PSYCHOMETRIKA—VOL. 43, NO. 4.
DECEMBER, 1978

A NEW COMPUTATIONAL METHOD TO FIT THE WEIGHTED EUCLIDEAN DISTANCE MODEL

JAN DE LEEUW

UNIVERSITY OF LEIDEN

SANDRA PRUZANSKY

BELL LABORATORIES

This paper describes a computational method for weighted euclidean distance scaling which combines aspects of an "analytic" solution with an approach using loss functions. We justify this new method by giving a simplified treatment of the algebraic properties of a transformed version of the weighted distance model. The new algorithm is much faster than INDSCAL yet less arbitrary than other "analytic" procedures. The procedure, which we call SUMSCAL (*su*bjective *m*etric *scal*ing), gives essentially the same solutions as INDSCAL for two moderate-size data sets tested.

Key word: multidimensional scaling.

Introduction

The weighted euclidean distance model originated independently by Horan [1969], Bloxom [Note 1], and Carroll and Chang [1970], is the basis for one of the most popular multidimensional scaling methods. The INDSCAL method of Carroll and Chang [1970] seems to provide a satisfactory approximation to important structural characteristics of the data in most of the cases in which it has been applied, and its feature of uniquely orienting the axes greatly facilitates interpretation. From the computational point of view the INDSCAL program has been quite successful, but convergence of the procedure is usually painfully slow. An alternative computational procedure has been suggested by Schönemann [1972]. His method gives a very simple "analytic" solution to the model equations in those cases in which perfect fit can be obtained. In most practical cases exact solutions will not be possible, and we have to be satisfied with approximate ones. Schönemann's methods have the disadvantage of making some arbitrary choices, which do not influence the solution in the perfect case, but which may greatly influence the solution in the fallible case. Carroll and Chang [Note 2] have proposed a modification of Schönemann's procedure in their IDIOSCAL program which seems considerably better from a statistical point of view, although the element of arbitrariness is still not eliminated completely. In this paper we combine aspects of Schönemann's "analytic" solution with an approach using loss functions. The result is an algorithm for weighted euclidean distance scaling which we call SUMSCAL (*su*bjective *m*etric *scal*ing). SUMSCAL is considerably faster than Carroll and Chang's INDSCAL procedure and considerably less arbitrary than Schönemann's.

Comments by J. Douglas Carroll and J. B. Kruskal have been very helpful in preparing this paper.

Requests for reprints should be sent to Sandra Pruzansky, Bell Laboratories, 2C-573, 600 Mountain Avenue, Murray Hill, New Jersey 07974.

1. The Model and Its Transformations

The basic model we are analyzing in this paper is

$$\delta_{ijk} \approx \left[\sum_{s=1}^{r} w_{ks}^2 (x_{is} - x_{js})^2 \right]^{1/2}. \tag{1}$$

Here δ_{ijk} is the (known) dissimilarity between *objects* i and j, as measured on *data source* k. The x's are the *coordinates* of the objects in *r-dimensional space* and the w's are weights or *saliences*. For an extensive discussion of the appropriateness of this model, and for its interpretation in several interesting applications, we refer to the basic Carroll and Chang [1970] paper, and to work described by Wish and Carroll [1974]. If we use d_{ijk} for the weighted euclidean distance defined in (1), then one obvious computational approach would be minimizing

$$\sum_{k=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} (\delta_{ijk} - d_{ijk})^2. \tag{2}$$

This, in fact, seems to be the logical extension of what Kruskal and Carroll [1969] refer to as "simple scaling". Elegant convergent iterative algorithms for minimizing the loss function (2) have been proposed by Heiser [Note 7] and De Leeuw and Heiser [1977]. Convergence of these algorithms, however, is usually very slow. As a consequence the algorithms are impractical if we do not have a very good starting point for the iterations. It does not seem to be possible to derive a "rational" initial configuration directly from the model equations (1).

It seems worthwhile, therefore, to apply a *transformation,* that is, to replace the model, interpreted as a system of nonlinear equations, by an equivalent system of equations. Of course models in data analysis are usually interpreted as "fuzzy" systems of equations, and transformations are applied as if the equations were exactly solvable. After the transformation we "fuzzify" the new system of equations, which now defines a new (but closely related) model. Thus a transformation actually changes the model, although it does not change the corresponding system of equations in an essential way. To make the discussion less awkward, however, we often call models equivalent if the corresponding systems of nonlinear equations are equivalent, and we call Model A derived from Model B if the system of equations corresponding with A is a consequence of the system corresponding with B.

