Behaviormetrika
1979, No.6, 45-56

NONMETRIC COMMON FACTOR ANALYSIS: AN ALTERNATING LEAST SQUARES METHOD WITH OPTIMAL SCALING FEATURES

Yoshio Takane,* Forrest W. Young**
and Jan de Leeuw***

We describe a convergent procedure for fitting the common factor analysis model to multivariate data whose variables may be nominal, ordinal or interval. Any mixture of measurement levels is permitted. There may be any pattern of missing data. As distinguished from previous work, the nonmetric relations (nominal or ordinal) are assumed on the *raw* observations (not on the correlations), and the model fitted is the *common* factor analysis model (not the principal components model) which isolates common from unique factor variation. The computational algorithm, based on the alternating least squares principle, is monotonically convergent and efficient. An illustrative example is presented.

1. Introduction

One of the most significant contributions of Shepard's [1962] landmark paper on nonmetric multidimensional scaling has been its profound impact on a very broad range of psychometric methods: The fact is, simply, that many quantitative models (not necessarily just distance models) have the potential of being fitted to qualitative (nonmetric) data, while totally respecting the qualitative measurement characteristics of the data. Thus started the "nonmetrization" of models originally designed for metric data, resulting in the current proliferation of nonmetric procedures.

Various linear and nonlinear models have been successfully fit to nonmetric data, including the additive model, the regression model, the principal components model, and individual differences and other variants of the distance model [Kruskal, 1965; Roskam, 1968; Young, 1972; Lingoes, 1973; Kruskal and Shepard, 1974; de Leeuw, Young and Takane, 1976; Young, de Leeuw and Takane, 1976a; Takane, Young and de Leeuw, 1977, 1978]. The most recent development, the alternating least squares (ALS) method, has established a general algorithmic construction scheme in which two steps, optimal model estimation and optimal data scaling, are performed independently [Young, de Leeuw & Takane, 1976b]; i.e., model parameters are estimated by methods which are independent of the assumed measurement (scale) level of data, and the optimal data transformation may be obtained regardless of the specific model being fitted.

Key Words and Phrases: common factor analysis, alternating least squares (ALS).

* Department of Psychology, McGill University, 1205 McGregor Ave. Montreal, Quebec, H3A 1B1, Canada.
** Psychometric Lab, University of North Carolina, Chapel Hill.
*** Department of Data theory, University of Leiden.
Requests for reprints should be addressed to Dr. Yoshio Takane, Dept. of Psychology, McGill University, 1205 McGregor Ave. Mtl, PQ, H3A 1B1, Canada

Among the most popular psychometric models the common factor analysis model has proven to be one of the most difficult models to extend to nonmetric data. The difficulty is undoubtedly related to the fact that, given a set of observations, there is no way to identify the common factor space (or equivalently the unique factor space). Only the correlations between factors and observed variables (factor loadings) and the variance of unique factors can be determined. Consequently a least squares criterion is typically defined in terms of correlations, and not in terms of the observed data themselves. If, however, the prescribed ordinal or nominal relations are assumed on entities (the data) which are not the direct defining components of the least squares loss function (the correlations), the simple monotonic regression algorithm [Kruskal, 1964] for ordinal variables or the least squares nominal transformation [de Leeuw, Young & Takane, 1976] for nominal variables does not apply. Thus, the optimal data transformations become difficult to obtain.

The history of "nonmetric factor analysis" suggests two ways out of this difficulty. Lingoes and Guttman [1967] proposed a "nonmetric factor analysis" in which monotonicity relations are assumed between the observed and predicted correlations. The proposal has one major disadvantage, as Kruskal and Shepard [1974] point out, in that the raw observations must be assumed to be measured at the interval level in order to calculate Pearsonian product moment correlation coefficients. It is not very satisfactory, then, to assume monotonicity on statistics derived on the basis of stronger measurement assumptions. One undesirable result is that invariance of the nonmetric procedure over admissible data transformations will not be obtained. An alternative approach has been taken by Roskam [1968], Young [1972] and Kruskal and Shepard [1974], who assumed monotonicity between raw data and their corresponding predictions, but employed the principal components model instead of the common factor model by discarding the idea of isolating common and unique factor variations. PRINCIPALS [Takane, Young and de Leeuw, 1978] is an extension of these methods to cover data measured at a variety of scale levels (any mixture of levels is permissible) using an alternating least squares approach. While nonmetric principal components analysis is interesting in its own sake, and certainly deserves serious attention, it is not "nonmetric factor analysis" as some have called it. Despite the unfortunate terminological confusion the principal components model is clearly distinct from the factor analysis model. We use the word "common" to emphasize the distinction.

