A^{-1} = \left(\sum_k W_k F' D_k^{-1} F W_k \right)^{-1} \left(\sum_k W_k F' D_k^{-1} Z_k^* X (X'X)^{-1} \right).$$

Non-negativity constraint. The solutions for the W_k matrices, from rearrangement of W in (27) or from W_k in (25), are independent for each individual. That is, the values of one row of W do not affect the values of other rows. Given this independence between individuals' weights, a non-negativity constraint will only affect those individuals' weights which are negative. However, the weights for one individual on each dimension are not independent. This means that one cannot set negative weights equal to zero without having to re-estimate the other weights for that individual. The re-estimation, with the offending weight set to zero, is equivalent to estimating the other weights ignoring that particular dimension. Re-estimation must be repeated in the same fashion until all weights are non-negative or have been set to zero.

Although the above procedure for constraining weights to be positive is easily implemented, the constraint may not be desirable. Unlike distance models, in the ALSCOMP3 weighted model there is a straightforward interpretation for a negative weight. This interpretation is that the subject reverses that dimension in making judgments. That is, a negative weight means that the subject treats objects with high loadings on that dimension as low loadings and low loadings as high.

Variance of the optimally scaled data equals zero. When the vector of optimally scaled data for any partition is a constant, its variance equals zero. This would appear to create a problem in the definition of squared stress (17), and for those expressions in the model estimation phase (19–26) which use the reciprocal of the variance. However, it is the case

that when the variance of the optimally scaled data for a partition is zero, stress equals one (for that partition) and the reciprocal of the variance used in the model estimation phase is zero. Under these conditions the data in that partition is totally ignored in the estimation of the model parameters. Although this is not intuitively obvious, it follows from the fact that the variance of the optimally scaled data which solves the unnormalized problem is also zero, and that the adjustment of this by a_p (18) should theoretically make the variance of the optimally scaled data which solves the normalized problem infinite. Sands [Note 9, Appendix B] gives a proof that under these conditions both stress and the corresponding elements in the model estimation phase are defined to be one and zero, respectively.

Evaluation

The ALSCOMP3 algorithm was evaluated via a small Monte Carlo study and the analysis of empirical data.

Monte Carlo Study

The Monte Carlo study had three main purposes. The first purpose was to evaluate the efficacy of ALSCOMP3's recovery of information when the correct measurement level was assumed and there was no random error present in the data. The second purpose was to evaluate the robustness of the algorithm in the presence of varying degrees of random error. The final purpose was to determine the effects of assuming inappropriately strong measurement characteristics.

Method. The data sets analyzed in the Monte Carlo study were generated according to a $3 \times 3 \times 5$ factorial design (Replications $\times$ Random Error $\times$ Systematic Distortion). First, three F, three W, and three X matrices were generated randomly. Then from these three sets of parameters, which we call true parameter values, three sets of errorfree data were generated according to the weighted model (9). Thus, this first three level factor was considered a replication factor.

Next, three levels of random error were added to each errorfree data set. These three levels of random error were 0%, 10%, and 25% of the standard deviation of the errorfree data set. At this point there were nine data sets. Five transformations were performed on each of the nine data sets in order to add systematic distortion. The five transformations and associated systematic distortion were (1) none, (2) monotonic distortion by a third power transformation, (3) monotonic distortion by a seventh power transformation, (4) non-monotonic distortion by a category preserving transformation [Young & Null, 1978], and (5) binary transformation.

This resulted in three data sets for each of the 15 conditions formed by the factorial combination of random and systematic error, so there was a total of 45 data sets. Each of these 45 data sets was analyzed under the correct and stronger-than-warranted measurement assumptions. The correct measurement assumptions are a function of the measurement characteristics associated with each level of systematic distortion. The appropriate measurement characteristics for each level of systematic distortion is, respectively, (1) interval, (2) ordinal, (3) ordinal, (4) nominal continuous, and (5) nominal continuous. The characteristics and analysis of the 15 data sets generated from each of the three parameter sets are given in Table 1 along with the purpose of each analysis.

Both the random error and systematic distortion were performed separately for each individual's data matrix. Therefore, all data sets were analyzed assuming partitions by subjects.

To avoid local optimal solutions in the non-monotonic transformation condition, the categories were coded in sequence except for one pair of adjacent categories which were

TABLE 1

Monte Carlo Study Conditions and Purposes

Error	Transformation	Assumed Measurement Characteristics			
		nominal continuous	nominal discrete	ordinal discrete	interval discrete
0%	interval				A*
	ordinal 1			A*	B
	ordinal 2			A	B
	nominal	A*	B	B	B
	binary	A*	B	B	B
10%	interval				C*
	ordinal 1			C*	D
	ordinal 2			C	D
	nominal	C*	D	D	D
	binary	C*	D	D	D
25%	interval				C*
	ordinal 1			C*	D
	ordinal 2			C	D
	nominal	C*	D	D	D
	binary	C*	D	D	D

Purposes:

A The recovery of information in data without random error when the correct measurement level is assumed

B The effects of assuming an incorrect (strong) measurement level with no error in the data

C Robustness to error when the correct measurement level is assumed

D Robustness to error when an incorrect (strong) measurement level is assumed

* Correct measurement characteristics for that transformation

reversed. This gives a strong correlation between the "true" order of the categories and the coded order as suggested by Young and Null [1978].

As stated previously, sets of true parameter values were generated randomly prior to adding random and systematic error. Each set of true parameters had 2 components, 10 scales, 15 objects, and 20 subjects. This made F a 10×2 matrix, X a 15×2 matrix, and W a 20×2 matrix. The elements of F and X varied between 2 and -2, and the elements of W between 0 and 1. This produced data between 8 and -8. The elements of F, W, and X were generated from uniform distributions within these intervals. After the data was generated, the configurations were normalized according to the constraints appropriate to subject conditional data (see Normalization). These normalized configurations were used as the original configurations in the measure of fit (see Results). After normalization the elements of both F and X had an expected value of 0.0 and an expected variance of 1.0. The elements of W had an expected value of .866 and an expected variance of 0.250. In all analyses iterations were continued until improvement in stress, the square root of the loss function (21), was less than .001.

Results. A measure of fit similar to those proposed by Carroll [cf., Shepard, 1966] and Lingoes and Schonemann [1974] was used to evaluate how well the configurations derived by ALSCOMP3 recovered the original structures. This measure of fit between the derived configuration A and the original configuration B is given by

$$r_c = \left(1 - \frac{\text{tr}(\hat{A} - B)'(\hat{A} - B)}{\text{tr}B'B}\right)^{1/2} \tag{28}$$

where $\hat{A}$ is related to A by a central dilation and/or a permutation which maximizes r_c. The measure of fit proposed by Carroll is identical to (28) except that $\hat{A}$ is an affine transformation of A which maximizes r_c. The measure proposed by Lingoes et al. allows a central dilation, a permutation, and an orthogonal rotation. Because the orientations of F, W, and X are fixed in the weighted model, neither measure is appropriate. However, the optimally scaled data, Z^*, are invariant to central dilations and joint permutations of F, W, and X.

