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

CANDELINC: A GENERAL APPROACH TO MULTIDIMENSIONAL ANALYSIS OF MANY-WAY ARRAYS WITH LINEAR CONSTRAINTS ON PARAMETERS

J. DOUGLAS CARROLL

SANDRA PRUZANSKY

JOSEPH B. KRUSKAL

BELL TELEPHONE LABORATORIES

Very general multilinear models, called CANDELINC, and a practical least-squares fitting procedure, also called CANDELINC, are described for data consisting of a many-way array. The models incorporate the possibility of general linear constraints, which turn out to have substantial practical value in some applications, by permitting better prediction and understanding. Description of the model, and proof of a theorem which greatly simplifies the least-squares fitting process, is given first for the case involving two-way data and a bilinear model. Model and proof are then extended to the case of N-way data and an N-linear model for general N. The case $N = 3$ covers many significant applications. Two applications are described: one of two-way CANDELINC, and the other of CANDELINC used as a constrained version of INDSCAL. Possible additional applications are discussed.

Key words: constrained least-squares, multilinear models, bilinear models, INDSCAL, multidimensional scaling, 3-mode factor analysis, CANDECOMP, LINCINDS, multivariate analysis.

A number of methods for multidimensional data analysis are special cases of a general procedure described by Carroll and Chang [1970] and now called CANDECOMP (for *CAN*onical *DECOMP*osition of N-way tables). Included are the approach to factor analysis often called "Eckart-Young decomposition" [Eckart & Young, 1936], the classical (two-way) metric version of multidimensional scaling [Torgerson, 1958] and the INDSCAL method of three-way, or individual differences, multidimensional scaling [Carroll & Chang, 1970] and the closely related procedure called PARAFAC proposed by Harshman [Note 6]. In fact, CANDECOMP was originally developed to implement a metric analysis in terms of the INDSCAL model.

In all of the methods named above, as well as in many other procedures for multidimensional analysis of behavioral science data, the problem of interpretation of dimensions or factors is of paramount importance. Even more important, in many practical applications, is the problem of predicting locations of new stimuli or other entities in the multidimensional space, and with the aid of these locations further predicting the relative preferability, judged similarities, or other judgmental attributes of the stimuli (or other entities).

In many cases, additional information about the stimuli (or other entities) may be available. For example, variables other than those used in forming the multidimensional space may have been measured. Such information can often be used to provide interpretations of the stimuli, and also to predict the locations of other stimuli in the multidimensional space (from their values on the other variables). In many cases, however, the

Requests for reprints should be sent to J. Douglas Carroll, Bell Laboratories, 2C-553, Murray Hill, New Jersey 07974.

0033-3123/80/0300-2824$00.75/0

variability that *can* be explained by these outside variables may be obscured by a considerable amount of variability that *cannot* be so explained. In such instances it may be desirable to constrain the dimensions of the multidimensional representation obtained to be completely explained by these outside variables. This may, and often will, cut down considerably on the variance accounted for in the original data. This increase in residual error should be well paid for by the much crisper meaning of dimensions which are obtained and by the ease of predicting the locations of new objects in the space from knowledge of their values on the outside variables.

Sometimes the "other variables" are not measured in the usual sense, but grow out of a factorial design underlying the stimuli or other entities. Specifically, if the stimuli form a factorial design, then the artificial (or dummy) variables which are conventionally used to code that design may be used as the outside variables. Since the two examples in this paper follow this approach, we shall explain it in detail later.

This paper has several purposes:

(i) We describe a general approach called CANDELINC for finding CANDECOMP and INDSCAL solutions which satisfy the constraints that various parameters are linearly related to prespecified variables.

(ii) We describe a practical method of least-squares fitting for the CANDELINC model by reducing this problem to fitting a much smaller CANDECOMP model. (Of course, least-squares fitting is already practical and in frequent use for CANDECOMP. In the two-way case, such fitting can be accomplished by the highly developed numerical methods for singular value decomposition or eigenvector extraction. For the three-way and many-way case, alternating least-squares methods have proved themselves practical, though no doubt subject to considerable improvement.)

(iii) We prove theorems, of some interest in themselves, which justify the reduction just described.

(iv) We illustrate the CANDELINC approach with two examples using real data, in which the constrained results are compared with the corresponding unconstrained results. One example is based on two-way CANDECOMP, the other on INDSCAL.

The CANDELINC approach was first described for the two-way case, and a corresponding theorem proved, by Carroll in Carroll, Green and Carmone [Note 1] and Green, Carroll, and Carmone [1976]. The many-way case of CANDELINC was first proposed by Carroll and Pruzansky [Note 2] (under the name MULTILINC, which has now been dropped). A theorem for the many-way case was conjectured by Carroll, Green, and Carmone [Note 1], but was not proved until a later time, by Carroll and Pruzansky [Note 2]. A more general theorem based on a more elegant proof for the many-way case has since been developed by Kruskal, and is included in the present paper. The theorem by Kruskal extends to a more general class of models, including Tucker's [1964, 1972] three-mode and multimode factor analysis and certain special cases. It thereby verifies an additional conjecture set forth by Carroll in Carroll, Green, and Carmone [Note 1] and Carroll and Pruzansky [Note 2].

The CANDECOMP model in its general form can be stated as

$$(1) \qquad y_{i_1 i_2 \cdots i_N} \cong \sum_{r=1}^{R} a_{i_1 r}^{(1)} a_{i_2 r}^{(2)} \cdots a_{i_N r}^{(N)},$$

where $y_{i_1 i_2 \cdots i_N}$ is the general entry in an N-way data table Y. The symbol "$\cong$" is used in this paper to indicate two things. First, it indicates that the data on the left are equal to the expression on the right except for an additive error term. Second, it indicates that least-squares estimates will be sought for the parameters on the right. While the distribution of the error term will not be specified in detail, the use of least-squares estimates does carry the implicit assumption that the standard deviation of the error term does not vary from one data value to another, i.e., the standard deviation is constant as a function of the subscripts. (It also carries the implicit assumption that the expected value of the error term is zero.) In practice, of course, the least squares fitting procedure will often be applied to situations in which these implicit assumptions are not likely to be met. The robustness of (ordinary) least-squares procedures makes this reasonable, so long as these implicit assumptions are not grossly violated.

Throughout this paper we deal only with ordinary least-squares fitting. In practice, due to the implicit assumptions this carries, it is often important to rescale the original variables in such a way that fitting errors of equal size in different variables have roughly equal "importance". The problem here of choosing the rescaling multipliers closely resembles the corresponding problem in principal components analysis.

Carroll and Chang [1970] describe a NILES procedure [Wold, 1966] or what has more recently come to be called an Alternating Least Squares (ALS) procedure, for least squares estimation of the parameters (the a's) of the general CANDECOMP model. In CANDELINC we seek to fit the CANDECOMP model, but with linear constraints on the a's. These constraints may require a particular set of a's to be an exact linear function of specified external variables, to be additively decomposable relative to a given factorial design, or may be defined in other ways described at a later point.

CANDELINC: The Two-Way Case

Let us first describe CANDELINC (*CAN*onical *DE*composition with *LIN*ear *C*onstraints) for the case where $N = 2$. This means that the data form a two-way array or matrix $\mathbf{Y}$, and that the model (as given below) is *bi*linear. Assume that $\mathbf{Y}$ is $J \times K$. The values in $\mathbf{Y}$ might entail preference judgements by J subjects or judges on K stimuli (or other objects). The bilinear CANDECOMP model (on which two-way CANDELINC rests) is given by

$$(2) \qquad \underline{\mathbf{Y}} \cong \mathbf{A}_1\mathbf{A}_2',$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ contain the parameters which are to be estimated. (Of course, this is also the Eckart-Young model.) $\mathbf{A}_1$ is $J \times R$, and each row will provide a description of one subject; $\mathbf{A}_2$ is $K \times R$, and each row will provide a description of one stimulus. The parameter R is the number of "dimensions" or "factors" which are assumed to explain the data. In some cases R may be determined on theoretical grounds, in others the data analyst must choose R, usually after trying several values, on the basis of some combination of statistical and "interpretability" criteria.

In order to constrain $\mathbf{A}_1$ and $\mathbf{A}_2$, we require them to have the form

$$(3) \qquad \mathbf{A}_1 = \mathbf{X}_1\mathbf{T}_1 \text{ and } \mathbf{A}_2 = \mathbf{X}_2\mathbf{T}_2$$

where

$$\left.\begin{matrix}\mathbf{X}_1, & J \times S_1 \\ \mathbf{X}_2, & K \times S_2\end{matrix}\right\} \begin{matrix}\text{are fixed known "design matrices"} \\ \text{(assumed to have full column-rank),}\end{matrix}$$

$$\left.\begin{matrix}\mathbf{T}_1, S_1 \times R \\ \mathbf{T}_2, S_2 \times R\end{matrix}\right\} \begin{matrix}\text{are unknown parameter matrices} \\ \text{which must be estimated,}\end{matrix}$$

with $S_1 \le J$ and $S_2 \le K$. The requirement that $\mathbf{X}_n$ have full column-rank is no real restriction, since if $\mathbf{X}_n$ does not satisfy the requirement we can replace it by some generating set of its columns, and exactly the same matrices $\mathbf{A}$ will satisfy the constraints. See Appendix A for an illustration of some design matrices.

