

A NEW SOLUTION TO THE ADDITIVE CONSTANT PROBLEM IN METRIC MULTIDIMENSIONAL SCALING*

LEE G. COOPER

UNIVERSITY OF CALIFORNIA, LOS ANGELES

A new solution to the additive constant problem in metric multidimensional scaling is developed. This solution determines, for a given dimensionality, the additive constant and the resulting stimulus projections on the dimensions of a Euclidean space which minimize the sum of squares of discrepancies between the formal model for metric multidimensional scaling and the original data. A modification of Fletcher-Powell style functional iteration is used to compute solutions. A scale free index of the goodness of fit is developed to aid in selecting solutions of adequate dimensionality from multiple candidates.

The additive constant problem was originally formulated as the problem of finding a constant, c , which converted the observed comparative interpoint distance between a pair of stimuli, h_{jk} , into an absolute interpoint distance, d_{jk} , in such a way as to minimize the dimensionality of the resulting Euclidean space in which the stimuli were to be represented.

$$(1) \quad h_{jk} = d_{jk} + c \quad j, k = 1, 2, \dots, n, j \neq k,$$

$$(2) \quad d_{jk} = \left[\sum_m^t (a_{jm} - a_{km})^2 \right]^{1/2},$$

where a_{jm} is the projection of the j th stimulus on the m th dimension of a t -dimensional Euclidean space. No solution has ever been found for this formulation of the problem.

The first systematic reformulation of this problem was by Messick and Abelson [1956]. They worked on the relation of the additive constant to the elements of a matrix of scalar products, B^* , among the stimulus projections. They consider the roots and vectors of B^* resulting from an Eckart-Young [1936] resolution. They note that in a "true" solution from this approach, there are a minimum number of large roots and the remaining roots are zero. With fallible data, this ideal is not achievable. Neither is it reasonable to assume that all errors would function so as to make the roots of B^* , which correspond to zero roots in the "true" solution, positive. Messick

* This research is based in part on the author's Ph.D. dissertation at the University of Illinois at Urbana-Champaign. Computer time was provided by the Campus Computing Network of the University of California, Los Angeles.

and Abelson, therefore, attempted to find the additive constant, for a given dimensionality, which makes the residual roots of B^* average zero. Their approach was well thought out, consistent and logical. There has, however, been some dissatisfaction with the results their procedure produces. The problem seems to be that quite often relatively large negative roots appear. If the absolute value of a negative root is larger than the value of a root corresponding to what is claimed to be a real perceptual dimension, then the investigator is in a very difficult position. He could claim that all negative roots, regardless of size, are treated as part of the error and not subject to interpretation. However, this would be an attempt to ignore the issue rather than deal with it. Such a claim would also be counter to the rationale underlying the Messick and Abelson solution, even though no violation of the mathematical formulation would be entailed. On the other hand, the investigator may accept the negative root as corresponding to an imaginary perceptual dimension. In such a case the investigator must somehow deal with a violation of the model proposed by Messick and Abelson and also attempt to interpret an imaginary perceptual dimension.

There is another possible formulation of the problem. It rests on the realization that metric multidimensional scaling provides a model for psychological reality. As such it could be in error. We could attempt to find a constant such that the lack of fit of the model to reality is minimized. This can be done by introducing a discrepancy term δ_{ik} ,

$$(3) \quad h_{ik} = d_{ik} + c + \delta_{ik},$$

and establishing an error function G ,

$$(4) \quad G = \frac{1}{2} \sum_k^n \sum_{i \neq k}^n \delta_{ik}^2.$$

The problem then reduces to determining, for a given dimensionality, the additive constant and the stimulus projections which will minimize G . If expressions can be found for the partial derivative of G with respect to c and each stimulus projection, then a solution can be developed using functional iteration.