The left side of Table 2 presents the results for the correct measurement assumption-no error conditions. The table gives the value of r_c for F, W, and X for each of the three replications. At first glance it appears that recovery of the original structure was poor except in the interval condition. However, these results are misleading. Closer inspection of Table 2 reveals that the variance of r_c is quite high for many conditions.

The cause for this condition-dependent degree of recovery is not immediately apparent. But, closer inspection of the data reveals that this results from poorly constructed initial configurations obtained by the algebraic solution outlined above. In fact, when there is systematic distortion in the data, the algebraic initialization procedure is likely to yield parameter values which are close to a local minimum (see Sands, Note 9, for details). In

TABLE 2

A Comparison of the Fit Between the Derived and Original Configurations for Random and Algebraic Starting Configurations for Those Conditions Where the Algebraic Start Resulted in Poor Fit

Transformation (correct measurement assumption)	Case	Starting Configuration: Algebraic					Starting Configuration: Random				
		ϕ_n^2	(Iter)	r_F	r_W	r_X	ϕ_n^2	(Iter)	r_F	r_W	r_X
Interval	1	.00	(1)	1.000	1.000	1.000					
(Interval	2	.00	(1)	1.000	1.000	1.000					
Discrete)	3	.00	(1)	1.000	1.000	1.000					
Ordinal 1	1	.01	(41)	.999	1.000	.999					
(Ordinal	2	.01	(67)	.450	.954	.617	.01	(19)	1.000	1.000	1.000
Discrete)	3	.01	(16)	1.000	1.000	1.000					
Ordinal 2	1	.02	(38)	.576	.904	.534	.01	(18)	1.000	1.000	1.000
(Ordinal	2	.02	(17)	.614	.995	.494	.01	(19)	1.000	1.000	.999
Discrete)	3	.01	(24)	1.000	1.000	1.000					
Nominal	1	.20	(14)	.656	.820	.522	.01	(4)	.995	.999	.994
(Nominal	2	.24	(9)	.684	.827	.615	.03	(49)	.441	.668	.153
Continuous)	3	.04	(69)	.284	.924	.276	.01	(5)	.997	.999	.998
Binary	1	.02	(13)	.935	.994	.947					
(Nominal	2	.07	(29)	.643	.904	.500	.02	(11)	.983	.997	.919
Continuous)	3	.01	(6)	.946	.989	.952					

addition, the presence of random error in the data had little effect on the initialization procedure and recovery of the original structures.

Because of the high variability in the accuracy of recovery of the original structures in numerous conditions, it is difficult to assess the effects of making incorrect measurement assumptions and the presence of random error. For this reason, and to further test the proposed explanation for the high variability, random initial configurations were used in the reanalysis of those cases which produced poor results in the conditions with high variability and/or large amounts of systematic distortion. Some of these cases were reanalyzed with random initial configurations in both the correct and incorrect measurement assumption conditions (three cases in the nominal condition and one case in the binary condition). The remaining cases were reanalyzed with random initial configurations *only* in the correct measurement assumption conditions (one case in the 3rd power ordinal condition and two cases in the 7th power ordinal condition). The ordinal data was not reanalyzed using random initial configurations in the incorrect measurement assumptions conditions for two reasons: the variability in the incorrect measurement assumption conditions was quite low; and preliminary analyses, using random initial configurations, resulted in recovery that was the same or worse than that found using the algebraic initialization procedure.

While the design of the reanalyses is somewhat ad hoc, the results to be reported are based on analyses using random initial configurations where performed. Thus, the results are partly based on random and partly based on algebraic initial configurations. We report such results because it seems most reasonable to use the best initial configuration procedure for each condition. This is particularly true since the user has the option to use either initialization procedure.

The results for the reanalyses in the correct measurement assumption-no error conditions are presented on the right side of Table 2. It is evident that the results of reanalyses based on random initial configurations are better than those based on the algebraic solution. This result supports the proposed explanation for the poor recovery in conditions with large amounts of systematic distortion. The same pattern of results is found for the 10% and 25% error conditions (reported in Sands, Note 9). In fact, on the basis of these results the user of ALSCOMP3 is strongly urged to try both a random initial configuration and the algebraic procedure when the data is assumed to be anything but interval.

Hopefully the two initialization procedures will result in the same final solutions. If not, it is recommended that the user use a number of random initial configurations as starting points. If the final solutions resulting from the different random initial configurations are similar, then one can assume that the solution associated with the algebraic start was in fact a local minimum. On the other hand, if the final solutions resulting from random starts are markedly different, the user should be cautious about any solution whose fit is markedly worse or better than the others. And in this case choice of solutions is best determined by theory underlying the particular study, with some preference for better fitting solutions.

Recall that the first purpose of the Monte Carlo study was to assess the accuracy of the recovered configurations from the analysis of error-free data with correct measurement assumptions. Table 3 presents the results of the no-error conditions. The results of interest are those farthest to the left in each row. For the interval and both ordinal data sets the recovered configurations were nearly identical to the originals. However, for nominal and binary data the recovery was less than perfect.

For the nominal data the standard deviation of the squared fit was quite high (sd = .460). The reason for the high variability in accuracy is apparent from inspection of the original data. In the analysis of one of the three cases the recovery of the original struc-

TABLE 3

Root Mean Square and Standard Deviation of Squared Fit Between the Derived and Original Configurations in the 0% Error Condition Using Some Random Initial Configurations

Configuration	Transformation	Assumed Measurement Characteristics: nominal continuous RMS	sd	nominal discrete RMS	sd	ordinal discrete RMS	sd	interval discrete RMS	sd
F	interval							1.0000	.000
	ordinal 1					.9995*	.001	.8830	.117
	ordinal 2					.9998**	.000	.7190	.126
	nominal	.8521***	.460	.8959***	.326	.6101	.082	.5984	.072
	binary	.9549*	.049	.9501*	.055	.9501*	.055	.9501*	.055
W	interval							1.0000	.000
	ordinal 1					.9999*	.000	.9690	.053
	ordinal 2					1.0000**	.000	.9563	.053
	nominal	.9024***	.319	.9277***	.239	.7793	.074	.8054	.112
	binary	.9932*	.008	.9952*	.008	.9952*	.008	.9952*	.008
X	interval							1.0000	.000
	ordinal 1					.9995*	.000	.8850	.172
	ordinal 2					.9996**	.000	.6971	.186
	nominal	.8180***	.559	.8283***	.528	.5840	.034	.5806	.029
	binary	.9395*	.033	.9318*	.040	.9318*	.040	.9318*	.030

*Includes one case using a random initial configuration
**Includes two cases using random initial configurations
***Includes three cases using random initial configurations

tures was very poor (r_f = .4412, r_w = .6676, r_x = .1526). For the other two cases the measures of fit between the derived configurations and the originals were always greater than .9900.