In this paper we show that the *common* factor model can be fit to qualitative data while assuming the nonmetric relations are on the *raw observations*. McDonald, Ishizuka and Nishisato [1977] present a procedure which fits the common factor model to multicategory data. However, the present formulation provides a much more general framework. It permits the common factor model to be fit to multivariate data whose variables may be defined at a mixture of measurement levels. There is no restriction on the mixture of levels: Some variables may be binary, some nominal, others ordinal, some may even be interval. Any mix will do.

2. Method

2.1 Preliminaries

We state the problem of fitting the common factor analysis model to nonmetric data as

$$Y^* \cong \hat{Y} = FA' + U\,, \tag{1}$$

where Y^* is an $N\times n$ matrix of optimally scaled data, $\hat{Y}$ is an $N\times n$ matrix of model estimates, F is an $N\times r$ matrix of common factor scores, A is an $n\times r$ matrix of common factor loadings, U is an $N\times n$ matrix of uniqueness variations, N is the sample size, n is the number of observed variables and r is the number of common factors. We assume the usual factor analytic assumptions,

$$U'\mathbf{1}_N = \mathbf{0}_n'\,, \tag{1-a}$$
$$F'\mathbf{1}_N = \mathbf{0}_r'\,, \tag{1-b}$$
$$F'U = {}_r\mathbf{0}_n\,, \tag{1-c}$$

and

$$U'U/N = D_u^2 \text{ (diagonal)} \tag{1-d}$$

hold where $\mathbf{1}$ and $\mathbf{0}$ are vectors of ones and zeroes, respectively, of subscripted orders, and D_u^2 is an $n\times n$ matrix of uniqueness variances. We make two additional assumptions,

$$F'F/N = I\,, \tag{1-e}$$

and

$$A'A = D \text{ (diagonal)} \tag{1-f}$$

for identification purposes. The "$\cong$" stands for the least squares fit in the sense defined below (Eq. (3)).

It will be useful in the following development to represent each column vector of Y^*, say $\boldsymbol{y}_j^*$ $(j=1, \cdots, n)$, using indicator matrix notation. Let an $N\times n_j$ (n_j is the number of distinct observations for variable j) indicator matrix be G_j, and an n_j-component vector of the optimally scaled (quantified) observation categories be $\boldsymbol{p}_j$, then

$$\boldsymbol{y}_j^* = G_j\boldsymbol{p}_j\,, \quad (j = 1, \cdots, n)\,. \tag{2}$$

The optimization criterion is stated as

$$Q = tr(R^*-\hat{R})^2 \tag{3}$$

where R^* is the matrix of correlations calculated from the optimally scaled data; i.e.,

$$R^* = Y^{*\prime}Y^*/N \tag{4}$$

and $\hat{R}$ is the matrix of predicted correlations; i.e.,

$$\hat{R} = AA' + D_u^2\,. \tag{5}$$

We optimize (3) (for a prescribed r) by alternately obtaining, first, the L.S. estimate of $\hat{R}$ (with the structure defined in (5)) for fixed R*, and, second, the L.S. estimate of R^* (with the structure defined in (4)) under the optional variablewise monotonicity restrictions on Y^* for fixed $\hat{R}$. The procedure is an alternating least squares

algorithm, a class of algorithms which has been proven to be monotonically convergent [Zangwill, 1969] under mild regularity conditions, and which has been successfully used to fit a variety of linear and quadratic models to qualitative data [de Leeuw, Young & Takane, 1976; Young, de Leeuw & Takane, 1976; Takane, Young & de Leeuw, 1977a, 1977b]. We call our procedure FACTALS (nonmetric common *fact*or analysis by the *a*lternating *l*east *s*quares method).

The critical notion to be observed in an ALS algorithm is that each phase of the iterative procedure obtains least squares estimates of a subset of the entire set of parameters, where the estimates are conditional upon the values of the parameters in the other subsets, and that the least squares estimates be derived from a single (common) optimization criterion. Within this general ALS framework, we have considerable freedom as to the choice of the specific numerical methods to solve each of the conditional least squares problems. We now turn to these specific methods.