One first transformation that may be applied involves squaring both sides of (1). This gives the model

$$\delta_{ijk}^2 \approx \sum_{s=1}^{r} w_{ks}^2 (x_{is} - x_{js})^2, \tag{3}$$

and the corresponding least squares loss function

$$\sum_{k=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} (\delta_{ijk}^2 - d_{ijk}^2)^2. \tag{4}$$

This loss function is minimized in the (ratio scale option of the) ALSCAL program of Takane, Young, and De Leeuw [1976]. Although (3) and (4) are simpler than (1) and (2), the minimization of (4) still involves fairly complicated iterative least squares procedures.

There is another transformation, which simplifies the model even more. It has been used by both Carroll and Chang [1970] and Schönemann [1972] in their analysis of the model (1). This transformation, defined in (5), converts the data from each data source into scalar products of vectors. It's history is reviewed in Torgerson's book [1958, Chapter 11]. Define

$$\beta_{ijk} = -\tfrac{1}{2}\{\delta_{ijk}^2 - \delta_{i\cdot k}^2 - \delta_{\cdot jk}^2 + \delta_{\cdot\cdot k}^2\}, \tag{5}$$

where dots replacing indices indicate averaging over the range of that index. In the same way

$$z_{ijk} = -\tfrac{1}{2}\{d^2_{ijk} - d^2_{i.k} - d^2_{.jk} + d^2_{..k}\}, \tag{6}$$

is applied to the right hand side, and the new model is

$$\beta_{ijk} \approx z_{ijk} = \sum_{s=1}^{r} w^2_{ks}(x_{is} - x_{.s})(x_{js} - x_{.s}). \tag{7}$$

Without loss of generality we can assume $x_{.s} = 0$ for all s, and use simply

$$z_{ijk} = \sum_{s=1}^{r} w^2_{ks} x_{is} x_{js}. \tag{8}$$

Here Carroll and Chang stop transforming the model, define the loss function

$$\sum_{k=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{n}(\beta_{ijk} - z_{ijk})^2, \tag{9}$$

and minimize it by using alternating least squares methods.

The advantage of (9) over the previous loss function (4) is that z_{ijk} is a much simpler function of the parameters, and more efficient methods for minimization become available. Nevertheless the Carroll-Chang CANDECOMP algorithm is still iterative. Convergence can be quite slow, and is not necessarily towards the global minimum. A very good initial configuration is needed. It turns out that an additional advantage of transforming model (1) into model (7) is that a 'rational' initial configuration can now be derived fairly easily.

2. *Algebra of Transformed Model*

We translate model (7) into matrix algebra. Define the $n \times n$ matrices Z_k by

$$Z_k = XW^2_kX' \tag{10}$$

with X an $n \times r$ matrix and W^2_k a non-negative diagonal $r \times r$ matrix. If we set B_k equal to the k^{th} slab of β_{ijk}, the basic model becomes

$$B_k \approx Z_k = XW^2_kX'. \tag{11}$$

In this section we are interested in the solvability of the corresponding system of equations

$$B_k = XW^2_kX', \tag{12}$$

or, more generally, in the structure of the solution set of this system. We make two very important simplifying assumptions. In the first place we are only interested in solutions for which W^2_+, the sum of the W^2_k, is nonsingular. Or, equivalently, for each $s = 1, \cdots, r$ there must be at least one k with $w^2_{ks} > 0$. This assumption can be made without any real loss of generality. If W^2_+ is singular, then a solution of our system is possible in $r' < r$ dimensions. In the second place we are only interested in solutions in which X has full column rank r. This excludes a considerable number of situations of theoretical, and possibly also of practical, interest. The assumption excludes, for example, cases in which the number of dimensions, r, is larger than the number of objects, n [cf. Kruskal, 1976]. It also excludes some situations with "oblique" dimensions discussed by Harshman [Note 6].

Schönemann [1972] is not very explicit about his assumptions, but his discussion indicates that he is only interested in solutions for which X has full column rank, and for which *all* W^2_k are positive definite (from now on we shall use the abbreviations pd and psd for positive definite and positive semi-definite).

Following Schönemann [1972], Tucker [1972], and Carroll and Chang [Note 2] we

first consider the more general system

$$B_k = XU_kX', \tag{13}$$

with U_k merely restricted to be symmetric and psd, and not necessarily diagonal. Again we are only interested in solutions for which U_* is *pd*, and for which X is of full column rank (abbreviated fcr). The major difference between (12) and (13) is that the requirement that X is fcr can be made without any real loss of generality for the system (13). More precisely, if (13) has a solution with rank $X = r' < r$, then (13) also has a solution in r' dimensions.

There is a fundamental indeterminacy in the system (13). If X, U_k is a solution, and T is any nonsingular matrix, then $\mathbf{X} = XT^{-1}$ and $\mathbf{U}_k = TU_kT'$ is another solution. For identification purposes we can require U_* to be any pd matrix. Following Schönemann we require $U_* = I$. With this restriction the same indeterminacy exists but with T restricted to orthogonality.