Note that in our applications interest focuses on the $A_n (n = 1, 2)$, not on the ultimate fitted parameters T_n nor on the ultimate data values $A_1 A_2'$. It is the A_n we examine in order to obtain information about the structure of the data. The constraints would perhaps have a clearer meaning (but would not be as easy to work with algebraically and computationally) if they were expressed in a different and more familiar dual form. In order to obtain the dual form, consider the space of all left null vectors z of the matrix X_n (i.e., the set of all row vectors z such that $zX_n = 0$). Select a generating set of such row vectors, and form a matrix Z_n with these as rows. Then $A_n = X_n T_n$ for some T_n if and only if $Z_n A_n = 0$. From this it is clear that the parameters T_n are merely a mathematical device (though interpretation of them is possible, as mentioned later).

For reasons to be discussed later, we may want to apply the constraints only to the stimuli. To accommodate this special case within the general framework, we can let X_1 be the identity matrix, so $A_1 = T_1$ is unconstrained. This will be referred to as "not using any design matrix for the subjects."

The matrix X_n might consist of S_n quantitative variables measured on the subjects or stimuli. Each column would contain the values of one variable. We are frequently interested, however, in the case where the columns of X_n contain a coding of categorical variables corresponding to treatments in an analysis of variance design. In this case, each column of X_n is an artificial (or what is often called a "dummy") variable. For example, the encoding might include "dummy" variables for main effects only, in which case the coordinates of A_n would be constrained to fit an additive model. On the other hand, some interaction terms could be included, or only certain one-degree-of-freedom contrasts (partitioning main effects, interactions, or both) could be included, if desired. An illustration of the construction of design matrices for both an additive model and a model incorporating a linear × linear interaction is given in Appendix A.

Without real loss of generality, it is possible to assume that $X_n' X_n = I_{s_n}$ (for $n = 1, 2$), that is, we assume that the X_n are all suborthogonal (sometimes called "orthogonal sections" or "column orthonormal" matrices). The fact that there is no loss of generality follows from two things: the fact that X_n has full column-rank, and the fact that our interest focuses on the A_n rather than the T_n. To see this, consider the constraint that $A_n = X_n T_n$ for some T_n. All this means is that the columns of A_n belong to the column space of X_n. Any other matrix X_n^* with the same column space yields the same set of constrained matrices A_n. If we choose any orthonormal set of vectors which generates the column space of X_n, and use these vectors as the columns of X_n^*, then X_n^* is suborthogonal and yields the same set of constrained matrices A_n. For theoretical purposes, this justifies the assumption that the X_n are suborthogonal.

For practical purposes, it is necessary to start with X_n and form X_n^*. For this purpose, a procedure such as the Gram-Schmidt orthogonalization procedure could be used. We have used the following procedure [Johnson, 1966] which finds X_n^* which is not only suborthogonal but which provides the least-squares suborthogonal fit to X_n. Decompose X_n into its singular value decomposition,

$$X_n = U_n \beta_n V_n' \tag{4}$$

(U_n and V_n are suborthogonal, β_n is diagonal). Then $X_n^* = U_n V_n'$ is suborthogonal and has the same column space as X_n, and among all such matrices it is the least-squares fit to X_n.

In applications other than the particular ones we discuss, there might be some interest in the parameters T_n themselves. For example, by looking at a given column of T_2 and examining which elements have large magnitude, we can interpret the meaning of the corresponding dimension of the solution. Similarly, though perhaps less useful, the rows of T_2 can be thought of as describing hypothetical idealized stimuli, of which the actual stimuli are linear combinations. In applications where we are interested in the values of $\mathbf{T}_n$ and not only in the values of $\mathbf{A}_n$, the argument above for assuming $\mathbf{X}_n$ suborthogonal is not valid. Nevertheless, the calculations for this case follow almost the same procedure described in this paper. The only differences are these. While forming the suborthogonal matrix $\mathbf{X}_n^*$ from $\mathbf{X}_n$ as described above, a matrix $\mathbf{S}_n$ should also be formed such that $\mathbf{X}_n^* = \mathbf{X}_n\mathbf{S}_n$. ($\mathbf{S}_n$ can be obtained as an easy byproduct while forming $\mathbf{X}_n^*$.) Following the procedures described in this paper then leads to $\mathbf{T}_n^*$ (not $\mathbf{T}_n$), but it is easy to show that $\mathbf{T}_n = \mathbf{S}_n\mathbf{T}_n^*$, so $\mathbf{T}_n$ is easily recovered. Note that in our notation henceforth we shall assume $\mathbf{X}_n^*$ has been substituted for $\mathbf{X}_n$. Thus, $\mathbf{X}_n$ itself is assumed to be suborthogonal, without the need to write the asterisk.

In the two-way case, then, we want to find transformation matrices, $\mathbf{T}_1$ and $\mathbf{T}_2$ such that $\hat{\mathbf{Y}} = \mathbf{A}_1\mathbf{A}_2$ provides a best least squares approximation to $\mathbf{Y}$, with the constraint that $\mathbf{A}_1 = \mathbf{X}_1\mathbf{T}_1$ and $\mathbf{A}_2 = \mathbf{X}_2\mathbf{T}_2$. We introduce the squared norm function, indicated by $\mathbf{N}$, which means the sum of squares of the elements (for any vector, matrix, array, or function). Then we want to find $\mathbf{T}_1$ and $\mathbf{T}_2$ which *minimize*

$$N(\mathbf{Y} - \hat{\mathbf{Y}}) \equiv \sum_i \sum_j (y_{ij} - \hat{y}_{ij})^2 \tag{5}$$

$$= N(\mathbf{Y} - \mathbf{A}_1\mathbf{A}_2') = N(\mathbf{Y} - \mathbf{X}_1\mathbf{T}_1\mathbf{T}_2'\mathbf{X}_2')\,.$$

Roughly speaking, the following theorem states that it is legitimate to multiply by $\mathbf{X}_1'$ on the left and $\mathbf{X}_2$ on the right, inside N.

Theorem: Minimizing (5) over $\mathbf{T}_1$ and $\mathbf{T}_2$ is equivalent to minimizing $N(\mathbf{X}_1'\mathbf{Y}\mathbf{X}_2 - \mathbf{T}_1\mathbf{T}_2')$ over $\mathbf{T}_1$ and $\mathbf{T}_2$. Thus if we define $\mathbf{Y}^* = \mathbf{X}_1'\mathbf{Y}\mathbf{X}_2$, the minimization problem reduces to minimizing $N(\mathbf{Y}^* - \mathbf{T}_1\mathbf{T}_2')$ over $\mathbf{T}_1$ and $\mathbf{T}_2$, which can be done in the two-way case by classical methods such as use of an Eckart-Young analysis.

Even though we prove a substantially more general theorem later by more elegant methods, it seems worthwhile to give a direct demonstration of this result.

Proof. For any matrix $\mathbf{Z}$, $N(\mathbf{Z}) = \text{tr}\,\mathbf{Z}\mathbf{Z}'$. We want to minimize

$$N(\mathbf{Y} - \mathbf{A}_1\mathbf{A}_2') = \text{tr}[(\mathbf{Y} - \mathbf{X}_1\mathbf{T}_1\mathbf{T}_2'\mathbf{X}_2')(\mathbf{Y}' - \mathbf{X}_2\mathbf{T}_2\mathbf{T}_1'\mathbf{X}_1')] \tag{6}$$

$$= \text{tr}[(\mathbf{Y}\mathbf{Y}' - \mathbf{Y}\mathbf{X}_2\mathbf{T}_2\mathbf{T}_1'\mathbf{X}_1' - \mathbf{X}_1\mathbf{T}_1\mathbf{T}_2'\mathbf{X}_2'\mathbf{Y}' + \mathbf{X}_1\mathbf{T}_1\mathbf{T}_2'\mathbf{X}_2'\mathbf{X}_2\mathbf{T}_2\mathbf{T}_1'\mathbf{X}_1')\,.$$

By well-known results concerning traces of square matrices, this can be written as:

$$N(\mathbf{Y} - \mathbf{A}_1\mathbf{A}_2') = \text{tr}(\mathbf{Y}\mathbf{Y}') - 2\text{tr}\,(\mathbf{X}_1'\mathbf{Y}\mathbf{X}_2\mathbf{T}_2\mathbf{T}_1') + \text{tr}(\mathbf{X}_1'\mathbf{X}_1\mathbf{T}_1\mathbf{T}_2'\mathbf{X}_2'\mathbf{X}_2\mathbf{T}_2\mathbf{T}_1'). \tag{7}$$

Since $\mathbf{X}_1'\mathbf{X}_1 = \mathbf{I}_{s_1}$ and $\mathbf{X}_2'\mathbf{X}_2 = \mathbf{I}_{s_2}$ those expressions both drop out of the last term, leaving

$$N(\mathbf{Y} - \mathbf{A}_1\mathbf{A}_2') = \text{tr}(\mathbf{Y}\mathbf{Y}') - 2\text{tr}(\mathbf{Y}^*\mathbf{T}_2\mathbf{T}_1') + \text{tr}(\mathbf{T}_1\mathbf{T}_2'\mathbf{T}_2\mathbf{T}_1')\,. \tag{8}$$

Now we expand out another expression and notice how similar it is to (8):

$$N(\mathbf{Y}^* - \mathbf{T}_1\mathbf{T}_2') = \text{tr}(\mathbf{Y}^*\mathbf{Y}^{*\prime}) - 2\text{tr}(\mathbf{Y}^*\mathbf{T}_2\mathbf{T}_1') + \text{tr}(\mathbf{T}_1\mathbf{T}_2'\mathbf{T}_2\mathbf{T}_1') \tag{9}$$

$$= N(\mathbf{Y} - \mathbf{A}_1\mathbf{A}_2') - C$$

where

$$C = \text{tr}(\mathbf{Y}\mathbf{Y}') - \text{tr}(\mathbf{Y}^*\mathbf{Y}^{*\prime})\,. \tag{10}$$

Since C is a constant that does not depend on $\mathbf{T}_1$ or $\mathbf{T}_2$, finding the matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ that minimize $N(\mathbf{Y}^* - \mathbf{T}_1\mathbf{T}_2')$ will also solve the problem of minimizing $N(\mathbf{Y} - \mathbf{A}_1\mathbf{A}_2')$. This ends the proof.