2.2 Model Estimation Phase

The least squares estimate of $\hat{R}$ (i.e., A and D_n^2) for fixed R^* may be obtained by at least two conventional methods. One is Thomson's [1951] refactoring method and the other is the MINRES method proposed by Harman and Jones [1966]. It is interesting to note that both of these methods are themselves ALS procedures. It is clear that MINRES (with its specific optimization algorithm) is explicitly intended to optimize (3) with fixed R^*. That Thomson's refactoring method is also an ALS procedure is less obvious, though this can be shown in a surprisingly simple way.

By differentiating (3) with respect to A and D_n^2 and setting the results to zero we obtain

$$(R^*-AA'-D_n^2)\,A = {}_n\mathbf{0}_r\,,$$

and

$$\operatorname{diag}(R^*-AA'-D_n^2) = {}_n\mathbf{0}_n\,.$$

Hence we have (by noting the identification restriction $A'A=D$)

$$(R^*-D_n^2)\,A = AD$$

and

$$D_n^2 = \operatorname{diag}(R^*-AA')\,.$$

The first equation is the eigen equation for obtaining the L.S. estimate of A for fixed D_n^2, and the second the L.S. estimate of D_n^2 for fixed A.

We employ Thomson's refactoring method to obtain model estimates because of its simplicity. Its reportedly slow convergence can be accelerated in ways which will be suggested later.

2.3 Optimal Scaling Phase

The optimization of Q with respect to the data transformations is somewhat more involved since, as stated earlier, a simplified quadratic programming procedure such as Kruskal's monotonic regression algorithm cannot be applied for ordinal variables, and the least squares optimal quantification for nominal variables cannot be obtained by simply averaging model predictions whose corresponding observations fall in the same

categories. However, by noting that diag $(\hat{R})=I_n$ and by explicitly imposing the normalization restriction that diag $(R^*)=I_n$, the problem of estimating R^* for fixed $\hat{R}$ may be reduced to successive applciations of a more standard type of quadratic programming procedure.

Let

$$\hat{\boldsymbol{r}}_j^{*\prime} = (r_{1j}^* \cdots r_{(j-1)j}^*,\ r_{(j+1)j}^* \cdots r_{nj}^*)\,,$$

and

$$\hat{\boldsymbol{r}}_j' = (\hat{r}_{1j} \cdots \hat{r}_{(j-1)j},\ \hat{r}_{(j+1)j} \cdots \hat{r}_{nj})\,,$$

then

$$Q = \sum_{j=1}^{n} (\boldsymbol{r}_j^* - \hat{\boldsymbol{r}}_j)'\,(\boldsymbol{r}_j^* - \hat{\boldsymbol{r}}_j) = \sum_{j=1}^{n} Q_j\,, \tag{6}$$

where

$$Q_j = (\boldsymbol{r}_j^* - \hat{\boldsymbol{r}}_j)'(\boldsymbol{r}_j^* - \hat{\boldsymbol{r}}_j)\,. \tag{7}$$

Define Y_j^* as Y^* with j' th column vector deleted, then $\boldsymbol{r}_j^*$ may be written as

$$\hat{\boldsymbol{r}}_j^* = Y_j^{*\prime}\boldsymbol{y}_j^*/N = Y_j^{*\prime}G_j\boldsymbol{p}_j/N = B_j\boldsymbol{p}_j\,,$$

where $B_j=Y_j^{*\prime}G_j/N$. Note that the rank of B_j is at most n_j-1, since we have $Y_j^{*\prime}\mathbf{1}_N=\mathbf{0}_{n-1}$ so that $B_j\mathbf{1}_{n_j}=\mathbf{0}_{n-1}$. The rank of B_j may be smaller than n_j-1, but we assume here that rank $(B_j)=n_j-1$. We assume further that n_j $(j=1, \cdots, n)$ does not exceed n. Q_j is, unfortunately, a function not only of $\boldsymbol{y}_j^*$ but also of Y_j^*. Thus the optimal scaling phase consists of n sub-phases which optimize Q through sequential optimizations of Q_j $(j=1, \cdots, n)$, with the newly estimated $\boldsymbol{p}_j$ promptly fed back into the optimization for the next variable.

In optimizing Q_j with respect to $\boldsymbol{p}_j$ (note that Q_j is a quadratic function of $\boldsymbol{p}_j$) we have to take the following three restrictions on $\boldsymbol{y}_j^*$ (and consequently on $\boldsymbol{p}_j$) into account:

1) Centering restriction

$$\boldsymbol{y}_j^{*\prime}\mathbf{1}_N = \boldsymbol{p}_j'D_j\mathbf{1}_{n_j} = 0\,, \tag{8}$$

where $D_j = G_j'G_j/N$.