The partial derivative of G with respect to the additive constant is relatively easy to obtain. Expressing G in a more workable form, we have

$$(5) \quad G = \frac{1}{2} \sum_k^n \sum_{i \neq k}^n \left((h_{ik} - c)^2 + \sum_m^i (a_{im} - a_{km})^2 - 2(h_{ik} - c) \left[\sum_m^i (a_{im} - a_{km})^2 \right]^{1/2} \right).$$

Then,

$$(6) \quad \frac{\partial G}{\partial c} = - \sum_k^n \sum_{i \neq k}^n h_{ik} + n(n-1)c + \sum_k^n \sum_{i \neq k}^n \left[\sum_m^i (a_{im} - a_{km})^2 \right]^{1/2}.$$

If the comparative interpoint distances are originally given a mean value of zero, setting the partial to zero and solving for the value of the additive constant which minimizes G for a given set of stimulus projections, results in the following expression for c :

$$(7) \quad c = \frac{-1}{n(n-1)} \sum_k^n \sum_{i \neq k}^n \left[\sum_m^i (a_{im} - a_{km})^2 \right]^{1/2}.$$

The partial derivative of G with respect to a particular stimulus projection, $a_{j \cdot m \cdot}$, will be obtained term by term.

The first term is a constant with respect to this derivative, so that

$$(8) \quad \frac{\partial \left[\frac{1}{2} \sum_k^n \sum_{i \neq k}^n (h_{ik} - c)^2 \right]}{\partial a_{j \cdot m \cdot}} = 0.$$

The derivative of the second term is

$$(9) \quad \begin{aligned} & \frac{\partial \left[\frac{1}{2} \sum_k^n \sum_{i \neq k}^n \left(\sum_m^i (a_{im} - a_{km})^2 \right) \right]}{\partial a_{j \cdot m \cdot}} \\ &= \frac{1}{2} \left(\sum_{i \neq j \cdot}^n 2(a_{im \cdot} - a_{j \cdot m \cdot})(-1) + \sum_{k \neq j \cdot}^n 2(a_{j \cdot m \cdot} - a_{km \cdot}) \right) \\ &= 2(n-1)a_{j \cdot m \cdot} - 2 \sum_{i \neq j \cdot}^n a_{im \cdot}. \end{aligned}$$

Since the distances are invariant with respect to the location of the origin of the space, we are free to aid the constraint that

$$(10) \quad \sum_{i=1}^n a_{im} = 0.$$

With this restriction we have

$$(11) \quad \sum_{i \neq j \cdot}^n a_{im \cdot} = -a_{j \cdot m \cdot}.$$

Consequently

$$(12) \quad \frac{\partial \left[\frac{1}{2} \sum_k^n \sum_{i \neq k}^n \left(\sum_m^i (a_{im} - a_{km})^2 \right) \right]}{\partial a_{j \cdot m \cdot}} = 2na_{j \cdot m \cdot}.$$

The partial of the third term is

$$\begin{aligned}
 \frac{\partial}{\partial a_{j^*m^*}} & \left[\sum_k^n \sum_{j \neq k}^n (h_{jk} - c) \left(\sum_m^i (a_{jm} - a_{km})^2 \right)^{1/2} \right] = \\
 (13) \quad & \left[\sum_{j \neq j^*}^n (h_{jj^*} - c) \left(\frac{1}{2} \right) \left(\sum_m^i (a_{jm} - a_{j^*m})^2 \right)^{-1/2} (-2)(a_{jm^*} - a_{j^*m^*}) \right. \\
 & \left. + \sum_{k \neq j^*}^n (h_{j^*k} - c) \left(\frac{1}{2} \right) \left(\sum_m^i (a_{j^*m} - a_{km})^2 \right)^{-1/2} (2)(a_{j^*m^*} - a_{km^*}) \right] \\
 & = 2 \sum_{k \neq j^*}^n (h_{j^*k} - c)(a_{j^*m^*} - a_{km^*}) \left(\sum_m^i (a_{j^*m} - a_{km})^2 \right)^{-1/2}.
 \end{aligned}$$

This leads to the final result

$$\begin{aligned}
 (14) \quad \frac{\partial G}{\partial a_{j^*m^*}} & = 2 \left[na_{j^*m^*} - \sum_{k \neq j^*}^n (h_{j^*k} - c) \right. \\
 & \left. \cdot (a_{j^*m^*} - a_{km^*}) \left(\sum_m^i (a_{j^*m} - a_{km})^2 \right)^{-1/2} \right].
 \end{aligned}$$

A particular type of functional iteration procedure, known as the Fletcher-Powell routine [Fletcher and Powell, 1963], is well suited to minimize G . Gruvaeus and Jöreskog [1970] have found their version of the Fletcher-Powell method to be most efficient when:

- i. the number of arguments is large
- ii. the function is complicated and time-consuming to compute
- iii. the additional time required to compute derivative values, after the function value has been computed, is relatively small
- iv. second-order derivatives are not available or are time-consuming to compute.