Theorem 1. The system (13) has a solution X, U_k, with X fcr, U_k psd for all k, and U_* pd iff B_k is psd for all k, rank $B_k \leq r$ for all k, and rank $B_* = r$.

Proof. We construct the general solution to (13) with $U_* = I$ to prove sufficiency. Necessity is obvious. If $U_* = I$ then $B_* = XX'$. Because rank $B_* = r$ and B_* is psd we can find an $n \times r$ matrix Y which is fcr, and satisfies $B_* = YY'$. Moreover the general solution to $B_* = XX'$ is given by $X = YT'$ for some orthonormal T. Define $C_k = Y^+B_k(Y^+)'$, with Y^+ denoting the Moore-Penrose inverse of Y. The general solution for U_k, if it exists, is of the form $U_k = TC_kT'$. Observe that indeed $U_* = I$. Now $U_k = TC_kT'$ is a solution if and only if $(I - YY^+)B_k = 0$ for all k. But $I - YY^+$ is the projector corresponding with the null space of B_*, and the null space of B_* is the intersection of the null spaces of the B_k. Thus $(I - YY^+)\,B_k = 0$ for all k, and the theorem is proved.

Our Theorem 1 simplifies the results in Section 3 of Schönemann. For completeness we also state the following corollary.

Corollary 2. Suppose (13) has a solution satisfying the conditions of Theorem 1. Then the solution is of the form

$$X = YT^{-1}, \tag{14a}$$

$$U_k = TC_kT' \tag{14b}$$

with T an arbitrary nonsingular matrix. The solutions satisfying the additional condition $U_* = I$ are of the same form, but with T orthonormal.

Now consider system (12). If X, W_k^2 is a solution, and T is any nonsingular diagonal matrix, then $X = XT^{-1}$ and $W_k^2 = TW_k^2T'$ is another solution. Again we can require without loss of generality that $W_*^2 = I$.

Theorem 3. The system (12) has a solution X, W_k^2, with X fcr, W_k^2 psd for all k, and W_*^2 pd iff B_k is psd for all k, rank $B_k \leq r$ for all k, rank $B_* = r$, and $B_kB_*^+B_h = B_hB_*^+B_k$ for all k, h.

Proof. It follows from Corollary 2 that (12) is solvable under the additional condition $W_*^2 = I$ iff there exists an orthonormal T such that TC_kT' is diagonal for all k. Such a T can be found iff the C_k commute in pairs, i.e., iff $C_kC_h = C_hC_k$ for all k, h. This transforms to the last condition mentioned in the theorem.

Our Theorem 3 simplifies the theorem given in Section 4 of Schönemann. It is closely related to a result of Bhimasankaram [1971, Section 4, Corollary 2].

Corollary 4. Suppose (12) has a solution satisfying the conditions of Theorem 3. Then the solution is of the form (14) with T a nonsingular matrix that diagonalizes all C_k. The solutions satisfying the additional condition $W_*^2 = I$ are of the same form, but now T is an orthonormal matrix that diagonalizes all C_k.

A final corollary gives a simple uniqueness theorem which is stronger than previous theorems proved by Harshman [Note 5, Note 6].

Corollary 5. Suppose X, W_k^2 is a solution of (12) with X fcr, W_k^2 psd for all k, and $W_*^2 = I$. Then this solution is the unique solution satisfying these conditions (up to a permutation of the dimensions) iff for each s, t there is at least one k such that $w_{ks}^2 \neq w_{kt}^2$.

Proof. The solution is unique iff the transformation matrix T is unique iff there is at least one linear combination of the C_k with different eigenvalues. But the eigenvalues of linear combinations of the C_k can be found by forming the same linear combinations of the W_k^2. Thus there must be at least one linear combination of the W_k^2 with all diagonal elements different. This is the condition mentioned in the corollary. Our results can be extended to the case in which we do not make the assumption that X is fcr [cf. also Kruskal, 1976]. This will be published separately.

3. *Two-Step Methods*

The algebra in the previous section suggests a two-step method to fit the model (11). In the first step we find an approximate solution of (13), and in the second step we transform this first solution in such a way that it becomes an approximate solution of (12). The first of these two-step methods was proposed by Schönemann (1972, Section 5). He first finds X up to a rotation by solving $B_* = XX'$ approximately, and he then finds the rotation matrix T by diagonalizing one of the C_k. The disadvantages of this procedure are clear, and already pointed out by Schönemann himself. The subject chosen in step two may be a particularly bad fitting subject, and the solution based on this subject's transformation matrix may be very misleading. We could, of course, compute a separate transformation matrix for all subjects, and compare or average the different solutions in some way. A major problem with this last approach is that each analysis only determines the solution up to a permutation of the dimensions, and consequently a comparison of different solutions is only possible by using a kind of matching algorithm that finds the optimal permutation matrices. Finding the optimal permutation matrices means solving a sequence of assignment problems, and is not at all trivial if r is, say, larger than two.