2) Normalization restriction

$$\boldsymbol{y}_j^{*\prime}\boldsymbol{y}_j^*/N = \boldsymbol{p}_j'D_j\boldsymbol{p}_j = 1\,. \tag{9}$$

3) Order (cone) restriction

$$H_j\boldsymbol{p}_j \geq \mathbf{0}_{n_j}\,, \tag{10}$$

where the $n_j \times n_j$ matrix H_j defines the ordinal relations on the elements of $\boldsymbol{p}_j$, as follows: If the elements of $\boldsymbol{p}_j$ are arranged in the desired ascending order (p_1 is the smallest category, p_2 the next smallest, etc.), then H_j has the following structure:

$$H_j = \begin{pmatrix} -1 & 1 & 0 & 0 \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 \cdots & 0 & 0 \\ 0 & 0 & -1 & 1 \cdots & 0 & 0 \\ & \cdots & & \cdots & \cdots & \\ 0 & 0 & 0 & 0 \cdots & -1 & 1 \\ 0 & 0 & 0 & 0 \cdots & 0 & 1 \end{pmatrix}.$$

The centering restriction states that the optimally scaled data should have zero means, and the normalization restriction (in conjunction with the centering restriction) states that they should have unit variances. The third restriction is optional, being required only when the variable is ordinal. If the variable is nominal the order restriction is not necessary and the data estimation is greatly simplified. If the variable is numerical (defined at the interval level) then the optimal scaling phase is skipped entirely.

In terms of a quadratic programming procedure the normalization restriction creates a difficulty, since it is not linear (but quadratic) and the standard quadratic programming procedures are designed to optimize a quadratic function under linear constraints. However, several methods have been suggested to deal with quadratic constraints similar to the present case; in these methods the constraints are introduced into an objective function by way of Lagrangean multipliers, which can be solved for by either reducing it to a special type of an eigen-problem [Golub and Saunders, 1969] or by a descent type method such as the Newton-Raphson method [Browne, 1967; Cramer, 1974].

The use of an iterative procedure to incorporate the normalization restriction would be formidable when the variable is ordinal since currently available quadratic programming procedures are iterative themselves. The eigen method would be very time consuming since the solution would have to be obtained on each iteration of a quadratic programming procedure, which in turn would have to be repeated for each ordinal variable on each ALS major iteration.

Fortunately we can reduce the normalization restriction to an essentially nonrestrictive form by a simple modification of the optimization criterion (3). Consider, for fixed $\hat{R}$,

$$Q^* = tr[D_s^{-1/2}(C^*-\hat{R})\,D_s^{-1/2}]^2 \tag{11}$$

where $C^*=Y^{*\prime}Y^*/N$ and $D_s=\mathrm{diag}(C^*)$ for unnormalized Y^*. Note that Q^* is equivalent to Q defined in (3) when the normalization restrictions happen to be met (i.e., if $D_s=I_n$ and $C^*=R^*$). Define $D_{s(j)}$ to be a diagonal matrix of order $n-1$ with the j'th diagonal element d_j of D_s delected. Assume for the moment that $D_{s(j)}=I_{n-1}$ (i.e., the normalization restrictions are satsified for the n–1 variables except for the j'th variable). The Q_j^* corresponding to Q_j in (7) can be written as

$$\begin{aligned} Q_j^* &= \frac{1}{d_j}(\boldsymbol{c}_j^*-\hat{\boldsymbol{r}}_j)'(\boldsymbol{c}_j^*-\hat{\boldsymbol{r}}_j) \\ &= (\boldsymbol{r}_j^*-\hat{\boldsymbol{r}}_j/d_j^{1/2})'(\boldsymbol{r}_j^*-\hat{\boldsymbol{r}}_j/d_j^{1/2}) \\ &= (\boldsymbol{r}_j^*-\tilde{\boldsymbol{r}}_j)'(\boldsymbol{r}_j^*-\tilde{\boldsymbol{r}}_j) \end{aligned}$$

where $\boldsymbol{c}_j^*$ is the j'th column vector of C^* with j' th element deleted and $\tilde{\boldsymbol{r}}_j=\hat{\boldsymbol{r}}_j/d_j^{1/2}$. Hence we may temporarily regard d_j as a fixed constant, obtain the unnormalized $\boldsymbol{y}_j^*$ (which gives $\boldsymbol{c}_j^*$), and then normalize it to meet the noramlization restriction. The dependent variable $\hat{\boldsymbol{r}}_j$ should be adjusted by the same factor. The corresponding j'th row vector of $\hat{R}$ should also be adjusted accordingly to retain the symmetry of $\hat{R}$. Note that the normalized $\boldsymbol{y}_j^*$ would be the unconstrained estimate if in fact $\tilde{\boldsymbol{r}}_j$ were the

dependent variable. Note also that the above solution is subtly different from what we have originally intended, but nonetheless it is consistent with the ALS framework. Since we can always ensure that $D_{s(j)}=I_{n-1}$ for any j and that $D_s=I_n$ throughout the model estimation phase, the model estimation can proceed as if we were optimizing (3).