To summarize, we may find the optimal $\mathbf{T}_1$ and $\mathbf{T}_2$ for our original problem (the two-way case of CANDELINC) by defining

$$\mathbf{Y}^* \equiv \mathbf{X}_1'\mathbf{Y}\mathbf{X}_2 ,$$

finding $\mathbf{T}_1$ and $\mathbf{T}_2$ such that $\hat{\mathbf{Y}}^* = \mathbf{T}_1\mathbf{T}_2'$ provides a least squares fit to $\mathbf{Y}^*$ in the required dimensionality (this may be done in the two-way case by use of the Eckart-Young decomposition), and finally forming [in accordance with (3)]

$$\mathbf{A}_1 = \mathbf{X}_1\mathbf{T}_1$$

$$\text{and } \mathbf{A}_2 = \mathbf{X}_2\mathbf{T}_2 .$$

Note that in general this two-way CANDELINC analysis requires finding the Eckart-Young decomposition of a much smaller matrix than would be necessary in the unrestricted analysis ($S_1 \times S_2$ rather than $J \times K$, where usually $S_1 \ll J$ and/or $S_2 \ll K$).

Extension to the Multi-Way Case

To reiterate, we seek to fit the model

$$(1') \qquad y_{i_1 i_2 \cdots i_N} \cong \sum_{r=1}^{R} a_{i_1 r}^{(1)} \; a_{i_2 r}^{(2)} \; \cdots \; a_{i_N r}^{(N)}$$

(where "$\cong$" implies a best least squares fit). In the multiway case we want to fit this general CANDECOMP model with linear constraints on the $a_r^{(n)}$'s which can be expressed in the equation

$$(3') \qquad \mathbf{A}_n \equiv \| a_{i_n r}^{(n)} \| = \mathbf{X}_n\mathbf{T}_n ,$$

where the $\mathbf{X}_n$'s are the design matrices for each way, and the $\mathbf{T}_n$'s are to be solved for.

In (1′) and (3′) the subscript i_n ranges from 1 to I_n. $\mathbf{X}_n$ is $I_n \times S_n$ (for some S_n) and $\mathbf{T}_n$ is $S_n \times R$, where R is the "dimensionality" of the decomposition ($r = 1, \cdots, R$). Stimulus dimensions, factors, subject weights, or whatever the n^{th} way of the design corresponds to, must be linear functions of some set of *a priori* independent variables described by the design matrices.

We now state the result that generalizes the two-way CANDELINC solution to the multiway case. Given an N-way array $\mathbf{Y}$ whose general entry is $y_{i_1 i_2 \cdots i_N} (i_n = 1, 2, \cdots, I_n)$, we wish to decompose it into a sum of products of the form given in (1′), with constraints on the $\mathbf{A}_n$'s as defined in (3′). We are given the additional technical condition on the $\mathbf{X}_n$'s that

$$(4') \qquad \mathbf{X}_n'\mathbf{X}_n = \mathbf{I}_{S_n}$$

(a condition that can always be satisfied, as we have seen, by appropriate definition of the $\mathbf{X}$'s). The solution is obtained by defining a new array $\mathbf{Y}^*$ with entries

$$(11) \qquad y^*_{s_1 s_2 \cdots s_N} \equiv \sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} y_{i_1 i_2 \cdots i_N} \, x_{i_1 s_1}^{(1)} x_{i_2 s_2}^{(2)} \cdots x_{i_N s_N}^{(N)}$$

(where $x_{i_n s_n}^{(n)}$ is the general entry of $\mathbf{X}_n$), and then finding the least squares solution to the problem

$$(12) \qquad y^*_{s_1 s_2 \cdots s_N} \cong \sum_{r=1}^{R} t_{s_1 r}^{(1)} t_{s_2 r}^{(2)} \cdots t_{s_N r}^{(N)} .$$

The matrices $\mathbf{T}_n \equiv \| t_{s_n r}^{(n)} \|$ then provide the desired solution. The matrices $\mathbf{A}_n$ satisfying the desired linear constraints can then be constructed via (3′). This result was first proved by Carroll in an unpublished paper by Carroll and Pruzansky [Note 2], but we present here only the more general result and more elegant proof due to Kruskal.

In this section we change notational conventions somewhat from those used earlier. Capital letters stand for N-way arrays and small Roman letters stand for matrices. We use square brackets to indicate elements: an element of a is $a[j, i]$, and an element of Z is $Z[i_1, \cdots, i_N]$. We generalize the product of two matrices in a simple way by defining the *product in position n* of a matrix a and an array Z. Informally, consider the array Z as made up of a large number of vectors or "columns" in the n^{th} direction, where each vector has the form

$$(Z[i_1, \cdots, 1, \cdots, i_N], \cdots, Z[i_1, \cdots, I_n, \cdots, i_N]).$$

Then the product in position n is formed by multiplying each of these vectors by the matrix a. More formally, $a \odot_n Z$ is the array such that

$$(a \odot_n Z)[i_1, \cdots j \cdots, i_N] = \sum_{i_n}^{I_n} a[j, i_n] Z[i_1, \cdots, i_n, \cdots, i_N]. \tag{13}$$

Note that products for different n commute. This is analogous to *associativity* of matrix multiplication.

Theorem: Suppose Y and D are given arrays, and x_n are given matrices such that $x_n' x_n$ = identity matrix. Suppose we wish to minimize the following expression over the matrices t_n:

$$N[Y - (x_1 t_1) \odot_1 \cdots (x_N t_N) \odot_N D]. \tag{14}$$

This can be accomplished by finding the t_n which minimize the following expression, since the two expressions differ by a constant:

$$N[x_1' \odot_1 \cdots x_N' \odot_N Y - t_1 \odot_1 \cdots t_N \odot_N D]. \tag{15}$$

This theorem reduces to Carroll's earlier theorem when D is the $R \times R \times \cdots \times R$ "identity" array, i.e., the array such that

$$D[r_1, \cdots, r_N] = \begin{cases} 1 \text{ if } r_1 = r_2 = \cdots = r_N, \\ 0 \text{ otherwise.} \end{cases} \tag{16}$$

Proof: From the condition on x_n we know that it consists of a set of orthonormal columns. Let $\tilde{x}_n$ consist of a complementary set of orthonormal columns, so the partitioned matrix $(x_n, \tilde{x}_n)$ is orthogonal. Then if ζ is a column vector, we have

$$\begin{aligned} N[\zeta] &= N[x_n' \zeta] + N[\tilde{x}_n' \zeta], \\ \tilde{x}_n' x_n &= 0. \end{aligned} \tag{17}$$

If we consider any array Z as made up of many columns in the n^{th} direction, then

$$N[Z] = \sum N[\text{column}]. \tag{18}$$

Therefore

$$N[Z] = N[x_n' \odot_n Z] + N[\tilde{x}_n' \odot_n Z]. \tag{19}$$

Now we apply this, for $n = 1$, to the first expression in the theorem [see (14)], and make use of the fact that $x_1' x_1$ = identity matrix and $x_1' \tilde{x}_1 = 0$ to obtain

$$\begin{aligned} &N[Y - (x_1 t_1) \odot_1 \cdots (x_N t_N) \odot_N D] \\ &= N[x_1' \odot_1 Y - t_1 \odot_1 (x_2 t_2) \odot_2 \cdots (x_N t_N) \odot_N D] + N[\tilde{x}_1' \odot_1 Y]. \end{aligned} \tag{20}$$

We apply the decomposition for $n = 2$ to the first term on the right, and then repeat in a similar manner for all n, ultimately obtaining this equation:

$$
\begin{aligned}
&N[Y - (x_1 t_1)\odot_1 \cdots (x_N t_N)\odot_N D] \\
&= N[x_1'\odot_1 \cdots x_N'\odot_N Y - t_1\odot_1 \cdots t_N \odot_N D] + \sum_{n=1}^{N} N[x_1'\odot_1 x_2'\odot_2 \cdots \tilde{x}_n'\odot_n Y].
\end{aligned} \tag{21}
$$

Note that only a single tilde occurs in each summand of the latter term. The latter term is constant, i.e., does not depend on the t_n, and so the proof is complete.