Carroll and Chang [Note 2] have generalized Schönemann's algorithms as follows. Suppose C_1 and C_2 are two linear combinations of the C_k, then we can find the transformation matrix T as that nonsingular matrix that diagonalizes both C_1 and C_2. If C_2 is the sum of the C_k, then $C_2 = I$, and T is the orthonormal matrix that diagonalizes C_1. If in addition $C_1 = C_k$ for some k, then we find Schönemann's proposal again. Carroll and Chang indicate that this choice of C_1 is not very satisfactory, and propose a different scheme. They divide the subjects into two groups by using a simple ad hoc clustering procedure, C_1 is the sum of the C_k in the first group, C_2 is the sum of the C_k in the second group. Because in this case $C_1 + C_2 = I$, it is clear that T can be found as the orthonormal matrix that diagonalizes either C_1 or C_2.

De Leeuw [Note 4] takes $C_2 = C_* = I$, and proposes a slightly more complicated procedure to find C_1. Because of the uniqueness theorem from the previous section it seems desirable to have the eigenvalues of C_1 as different as possible. Thus de Leeuw chooses C_1 as that linear composite for which the eigenvalues have maximum variance. This leads to finding the eigenvector corresponding with the dominant eigenvalue of a certain $m \times m$ matrix.

Although the proposals of Carroll and Chang and that of de Leeuw improve Schönemann's second step, there is still some degree of arbitrariness in the choice of at least one of the two composites. Moreover T always diagonalizes the two composites exactly, but we are really interested in how well T diagonalizes all individual C_k. This leads directly to the idea of minimizing a loss function that measures how well T diagonalizes all

C_k. Several different approaches have been discussed, independently and almost simultaneously, by MacCallum [1976], Cohen [Note 3], and by us. We translate the two alternative proposals into our notation.

MacCallum uses the loss function

$$\sum_{k=1}^{m} \text{tr}\,(TC_kT' - W_k^2)^2, \tag{15}$$

which must be minimized over all nonsingular T, with certain normalizing constraints on T, and over all diagonal W_k^2. This is clearly equivalent to minimizing the sum of squares of all off-diagonal elements of the matrices TC_kT'. MacCallum's algorithm uses a gradient method. Cohen proposes to minimize

$$\sum_{k=1}^{m} \text{tr}\,(C_k - SW_k^2S')^2, \tag{16}$$

over all nonsingular S and all diagonal W_k^2. The algorithm he uses for this minimization is simply CANDECOMP. After convergence we set $T = S^{-1}$.

We propose to use the same loss function as MacCallum, but to minimize this over all orthogonal T and over all diagonal W_k^2. Because in our case C_k is constructed in such a way that $C_* = I$, and we are assuming $W_*^2 = I$, the orthogonality restrictions follow from the fact that $TC_*T' = W_*^2$. Working with orthonormal matrices has a number of useful features. In the first place the number of free parameters in a nonsingular matrix is approximately twice the number of free parameters in an orthonormal matrix, in the second place we know that the algorithm we propose has a quadratic convergence rate if there is only a single C_k, in the third place working with orthonormal matrices usually guarantees good behavior with respect to rounding error, and finally orthonormality of T can be used to simplify the loss function. In fact it follows from orthonormality that we minimize (15) if we maximize the sum of squares of the diagonal elements of the TC_kT'. Considering diagonal elements only is much more efficient than considering off-diagonal elements only. We also think that gradient methods in problems like this are usually unnecessarily slow [Harshman, Note 5], and that using CANDECOMP in the algorithm somehow defeats the purpose of finding a substitute for INDSCAL. Nevertheless a more detailed comparison of the three approaches is needed.

The algorithm we propose can be summarized as follows. We have two explicit loss functions that must be minimized. The first one is

$$\text{tr}\,(B_* - YY')^2, \tag{17}$$

which must be minimized over Y. It tests the fit of the model corresponding with equations (13). The solution $\hat{Y} = K_r\Lambda_r^{1/2}$ is found by truncating the eigen-decomposition $B_* = K\Lambda K'$. The corresponding solution for the parameters of (13) is

$$\hat{X} = \hat{Y}, \tag{18a}$$

$$\hat{U}_k = C_k. \tag{18b}$$

Observe that $\hat{U}_* = I$. The second loss function we use is (15), which must be minimized over all orthonormal T and over all diagonal W_k^2. It tests the fit of the model (11) *within* the more general model corresponding with (13). We can minimize (15) by maximizing the sum of squares of the diagonal elements of the TC_kT'. An explicit algorithm is given in the next section. If $\hat{T}$ is the solution of this second minimization problem, then we can find our estimates of the parameters of (11) as follows

$$\hat{X} = \hat{Y}\hat{T}', \tag{19a}$$

$$\hat{W}_k^2 = \text{diag}\,\hat{T}C_k\hat{T}'. \tag{19b}$$

Again these estimates have the normalization property $\hat{W}_*^2 = I$.