Poor recovery for this one case of nominal data is most probably a result of the low ratio of data points to model parameters for this condition. A nominal discrete transformation requires the estimation of a model parameter for each category for each data partition. Given that the data are matrix conditional and there are five categories, the estimation of 100 additional model parameters is required for the nominal transformation. On the other hand, an interval transformation requires the estimation of two model parameters for each data partition. This means that the data-parameter ratio for nominal assumptions was approximately 15:1, whereas the ratio for interval assumptions was approximately 27:1.

For binary data, the standard deviations of squared fit were low. In fact, although recovery was not perfect, recovery for the three cases was above .9300 (see Table 3).

The second question addressed by the Monte Carlo study concerned the effect of assuming inappropriately strong measurement characteristics with error-free data. A general inspection of Table 3 strongly suggests that the accuracy of recovery of the original configurations was considerably worse with inappropriately strong measurement level assumptions. Also, from comparing different levels of systematic distortion, it appears that the consequences of incorrect measurement level assumptions were more severe with increasing levels of systematic distortion (interval-ordinal 1, ordinal 2).

However, inspection of the results for both nominal and binary data reveal a quite interesting finding. In both conditions incorrect process assumptions had little effect.

Whereas, for nominal data, incorrect level assumptions produced considerable deterioration in recovery of the original configurations, for binary data incorrect level assumptions had no effect (see Table 3). This reflects the fact that binary data is a special case of nominal, ordinal, and interval data.

With the availability of a general procedure such as ALSCOMP3 it is unlikely that a researcher will have to assume inappropriately strong measurement characteristics. However, real data always contains a certain amount of random error. This emphasizes the importance of the third question addressed by this study: How robust is the ALSCOMP3 algorithm to the presence of random error when the correct measurement characteristics are assumed? Tables 4 and 5 present the relevant results from the 10% and 25% random error conditions.

The results from these conditions are quite encouraging. It appears that there was little, if any, effect of these levels of random error on the accuracy of recovery of the original configurations (see Tables 3, 4, and 5). For nominal data it appears that the presence of random error increased the accuracy of recovery. However, inspection of the original data revealed that this was not a reliable effect. As mentioned previously, in the no-error condition recovery was poor for one of the three replications. In the 10% error condition recovery for that case was nearly perfect ($r_c > .9900$), but recovery for a different case was quite poor ($r_f = .0467$, $r_x = .9573$, $r_x = .2972$). Even more interesting, in the 25% error condition recovery was nearly perfect for both of the above mentioned cases ($r_c > .9900$). However, in this condition recovery of the original configurations for the third case was poor ($r_f = .9052$, $r_w = .9328$, $r_x = .5432$). These results provide further support for the contention that with a low data-parameter ratio (15:1) recovery of the original configurations

TABLE 4

Root Mean Square and Standard Deviation of Squared Fit Between the Derived and Original Configurations in the 10% Error Condition Using Some Random Initial Configurations

Configuration	Transformation	Assumed Measurement Characteristics: nominal continuous RMS	sd	nominal discrete RMS	sd	ordinal discrete RMS	sd	interval discrete RMS	sd
F	interval							.9999	.000
	ordinal 1					.9992*	.002	.9160	.117
	ordinal 2					.9996**	.000	.7168	.157
	nominal	.8156***	.574	.8160***	.569	.6148	.092	.6040	.082
	binary	.9601*	.046	.9519*	.053	.9519*	.053	.9519*	.053
W	interval							.9999	.000
	ordinal 1					.9999*	.000	.9911	.014
	ordinal 2					.9999**	.000	.9602	.036
	nominal	.9855***	.047	.9981***	.004	.7761	.071	.8016	.108
	binary	.9929*	.009	.9892*	.010	.9892*	.010	.9892*	.010
X	interval							.9999	.000
	ordinal 1					.9992*	.001	.8991	.129
	ordinal 2					.9994**	.001	6963	.198
	nominal	.8325***	.524	.8407***	.491	.5845	.034	.5788	.024
	binary	.9508*	.033	.9376*	.044	.9376*	.044	.9376*	.044

*Includes one case using a random initial configuration
**Includes two cases using random initial configurations
***Includes three cases using random initial configurations

TABLE 5

Root Mean Square and Standard Deviation of Squared Fit Between the Derived and Original Configurations in the 25% Error Condition Using Some Random Initial Configurations

Configuration	Transformation	Assumed Measurement Characteristics: nominal continuous RMS	sd	nominal discrete RMS	sd	ordinal discrete RMS	sd	interval discrete RMS	sd
F	interval							.9996	.000
	ordinal 1					.9984*	.003	.9192	.076
	ordinal 2					.9987**	.002	.7243	.168
	nominal	.9685***	.103	.9664***	.109	.6188	.088	.6097	.078
	binary	.9710*	.035	.9610*	.023	.9610*	.023	.9610*	.023
W	interval							.9998	.000
	ordinal 1					.9996*	.001	.9873	.017
	ordinal 2					.9996*	.001	.9352	.045
	nominal	.9777***	.074	.9815***	.061	.7613	.066	.7952	.119
	binary	.8556*	.452	.8823*	.363	.8823*	.363	.8823*	.363
X	interval							.9995	.001
	ordinal 1					.9984*	.002	.9166	.121
	ordinal 2					.9989**	.003	.7118	.200
	nominal	.8736***	.405	.8707***	.413	.5933	.040	.5938	.025
	binary	.9747*	.027	.9586*	.049	.9586*	.049	.9586*	.049

*Includes one case using a random initial configuration
**Includes two cases using random initial configurations
***Includes three cases using random initial configurations

underlying nominal data is unreliable and dependent on unknown qualities of the particular data set.

In only one situation did the presence of random error seem to have a detrimental effect on the recovery of the original information. In the 25% error condition, for binary data, recovery of W was considerably worse than in the zero and 10% error conditions. However, inspection of Table 5 shows a high standard deviation for the squared measure of fit over the three replications. This apparent deterioration in recovery resulted from poor recovery ($r_w = .4386$) in one of the three replications. Again, recovery for the other two was nearly perfect ($r_c > .9950$).

The final question addressed by this Monte Carlo study concerned the worst possible situation, when one assumes inappropriately strong measurement characteristics and the data contains random error. From the results presented in Tables 3–5 it is apparent that there was no interaction between incorrect measurement assumptions and the presence of random error. All inaccuracy in the recovery of the original information could be accounted for by the incorrect measurement assumptions. Even in this situation the presence of 10% or 25% random error had no effect.