These four characteristics well describe the circumstances surrounding the function of concern here.

This iterative scheme approximates the error function with a generalized quadratic equation. It uses the first derivatives and develops estimates of the inverse of the matrix of second-order derivatives to calculate optimal increments at each iteration. Since no attempt was made to derive more appropriate estimates of the inverse of the matrix of second-order derivatives, the identity matrix is used as an initial approximation. This makes the early iterations like those in a standard steepest descent procedure. These estimates are improved internally to provide for very rapid convergence. The solution using this procedure will initially be a centroid solution, but the origin may be translated to a psychologically more meaningful position, if one can be found, without altering the value of the error function.

A program, COSCAL, which incorporates the Fletcher-Powell procedure, was developed for computing the solution for the additive constant and the coordinates of the stimulus-objects on the dimensions of a Euclidean space. It was written in FORTRAN IV for the IBM OS 360/91 computer and the FORTRAN H compiler. This program provides the user with the option of using his own starting points or using starting points defined by the program. This initial configuration is determined by estimating an additive constant such that the smallest absolute distance is slightly larger than zero. From these absolute distances a matrix of scalar products is developed according to the formula given by Torgerson [1958, p. 258]. This matrix is resolved into its principal axes and the coordinates of the objects on the first t principal axes are used as the initial configuration for a t -dimensional solution.

Initial testing of this program was centered around two examples. The first example involved errorless data to test the accuracy of the solution. It was based on a triangular pattern of five stimuli in two dimensions. The solution produced by the program is plotted in Figure 1. The routine converged after 11 iterations within an orthogonal rotation of the exact solution. The value of the error function, G , was zero to double precision accuracy.

The second example used real data to investigate the identifiability of adequate solutions from multiple candidates. To aid in this an index of the goodness of fit was developed. It is based on the ratio of the sum of squares of discrepancies to the sum of squares of the original comparative interpoint distances adjusted to have a mean of zero.

$$(15) \quad FIT = 1 - \frac{\sum_{i < k}^n \delta_{ik}^2}{\sum_{i < k}^n (h_{ik} - h_{..})^2}.$$

This index eliminates the scale dependency of the function, G . When the solution is perfect $G = 0$ and $FIT = 1$. At convergence the index of the goodness of fit can never be less than zero.

Since a solution in zero dimensions would have all stimuli at a single point; the additive constant which would minimize the value of G would be zero according to (7); all absolute distances would be zero; and thus, the sum of squares of discrepancies would be the sum of squares of the original comparative interpoint distances, when they are set to have a mean of zero. Since the denominator in (15) is really G_0 , the error value for a zero-dimensional solution, and the numerator is G_t , the error value for a t -dimensional solution, the index may be rewritten as