Extension of the Multiway CANDELINC Result to Tucker's Three-Mode Factor Analysis and Scaling Models.

Since the above proof did not rely in any way on the D array being fixed, it follows that the theorem holds for *D variable* as well, that is, for the case in which the t_i *and* the D array are to be solved for so as to minimize the expression in (14). The model implied in this case (in particular if the t_i's are allowed to have arbitrary, and possibly different, column orders) is equivalent to Tucker's [1964] three-mode or multimode factor analysis model (with D being the "core matrix," or 'core box" as it is sometimes called). Therefore, the result of this theorem extends to those models. It also extends, by implication, to the special case of three-mode factor analysis applied to three-way arrays symmetric in two subscripts which is often called "three-mode scaling" [Tucker, 1972].

It is important to state precisely what this result says about the constrained version of three or multimode factor analysis or scaling (henceforth to be referred to generally as multimode f.a.). First of all, the result relates to *least squares* fitting of the appropriate model. It says that a least squares solution for the linearly constrained model can be found by finding a least squares solution for the reduced model. Since the standard solution for multimode f.a. is not, strictly speaking, a least squares solution (although it tends to provide a very good approximation to it) the result does not strictly apply to that standard solution. However, this does suggest that, if an *approximate* least squares solution, such as the standard solution for the three-mode factor analysis model, is applied to the reduced matrix, the result should produce an *approximate* solution for the linearly constrained model. Incidentally, Tucker and MacCallum [Note 11] have produced an iterative algorithm leading to an exact least squares solution (barring local minimum problems) for the three-mode f.a. case. The result should apply *strictly* to the solution produced by this algorithm.

Linear Constraints: Their Geometrical Meaning

To illustrate the geometrical meaning of linear constraints, we happen to use an application of CANDELINC to INDSCAL. We use a tiny data set consisting of 9×9 matrices of dissimilarities provided by three subjects. The numbering of the nine stimuli from a 3×3 factorial design is shown in Appendix A along with some constraint matrices. Figure 1A shows the stimulus configuration from an unconstrained two-dimensional solution. Figure 1B shows the solution which is constrained to satisfy the additive model. Geometrically, this constraint is equivalent to requiring that the three sets 1-2-3, 4-5-6, and 7-8-9, must be the same except for translation. More graphically though less precisely, the three solid "curves" must be "parallel" to one another. Alternatively, the constraint is also equivalent to requiring that the three dashed curves are parallel to one another.

Figure 1C shows the solution which is constrained to satisfy the model which has an additive part plus a linear $\times$ linear interaction term. Figure 1D shows the same solution decomposed into these parts. The interaction term only contributes to the positions of the four corner points. It is important to note one way in which this situation differs from the corresponding situation in ordinary ANOVA. Even though the linear × linear term is orthogonal to the additive terms, the least-squares fitted values for the additive terms change slightly when the interaction term is incorporated, as can be seen by a careful comparison of Figures 1B and 1D. This is because the subject weights, which have no parallel in the ordinary ANOVA situation, are affected when the interaction term is incorporated, and they in turn affect the additive terms.

Application of CANDELINC to Two-Way Data

We shall discuss one application of two-way CANDELINC to data from Wish [1975] concerning ways that people in various role relations communicate with each other in different situations. The stimuli were constructed according to an 8 × 8 factorial design.

Subjects in the study were asked to rate, on several different bipolar adjective scales, their own communication with people in eight different role relations in eight different situational contexts. The following statement is an example of an item to be rated.

> On a 9 point scale ranging from very hostile (1) to very friendly (9) rate how friendly you think the interaction is between you and a co-worker when having a brief exchange about a minor detail.

The eight situations and eight relationships used in the study are listed in Table 1. Each situation was paired with each relationship for a total of 64 stimuli. For the analyses that will be described we selected a subset of 24 subjects making judgements on three scales related to cooperativeness, for a total of 72 judge-scale combinations. The scales were:

cooperative-competitive,
no conflict-constant conflict,
friendly-hostile.

We were interested in constraining the stimulus configuration from an Eckart-Young decomposition so that it can be completely accounted for in a simple way from the situation and the relationship. For this purpose we used the additive model, which yields par-

TABLE 1
Interpersonal Relationships and Situational Contexts that Served as Stimuli in Study for Two-way CANDELINC Example

Relationships	
1.	Spouse or Best Friend
2.	College Advisor
3.	Casual Acquaintance
4.	Co-worker
5.	Father
6.	Sibling
7.	Supervisor
8.	Person You Dislike Most
Situational Contexts	
A.	Talking to each other at a large social gathering.
B.	Expressing anxiety about a national crisis that is affecting both of you personally.
C.	Having a brief exchange about a minor technical detail.
D.	Pooling knowledge and skills to solve a difficult problem.
E.	Working for a common goal with one of you directing the other.
F.	Discussing a controversial social issue on which your opinions differ.
G.	Attempting to work out a compromise when your goals are strongly opposed.
H.	Blaming one another for a serious error that was made.

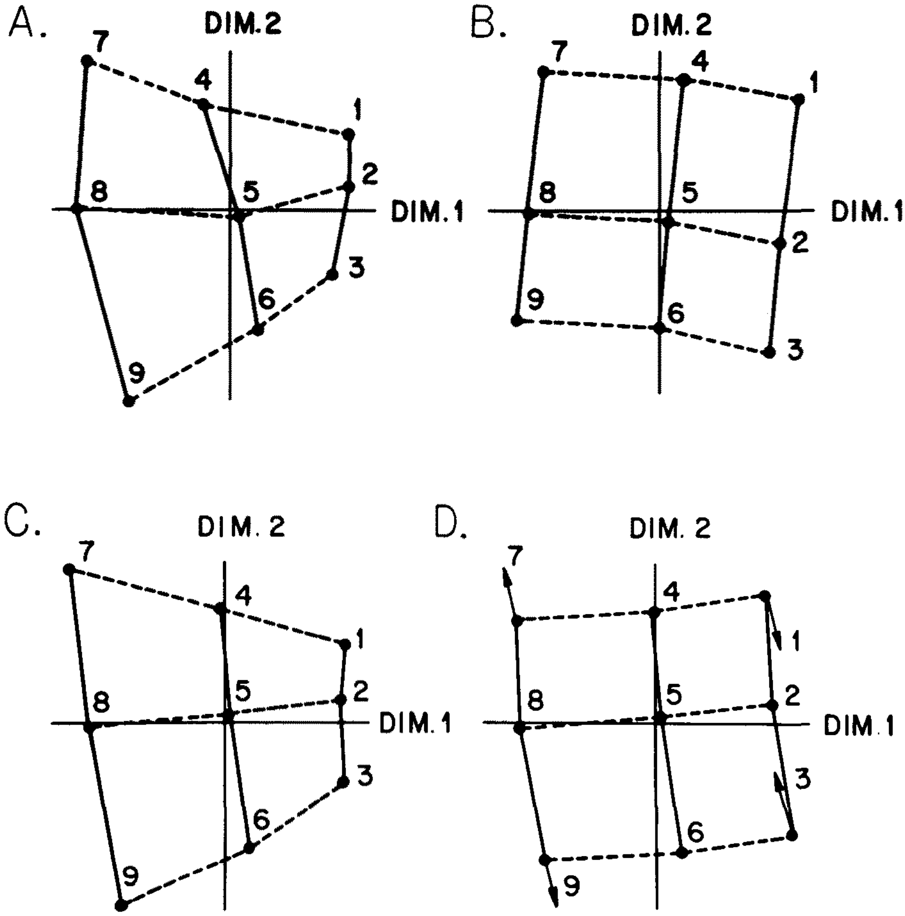

FIGURE 1

Unrotated stimulus configurations from three different two-dimensional analyses, one with no constraints and two with different linear constraints. The data set consisted of 9 stimuli formed by a 3 × 3 factorial design. The solid lines connect the 3 levels of Factor One and the dashed lines connect the 3 levels of Factor Two.

A. an unconstrained INDSCAL solution
B. a solution constrained to fit an additive model
C. a solution constrained to fit a model which contains the additive terms and a linear × linear interaction
D. the same solution as Figure 1C showing the additive portion (connected by the solid and dashed lines) and the effects of the interaction component (represented by arrows)

allel curves like those in Figure 1. We wanted to compare the results from this analysis with an unconstrained Eckart-Young analysis to see how similar the two solutions would be. We used the two-way CANDELINC method, described above, to perform the constrained analysis and the MDPREF procedure [Carroll, 1972] to perform the unconstrained Eckart-Young analysis.

The data matrix used as input to both procedures consists of a 72 × 64 matrix. Each row represents a combination of one subject (from 24) with one scale (from 3). Each column represents a combination of one situation (from 8) with one relationship (from 8).

In addition, a design matrix for the 64 stimuli was provided as input to CANDELINC. (No design matrix was used for the subjects.) The 64 × 14 design matrix embodied the familiar additive model for the 8 × 8 design. There are 7 effects for the 8 situations, and 7 effects for the 8 relationships. The first 7 columns of the design matrix correspond to situation effects, the last 7 to relationship effects. No column is needed to correspond to the grand mean, since this is constrained to be 0.

