TRIGONOMETRIC CONVERSION OF FALLIBLE DISTANCES INTO COORDINATES IN MDS

Purnell H. Benson, Rutgers Graduate School of Business Administration

In recent years, the problem of determining coordinates of points when interpoint distances are given has claimed attention of statisticians in the field of psychological measurement. The problem arises from a practical standpoint because human subjects are frequently able to report similarities between stimuli when they are unable to describe characteristics of the stimuli. From judgments of proximities between stimuli are calculated the coordinates of stimulus points. These coordinates, after suitable rotation and translation of axes, are in effect measurements of characteristics of the stimuli. The multidimensional distance scaling procedures of Torgerson (1952, 1958) and of Shepard (1962a, 1962b) and Kruskal (1964a, 1964b) have sought solutions to systems of quadratic equations with many sets of roots.

The problem of ascertaining the globally correct solution has not been resolved to the satisfaction of all concerned. Trying many different solutions to find the one with minimum stress upon the inputted distances is one answer. Another is to use the principal components procedure of Torgerson. Shepard (1974) expresses reservations about the adequacy of any existing procedures to furnish the globally correct solution except by repeated trial and error to find that set of coordinates which most closely fit the interpoint distances.

The present author has proposed a trigonometric solution for obtaining point coordinates from exact interpoint distances (1976b). The formulas for calculating coordinates of points from inputted interpoint distances are given in Table 1 for 4 points. The pairs of numbers in parentheses indicate interpoint distances. The triads of numbers designate angles. The hinge of the angle is an italicised number. For dihedral angles, the hinge involves more than one point.

Generalizing the formulas in Table 1 to larger numbers of points in more than 3 dimensions, letting $C_{M,N}$ be the $\underline{M}$th coordinate of the $\underline{N}$th point, $M<(N-1)$, we have

$$C_{M,N} = (1N)\ \sin(2\ \underline{1}\ N)\ \sin(3\ \underline{12}\ N)\ \sin(4\ \underline{123}\ N)\ldots\cos(M\ \underline{123..M-1}\ N).$$

If M = N-1 for the last coordinate of the $\underline{N}$th point in the simplex, the terminal cosine term in the product is replaced by the sine of the same dihedral angle.

Also we have, with $L<M<N$

$$\cos(M\ \underline{123..L}\ N) = [\cos(M\ \underline{123..L-1}\ N) - \cos(L\ \underline{123..L-1}\ M)\ \cos(L\ \underline{123..L-1}\ N)] / [\sin(L\ \underline{123..L-1}\ M)\ \sin(L\ \underline{123..L-1}\ N)]$$

These formulas have been tested for randomly selected coordinates of 20 points in 14 dimensions and have been found to be correct.

Application to Fallible Data

The formulas given can be directly applied to a network of error-free interpoint distances. The usual problem in scaling characteristics of stimuli is that the respondents give inaccurate proximity judgments which do not form an exact system. Depending upon which interpoint distances are used, different sets of coordinates result, if indeed any calculation at all is mathematically possible.

A plausible procedure is to obtain exact solutions to alternative sets of points and then to average the resulting coordinates transformed to the same axes of orientation. A machine program was prepared which takes simplexes of

Table 1

COORDINATES FOR 4 POINTS IN 3 DIMENSIONS

	Coordinate		
Point	1	2	3
1	0	0	0
2	(12)	0	0
3	$(13)\cos(2\underline{1}3)$	$(13)\sin(2\underline{1}3)$	0
4	$(14)\cos(2\underline{1}4)$	$(14)\sin(2\underline{1}4)\cos(3\underline{12}4)$	$(14)\sin(214)\sin(3\underline{12}4)$

points in all possible sets to calculate average coordinates from the different possible exact solutions. The program, TRIVCOR, is available.

First is selected the largest simplex which can be formed from the points in the number of dimensions for which a solution is pursued. Then each possible simplex in that number of dimensions has its coordinates calculated and referred to the largest simplex for the axes on which the averaging of coordinates is done.