$$(16) \quad FIT = 1 - \frac{G_t}{G_0}.$$

Therefore, this index may be interpreted as showing the relative improvement

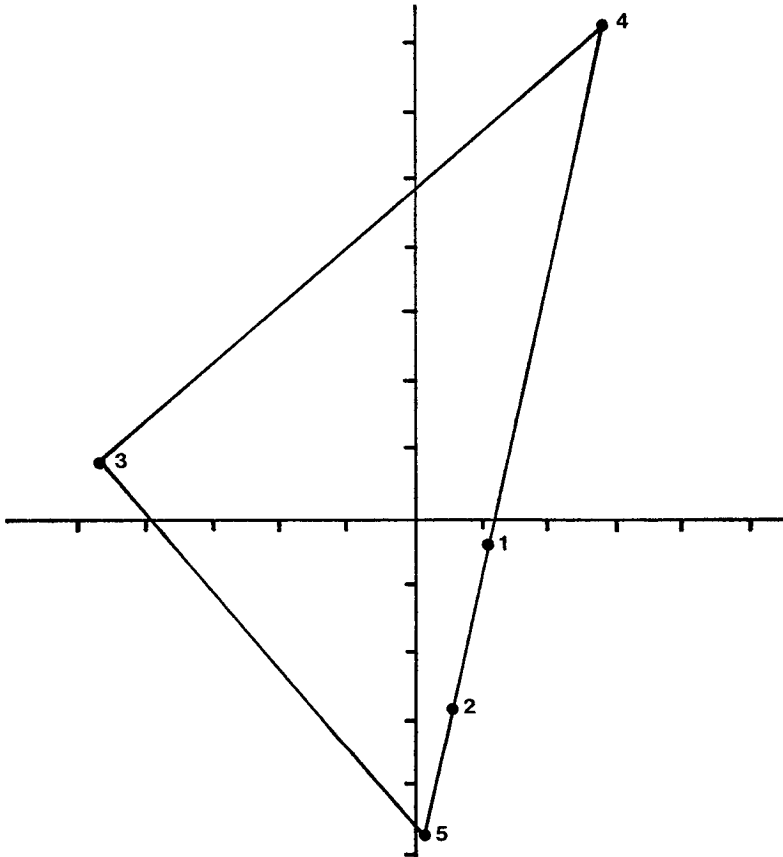


FIGURE 1. SOLUTION TO THE FIRST EXAMPLE. A TRIANGULAR PATTERN OF FIVE STIMULI IN TWO DIMENSIONS.

of a t -dimensional solution over a zero-dimensional solution. It can be used in a "root staring" type of procedure to aid in the selection of the proper number of dimensions in a solution.

The comparative interpoint distances analyzed in the second example are taken from Torgerson, *Theory and methods of scaling*, p. 286. They involve judgments collected by the complete method of triads on nine Munsell colors. All nine were of the same red hue. They vary in brightness and saturation only. The original pattern of brightness, in value units, and saturation, in chroma units, appears as the dots in Figure 2. Solutions were sought for one dimension to eight dimensions. The index of the goodness of fit, the number of iterations required for convergence, and the execution time appear in Table 1.

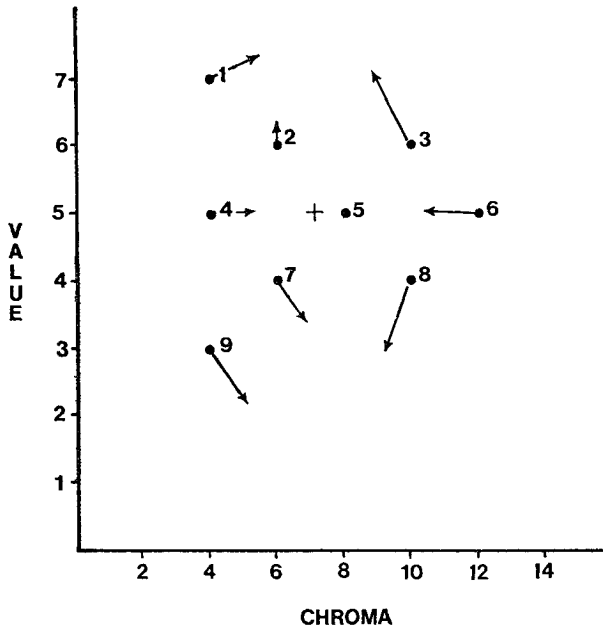


FIGURE 2. BRIGHTNESS IN VALUE UNITS AND SATURATION, IN CHROMA UNITS, FOR THE NINE MUNSELL COLORS. FROM TORGERSON (1958).

A plot of the fit index values versus dimensions in solution appears in Figure 3. On the basis of scanning this plot a two-dimensional solution was determined to be adequate. The contribution of the third dimension is extremely minor. The two-dimensional solution is listed in Table 2.

Table 1
Summary of solution characteristics
for nine Munsell colors -- Example 2.

Dimensions	FIT	Iterations	Time in seconds
1	0.44	8	0.36
2	0.98	15	0.51
3	0.99	27	0.93
4	1.00-	38	1.67
5	1.00-	38	2.46
6	1.00-	41	3.05
7	1.00-	58	5.73
8	1.00	62	7.26

Table 2
Solution for nine Munsell colors--Example 2.