Both analyses yield a set of stimulus coordinate values as well as a set of coordinates of termini of judge vectors. Figure 2A shows a two-dimensional CANDELINC solution rotated 45 degrees for interpretability. Figure 2B shows the two-dimensional MDPREF solution rotated to best fit the dimensions of Figure 2A. (The rotational orientation of MDPREF solutions are arbitrary, so both rotations are legitimate.) The coding of the stimuli are as in Table 1; the letters refer to the situational context and the numbers to the relationship. Dimension One, for both solutions, can be interpreted as friendliness, from spouse or best friend (1) on the right to person you dislike most (8) on the left. Dimension Two, for both solutions, can be labelled amount of conflict, with situation A, talking to each other at a social gathering, on the top and situation H, blaming one another for a serious error, on the bottom.

The two solutions are very similar, as we see from the very high correlations in Table 2 between corresponding dimensions of the CANDELINC and the MDPREF solutions. The amount of variability accounted for in the constrained solution, 0.539 is almost as great as for the unconstrained solution, 0.555. The degrees of freedom for the data and both the unconstrained and constrained solutions are given in Appendix B. These results indicate that these data can be accounted for well by an additive model for stimuli.

LINCINDS: The Application of CANDELINC to INDSCAL

INDSCAL with Linear Constraints

One principal application of multi-way CANDELINC could be to provide a constrained version of INDSCAL [Carroll & Chang, 1970]. We call this procedure LINCINDS (for *LIN*early Constrained *INDS*CAL). The unconstrained version of INDSCAL utilizes a symmetric version of the Carroll-Chang [1970] CANDECOMP procedure for decomposition of N-way tables via a model of the form given in (1) (but without constraints). Symmetric CANDECOMP consists simply of application of (ordinary) CANDECOMP to a three-way array that is symmetric in two of its indices, say indices 2 and 3. In the case of INDSCAL, the three-way array is a subjects × stimuli × stimuli array comprising a matrix of derived "scalar products" between stimuli for each subject. The "scalar products" are derived from similarities or dissimilarities data by the procedures described in Torgerson [1958] in his chapter on what is now called "classical" (two-way) multidimensional scaling. It can be shown that if CANDECOMP is applied to an array exhibiting such symmetry, the solution will also display this symmetry, up to appropriate equivalence relations. Specifically, suppose the array is symmetric in indices 2 and 3. Then matrices A_2 and A_3 in the solution will be related by a nonsingular *diagonal* transformation matrix. A final step in symmetric CANDECOMP entails normalizing A_2 and A_3 to make sure they are *actually* equal (i.e., that the diagonal transformation is an identity).

The constrained version of INDSCAL, then, entails applying three-way CANDELINC instead of CANDECOMP to the derived scalar products array. In this case, X_2 and X_3 would, of course, be the same. In addition, where CANDECOMP is applied to a reduced array, for the INDSCAL case *symmetric* CANDECOMP would be used.

LINCINDS Example

We illustrate the use of LINCINDS by applying it to a study by Wish and Kaplan [1977] that is very similar to the one described above. In the study 72 subjects were asked to rate, on 14 bipolar adjective scales, the communication of *typical* interpersonal relationships in different situational contexts, rather than their own relationships. Again, subjects

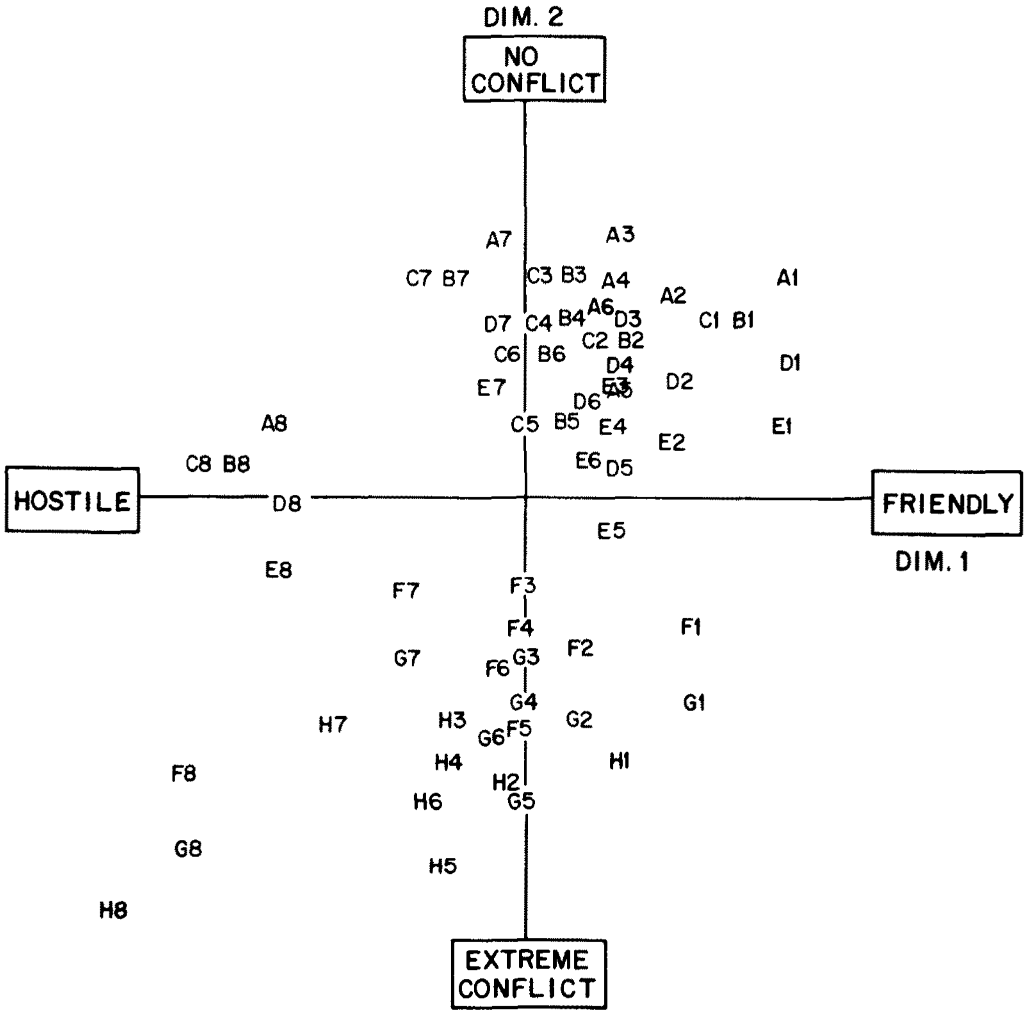

FIGURE 2

Stimulus configurations from two different two-dimensional analyses, one with linear constraints and one with no constraints. The data set consisted of 64 stimuli formed by an 8 × 8 factorial design.

A. a CANDELINC solution, constrained to fit an additive model, rotated 45 degrees for interpretability

were presented all combinations of eight situations and eight relationships for a total of 64 stimuli. The situations were the same as in the previous experiment (see Table 1), but the relationships were somewhat different. Table 3 shows the eight relationships and the 14 bipolar scales. Thus the raw data array Z is 72 subjects (i) × 14 bipolar scales (j) × 64 stimuli (k).

Wish and Kaplan did an INDSCAL analysis to determine the dimensions people implicitly used in making judgements about interpersonal communication based on all 14 of the rating scales. (The purpose of this analysis differs from the purpose of the MDPREF analysis described above. The MDPREF analysis was concerned with the dimensionality for one particular set of *related* adjectives.) The INDSCAL procedure was applied to an array $y_{kjj'}$ which is 14 bipolar scales (k) × 64 stimuli (j) × 64 stimuli (j') as computed from