4. *Simultaneous Diagonalization of Symmetric Matrices*

One classic method for diagonalizing a symmetric matrix is the method of Jacobi [1846]. The Jacobi procedure finds an orthonormal matrix T which transforms a real symmetric matrix C into a diagonal matrix D. The diagonal matrix is found by zeroing, in turn, selected off-diagonal elements of C by "elementary" orthogonal rotations. The elementary rotation matrix $E(i, j; \theta)$ is equal to the identity, except for elements (i, i), (j, j), (i, j) and (j, i), which are given by

$$\begin{aligned} e_{jj} &= e_{ii} = \cos\theta \\ e_{ij} &= -e_{ji} = \sin\theta \end{aligned} \tag{20}$$

where θ is an angle of rotation to be determined.

An updated matrix C^+ (note, the superscript $^+$ in this context denotes an updated matrix not the Moore-Penrose inverse) is the result of the transformation $E'CE$. In the Jacobi procedure for a single matrix C, θ is chosen in such a way that the selected off-diagonal element, or pivot element, $c_{ij}^+ = 0$. Each off-diagonal element is used as a pivot element, θ is determined so as to zero the pivot element and the matrix is again updated. Since each of these elementary rotations affects more than just the pivot element a single sweep through all of the off-diagonal elements will, generally, not result in a diagonal matrix. Several iterations are needed to obtain the desired accuracy.

We generalize the Jacobi procedure to the case where there are m real symmetric matrices C_k. In particular, we want to minimize the sums of squares of the off-diagonal elements of the matrices TC_kT'. An updated matrix C_k^+ is computed for every k by applying the elementary rotation matrix $E(p, q; \theta)$ to each C_k. The angle θ is selected such that the off-diagonal elements are minimized by minimizing

$$\sum_{k=1}^{m} (c_{pqk}^+)^2.$$

We carry out the minimization to find an optimal θ because if $m \geq 2$ it will not be possible, in general, to make

$$\sum_{k=1}^{m} (c_{pqk}^+)^2 = 0.$$

The first step in our procedure is to minimize

$$\sum_{k=1}^{m} (c_{pqk}^+)^2$$

to determine the optimal θ. For a particular pivot location p, q the result of the transformation $E'C_kE$ is

$$\begin{aligned} c_{pqk}^+ &= 2u_{pqk} \sin\theta \cos\theta + c_{pqk}(\cos^2\theta - \sin^2\theta) \\ &= u_{pqk} \sin 2\theta + c_{pqk} \cos 2\theta \end{aligned} \tag{21}$$

with

$$u_{pqk} \triangleq \tfrac{1}{2}(c_{ppk} - c_{qqk}). \tag{22}$$

Consequently

$$\begin{aligned} S^+ \triangleq \sum_{k=1}^{m} (c_{pqk}^+)^2 &= f \sin^2 2\theta + g \cos^2 2\theta + 2h \sin 2\theta \cos 2\theta \\ &= g + (f - g) \sin^2 2\theta + 2h \sin 2\theta \cos 2\theta \\ &= g + (f - g) \sin^2 2\theta + h \sin 4\theta. \end{aligned} \tag{23}$$

With

$$f \triangleq \sum_{k=1}^{m} u_{pqk}^2 , \tag{24}$$

$$g \triangleq \sum_{k=1}^{m} c_{pqk}^2 , \tag{25}$$

$$h \triangleq \sum_{k=1}^{m} u_{pqk} c_{pqk} . \tag{26}$$

Differentiating we find the necessary condition for an extreme value

$$\tfrac{1}{2}(f - g) \sin 4\theta + h \cos 4\theta = 0 . \tag{27}$$

In the cases where both $h \neq 0$ and $f - g \neq 0$, the solution is given by

$$\tan 4\theta = -\frac{2h}{f - g} . \tag{28}$$

Letting

$$\alpha \triangleq \arctan\left(-\frac{2h}{f - g}\right) \tag{29}$$

we find

$$\theta = \tfrac{1}{4}\alpha + \tfrac{1}{4}k\pi \tag{30}$$

with k any integer, and arctan having its principal value between $\pm\frac{1}{2}\pi$. The second derivative of S^+ has the same sign as

$$\tfrac{1}{2}(f - g) \cos 4\theta - h \sin 4\theta . \tag{31}$$

This means that we want to choose the integer k in such a way that $\cos 4\theta$ has the same sign as $f - g$ (or, equivalently, that $\sin 4\theta$ and h have opposite signs).