Stimulus Number	Dimension	
	I	II
1	2.11	2.19
2	1.08	1.44
3	2.51	-0.80
4	-0.59	1.58
5	0.19	-0.84
6	0.94	-2.78
7	-1.65	0.17
8	-1.38	-2.02
9	-3.21	1.06

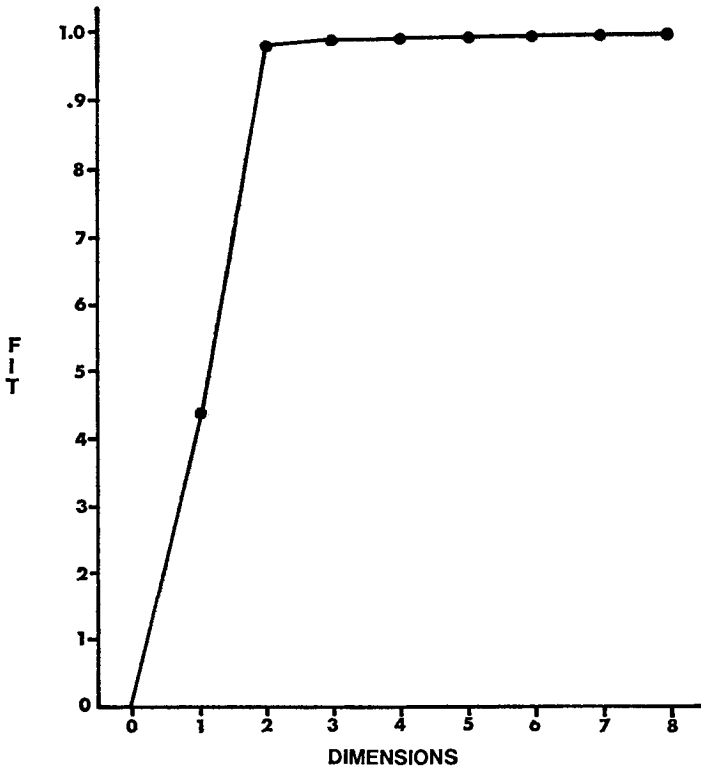


FIGURE 3. PLOT OF THE GOODNESS OF FIT VERSUS DIMENSIONS FOR NINE MUNSELL COLORS--EXAMPLE 2.

For the purpose of comparing the solution obtained by this procedure with the solution obtained using the Messick and Abelson approach, both solutions were rotated to maximal congruences with the original theoretical configuration. The rotational system was the Schönemann and Carroll [1970] extension of orthogonal Procrustes [Schönemann, 1966; and Cliff, 1966] which allows for translation of origin, central dilation, and orthogonal rotation to a least squares criterion. The average squared error for the Messick and Abelson solution given by Torgerson [1958, p. 289] was .99 and for this new approach was .95. This indicates that the new solution is at least somewhat superior to the Messick and Abelson solution in its ability to retrieve the theoretical configuration from the real data. The transformation matrices, the dilation factors, the translation vectors, and the rotated solutions are listed in Table 3 for both solutions.

Table 3
Rotation of Solutions to the Theoretical Configuration

Coscal Solution	Messick and Abelson Solution
Transformation Matrix	
0.95 0.31 0.31 -0.95	1.00- .07 - .07 1.00-
Dilation Factor	
1.04	1.11
Translation Vector	
5.00 7.11	5.00 7.11
Rotated Solution	
7.79 5.64 6.53 6.04 7.22 8.71 4.93 5.36 4.92 7.99 5.02 10.16 3.43 6.42 2.98 8.66 2.18 5.02	8.12 5.60 6.79 5.78 7.48 9.02 4.84 5.03 4.97 8.21 4.86 10.12 3.19 6.30 2.88 8.92 1.90 5.02

The rotated configuration using the new solution is indicated in Figure 2 as the tips of the arrows originating from the theoretical position. If we consider the centroid of the theoretical configuration as a "psychological origin" the real data points indicate that the subjects consistently underestimate the magnitude on the "chroma" dimension and consistently overestimate the magnitude on the "value" dimension. This is reflected numerically in a .99 correlation of the real data on the "chroma" dimension with the theoretical data on this dimension, and a .99 correlation of the real data on the "value" dimension with the theoretical data on this dimension.

Discussion

With the development of nonmetric multidimensional scaling, a question has arisen as to the practical and theoretical value of metric multidimensional scaling. On the theoretical side of the issue, data meeting metric assumptions should be scaled using a metric procedure rather than a nonmetric one. This assertion does not have the same force when applied to an exploratory methodology such as multidimensional scaling as it does with the confirmatory methods of statistics. A parametric statistic is more powerful than a corresponding nonparametric statistic [Siegel, 1956, p. 31], but no criteria parallel to "power" exists for multidimensional scaling. It can be easily demonstrated [Cooper, 1970a] that in theory as the number of stimuli becomes smaller, a metric multidimensional scaling criterion allows for far less distortion of the underlying stimulus configuration than does a nonmetric criterion such as "Stress" [Kruskal 1964a and b]. In practice this advantage may not be as dramatic. This is because the numerical algorithm widely used for nonmetric multidimensional scaling [Young and Torgerson, 1967] generates starting points for the iterative procedure in the same way as this metric procedure. Even though it is theoretically inappropriate to use such a procedure on nonmetric data, the practical ramification of this for metric data would most likely be some reduction in the advantage of metric multidimensional scaling.