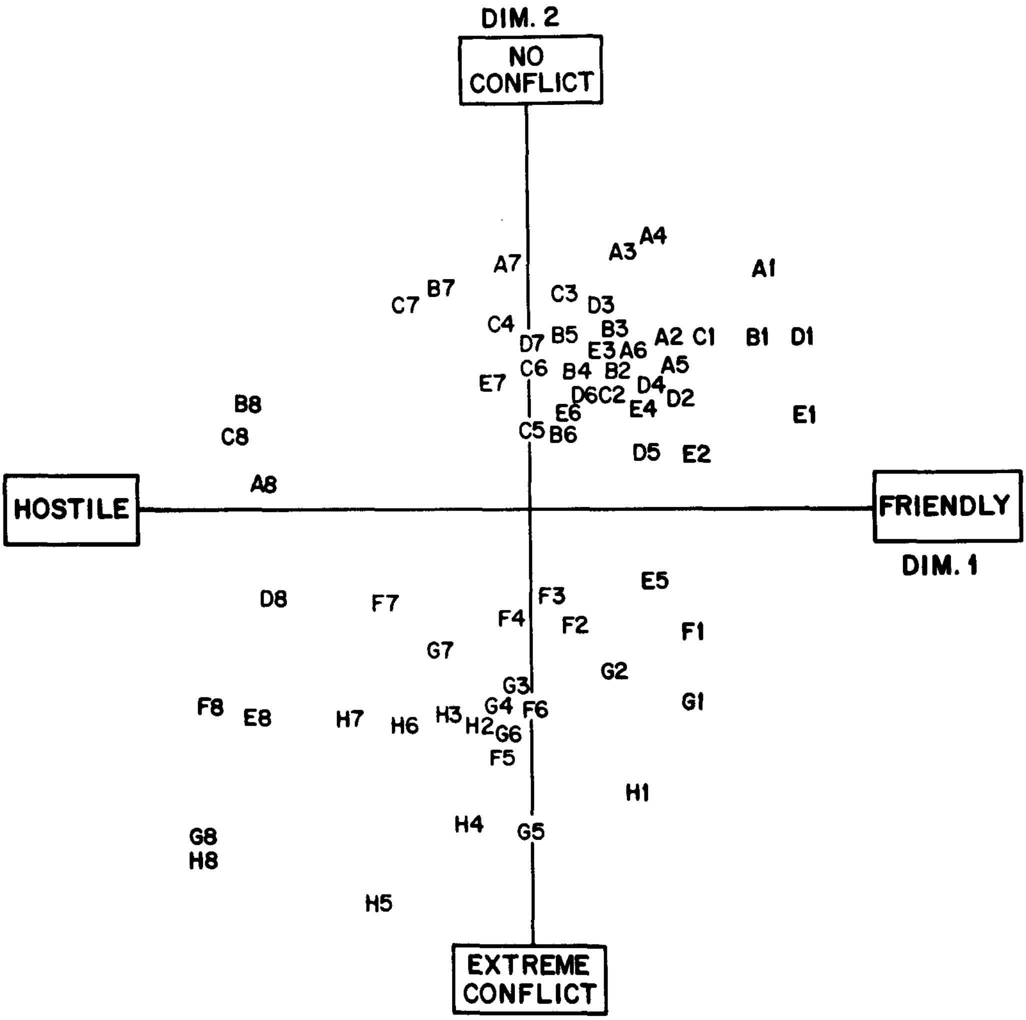

FIGURE 2 (cont)

B. an unconstrained MDPREF solution rotated to best fit the dimensions of the rotated CANDELINC solution

TABLE 2
Comparison of 2-Dimensional Solutions from MDPREF and Two-way CANDELINC with Stimulus Design Matrix for an Additive Model

	Variance Accounted For	
MDPREF Solution	.555	
CANDELINC—Additive Model	.539	
Correlation Between Corresponding Dimensions		
	DIM. 1	DIM. 2
Judge Vectors	.977	.966
Stimulus Coordinates	.987	.978

TABLE 3
Interpersonal Relationships and Bipolar Rating Scales Used in Study for LINCINDS Example

Relationships		
	1.	Bitter Enemies
	2.	Business Partners
	3.	Casual Acquaintances
	4.	Husband and Wife
	5.	Marine Sergeant and Private
	6.	Parent and Teenager
	7.	Political Rivals
	8.	Supervisor and Employee
Bipolar Rating Scales		
	1.	Hostile vs. Friendly
	2.	Cooperative vs. Competitive
	3.	No Conflict vs. Constant Conflict
	4.	Superficial vs. Intense
	5.	Uninterested & Uninvolved vs. Completely Engrossed
	6.	Unemotional vs. Emotional
	7.	Impersonal vs. Personal
	8.	Very different Roles vs. Very Similar Roles
	9.	Treat Each Other Equally vs. One Totally Dominant
	10.	Democratic vs. Autocratic
	11.	Informal vs. Formal
	12.	Reserved and Cautious vs. Frank & Open
	13.	Not at All Task Oriented vs. Entirely Task Oriented
	14.	Unproductive vs. Productive

the raw data by the formula

$$y_{kjj'} = \left[\sum_i (z_{ijk} - z_{ij'k})^2 \right]^{1/2}.$$

The array $y_{kjj'}$, can be thought of as consisting of 14 symmetric 64 × 64 matrices, whose entries are dissimilarities derived from the raw data. Kaplan and Wish interpreted the five dimensions of their solution respectively as cooperation, intensity, dominance, formality, and task orientation. They also reported several small but statistically significant interactions between situations and relationships for each dimension, and provided interpretation for these interactions.

The LINCINDS analysis is based on the same 14 × 64 × 64 array used above and is carried out in five-dimensional space for comparability with the INDSCAL analysis. All interactions between situations and relationships in the stimulus configuration were constrained to be 0, i.e., the stimulus configuration was constrained to be additive with respect to the 8 × 8 structure of the stimuli. (The design matrix used for this purpose is the same 64 × 14 matrix used in the two-way CANDELINC analysis and described in that connection.) It is interesting to see what effect is achieved by eliminating the interactions found by Wish and Kaplan. More interesting, perhaps, would be constraints which eliminate all the interactions except those found by Wish and Kaplan. This analysis would be expected to yield cleaner results than the unconstrained analysis.

INDSCAL solutions, and hence also LINCINDS solutions, are not subject to free rotation, but only to permutation and reversal (reflexion) of axes. Thus when measuring how similar two solutions are it is appropriate merely to correlate corresponding dimensions. Table 4 shows this comparison. As it happened correspondingly numbered dimensions of the INDSCAL and LINCINDS solutions correspond to each other and no axes

TABLE 4
Comparison of 5-Dimensional Solutions from INDSCAL and LINCINDS with Stimulus Design Matrix for an Additive Model

	Variance Accounted For				
INDSCAL Solution	.862				
LINCINDS—Additive Model	.798				
	Correlation Between Corresponding Dimensions				
	DIM.1	DIM.2	DIM.3	DIM.4	DIM.5
Rating Scale Weights	.999	.999	.999	.998	.999
Stimulus Coordinates	.982	.986	.979	.970	.979

are reversed, as we see from the high positive correlations. (This is not so surprising as it might at first appear, since each program numbers the dimensions according to variance accounted for, and since the programs use closely related numerical procedures.)

The amount of variance accounted for in the constrained solution is 0.798, slightly smaller than for the unconstrained solution, which is 0.862. The degrees of freedom for the data and both the unconstrained and constrained solutions are given in Appendix B. The correlations for the rating scale weights are all 0.998 or above, quite a bit higher than the correlations for the stimulus coordinates. Presumably this is because the constraints affect the stimulus coordinates directly, and affect the rating scale weights only indirectly.

LINCINDS as an Aid for Interpreting Dimensions

Another application of the CANDELINC method, to be reported in detail elsewhere, shows how LINCINDS was used as an aid in interpreting the dimensions of an INDSCAL solution. In this example, acoustic stimuli consisting of 24 frequency modulated tones were constructed according to a 4×3×2 factorial design. By appropriate selection of design matrices we constrained the dimensions to fit several different models incorporating several different interaction terms, and compared the solutions of the various models with the unconstrained INDSCAL solution. We chose the model that fit the data nearly as well as the INDSCAL solution and used it to explain the dimensions of the original scaling solution.

Application of LINCINDS to Define a Rapid Approximate INDSCAL Solution

It was recently realized by the first two authors that one interesting and potentially very useful application of LINCINDS is to provide a computationally highly efficient procedure for an approximate INDSCAL solution. This solution may in many cases be sufficiently good in itself to be used instead of a full INDSCAL analysis. Otherwise it can be used to define an excellent "rational" starting configuration for the full least squares INDSCAL solution. The details of this procedure will be published separately.

A brief description of the procedure is given here, although a full account is being published separately [Carroll & Pruzansky, Note 3]. First, the Young-Householder procedure is applied to convert dissimilarity matrices for each subject into "scalar product" form. Then these matrices are averaged to provide one single matrix. The first several eigenvectors of this matrix are calculated, and used as the columns of a matrix $\mathbf{X}_2$. Let matrix $\mathbf{A}_2$ be the INDSCAL stimulus configuration of the same dimensionality as $\mathbf{X}_2$. It can be proved, for error-free data satisfying the INDSCAL model, that the columns of $\mathbf{X}_2$ differ from the columns of $\mathbf{A}_2$ only by a linear transformation; that is, $\mathbf{A}_2 = \mathbf{X}_2\mathbf{T}_2$ where $\mathbf{T}_2$ is a square matrix (usually, but not necessarily nonsingular). This relationship is also observed to hold approximately for real data. LINCINDS is used, with $\mathbf{X}_2$ serving as the "design matrix" for stimuli, to find $\mathbf{T}_2$ and $\mathbf{A}_2$. By this means, a good approximate solution to a large INDSCAL problem can be found by solving a small INDSCAL problem. This procedure is exactly equivalent to the TRAIS procedure proposed by Cohen [Note 4] for rotating the "group stimulus space" from a three-mode scaling [Tucker, 1972] solution to approximate GNDSCAL structure to aid interpretability.

A further extension is possible in which an internally generated design matrix is used for subjects as well as for stimuli. This leads to finding an approximate INDSCAL solution by solving an even smaller INDSCAL problem.

Linear Constraints Are Not Desirable for Subject Vectors

We have tried, at various times, to put constraints on subject vectors, either in the two-way constrained MDPREF-like metric analysis in terms of a vector model for preference data, or in the case of subject weight vectors in the LINCINDS approach to constrained INDSCAL analysis. In both cases the results were very poor. The stimulus space was very severely distorted and the corresponding fit measure declined precipitously. The first instance in which this precipitous decline in variance accounted for was observed was the analysis reported by Carroll, Green and Carmone [Note 1; also see Green, Carroll & Carmone, 1976].