The ALSCOMP3 algorithm appears to be robust to the presence of random error of these magnitudes. In fact, the algorithm seems to be so robust that one might suspect there was no random error in the data. However, this possibility can be rejected by examining the stress values for the various conditions of the Monte Carlo study (see Table 6). Recall that squared stress can be interpreted as the amount of variance unaccounted for in the optimally scaled data. Thus, the value of stress indicates the standard deviation of the estimated error. Notice that for interval data with interval assumptions stress is quite

TABLE 6

Root Mean Square and Standard Deviation of STRESS Over Three Replications Using Some Random Initial Configurations

Error	Transformation	Assumed Measurement Characteristics: nominal continuous RMS	sd	nominal discrete RMS	sd	ordinal discrete RMS	sd	interval discrete RMS	sd
0%	interval							.000	.000
	ordinal 1					.009*	.000	.309	.032
	ordinal 2					.007**	.000	.324	.090
	nominal	.022***	.001	.375***	.127	.618	.005	.670	.027
	binary	.018*	.000	.485*	.049	.485*	.049	.485*	.049
10%	interval							.094	.000
	ordinal 1					.067*	.000	.310	.027
	ordinal 2					.066**	.001	.393	.158
	nominal	.047***	.003	.408***	.104	.631	.011	.719	.015
	binary	.032*	.000	.502*	.016	.502*	.016	.502*	.016
25%	interval							.228	.002
	ordinal 1					.182*	.001	.446	.046
	ordinal 2					.181**	.001	.596	.183
	nominal	.143***	.027	.410***	.074	.673	.021	.766	.022
	binary	.048*	.000	.572*	.006	.572*	.006	.572*	.006

*Includes one case using a random initial configuration
**Includes two cases using random initial configurations
***Includes three cases using random initial configurations

close to the standard deviation of the actual error in the data. This suggests that the ALSCOMP3 algorithm does not fit the model to the random error.

Considering recent Monte Carlo studies of non-metric principal component algorithms for two-way data [Kruskal & Shepard, 1974; Sands, Note 9, footnote 21], it is quite surprising that ALSCOMP3 appears not to fit the model to random error. However, the data-parameter ratios in this Monte Carlo study are much higher than those in previous studies. For example, for interval assumptions this ratio is approximately 27:1 as compared to 4:1 in the Kruskal and Shepard Monte Carlo study. Our highly over-determined situation is obtained by the addition of the third way in the data with relatively few additional model parameters.

Given the higher data-parameter ratios obtained with three-way models, it appears that random error of the levels employed in this study will have little effect on recovery. However, if higher levels of error were used or the data-parameter ratio reduced, we expect that the recovery of the original structure would be detrimentally affected. This contention is supported by the error dependent variability in recovery for the nominal data sets. As previously mentioned, the data-parameter ratio under nominal assumptions is 15:1. This reduction appeared to result in highly variable recovery for each of the three replications as a function of the level of random error.

Summary. This small Monte Carlo evaluation revealed several important points. First, when there is systematic distortion in the data it is advisable to perform analyses using both the random and algebraic initial configurations. In fact, the results of this study suggest that more reliable results may be obtained with the random initial configurations. Second, ALSCOMP3 recovers known configurations for interval and ordinal data when the correct measurement characteristics are assumed. This is true for data with or

without random error. Third, when the data is nominal or binary, recovery of original information is generally good, but less reliable than with interval or ordinal data. However, increasing the data-parameter ratio should remedy this problem. Fourth, assuming inappropriately strong measurement characteristics has a detrimental effect on the recovery of the original configurations. This effect appears to be most prominent in the analysis of data containing large amounts of systematic distortion. Finally, ALSCOMP3 is robust to the presence of random error. This was true under all conditions. The presence of 10% or 25% random error had little effect on the recovery of the original configurations for binary, ordinal, or interval data analyzed with correct or incorrect measurement assumptions.

Real Data and the Weighted Model

The data analyzed consisted of 28 subjects' ratings of 4 political roles (Ideal President, Nixon, McGovern, Successful Lawyer) on 77 adjectives (see Sands, Note 9, for list of adjectives). This data is a subset of 93 subjects' ratings of 15 political roles on 77 adjectives (the data were collected at Rutgers University by Drs. Charles Schmidt and Andrea Sedlak). The subjects can be classified by two factors, time and presidential choice. Half the subjects completed the rating task in November, 1972; and half in August, 1973. The first time was prior to major publicity concerning the Watergate scandal; while the second was after the peak of this publicity. Both groups of subjects were asked to indicate their presidential choice between Nixon and McGovern. Each subject filled in the 77×4 trait by role matrix by placing a check when they felt a trait described a particular role. This gave a $28 \times 77 \times 4$ three-way, binary data matrix.

This particular data set was selected for a number of reasons. First, the relationship among traits used to describe others has been studied with factor analysis and MDS techniques. In general, the results of these studies are consistent. The structural representation of the trait associations have two or three dimensions. These are interpretable by some combination of evaluative, activity, and potency dimensions [Rosenberg & Sedlak, 1972]. This allows the comparison of the common trait space from the ALSCOMP3 analysis with those of previous studies.

Second, there is reason to believe that there might be individual differences in the perception of the political roles. Sherman and Ross [1972] reported systematic individual differences in the perception of political figures. In addition, if there are individual differences in role perception, they would likely be related to the Watergate publicity and individual political orientations. Furthermore, the weighted model can be interpreted as describing how individuals differentially represent roles in a common trait space. This interpretation is consistent with previous research which indicates few systematic individual differences in trait associations [Messick & Kogan, 1966; Pederson, 1965; Sherman, 1972; Walters & Jackson, 1966].

Finally, data of this nature typically present problems for other algorithms. The data is binary, subject conditional, and three-way. The use of other multidimensional techniques would involve indirect similarity measures, overly stringent measurement assumptions, and/or ignoring the conditional nature of the data. Any of these can cause distortion and loss of information. On the other hand, the results of the Monte Carlo evaluation of ALSCOMP3 suggest that, with the correct measurement assumptions, recovery of the structure underlying binary, subject conditional data is quite good.

Results. Solutions, using continuous measurement process assumptions, were obtained for one to three dimensions. Both algebraic and random initial configurations were used. The derived configurations and stress values from both sets of analyses were nearly identical. The results presented are from the solutions using the algebraic initial configu-

ration. The squared stress for these solutions were 0.107 for the one dimensional, 0.078 for the two dimensional, and 0.041 for the three dimensional solution. These stress values do not help to decide on the dimensionality. However, informal inspection of the three solutions led us to select the two-dimensional solutions for further interpretation. In particular, the weight space for this solution has a very simple interpretation (as will be seen), whereas the corresponding one- and three-dimensional spaces did not.