Furthermore, as it will be clear shortly, the centering restriction in the present case is also nonrestrictive. Since the $n_j \times n_j$ matrix B_j has rank of at most n_j-1, the solution to (7) is not unique without additional restrictions. The centering restriction can be used to uniquely identify the solution. Again the solution can be obtained by first obtaining an unrestricted solution and then by adjusting it to meet the restriction.

What we have described so far suffices for nominal variables. However, if a variable is ordinal then we must impose the order restriction given in Equation 10 in addition to the centering and normalization restrictions given by Equations 8 and 9. Thus, we now turn to the ordinal (cone) restrictions for ordinal variables, which produce a quadratic programming problem.

There are various ways to solve a quadratic programming problem, that is, an optimization problem of a quadratic function under linear (inequality) constraints. These include the conventional complementary pivoting algorithm [Cottle and Danzig, 1968], and the manifold suboptimization methods [Stoer, 1971; Lawson and Hanson, 1974] among others. These methods are all iterative, but are assured of convergence to an exact solution in a finite number of steps. However, considering the particularly simple type of linear restrictions in the present case, we may take yet another approach to the quadratic programming problem. The matrix H_j, which defines the order relations among the elements of $\boldsymbol{p}_j$, assumes a special form in our case, which, together with the fact that the restrictions are a set of linear homogeneous inequalities (i.e., no constant terms exist), permits a reparametrization of $\boldsymbol{p}_j$ into an easily manageable form. Suppose, for example, that there are only three distinct observation categories, and that their prescribed order is $p_{j3} \geq p_{j2} \geq p_{j1}$ where $\boldsymbol{p}_j'=(p_{j1}, p_{j2}, p_{j3})$. Then H_j will be of the form,

$$H_j = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{bmatrix},$$

and

$$H_j \boldsymbol{p}_j = \begin{bmatrix} p_{2j} - p_{1j} \\ p_{3j} - p_{2j} \\ p_{3j} \end{bmatrix}$$

(Without loss of generality p_{3j} may be assumed to be nonnegative.) Define the order restrictions

$$\boldsymbol{q}_j = H_j \boldsymbol{p}_j \geq \boldsymbol{0}, \tag{12}$$

then we can write Q_j as

$$Q_j = (\hat{B}_j \boldsymbol{q}_j - \hat{\boldsymbol{r}}_j)'(\hat{B}_j \boldsymbol{q}_j - \hat{\boldsymbol{r}}_j) \tag{13}$$

where $\hat{B}_j = B_j H_j^{-1}$. Note that the order restrictions on the elements of $\boldsymbol{p}_j$ have been transformed into nonnegativity restrictions on the elements of $\boldsymbol{q}_j$. The important

consequence of this modification is that the constraints are now separable (i.e., imposed separately on each element of $\boldsymbol{q}_j$). A similar reparametrization technique has been used for INDISCAL, an indicator method of multidimensional scaling (de Leeuw, Takane and Young, in preparation).

It can be easily verified that H_j as defined above is always nonsingular for any order and has a regular inverse of the form,

$$H_j^{-1} = \begin{bmatrix} -1 & -1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{bmatrix},$$

in case of the 3×3 matrix. $\hat{B}_j$, however, is still rank deficient. We define B_j^* as $\hat{B}_j$ without the last column, which is all zeroes for $B_j \mathbf{1}_{n_j} = \mathbf{0}$, and $\boldsymbol{q}_j^*$ as $\boldsymbol{q}_j$ with the last element deleted. Then an unconstrained L.S. estimate of $\boldsymbol{q}_j$ is given by (as is well known)

$$\hat{\boldsymbol{q}}_j^* = (B_j^{*\prime} B_j^*)^{-1} B_j^{*\prime} \hat{\boldsymbol{r}}_j . \tag{14}$$