It has now become clear to us that this empirically observed degeneration of configuration and fit measures can be related to a theoretical argument against putting constraints of this type on entities which are interpreted as vectors, rather than as points. Recall that in INDSCAL the length of the subject vectors has only a secondary meaning, based on the degree of fit of that subject's data matrix, and is not intrinsic to the model. Also note that while the length of the subject vector in MDPREF is intrinsic to the model, it depends only on the degree of fit and the overall size of that subject's data vector. Since the directional information is of primary interest in the case of such vector-like entities, it would seem that only constraints should be considered that are invariant under changes in length of individual vectors. In particular, it should be clear that the linear constraints imposed in CANDELINC are *not* invariant under such length altering transformations. An easy way to see this is to imagine subject points arrayed in a regular square or rectangular lattice arrangement in the positive quadrant of a two-dimensional space. If we now change the lengths, say by normalizing all vectors to unit length, this lattice structure will be almost wholly destroyed.

This theoretical objection to linear constraints on subject vectors is closely related to MacCallum's [1977] argument, on very similar grounds, that no procedures based on a general linear model should be used as aids in interpreting INDSCAL subject spaces. We partially agree with MacCallum, although we disagree with his flat prohibition against use of such techniques. Our criticism of MacCallum's argument will be discussed elsewhere.

Discussion

Constrained MDS seems to be part of the "Zeitgeist". Approaches include those by Bentler and Weeks [1978], Bloxom [1978], Noma and Johnson [Note 10], deLeeuw and Heiser [Note 5], and Borg and Lingoes [1980].

The approach of Bentler and Weeks [1978] allows specific parameters (in a two-way MDS context) to be constrained either to equality or proportionality with some specified values. One useful special case of this approach would allow constraining a particular dimension of the stimulus space in a two-way MDS analysis to be proportional to some specified outside variable. This is clearly a "linear constraint", but of a quite different type than allowed in CANDELINC. This amounts to assuming a one-one relationship between specific outside variables and dimensions. In the language of our paper, the Bentler and Weeks constraints come to requiring that the stimulus configuration **A** satisfy **A** = **XT** where **X** and **T** have the following properties:

(i) **T** is *diagonal*;

(ii) some columns of **X** are fixed (corresponding to constrained columns of **A**) and some columns of **X** are free, that is, have values chosen during the optimization (corresponding to unconstrained columns of **A**);

(iii) some diagonal elements of **T** are fixed (corresponding to equality constrained columns of **A**), some are free (corresponding to proportionality constrained columns of **A**), and some may be either (corresponding to unconstrained columns of **A**).

CANDELINC entails a much more general relationship between variables and dimensions. All dimensions are assumed to be different linear combinations of the same set of variables. An intermediate case that might be of interest is one in which each dimension, for example, is assumed to be a linear combination of a set of variables uniquely associated with that dimension.

Noma and Johnson [Note 10] allow a case similar to the special case of Bentler and Weeks' approach, just described, in which there is assumed to be a one-one relationship between variables and dimensions, but this is defined only ordinally. That is, the outside variable is assumed to be defined on a merely ordinal scale, or, equivalently, the rank order of projections on each dimension is constrained, but not the exact values (up to a constant of proportionality) as would be true in Bentler and Weeks.

Bloxom's [1978] approach entails a more general model than the others. The model is the one for individual differences MDS involving generalized euclidean metrics called IDIOSCAL by Carroll [see Carroll & Wish, 1974] whose special cases include Tucker's [1972] three-mode scaling, the INDSCAL model [Horan, 1969; Carroll & Chang, 1970], Harshman's [Note 6] PARAFAC and PARAFAC-2 [Harshman, Note 7] as well as the standard two-way MDS model. The constraints that Bloxom considers, however, are limited to strict equality constraints. That is, each parameter is constrained to be equal to some specified value a priori, or pairs or subsets of parameters can be constrained to be equal to one another. Since, however, these constraints can be placed on many different components of the general model, this allows, in fact, a fairly wide class of constraints to be imposed. Bloxom [1978] provides a good discussion of these procedures for constrained solutions, as well as the relationship of CANDELINC to his own approach. Bloxom also relates these procedures to an older procedure proposed by McGee [1968] for constraining the degree of relatedness of different MDS configurations.

DeLeeuw and Heiser [Note 5] describe a very general "algorithm model" (that is, an incompletely specified algorithm) for fitting MDS models with any type of constraint for which an associated "metric projection" problem can be solved. This metric projection problem is essentially that of projecting the current set of parameter estimates into the constraint region (the region of parameter space satisfying the constraints) in a least squares fashion (i.e., to find the point in the constraint region closest to the current parameter point in euclidean distance.) For some types of constraints the solution to this metric projection problem is "nice" and "easy"; for others it is very hard indeed. DeLeeuw and Heiser discuss the general solution for some cases of interest, and also provide a discussion relating their approach to some of the others mentioned here, as well as to CANDELINC.

Finally, the approach of Borg and Lingoes [1980] constrains certain distances in the obtained configuration to satisfy specified ordinal constraints. This is equivalent to providing two proximity matrices as input, one corresponding to data, and a second specifying the ordinal constraints. This second matrix will generally have two special characteristics; 1) it will have a large number of missing entries and 2) the order constraints implied by the nonmissing entries can be satisfied exactly (but, typically, in more than one way) in the dimensionality in which the solution is to be obtained. Borg and Lingoes use a penalty function approach to require the constraints defined by the "second" matrix to be satisfied exactly, while the "first" matrix is fit as well as possible within the constraints imposed by the "second."

The Borg and Lingoes approach is actually very similar to a possibility that has existed for some time within the scope of the MDSCAL-5 program of Kruskal [Note 8] and the KYST and KYST-2 programs of Kruskal, Young and Seery [Note 9]. In these procedures it is possible to provide two or more matrices as input (each with data values permitted to be missing), with different weights associated with each matrix. The overall loss function is a root mean square of the weighted stress values associated with the individual matrices. Constraints essentially identical to those of Borg and Lingoes can be achieved in the earlier programs by using two matrices and associating a very large weight with the second matrix relative to the weight associated with the first. The Borg and Lingoes computational procedure is equivalent to allowing the second weight, asymptotically, to grow "infinitely large." In practice, however, their penalty function approach will simply increase the second weight in a sequence of steps until it grows "sufficiently large" to force exact satisfaction of the ordinal constraints. When imposing Borg-Lingoes type constraints with MDSCAL or KYST, a result in principle identical to theirs can be achieved by using a weight which is sufficiently large. Furthermore, it might be noted that the MDSCAL or KYST programs permit use of more than two matrices. Thus, for example, several different sets of ordinal constraints could be imposed simultaneously, by allowing "sufficiently large" weights for each of a set of matrices defining these constraints. This can be done if the constraints are mutually consistent and can all be satisfied in the dimensionality specified for the solution.

Appendix A: Design Matrices

As illustration we use a tiny data set consisting of 9 × 9 matrices of dissimilarities provided by three subjects. The nine stimuli form a 3 × 3 factorial design, as shown here,

	Factor 2 Levels 1	2	3
Factor 1 Levels 1	1	2	3
2	4	5	6
3	7	8	9

where the cell entry indicates stimulus number, or here,

Stimulus Number	Factor 1	Factor 2
1	1	1
2	1	2
3	1	3
4	2	1
5	2	2
6	2	3
7	3	1
8	3	2
9	3	3

First consider the additive model constraint on the 9 stimulus vectors, which are the rows of a 9×2 stimulus matrix $\mathbf{A}_2$. We shall write these vectors in the form $a_{i\to} = (a_{i1},a_{i2})$, omitting the superscript "(2)" for simplicity. (The horizontal subscript arrow is used to indicate a row vector.) Algebraically, the additive model means that the stimulus vector $a_{i\to}$ for stimulus (m,n) must have the form

$$a_{i\to} = u_{m\to} + v_{n\to},$$

where $u_{m\to} = (u_{m1},u_{m2})$ with $m = 1,2,3$ are the three "effect" vectors for the first factor, and $v_{n\to} = (v_{n1}, v_{n2})$ for $n = 1,2,3$ are the three "effect" vectors for the second factor. There is no "grand mean" or "constant term" in the above expression INDSCAL along with many similar methods imposes the condition that the mean of the stimulus vectors be 0. This condition is desirable in order to remove translational indeterminacy. To achieve this condition it is also necessary to assure that the sum of the effects for each factor is 0, namely, that

$$\sum_{m=1}^{3} u_{m\to} = 0 \text{ and } \sum_{n=1}^{3} v_{n\to} = 0.$$

The most obvious way to express the additive model constraint on $\mathbf{A}_2$ in the form $\mathbf{A}_2 = \mathbf{X}_2\mathbf{T}_2$ is to set

$$\mathbf{X}_2 = \left[\begin{array}{ccc|ccc} 1 & & & 1 & & \\ 1 & & & & 1 & \\ 1 & & & & & 1 \\ \hline & 1 & & 1 & & \\ & 1 & & & 1 & \\ & 1 & & & & 1 \\ \hline & & 1 & 1 & & \\ & & 1 & & 1 & \\ & & 1 & & & 1 \end{array}\right], \quad \mathbf{T}_2 = \begin{bmatrix} u_{1\to} \\ u_{2\to} \\ u_{3\to} \\ v_{1\to} \\ v_{2\to} \\ v_{3\to} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \\ v_{11} & v_{12} \\ v_{21} & v_{22} \\ v_{31} & v_{32} \end{bmatrix}$$

where the blank entries indicate zeros. However, this approach does not constrain the sum of the factor one effects and the sum of the factor two effects to be 0. To do so, we eliminate $u_{3\rightarrow}$ and $v_{3\rightarrow}$ by using $u_{3\rightarrow} = -u_{1\rightarrow} - u_{2\rightarrow}$ and $v_{3\rightarrow} = -v_{1\rightarrow} - v_{2\rightarrow}$. This leads to

$$X_2 = \left[\begin{array}{cc|cc} 1 & & 1 & \\ 1 & & & 1 \\ 1 & & -1 & -1 \\ \hline & 1 & 1 & \\ & 1 & & 1 \\ & 1 & -1 & -1 \\ \hline -1 & -1 & 1 & \\ -1 & -1 & & 1 \\ -1 & -1 & -1 & -1 \end{array}\right], \quad T_2 = \begin{bmatrix} u_{1\rightarrow} \\ u_{2\rightarrow} \\ v_{1\rightarrow} \\ v_{2\rightarrow} \end{bmatrix}.$$