Table 7 presents the average, standard deviations, and angle with respect to dimension 1 of the subjects' weights for the four groups. (The averages are weighted averages. Each subject's loadings are weighted by the variance accounted for in that subject's data before averaging. This provides a least squares estimate for each group.) As with MDS models, for conditional data, weights for individuals are not directly comparable. For this reason, the following discussion is based on the relative weighting of the two dimensions by each group. The relative weighting of dimensions is indicated by the angle of the weight vectors given in Table 7 and Figure 1. It is evident from the angles of the weight vectors in Table 7 that there are systematic group differences in the relative weighting of the two dimensions. Inspection of the original configuration showed little overlap between the angles of the weight vectors of the subjects in the four groups (see Figure 1). It appears that McGovern subjects weighted Dimension 2 relative to Dimension 1 more heavily than Nixon subjects. Also, the subjects from 1973 weighted Dimension 2, relative to Dimension 1, more heavily than those from 1972. In fact, the 1972 Nixon group did not use Dimension 2 at all in their trait ascriptions (see Table 7).

Figure 1 presents the loadings of the roles in the two dimensions. Dimension 1 differentiates McGovern from the other three roles. On this dimension, ideal president, Nixon, and successful lawyer are nearly identical (see Figure 1). On the other hand, all four roles are differentiated by Dimension 2. The order of the loadings on this dimension is Ideal president, McGovern, Successful lawyer, and Nixon (from high to low—see Figure 1). So Nixon is closest to ideal president on Dimension 1, and McGovern is closest to ideal president on Dimension 2; while Nixon is farthest from ideal president on Dimension 2, and McGovern is farthest from ideal president on Dimension 1.

Interpretation of these dimensions requires locating axes corresponding to properties which most closely coincide with the two dimensions of the trait configuration. Because the dimensions were correlated ($r = .535$, $\alpha = 58°$), multiple regression was employed to fit the ratings of the 77 traits on each of nine properties, used to interpret previous trait configurations [Rosenberg & Olshan, 1970; Rosenberg & Sedlak, 1972]. These ratings

TABLE 7

Weighted Average and Standard Deviations of Subjects' Weights by Groups

	Dimension 1		Dimension 2		$1-\phi^2$	Angle between weight vector
	Mean	s.d.	Mean	s.d.	(VAF)	and dimension 1
Nixon 1972	1.411	.044	.098	.122	.903	3.3°
Nixon 1973	1.368	.178	.361	.177	.927	11.1°
McGovern 1972	1.300	.047	.557	.091	.937	16.5°
McGovern 1973	1.246	.065	.670	.148	.935	19.5°

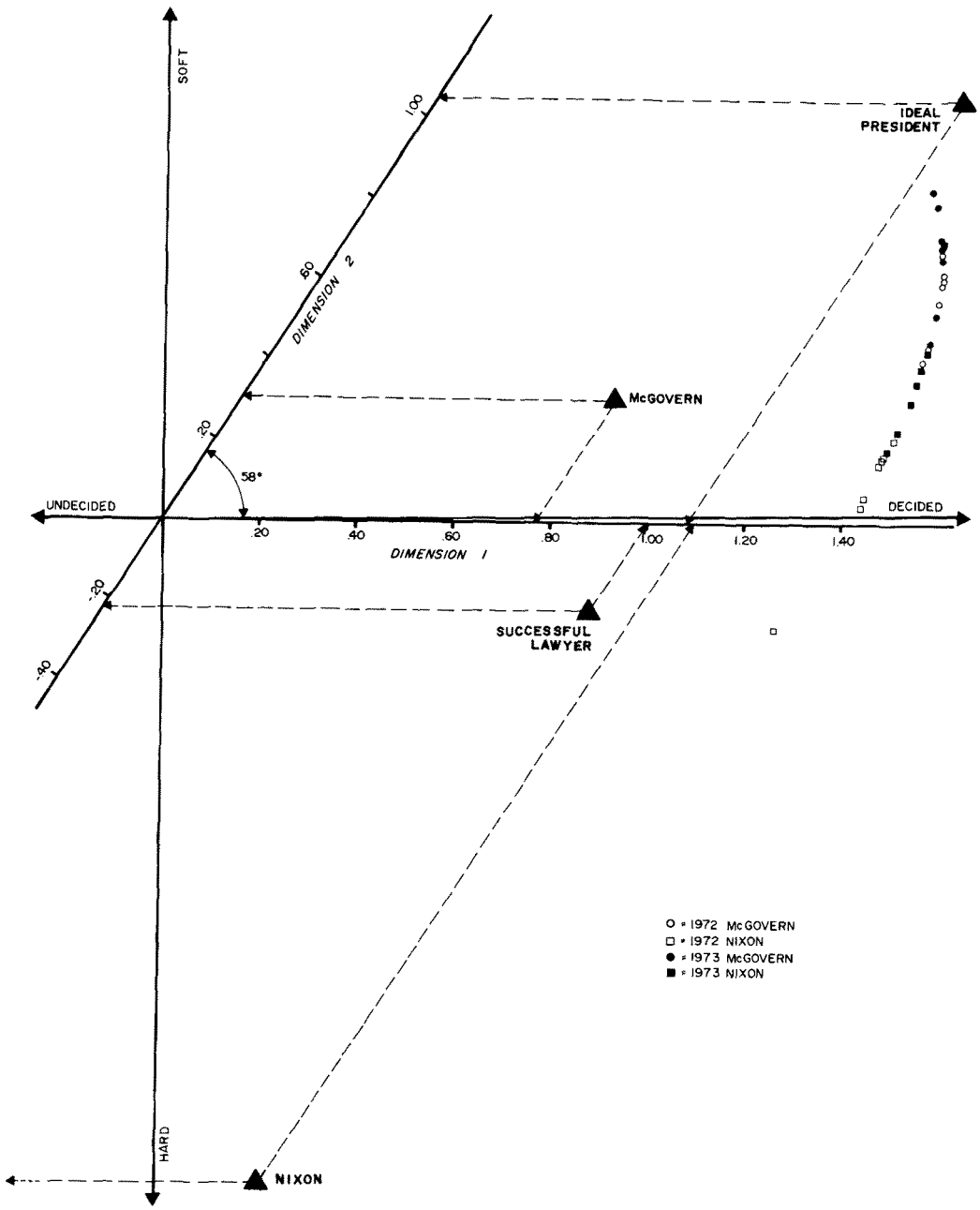

FIGURE 1
Group Differences in the Perception of Political Roles

were obtained in 1975 by Dr. Andrea Sedlak and the first author. One hundred twenty-four subjects at the University of North Carolina rated the 77 traits on one of the nine properties. Median ratings on each property for all 77 traits are used in all subsequent analyses. These median ratings were based on no fewer than 13 subjects. Multiple correlations were obtained between each of the nine properties and the 2-dimensional trait configuration. The multiple correlations were not significant for the two properties *introverted-extroverted* and *impulsive-inhibited.* All the other multiple correlations were

highly significant ($p < .001$; except *active-passive*, $p < .05$). The direction of the *decided-undecided* property coincided closely with Dimension 1. When this property was fit to the two-dimensional solution the angle between this property and Dimension 1 was less than 1° (multiple $r = .600$; univariate $r = .596$ with Dimension 1, and univariate $r = .380$ with Dimension 2). The direction of the *hard-soft* property coincided most closely with Dimension 2 (multiple $r = .467$). The angle between this property and Dimension 2 was 32° (univariate $r = .429$) and 90° to Dimension 1 (univariate $r = .074$). The fact that the directions of these two properties closely coincided with the dimensions of the trait configuration supports the assertion that the orientation of the dimensions of the solution from the weighted model is meaningful.