If the elements of $\hat{\boldsymbol{q}}_j^*$ as defined above are all nonnegative, $\hat{\boldsymbol{q}}_j^*$ is also the constrained L. S. estimate under the nonnegativity restriction (3). If, on the other hand, any elements of $\hat{\boldsymbol{q}}_j^*$ are negative, we switch to the following elementwise estimation procedure. Suppose the k' th element of $\hat{\boldsymbol{q}}_j^*$, namely $\hat{q}_{jk}^*$, is found to be negative. We set $\hat{q}_{jk}^* = 0$, and see what effect this replacement would incur on other estimates. We estimate the $k+1'$ st element of $\hat{\boldsymbol{q}}_j^*$ by

$$\hat{q}_{j(k+1)}^* = \boldsymbol{b}_{k+1}^{(j)\prime} (\hat{\boldsymbol{r}}_j - \sum_{j \neq k+1} \boldsymbol{b}_i^{(j)} \hat{\boldsymbol{q}}_{ji}^*) / \boldsymbol{b}_{k+1}^{(j)\prime} \boldsymbol{b}_{k+1}^{(j)} \tag{15}$$

where $\boldsymbol{b}_i^{(j)}$ is the i' th column vector of B_j^*. If $\hat{q}_{j(k+1)}^*$ as defined in (15) is nonnegative, we pass on to the estimation of the next element. However, if it is negative, we set $\hat{q}_{j(k+1)}^* = 0$, and turn to the next element. We repeat this procedure until the elements of $\hat{\boldsymbol{q}}_j^*$ are all stabilized. (The rationale of this procedure for the quadratic programming problem has been established by Hildreth [1957], among others. In fact it is also a special case of an alternating least squares procedure.) After $\hat{\boldsymbol{q}}_j^*$ has converged, we obtain $\hat{q}_{jn}^*$ and adjust $\boldsymbol{q}_j^*$ by restriction (8).

Finally we normalize $\hat{\boldsymbol{q}}_j^*$ so that

$$\hat{\boldsymbol{q}}_j^{*\prime} (H_j^{-1})' D_j H_j^{-1} \hat{\boldsymbol{q}}_j^* = 1 ,$$

and adjust the length of $\hat{\boldsymbol{r}}_j$ and the corresponding j' th row of $\hat{R}$.

Once the optimal scaling of the j' th variable is obtained, we repeat the same procedure successively for all variables. We then check convergence, and if not converged go back to the model estimation phase. Note, however, that one sequential sweep through the optimal scaling of all variables will not completely minimize (3) or (11) for fixed $\hat{R}$, since, as noted earlier, Q_j optimized with respect to the j' th variable depends on all other $n-1$ variables. We could repeat the optimal scaling phase and iterate it until the absolute minimum of Q for fixed $\hat{R}$ is obtained before going back to the model estimation phase. However, our experience with ALS procedures indicates that convergence is obtained more quickly by returning to the model estimation phase after one sweep of all variables.

2.4 Some acceleration techniques

Various acceleration techniques can be suggested for the phases and subphases of the iterative scheme presented above.

The refactoring procedure may be very slow in convergence and may be very time consuming; it has to solve the r dominant eigenvalues and vectors of the matrix $R^*-D_u^2$ with slightly different D_u^2' s repeatedly. We recommend using the Clint and Jennings [1970] simultaneous power method since the successive eigenvectors are most often only slightly different and the set of eigenvectors from the previous iteration may be used as initial estimates of those on the next iteration. The method tends to give faster and faster convergence as iterations proceed since the change in D_u^2 will typically diminish.

In addition to the use of a special eigen routine, a relaxation process has been introduced in the updating scheme of D_u^2 to further accelerate convergence. In the relaxation process the new estimate of a parameter vector θ_r is given by

$$\theta_r^{(i+1)} = (1-\alpha)\theta^{(i)} + \alpha\theta_u^{(i+1)}$$

where the parenthesized superscript indicates an iteration number and $\theta_u^{(i+1)}$ is the estimate of θ from the unrelaxed process. We set $\theta^{(i+1)}=\theta_r^{(i+1)}$ for the $i+2$' nd iteration. Ramsay [1975] has given an ingeneous method for choosing an appropriate value for α, and reports that the relaxation process significantly cuts down the number of iterations to convergence in the refactoring method.