Coordinates between -1.0 and 1.0 were randomly chosen for 8 points in 1, 2, 3, 4, and 5 dimensions. The distances between the points were calculated and then degraded by adding error quantities whose absolute averages are respectively .1, .2, .4 and .8, in separate computations. The 4 levels of error for 5 different levels of dimensions were pursued in 20 different computations to fit coordinates to simulated data. The coordinates obtained were then reconverted to interpoint distances whose errors are then compared with the original distances before their degradation by random errors.

In some cases, the error degradation of the interpoint distances led to impossible configurations, such as one side of a triangle being greater than the sum of the other two sides. An adjustment computation was programmed and performed which modified the system of interpoint distances by small increments until a consistent set of distances was obtained. Starting with the largest interpoint distance, each set of 3 was tested for consistency with a tolerance of .01. If the tolerance condition was not met, the longer side was reduced by .005 and the shorter sides were increased by .0025 each. The consistency adjustment was repeated iteratively until a consistent system of all of the interpoint distances was obtained. If the distances are not consistent, the trigonometric calculation gives defective results.

This consistent (but not exact) set of interpoint distances then formed the starting point for the calculation and averaging of sets of coordinates for all possible simplexes. With 8 points, the number of simplexes is 28 for 1 dimension, 56 for 2 dimensions, 70 for 3 dimensions, 56 for 4 dimensions, and 28 for 5 dimensions. These are the numbers of separate calculations of coordinates made before averaging coordinates for the different numbers of dimensions.

The results from the 20 different computations are given in Table 2. The coordinates for interpoint distances with small random errors, of the order of 10 per cent, can be satisfactorily estimated for points in one or two dimensions. For larger errors or larger numbers of dimensions, the estimation of coordinates becomes unsatisfactory.

Table 2

AVERAGE ERRORS IN OUTPUTTED INTERPOINT DISTANCES RECOVERED FROM COORDINATES CALCULATED FROM DISTANCES BETWEEN RANDOM POINTS

Number of dimensions	Average error introduced in distances at random	Average original distance	Average error in inputted distance	Average error in outputted distance recovered
1	.10	.79	.09	.09
	.20	.52	.19	.13
	.40	.69	.35	.24
	.80	.71	.63	.58
2	.10	1.16	.10	.12
	.20	1.08	.19	.20
	.40	1.09	.35	.23
	.80	1.16	.72	.32
3	.10	1.41	.10	.66
	.20	1.27	.15	.12
	.40	1.06	.36	.56
	.80	1.44	.78	.48
4	.10	1.49	.10	.26
	.20	1.86	.20	1.38
	.40	1.65	.50	1.67
	.80	1.43	.65	.62
5	.10	1.81	.10	.43
	.20	1.88	.23	.89
	.40	2.04	.40	1.22
	.80	1.96	.75	1.64

Possibly the results from averaging coordinates of alternative simplexes can form the starting solution for the iterative solution of the quadratic system of equations. The most dependable procedure appears to be to use the method of descent described by the author (1976a). In this method of descent, the exact calculation of coordinates from alternative simplexes may be expected to improve the computational efficiency.

REFERENCES

Benson, P. H. Using simplex relationships in multidimensional scaling. Working paper, presented at the program of the Psychometric Society and the Mathematical Psychology Group April 2, 1976. (a)

Benson, P. H. A direct procedure for calculating point coordinates from interpoint distances. Publication pending, expected 1976 or 1977. (b)

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

Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. Psychometrika, 1964, 29, 28-42. (b)

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika, 1962, 27, 125-140. (a)

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. Psychometrika, 1962, 27, 219-246. (b)

Shepard, R. N. Representation of structure in similarity data: Problems and prospects. Psychometrika, 1974, 39, 373-421.

Torgerson, W. S. Multidimensional scaling: I. theory and method. Psychometrika, 1952, 17, 401-419.

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