We must also consider the special cases where $h = 0$ or $f - g = 0$. If $h = 0$ and $f - g = 0$, then we do not perform any rotation. If $h = 0$ and $f - g \neq 0$, we find the solution to $\sin 4\theta = 0$, that is, $\theta = \frac{1}{4}k\pi$, with k equal to any integer. If $f > g$ we perform no rotation; if $f < g$ we choose $\theta = \pm\frac{1}{4}\pi$, giving $S^+ = f$. If $h \neq 0$ and $f - g = 0$ we find the solution to $\cos 4\theta = 0$, that is $\theta = \frac{1}{8}\pi + \frac{1}{4}k\pi$, with k equal to any integer. If $h > 0$ we choose $\theta = -\frac{1}{8}\pi$ making $S^+ = g - h$; if $h < 0$ we choose $\theta = +\frac{1}{8}\pi$ making $S^+ = g + h$.

The determination of θ, depending on the signs of h and $f - g$, is summarized in the following table:

h	$f - g$	θ
+	+	$\frac{1}{4}\alpha$
+	−	$\frac{1}{4}\alpha - \frac{1}{4}\pi$
−	+	$\frac{1}{4}\alpha$
−	−	$\frac{1}{4}\alpha + \frac{1}{4}\pi$
0	0	no rotation
0	+	no rotation
0	−	$\pm\frac{1}{4}\pi$
+	0	$-\frac{1}{8}\pi$
−	0	$\frac{1}{8}\pi$

The next step in our procedure is to minimize the off-diagonal elements of the C_k matrices and to update the C_k^+ matrices. For a particular pivot location p, q, the elements

of C_k^+, where $C_k^+ = E'C_kE$, are as follows [cf., Greenstadt, 1960]:

$$
\begin{aligned}
c_{pjk}^+ &= c_{pjk}\cos\theta - c_{qjk}\sin\theta,\\
c_{qjk}^+ &= c_{pjk}\sin\theta + c_{qjk}\cos\theta,\\
c_{ipk}^+ &= c_{ipk}\cos\theta - c_{iqk}\sin\theta, \qquad i,j \neq p,q\\
c_{iqk}^+ &= c_{ipk}\sin\theta + c_{iqk}\cos\theta,\\
c_{ijk}^+ &= c_{ijk},
\end{aligned} \tag{32}
$$

and

$$c_{ppk}^+ = c_{ppk}\cos^2\theta + c_{qqk}\sin^2\theta - 2c_{pqk}\sin\theta\cos\theta\,, \tag{33}$$

$$c_{qqk}^+ = c_{ppk}\sin^2\theta + c_{qqk}\cos^2\theta + 2c_{pqk}\sin\theta\cos\theta \tag{34}$$

$$c_{pqk}^+ = (c_{ppk} - c_{qqk})\sin\theta\cos\theta + c_{pqk}(\cos^2\theta - \sin^2\theta)\,. \tag{35}$$

We repeat the above two steps for each off-diagonal location selected as a pivot location, in turn, and perform an elementary rotation by applying (32)–(35). We continue to build up these transformations until the transformed C_k matrices are as nearly diagonal as possible. We consider the iterative process complete when the minimum $\cos\theta$ in one complete sweep over all pivot locations is $>1 - 10^{-10}$ and <1.

For each elementary rotation in the iterative process, we also update the orthonormal transformation matrix T. We have

$$T^{(1)} = E'_{(1)} \tag{36}$$

$$T^+ = E'_+T\,. \tag{37}$$

The elements of the updated matrix T^+, then, are:

$$t_{ij}^+ = t_{ij} \qquad (i \neq p,q) \tag{38}$$

$$t_{pj}^+ = t_{pj}\cos\theta - t_{qj}\sin\theta, \tag{39}$$

$$t_{qj}^+ = t_{pj}\sin\theta + t_{qj}\cos\theta\,, \tag{40}$$

where θ is the angle of rotation for the current pivot location.

Let $C_k^{(s)}$ be the updated C_k after s sweeps, and $T^{(s)}$ the cumulated transformation matrix. The diagonals of the transformed $C_k^{(s)}$ converge to our estimates of W_k^2 and $T^{(s)}$ converges to our estimate of T.