3. Analysis of interest inventory data

We present just one example of FACTALS analysis with a real set of data. The original data were collected as part of a large scale vocational aptitude study conducted by Hidano, et. al. (1971). The data consist of subjects' responses to several hundred items from over 2000 subjects. Ten items in the interest domain, which have been found to have a two-common-factor structure from a preliminary analysis, were chosen for the purpose of the present study. The ten items are:

A.	Building a ship	(1)
B.	Persuading people	(2)
C.	Reparing second hand cars	(1)
D.	Building bridges	(1)
E.	Discussing the recent world situation	(2)
F.	Participating in conferences	(2)
G.	Constructing new railways	(1)
H.	Reparing tools	(1)
I.	Reading articles on political matters	(2)
J.	Talking about politics	(2)

(The numbers in parentheses are designated factors: (1) mechanical interest, (2) political interests.) Subjects indicated the extent to which they are interested in doing things described in the items on five-point scales: 1. not at all interested, 2. not very much interested, 3. undecided, 4. slightly interested, and 5. very interested.

Hidano, et. al. applied the usual (metric) common factor analysis to the original

data assuming that preassigned category values (one to five) constitute interval scales. We, on the other hand, applied FACTALS to the selected sample of items assuming that categories of rating scales are merely ordinal. Preassigned category values and their squares were used as initial category values, and these two monotonically related data sets were analyzed under the interval and ordinal assumptions.

The derived factor loadings are plotted in Fig. 1. It can be clearly seen that the derived factor loading patterns are remarkably similar in all cases. Implications are twofold. First, FACTALS can obtain monotone-invariant results; solutions obtained from the two monotonically related data sets by FACTALS are virtually indistinguishable (so much so that the two solutions are indicated as if they were a single solution in the figure). Thus, FACTALS satisfies the basic condition required of a nonmetric procedure. Second, although this particular example fails to indicate that the nonmetric analysis gives more sensible results than its metric counterpart, it provides a direct support for the metric assumption tacitly made by Hidano, et. al. in their analyses. If nonmetric solutions are not very much different from metric solutions, there seems to be no reason why we should apply more expensive nonmetric procedures.

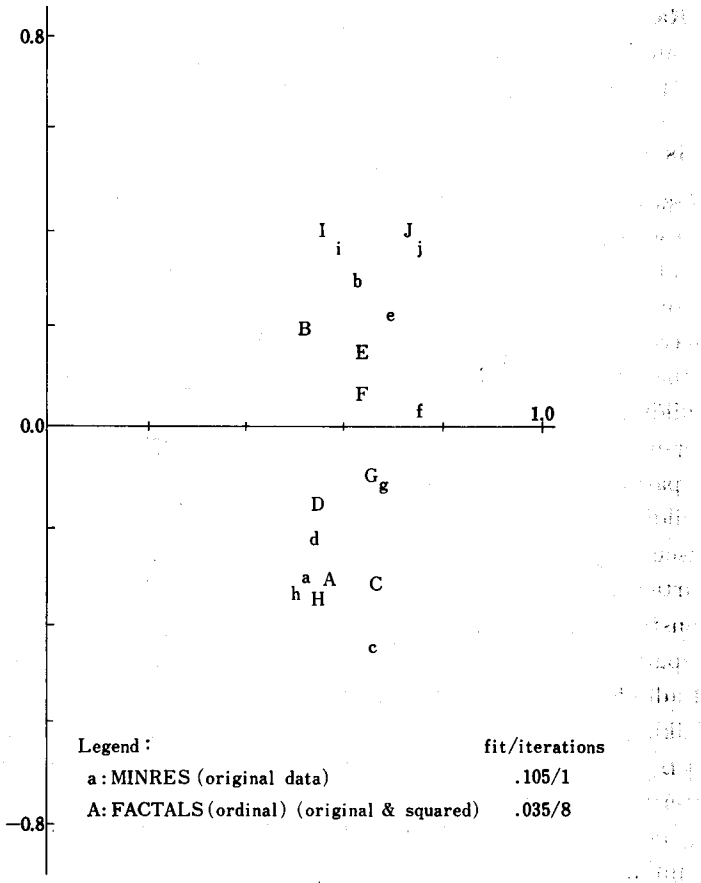

Fig. 1 Plot of factor loagings: Interest Inventory data

However, it is not until we find the two solutions are essentially the same that we can justify the stronger measurement assumption. That we can empirically test the plauisbility of the metric assumption is one of the major advantages of nonmetric procedures.

4. Discussion

We have seen one use of FACTALS through the analysis of example data. Once by virtue of nonmetric analyses it is established that rating scales consist of equally spaced categories, we may apply the metric procedure with some confidence. Or we may in a shooting-in-the-dark manner find the nonmetric analysis by FACTALS uncover a truly meaningful data structure which would never be discovered otherwise, though, of course, we cannot expect this kind of luck to happen so frequently.