The practical advantages of metric over nonmetric multidimensional scaling have not yet been fully investigated. However, one of the greatest advantages is that metric multidimensional scaling provides the investigator with the clearest and most traceable relations between the original data and the final solution.

This property becomes especially important when multidimensional scaling is used in combination with other multivariate techniques such as an individual differences model for multidimensional scaling [Tucker and Messick, 1963; Tucker, 1970]. These models are extremely powerful in their ability to represent systematic individual differences from the variability in the original responses. The proper data with which to enter these models are the scalar products between stimuli [c.f. Cooper, 1970b; Tucker, 1970].

Scalar products can, of course, be obtained from either metric or nonmetric multidimensional scaling. But metric multidimensional scaling has a special advantage here. The investigator can use the metric scaling criterion to solve for the appropriate or adequate number of dimensions to include in the scaling solution for each subject. The stimulus configurations can be saved for later between-individual comparisons, while the additive constant can be used with the original comparative interpoint distances to form scalar products. These scalar products contain variability which may be common over subjects, but which might well have been lost if the scalar products had been computed from the stimulus configuration. The loss of this variability would occur because of a reduction in the dimensionality of the stimulus configuration. Since the nonmetric method can only obtain scalar products from the stimulus configuration, the only way to preserve this variability is to obtain a configuration with zero Stress. This would usually entail obtaining an $(n - 1)$ -dimensional solution for n stimuli, which is very impractical. Further, even if the $(n - 1)$ -dimensional solution is obtained it is not very useful for later, between-individual comparison.

While further research is needed to compare directly metric and nonmetric multidimensional scaling solutions from metric data, there seem to be two practical advantages for metric multidimensional scaling. First, metric scaling maintains a simple linear relation between the original data and the final solution. Second, metric scaling is more readily amenable for usage with the individual differences model for multidimensional scaling.

REFERENCES

- Cliff, N. Orthogonal rotation to congruence. *Psychometrika*, 1966, 31, 33-42.
- Cooper, L. G. Metric and nonmetric multidimensional scaling—a confrontation. Paper presented to the Pacific Association for Public Opinion Research, January, 1970a.
- Cooper, L. G. Metric multidimensional scaling and the concept of preference. Paper No. 163. Western Management Science Institute, University of California, Los Angeles, 1970b.
- Eckart, C. and Young, G. The approximation of one matrix by another of lower rank. *Psychometrika*, 1936, 1, 211-218.
- Fletcher, R. and Powell, M. J. D. A rapidly converging descent method for minimization. *Computer Journal*, 1963, 2, 163-168.
- Gruvæus, G. T. and Jöreskog, K. G. A computer program for minimizing a function of several variables. *Educational Testing Service Research Bulletin*. RB-70-14, 1970.
- Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964a, 29, 1-27.
- Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 1964b, 29, 115-128.
- Messick, S. J. and Abelson, R. P. The additive constant problem in multidimensional scaling. *Psychometrika*, 1956, 21, 1-15.
- Ralston, A. *A first course in numerical analysis*. New York: McGraw-Hill Inc., 1965.
- Schönemann, P. H. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 1966, 21, 1-10.

- Schönemann, P. H. and Carroll, R. Fitting one matrix to another under choice of central dilation and rigid motion. *Psychometrika*, 1970, 35, 245-255.
- Siegel, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.
- Tucker, L. R. Relations between multidimensional scaling and three-mode factor analysis. Technical Report, University of Illinois, 1970.
- Tucker, L. R. and Messick, S. An individual difference model for multidimensional scaling. *Psychometrika*, 1963, 28, 333-367.
- Young, F. W. and Torgerson, W. S. TORSCA, A FORTRAN IV program for Shepard-Kruskal multidimensional scaling analysis. *Behavioral Science*, 1967, 12, 498.

Manuscript received 9/10/70

Revised manuscript received 8/6/71