The interpretation of Dimension 1 in terms of the *undecided-decided* property and Dimension 2 in terms of *hard-soft* fits quite nicely with the group weights and the role loadings (see Table 7 and Figure 1). The 1972 Nixon group did not use the *hard-soft* dimension. Their judgments were based on their perception of McGovern as more *undecided* than an ideal president, Nixon, or a successful lawyer. On the other hand, the McGovern subjects used both dimensions. They perceived an ideal president to be *decisive* but not *hard*. They also perceived McGovern as less *decisive* than Nixon or an ideal president, but saw Nixon as being extremely *hard*. The Watergate publicity resulted in an increase in the salience of the *hard-soft* dimension for both Nixon and McGovern supporters. One possible explanation for this finding is that the media's presentation of Nixon's actions which brought about Watergate and his public reactions to the uncovering of the scandal accentuated his *hardness*. Thus, in 1973 this dimension became more salient to everyone in the country regardless of their political orientation.

Conclusions

The results of both the Monte Carlo evaluation and the analysis of real data suggest that ALSCOMP3 may be a viable algorithm for nonmetric individual differences component analysis.

ALSCOMP3 is flexible in numerous ways. First, it is flexible with regard to the types of data which may be analyzed without violation of measurement level, process, or conditionality assumptions. Second, it is flexible with regard to the models which can be realized within the present framework. By combining the replications and weighted models with the three types of conditionality, three distinct individual differences models are possible. Individual differences can be allowed only in the response process, only in the judgmental process, or in both [cf. Takane et al., 1977]. And any of these models can be incorporated in external analyses. Furthermore, with minor modifications ALSCOMP3 can incorporate the other individual differences models discussed in the introductory sections, except Tucker's TMFA. That is, the generalized subjective metric model (1) would necessitate a stage in the ALS algorithm that found the LS oblique procrustean transformation for each individual [cf. Mulaik, 1972, pp. 293–321].

The generalized subjective metric model is also discussed by Kroonenberg and de Leeuw [1980]. Their approach compliments ours in a number of respects. First, they discuss the general three mode component model [their Tucker 3 model, our TMFA in (1)] in detail, in addition to their consideration of the generalized subjective metric model (Tucker 2 model). In contrast, we are concerned with further restrictions of the generalized subjective metric model which are not considered by Kroonenberg and de Leeuw. Second, the programs presented by Kroonenberg and de Leeuw are only appropriate for metric data, while our algorithm is appropriate for data with a wide variety of measurement characteristics. Finally, while we concentrate on the structural details of the ALSCOMP3 algorithm necessary to accomodate such a wide variety of data, Kroonen-

berg and de Leeuw focus on the mathematical properties of the models and estimation procedures.

ALSCOMP3 is robust to the presence of random error of the magnitude employed in the Monte Carlo study. In fact, under no circumstances in this study did the presence of random error have an appreciable effect on the recovery of the original configurations. With correct measurement assumptions the recovery of these configurations was excellent for interval and ordinal data, and quite good for nominal and binary data. As expected, incorrect (strong) measurement assumptions are detrimental to the recovery of the structure underlying the data.

The analysis of the role data suggests that the model does fit real world data. The analysis of the four roles revealed meaningful systematic individual differences. Also, this application of the ALSCOMP3 model demonstrates that the adjective check-list is a feasible data collection method.

There does appear to be one problem with the ALSCOMP3 algorithm. That is, the algebraic initialization procedure may lead to non-optimal solutions when the data has severe monotonic distortion or is nominal. However, in all conditions where this occurred, the use of random initial configurations alleviated the problem.

The present evaluations focused on subject conditional data. However, the most interesting applications may be to the row conditional case. In addition, row conditional assumptions appear to be the most realistic for much of the data collected in the behavioral sciences. The behavior of the ALSCOMP3 algorithm with this assumption is not known. Hopefully, ALSCOMP3 will not be prone to degeneracies similar to those which occur with the (unfolding) distance model.

Appendix A: Proof of the Relation Between the Normalized and Unnormalized Loss Function

The following proof of the relation between the values which minimize the normalized loss function (17) and those which minimize the unnormalized function (4) was suggested by Yoshio Takane.

Define four different loss functions as

$$\phi_1^2(X) = (X - Y)'(X - Y),$$

$$\phi_2^2(X) = [X - (Y - \bar{Y})]'[X - (Y - \bar{Y})],$$

$$\phi_3^2(X) = \frac{[X - (Y - \bar{Y})]'[X - (Y - \bar{Y})]}{X'X},$$

$$\phi_4^2(X) = \frac{(X - Y)'(X - Y)}{(X - \bar{X})'(X - \bar{X})};$$

where X and Y are column vectors whose mean constant vectors are, respectively, $\bar{X}$ and $\bar{Y}$. Now define vectors X_1, X_2, X_3, and X_4 such that each minimizes the loss function with the same subscript, while respecting the measurement characteristics of the original observations X_0.

The vector X_1 will be some transformation of X_0 which is determined by the assumed measurement characteristics of X_0 [Young, de Leeuw, & Takane, 1976]. For any assumed measurement characteristics of X_0 and associated transformations we know that

(A1) $$\bar{X}_1 = \bar{Y}$$

and

(A2) $$X_2 = X_1 - \bar{Y}.$$

Substituting (A1) into (A2) and averaging gives

$$\bar{X}_2 = 0, \tag{A3}$$

where 0 is a vector of zeros. Furthermore, de Leeuw [Note 10] and Kruskal and Carroll [1969] prove

$$\frac{(Y - \bar{Y})'(Y - \bar{Y})}{X_2'X_2}. \qquad X_3 = X_2 \tag{A4}$$

Now if we define

$$X^* = X_3 + \bar{Y}, \tag{A5}$$

then

$$\phi_4^2(X^*) = \frac{[X_3 - (Y - \bar{Y})]'[X_3 - (Y - \bar{Y})]}{(X^* - \bar{X}^*)'(X^* - \bar{X}^*)}. \tag{A6}$$