5. *Applications to Real Data*

Implementation of the SUMSCAL procedure described above, consists of [1] factoring B_*, an $n \times n$ matrix of scalar products, averaged over all k data sources, [2] diagonalizing all C_k matrices and computing the transformation matrix T according to the procedure described in Section 4, and finally, [3] computing $X = K_r\Lambda_r^{1/2}T'$, where Λ_r contains the r largest eigenvalues and K_r the r largest eigenvectors of matrix B_*. The elements of X are the SUMSCAL estimates of the coordinates of the *n objects* in the r-dimensional stimulus space and each diagonalized C_k matrix contains the *weights* or saliences for each *data source k*.

This SUMSCAL algorithm was applied to two sets of data whose INDSCAL solutions have been reported previously. Comparisons are made between the stimulus and weights spaces from SUMSCAL and INDSCAL and the CPU times required by the two procedures to analyze the data.

The first set of data analyzed by SUMSCAL is from Helm's [1964] study on judged psychological distances among ten colors. A reanalysis of these data using INDSCAL has

been reported by Wish and Carroll [1974]. Sixteen matrices of perceived dissimilarities among 10 colors served as input to SUMSCAL. The data were analyzed in two dimensions only since Wish and Carroll [1974] showed that a two-dimensional solution was adequate. The SUMSCAL weights and stimulus spaces were virtually identical to the INDSCAL configuration shown in Wish and Carroll [1974]. The product moment correlations between the corresponding unrotated dimensions of SUMSCAL and INDSCAL were:

	Dimension 1	Dimension 2
stimulus space	.9999	.9999
weight space	.9999	.9999

.

The following table shows the CPU times required for a two-dimensional solution of the Helm data for both SUMSCAL and INDSCAL:

	Iterations to converge	CPU time in sec
INDSCAL*	25	10.08
SUMSCAL		.738

.

The second set of data which we analyzed with SUMSCAL is from a study by Bricker and Pruzansky of judged dissimilarities of auditory tones generated by varying three physical properties. The INDSCAL results from this study were reported in Carroll and Chang's original INDSCAL paper [Carroll & Chang, 1970]. This set of data was selected because it took many INDSCAL iterations to converge and it is relatively noisy data.

Twenty matrices of perceived dissimilarities among 24 tones served as input to SUMSCAL. A three-dimensional SUMSCAL solution was obtained so that it could be compared with the three-dimensional INDSCAL solution. The SUMSCAL weights and stimulus spaces were, for all practical purposes, the same as the INDSCAL solution. The product moment correlations between corresponding dimensions of SUMSCAL and INDSCAL were:

	Dimension 1	Dimension 2	Dimension 3
stimulus space	.9966	.9989	.9986
weights space	.9978	.9991	.9989

.

The following table shows the CPU times required for a three-dimensional solution of the Bricker and Pruzansky data for both SUMSCAL and INDSCAL:

	Iterations to convergence	CPU time in sec
INDSCAL	130	341.64
SUMSCAL		4.48

.

* All of the INDSCAL results reported in this paper including CPU times, were actually obtained from a new program SINDSCAL (Pruzansky, 1975) which was written to handle only symmetric data. It uses the INDSCAL algorithm and produces identical results but runs approximately 30% faster than the INDSCAL program that is currently being distributed.

6. *Discussion*

It is now possible to compare the main advantages of our approach and the other approaches. INDSCAL captures all the properties of its model in a single loss function. In analysis of variance, stepwise regression, and contingency table analysis, it has proved to be useful to partition the loss into components. Of course in our situation, the partitioning does not have any nice additivity, orthogonality, or independence properties. Still the magnitude of the individual loss components can indicate what causes a bad fit of the overall model.

The fact that we apply more transformations before defining the loss is a possible disadvantage. Although INDSCAL also supplies at least one transformation, it defines the loss at a much earlier stage. As a consequence, treatment of missing data and of nonmetric information is much more difficult in our framework, because all the incomplete information in these cases also has to be transformed in the same way as the numerical information is transformed. While our transformations do simplify matters in the complete case, they complicate matters in these other cases. It will also be clear, however, that our method does improve Schönemann's in most respects. Although the computations are slightly more complicated, we introduce a satisfactory loss function at the precise point where the Schönemann algorithm becomes unsatisfactory (for fallible data).

The ultimate success of the method will depend on its numerical behavior (speed, local minima) and on its successfully displaying structure in data in the same way as INDSCAL. Although this new method is closer to Schönemann's method than to the Carroll-Chang-Harshman method in many respects, the main inspiration for doing this work is the success the INDSCAL program has had in many practical applications. Extensive numerical investigations are now being carried out to find out if our new method can compete with INDSCAL in this respect. As the examples in Section 5 show we seem to find solutions that are identical for all practical purposes, and we are certainly getting them much cheaper.