One cautionary remark is in order against routine applications of FACTALS, or more broadly against all kinds of routine applications of common factor analysis in general. The common factor analysis model is a falsifiable model, unlike the principal components model, imposing rather stringent assumptions on the data. Indeed as Guttman once noted, common factor analysis should not be applied off hand unless one has a reasonable degree of confidence about the plausibility of its assumptions. The common factor model is *not* a general data analytic model as has been predominantly used or rather abused, and as such it has been enjoying its popularity, *but* a very specific model of behavior as Spearman originally conceptualized it as a structural model of intelligence. A small Monte Carlo study has revealed that FACTALS is not very robust against violations of the linear local independence assumption (1-d). It seems necessary that we be very cautious about the plausibility of the model assumptions when we wish to apply FACTALS and evaluate the empirical meaning of the results.

References

Browne, M.M. (1967) On oblique procrustes rotation. *Psychometrika*, **32**, 125–132.

Clint, M., and Jennings, A. (1970) The evaluation of eigenvalues and eigenvectors of real symmetric matrices by simultaneous iteration. *The Computer Journal*, **13**, 76–80.

Cottle, R.W., and Danzig, T.B. (1968) Complementary pivot theory of mathematical programming. In G.B. Danzig and A.F. Veinott (Eds), *Mathematics of decision sciences*, Part I. Providence, R.I.: American Mathematical Society.

Cramer, E.M. (1974) On Browne's solution for oblique procrustes rotation. *Psychometrika*, **39**, 159–163.

de Leeuw, J., Takane, Y., and Young, F.W. (1978) INDISCAL: An indicator method of multidimensional scaling (in preparation).

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

Golub, G.H., and Saunders, M.A. (1969) Linear least squares and quadratic programming. Technical Report No. CS134, Stanford University.

Harman, H.H., and Jones, W.H. (1966) Factor analysis by minimizing residuals (minres). *Psychometrika*, **31**, 351–368.

Hidano, T., et al. (1971) A methodological study of test battery construction. *Japanese Journal of Educational Psychology* (in Japanese with English abstract), **19**, 37–51.

Hildreth, C.A. (1957) A quadratic programming procedure. *Naval Research Logistics Quarterly*, **4**, 79–85.

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

Kruskal, J.B. (1965) Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society, Series B*, **27**, 251–263.

Kruskal, J.B., and Shepard, R.N. (1974) A nonmetric variety of linear factor analysis. *Psychometrika*, **39**, 123–157.

Lawson, C.L., and Hanson, R.J. (1974) *Solving least squares problems*. Englewood Cliffs, N.J.: Prentice-Hall.

Lingoes, J.C. (1973) *The Guttman-Lingoes nonmetric program series*. Ann Arbor, Michigan: Mathesis Press.

Lingoes, J.C., and Guttman, L. (1967) Nonmetric factor analysis: A rank reducing alternative to linear factor analysis. *Multivariate Behavioral Research*, **2**, 485–505.

McDonald, R.P., Ishizuka, T., and Nishisato, S. (1977) Weighting multicategory data to the common factor model. Paper presented at the Psychometric Society meeting.

Ramsay, J.O. (1975) Solving implicit equations in psychometric data analysis. *Psychometrika*, **40**, 337–360.

Roskam, E.E. (1968) *Metric analysis of ordinal data in psychology*. Voorschoten.

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

Stoer, J. (1971) On the numerical solution of constrained least squares problems. *SIAM Journal of Numerical Analysis*, **8**, 382–411.

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

Takane, Y., Young, F.W., and de Leeuw, J. (1978) The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika*, in press.

Thomson, G.H. (1951) *The factorial analysis of human ability*. New York: Houghton Mifflin.

Young, F.W. (1972) A model for polynomial conjoint analysis algorithms. In R.N. Shepard et. al. (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences*. Vol. I. New York: Seminar Press.

Young, F.W., de Leeuw, J., and Takane, Y. (1976a) Multiple and canonical regression with a mix of qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika*, **41**, 505–529.

Young, F.W., de Leeuw, J., and Takane, Y. (1976b) Quantifying qualitative data. Psychometric Lab Report #149, Univ. of North Carolina.

Young, F.W., de Leeuw, J., and Takane, Y. (1979) *Multidimensional scaling: Theory and methods*. (in preparation).

Zangwill, W.I. (1969) *Nonlinear programming: A unified approach*. New York: Prentice-Hall.

(Received September, 1978)