From (A3) and (A4) we know $\bar{X}_3 = 0$, then from (A5) $\bar{X}^* = \bar{Y}$, which implies

$$X_3 = X^* - \bar{X}^* \tag{A7}$$

and substituting (A7) into (A6) gives

$$\phi_4^2(X^*) = \frac{[X_3 - (Y - \bar{Y})]'[X_3 - (Y - \bar{Y})]}{X_3'X_3} = \phi_3^2(X_3).$$

By definition X_3 minimizes ϕ_3^2. However, because $\phi_3^2(X_3) = \phi_4^2(X^*)$, X^* must minimize ϕ_4^2, hence

$$X_4 = X^* = X_3 + \bar{Y}. \tag{A8}$$

Now we must express X_4 in terms of X_1. To do this, first substitute (A4) into (A8), which gives an equation expressing X_4 in terms of X_2. By substituting (A2) into this equation we arrive at

$$X_4 = (X_1 - \bar{Y}) \frac{(Y - \bar{Y})'(Y - \bar{Y})}{(X_1 - \bar{X}_1)'(X_1 - \bar{X}_1)} + \bar{Y}.$$

This completes the proof.

Appendix B: Proof of the Maximum Value of Stress

The following is a proof that when the variance of the optimally scaled data is zero, stress (17) is unity and the corresponding elements of D_1^{-1} or d_{in}^{-1} (19–25) are zero.

The variance of the adjusted optimally scaled data [Z^* in (17)] for any partition can only be zero if it is a vector of constants. This can only occur if the vector of optimally scaled data which minimizes the unnormalized problem is also a vector of constants. Under these conditions a_p (18) is undefined and the appropriate adjustment cannot be performed. Therefore, it is necessary to express normalized stress (17) in terms of the unadjusted optimally scaled data which minimizes unnormalized stress.

Definitions

Define Y as a vector of model estimates, X_1 as the vector of optimally scaled data which minimizes the unnormalized loss function $\phi_1^2(X)$ (defined in Appendix A), and X_4 as the vector of optimally scaled data which minimizes the normalized loss function $\phi_4^2(X)$ (defined in Appendix A). Appendix A proves that $X_4 = aX_1 + B$; where

$$a = \frac{(Y - \bar{Y})'(Y - \bar{Y})}{(X_1 - \bar{X}_1)'(X_1 - \bar{X}_1)},$$

and

$$B = (1 - a)\bar{Y}.$$

Proof

First, define the normalized loss function $\phi_4^2(X_4)$ in terms of variances. Because $\bar{X}_4 = \bar{Y}$ (see Appendix A)

(B1) $$\phi_4^2(X_4) = \frac{\sigma^2_{(X_4 - Y)}}{\sigma^2_{X_4}}.$$

But

(B2) $$\sigma^2_{(X_4 - Y)} = \sigma^2_{X_4} - 2\sigma_{X_4 Y} + \sigma^2_Y,$$

where $\sigma_{X_4 Y}$ is the covariance between X_4 and Y. Recall that $X_4 = aX_1 + B$. Thus

(B3) $$\sigma^2_{X_4} = a^2\sigma^2_{X_1} \text{ and } \sigma_{X_4 Y} = a\sigma_{X_1 Y};$$

and substituting (B3) into (B2) gives

(B4) $$\sigma^2_{(X_4 - Y)} = a^2\sigma^2_{X_1} - 2a\sigma_{X_1 Y} + \sigma^2_Y.$$

Substituting (B3) and (B4) into (B1) gives

(B5) $$\phi_4^2(X_4) = \frac{a^2\sigma^2_{X_1} - 2a\sigma_{X_1 Y} + \sigma^2_Y}{a^2\sigma^2_{X_1}},$$

(B6) $$\phi_4^2(X_4) = \frac{a^2\sigma^2_{X_1}}{a^2\sigma^2_{X_1}} - \frac{2a\sigma_{X_1 Y}}{a^2\sigma^2_{X_1}} + \frac{\sigma^2_Y}{a^2\sigma^2_{X_1}},$$

(B7) $$\phi_4^2(X_4) = 1 - \frac{2\sigma_{X_1 Y}}{a\sigma^2_{X_1}} + \frac{\sigma^2_Y}{a^2\sigma^2_{X_1}}.$$

However, a can be expressed as

$$a = \frac{\sigma^2_Y}{\sigma^2_{X_1}}.$$

Substituting this expression for a into (B5) gives

$$\phi_4^2(X_4) = 1 - \frac{2\sigma_{X_1 Y}}{\sigma^2_Y} + \frac{\sigma^2_{X_1}}{\sigma^2_Y}.$$

Finally, as $\sigma^2_{X_1} \to 0$ so does $\sigma_{X_1 Y} \to 0$. Therefore,

(B8) $$\lim_{\substack{\sigma^2_{X_1} \to 0 \\ \sigma^2_{X_1 Y} \to 0}} \phi_4^2(X_4) = 1 - 0 - 0 = 1.$$

This completes the proof.

This proof assumes the σ^2_Y is greater than zero. Under certain conditions σ^2_Y can be zero. In this case all terms in (B6) are undefined and the value of the expression in the

limit is also undefined. This suggests that even with the use of the normalized loss function certain degenerate solutions may occur. Whether or not a degenerate solution occurs will depend on the relative rates that σ_{X_1Y}, σ_Y^2, and $\sigma_{X_1}^2$ approach zero. These determine whether the loss function is decreasing or increasing as σ_Y^2 approaches 0.

To prove the corresponding element of D_1^{-1} or d_{in}^{-1} is zero when the variance of the unadjusted optimally scaled data is zero it is necessary to express d^{-1} in terms of the variance of X_1. Recall that

$$d^{-1} = \frac{1}{\sigma_{X_4}^2}$$

and

$$\sigma_{X_4}^2 = a^2\sigma_{X_1}^2 .$$

Combining these two expressions with the expression for a gives

$$d^{-1} = \frac{\sigma_{X_1}^2}{(\sigma_Y^2)^2}$$

and

$$\lim_{\sigma_{X_1}^2 \to 0} d^{-1} = 0 .$$

This completes the proof.