Although Schönemann's algebraic solution has been a very important contribution to the methodology in this paper, we do not agree with his ideas about this class of methods in one very important respect. It is not true that the rotational indeterminacy is resolved in these models in the same way as in principal component analysis, by imposing some "mathematically convenient constraint". The fact that rotations are not permitted here is a consequence of the model itself, and it provides, as Carroll-Chang and Harshman have observed, a very powerful way to facilitate interpretation. Harshman, in particular, has emphasized how this also provides an alternative way to eliminate rotational indeterminacy in factor analysis. The other easily available method of doing this is the "confirmatory" factor analysis method, which postulates certain patterns of zeroes [e.g., Jöreskog, 1969]. Postulating just enough zeroes to eliminate indeterminacy may be considered as merely a mathematical convenience, but postulating more zeroes than needed clearly introduces extra constraints into the model. In our subjective metrics models the situation is the same. Requiring that W_k^2 be merely positive semidefinite leaves room for indeterminacy. Postulating patterns of zeroes in W_k^2 increases the number of constraints, and postulating diagonality of all W_k^2 certainly introduces extra constraints, which define a different model and not simply the same model with different identification conditions. Of course the methods outlined in Section 4 can, with slight modifications, also be used to estimate parameters in models with more general patterns of zeros in W_k^2.

REFERENCE NOTES

1. Bloxom, B. Individual differences in multidimensional scaling (*Res. Bull.* 68-45). Princeton, New Jersey: Educational Testing Service, 1968.

2. Carroll, J. D., & Chang, J. J. *IDIOSCAL: A generalization of INDSCAL allowing IDIOsyncratic reference systems as well as an analytic approximation to INDSCAL.* Paper presented at the Spring Meeting of the Psychometric Society, March 30–31, 1972, Princeton, New Jersey.
3. 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.
4. de Leeuw, J. *An initial estimate for INDSCAL.* Unpublished paper, Bell Telephone Laboratories, Murray Hill, June, 1974.
5. Harshman, R. A. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-model factor analysis. *UCLA Working Papers in Phonetics,* 1970, *16,* 1–84.
6. Harshman, R. A. PARAFAC2; mathematical and technical notes. *UCLA Working Papers in Phonetics,* 1972, *22,* 30–44.
7. Heiser, W. J. *Individual differences scaling* (Report MT 001-75). Leiden, Holland; Psychology Department, University of Leiden, 1975.
8. Pruzansky, S. *SINDSCAL: A Computer Program for Individual Differences in Multidimensional Scaling,* Unpublished manuscript, Bell Laboratories, September 1975.

REFERENCES

Bhimasankaram, P. Simultaneous reduction of several Hermitian forms. *Sankhya,* A, 1971, *33,* 417–422.

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.

de Leeuw, J., & Heiser, W. J. Convergence of correction matrix algorithms for multidimensional scaling. In J. Lingoes (Ed.), *Geometric representations of relational data.* Ann Arbor: Mathesis Press, 1977.

Greenstadt, J. The determination of the characteristic roots of a matrix by the Jacobi method. In A. Ralston & H. Wilf (Eds.), *Mathematical Methods for Digital Computers,* (Vol. I) New York: Wiley, 1960, pp. 84–91.

Helm, C. E., A multidimensional ratio scaling analysis of perceived color relations. *Journal of the Optical Society of America,* 1964, *54,* 256–262.

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

Jacobi, C. G. J., Über ein leichtes Verfahren, die in der Theorie der Säkularstörungen vorkommenden Gleichungen numerisch aufzulösen. *J. Reine Agnew. Math,* 1846, *30,* 51–95.

Jöreskog, K., A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika,* 1969, *34,* 183–202.

Kruskal, J. B. More factors than subjects, tests and treatments: An indeterminacy theorem for canonical decomposition and individual differences scaling. *Psychometrika,* 1976, *41,* 281–293.

Kruskal, J. B., & Carroll, J. D. Geometric Models and Badness-of-Fit Functions. In P. R. Krishnaiah (Ed.), *Multivariate Analysis,* (Vol. II) New York: Academic Press, 1969, pp. 639–671.

MacCallum, R. C. Transformation of a three-mode multidimensional scaling solution to INDSCAL form. *Psychometrika,* 1976, *41,* 385–400.

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

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

Torgerson, W. S. *Theory and methods of scaling,* New York: Wiley, 1958.

Wish, M., & Carroll, J. D. Applications of Individual Differences Scaling to Studies of Human Perception and Judgement. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of Perception,* (Vol. II) *Psychophysical Judgment and Measurement,* New York: Academic Press, Inc., 1974, pp 449–491.

Manuscript received 3/8/76
Final version received 4/26/78