Now consider the model incorporating a linear × linear interaction term. Algebraically, for the 3 × 3 case, this means that the stimulus vector for stimulus (m,n) has the form

$$u_{m\rightarrow} + v_{n\rightarrow} + e_{m,n} w_{\rightarrow}$$

where $w_{\rightarrow}$ is the interaction vector and

$$e_{m,n} = \begin{cases} 1 & \text{if } m = n = 1 \text{ or } 3, \\ -1 & \text{if } m = 1 \text{ and } n = 3 \text{ or vice versa}, \\ 0 & \text{otherwise.} \end{cases}$$

This extra term is incorporated into $\mathbf{X}_2$ and $\mathbf{T}_2$ as follows:

$$\mathbf{X}_2 = \left[\begin{array}{cc|cc|c} 1 & & 1 & & 1 \\ 1 & & & 1 & \\ 1 & & -1 & -1 & -1 \\ \hline & 1 & 1 & & \\ & 1 & & 1 & \\ & 1 & -1 & -1 & \\ \hline -1 & -1 & 1 & & -1 \\ -1 & -1 & & 1 & \\ -1 & -1 & -1 & -1 & 1 \end{array}\right], \quad \mathbf{T}_2 = \begin{bmatrix} u_{1\rightarrow} \\ u_{2\rightarrow} \\ v_{1\rightarrow} \\ v_{2\rightarrow} \\ w_{\rightarrow} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ v_{11} & v_{12} \\ v_{21} & v_{22} \\ w_1 & w_2 \end{bmatrix}.$$

Appendix B: Degrees of Freedom for Constraints

In response to concerns expressed by one referee about whether there are adequate degrees of freedom to determine how well the constraints fit, it seems worthwhile to count the degrees of freedom explicitly. In the 2-way CANDELINC application, the data consist of a 72 × 64 matrix, the unconstrained solution consists of 72 × 2 subject-scale parameters plus 64 × 2 stimulus parameters, and the constrained solution consists of 72 × 2 subject-scale parameters plus 14 × 2 stimulus parameters. The data have 4536 = 72 × 64 − 72 degrees of freedom, (where 72 is subtracted because each subject-scale variable is centered). The unconstrained solution has 266 = 72 × 2 + 64 × 2 − 6 degrees of freedom, and the constrained solution has 166 = 72 × 2 + 14 × 2 − 6 degrees of freedom (where the 6 which are subtracted in each case represent 2 for centering the stimuli at the origin and 4 for the arbitrary 2 × 2 matrix of the fundamental indeterminacy).

In the LINCINDS application, the data consist of a 14 × 64 × 64 symmetric array, the unconstrained solution consists of 64 × 5 stimulus parameters plus 14 × 5 scale weight parameters, and the constrained solution consists of 14 × 2 stimulus parameters plus 14 × 2 scale weight parameters. The data have 28224 = 14 × (64 × 63)/2 degrees of freedom, the unconstrained model has 385 = 64 × 5 + 14 × 5 − 5 degrees of freedom, and the constrained model has 135 = 14 × 5 + 14 × 5 − 5 degrees of freedom.

From these calculations, it would appear that in these examples at least there is no lack of degrees of freedom for any relevant purpose.

REFERENCE NOTES

1. Carroll, J. D., Green, P. E., & Carmone, F. J. *CANDELINC: A new method for multidimensional analysis with constrained solutions.* Paper presented at the meeting of the International Congress of Psychology, Paris, France, July 1976.
2. Carroll, J. D., & Pruzansky, S. *MULTILINC: Multiway CANDELINC.* Paper presented at the meeting of the American Psychological Association, San Francisco, August 1977.
3. Carroll, J. D. & Pruzansky, S. *Use of LINCINDS as a rational starting configuration for INDSCAL.* Unpublished memorandum, Bell Telephone Laboratories, 1979.
4. Cohen, H. S. *Three-mode rotation to approximate INDSCAL structure.* Paper presented at the annual meeting of the Psychometric Society, Palo Alto, California, March, 1974.
5. DeLeeuw, J. & Heiser, W. *Multidimensional scaling with restrictions on the configuration.* Proceedings of the Fifth International Symposium on Multivariate Analysis, Pittsburgh, June 1978.
6. Harshman, R. A. Foundations of the PARAFAC procedure; Models and conditions for an "explanatory" multi-modal factor analysis, *UCLA Working Papers in Phonetics,* 1970, *16,* 1–84.
7. Harshman, R. A. PARAFAC-2; mathematical and technical notes, *UCLA Working Papers in Phonetics,* 1972, *22,* 30–44.
8. Kruskal, J. B. & Carmone, F. *How to use M-D-SCAL (Version 5M) and other useful information.* Unpublished memorandum, Bell Telephone Laboratories, 1969.
9. Kruskal, J. B., Young, F. W., & Seery, J. B. *How to use KYST, a very flexible program to do multidimensional scaling and unfolding.* Unpublished memorandum, Bell Telephone Laboratories, 1973.
10. Noma, E., & Johnson, J. *Constraining nonmetric multidimensional scaling configurations* (Tech. Rep. 60). Ann Arbor: The University of Michigan, Human Performance Center, August 1977.
11. Tucker, L. R. & MacCallum, R. C. Personal communication, 1974.

REFERENCES

Bentler, P. M., & Weeks, D. G. Restricted multidimensional scaling models, *Journal of Mathematical Psychology,* 1978, *17,* 138–151.

Bloxom, B. Constrained multidimensional scaling in N spaces. *Psychometrika,* 1978, *43,* 397–408.

Borg, I. & Lingoes, J. C. A model and algorithm for multidimensional scaling with external constraints on the distances. *Psychometrika,* 1980, *45,* 25–38.

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* (Vol. I). New York, Academic 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,* 283–320.

Carroll, J. D. & Wish, M. Multidimensional perceptual models and measurement methods. In E. C. Carterette & M. P. Friedman (Eds.). *Handbook of perception. (Vol. II) Psychophysical judgement and measurement.* New York: Academic Press, Inc., 1974, 391–447.

Eckart, C. & Young, G. The approximation of one matrix by another of lower rank, *Psychometrika,* 1936, *1,* 211–218.

Green, P. E. Carroll, J. D. & Carmone, F. J. Superordinate factorial designs in the analysis of consumer judgments. *Journal of Business Research,* 1976, *4,* 281–295.

Horan, C. B. Multidimensional scaling: Combining observations when individuals have different perceptual structures, *Psychometrika,* 1969, *34,* 139–165.

Johnson, R. M. Minimal transformation to orthonormality, *Psychometrika,* 1966, *31,* 61–66.

MacCallum, R. C. Effects of conditionality on INDSCAL and ALSCAL weights, *Psychometrika,* 1977, *42,* 297–305.

McGee, V. E. Multidimensional scaling of N sets of similarity measures: A nonmetric individual differences approach, *Multivariate Behavioral Research,* 1968, *3,* 233–248.

Torgerson, W. S. *Theory and Methods of Scaling,* New York: Wiley, 1958.

Tucker, L. The extension of factor analysis to three-dimensional matrices, In Frederiksen, N. & Gulliksen, H. (Eds.) *Contributions to mathematical psychology.* New York: Holt, Rinehart & Winston, 1964, 109–127.

Tucker, L. R. Relations between multidimensional scaling and three-mode factor analysis, *Psychometrika,* 1972, *37,* 3–27.

Wish, M. Subjects' expectations about their own interpersonal communication: A multidimensional approach, *Personality and Social Psychology Bulletin,* 1975, *1,* 501–504.

Wish, M. & Kaplan, S. J. Toward an implicit theory of interpersonal communication, *Sociometry,* 1977, *40,* 234–246.

Wold, H. Estimation of principal components and related models by iterative least squares, In Krishnaiah, P. R. (Ed.) *Multivariate Analysis,* New York: Academic Press, 1966, 391–420.

Manuscript received 1/30/79
Final version received 10/24/79