REFERENCE NOTES

1. Harshman, R. A. *Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimode factor analysis* (Working Papers in Phonetics No. 16). Los Angeles: University of California, 1970.
2. Jennrich, R. A. *A generalization of the multidimensional scaling model of Carroll and Chang* (Working Papers in Phonetics No. 22). Los Angeles: University of California, 1972.
3. Young, F. W., de Leeuw, J., & Takane, Y. *Quantifying qualitative data* (L. L. Thurstone Psychometric Laboratory Report, No. 149). Chapel Hill, North Carolina: The L. L. Thurstone Psychometric Laboratory, July 1978.
4. Young, F. W., de Leeuw, J., & Takane, Y. *Multidimensional scaling: Theory and method.* Book in preparation, 1981.
5. Young, F. W. *Conjoint scaling* (L. L. Thurstone Psychometric Laboratory Report, No. 118). Chapel Hill, North Carolina: The L. L. Thurstone Psychometric Laboratory, April, 1973.
6. Schönemann, P. H., Carter, F. S., & James, W. L. Contributions to subjective metrics scaling: I. *COSPA, a fast method for fitting and testing Horan's model and an empirical comparison with INDSCAL and ALSCAL* (Institute for Research in the Behavioral, Economic, and Management Sciences, Report No. 587). West Lafayette, Indiana: Krannert Graduate School of Management, Purdue University, December, 1976.
7. *EISPACK, an Eigensystem Subroutine.* Argonne National Laboratory, Stanford University and University of Texas Center for Numerical Analysis, 1974.
8. Roskam, E. E. *Data theory and algorithms for nonmetric scaling* (parts 1 and 2). Unpublished manuscript, Catholic University, Nijmegen, The Netherlands, 1969.
9. Sands, R. *Component models for three-way data: ALSCOMP3, an alternating least squares algorithm with optimal scaling features.* Unpublished Masters Thesis, University of North Carolina at Chapel Hill, 1978.
10. de Leeuw, J. *A normalized cone regression approach to alternating least squares algorithms.* Unpublished paper, Department of Data Theory, University of Leiden, The Netherlands, 1977.

REFERENCES

Abelson, R. P. Scales derived by considerations of variance components in multi-way tables. In H. Gulliksen & S. J. Messick (Eds.), *Psychological scaling: Theory and applications.* New York: Wiley, 1960, pp. 169–186.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multi-trait multi-method matrix. *Psychological Bulletin,* 1959, *56,* 81–105.

Carroll, J. D. Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences.* New York: Seminar Press, 1972.

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.

Cattell, R. B. The three basic factor-analysis research designs: Their interrelations and derivatives. *Psychological Bulletin,* 1952, *49,* 499–500.

de Leeuw, J., Young, F. W., & Takane, Y. Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika,* 1976, *41,* 471–503.

Eckart, C., & Young, G. The approximation of one matrix by another of lower rank. *Psychometrika,* 1936, *3,* 211–218.

Guttman, L. What lies ahead of factor analysis. *Education and Psychological Measurement,* 1958, *18,* 497–515.

Israelsson, A. Three-way (or second order) component analysis. In H. Wold & E. Lyttkens (Eds.), Nonlinear iterative partial least squares (NIPALS) estimation procedures. *Bulletin of International Statistical Institute,* 1969, *43,* 29–47.

Jones, L. E., & Young, F. W. The structure of a social environment: A longitudinal individual differences scaling of an intact group. *Journal of Personality and Social Psychology,* 1972, *24,* 108–121.

Jöreskog, K. G. Statistical analysis of sets of congeneric tests. *Psychometrika,* 1971, *36,* 109–133.

Kroonenberg, P., & de Leeuw, J. Principal components analysis of three mode data by means of alternating least squares algorithms. *Psychometrika,* 1980, *45,* 69–97.

Kruskal, J. B. Nonmetric multidimensional scaling. *Psychometrika,* 1964, *29,* 1–27; 115–129.

Kruskal, J. B., & Carroll, J. D. Geometric models and badness-of-fit functions. In P. R. Krishnaiah, *Multivariate analysis,* (Vol. 2). New York: Academic Press, 1969.

Lingoes, J. C., & Schonemann, P. H. Alternative measures of fit for the Schonemann-Carroll Matrix Fitting Algorithm. *Psychometrika,* 1974, *39,* 423–427.

Levin, J. Three-mode factor analysis. *Psychological Bulletin,* 1965, *64,* 442–452.

Messick, S., & Kogan, N. Personality consistencies in judgment: Dimensions of role constructs. *Multivariate Behavioral Research,* 1966, *1,* 165–175.

Mulaik, S. A. *The foundations of factor analysis.* New York: McGraw Hill, 1972.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning.* Urbana: University of Illinois Press, 1957.

Pederson, D. M. The measurement of individual differences in perceived personality-trait relationships and their relation to certain determinants. *Journal of Social Psychology,* 1965, *65,* 233–258.

Rosenberg, S., & Olshan, K. Evaluative and descriptive aspects in personality perception. *Journal of Personality and Social Psychology,* 1970, *16,* 619–626.

Schönemann, P. H. An algebraic solution for a class of subjective metric models. *Psychometrika,* 1972, *37,* 441–451.

Shepard, R. N. Metric structures in ordinal data. *Journal of Mathematical Psychology,* 1966, *3,* 287–315.

Sherman, R. Individual differences in perceived trait relationships as a function of dimensional salience. *Multivariate Behavioral Research,* 1972, *7,* 109–129.

Sherman, R., & Ross, L. B. Liberal-conservatism and dimensional salience in the perception of political figures. *Journal of Personality and Social Psychology,* 1972, *23,* 120–127.

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

Takane, Y., Young, F. W., & de Leeuw, J. The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika,* 1978, *43,* 279–281.

Tucker, L. R. Implications of factor analysis of three way matrices for measurement of change. In C. W. Harris (Eds.), *Problems in measuring change.* Madison: University of Wisconsin Press, 1963, pp. 122–137.

Tucker, L. R The extension of factor analysis to three dimensional matrices. In H. Gulliksen & N. Frederiksen (Eds.), *Contributions to mathematical psychology.* New York: Holt, Rinehart, and Winston, 1964, pp. 110–119.

Tucker, L. R. Some mathematical notes on three mode factor analysis. *Psychometrika,* 1966, *31,* 279–312.

Walters, H. A., & Jackson, D. N. Group and individual regularities in trait inference: A multidimensional scaling analysis. *Multivariate Behavioral Research,* 1966, *1,* 145–163.

Wish, M., Deutsch, M., & Biener, L. Differences in perceived similarity of nations. In A. K. Romney, R. N.

Shepard, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications to the behavioral sciences* (Vol. 2). New York: Seminar Press, 1972, pp. 290–314.

Wish, M., Deutsch, M., & Kaplan, S. J. Perceived dimensions of interpersonal relations. *Journal of Personality and Social Psychology,* 1976, *33,* 409–420.

Young, F. W. Nonmetric multidimensional scaling: Recovery of metric information. *Psychometrika,* 1970, *35,* 455–474.

Young, F. W. A model for polynomial conjoint analysis algorithms. In R. N. Shepard, A. K. Romney, & S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences.* New York: Academic Press, 1972.

Young, F. W., de Leeuw, J., & Takane, Y. Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika,* 1976, *41,* 505–529.

Young, F. W., & Null, C. H. Multidimensional scaling of nominal data: The recovery of metric information with ALSCAL. *Psychometrika,* 1978, *43,* 367–379.

Manuscript received 11/16/78
First revision received 4/27/79
Final version received 10/1/79