

MULTIDIMENSIONAL SCALING

J. DE LEEUW

The term ‘Multidimensional Scaling’ or MDS is used in two essentially different ways in statistics (de Leeuw & Heiser 1980a). *MDS in the wide sense* refers to any technique that produces a multidimensional geometric representation of data, where quantitative or qualitative relationships in the data are made to correspond with geometric relationships in the representation. *MDS in the narrow sense* starts with information about some form of dissimilarity between the elements of a set of objects, and it constructs its geometric representation from this information. Thus the data are *dissimilarities*, which are distance-like quantities (or *similarities*, which are inversely related to distances). This chapter concentrates on narrow-sense MDS only, because otherwise the definition of the technique is so diluted as to include almost all of multivariate analysis.

MDS is a descriptive technique, in which the notion of statistical inference is almost completely absent. There have been some attempts to introduce statistical models and corresponding estimating and testing methods, but they have been largely unsuccessful. I introduce some quick notation. Dissimilarities are written as δ_{ij} , and distances are $d_{ij}(X)$. Here i and j are the objects we are interested in. The $n \times p$ matrix X is the *configuration*, with coordinates of the objects in \mathcal{R}^p . Often, we also have as data *weights* w_{ij} reflecting the importance or precision of dissimilarity δ_{ij} .

1. SOURCES OF DISTANCE DATA

Dissimilarity information about a set of objects can arise in many different ways. I review some of the more important ones, organized by scientific discipline.

1.1. Geodesy. The most obvious application, perhaps, is in sciences in which distance is measured directly, although generally with error. This happens, for instance, in triangulation in geodesy, in which measurements are made which are approximately equal to distances, either Euclidean or spherical, depending on the scale of the experiment.

In other examples, measured distances are less directly related to physical distances. For example, one could measure airplane or road or train travel distances between different cities. Physical distance is usually not the only factor determining these types of dissimilarities.

1.2. Geography/Economics. In economic geography, or spatial economics, there are many examples of input-output tables, where the table indicates some type of interaction between a number of regions or countries. For instance, we may have n countries, and entry f_{ij} indicates the number of tourists travelling, or the amount of grain exported, from i to j . It is not difficult to think of many other examples of these square (but generally asymmetric) tables. Again, physical distance may be a contributing factor to these dissimilarities, but certainly not the only one.

1.3. Genetics/Systematics. A very early application of a scaling technique was Fisher (1922). He used crossing-over frequencies from a number of loci to construct a (one-dimensional) map of part of the chromosome. Another early application of MDS ideas is in Boyden (1931), where reactions to sera are used to give similarities between common mammals, and these similarities are then mapped into three-dimensional space.

In much of systematic zoology distances between species or individuals are actually computed from a matrix of measurements on a number of variables describing the individuals. There are many measures of similarity or distance which have been used, and not all of them have the usual metric properties. The derived dissimilarity or similarity matrix is analyzed by MDS, or by cluster analysis, because systematic zoologists show an obvious preference for tree representations over continuous representations in \mathcal{R}^p .

1.4. Psychology/Phonetics. MDS, as a set of data analysis techniques, clearly originates in psychology. There is a review of the early history, which starts with Carl Stumpf around 1880, in de Leeuw & Heiser (1980a). Developments in psychophysics concentrated on specifying the shape of the function relating dissimilarities and distances, until Shepard (1962) made the radical proposal to let the data determine this shape, requiring this function only to be increasing.

In psychophysics one the basic forms in which data are gathered is the *confusion matrix*. In such a matrix we record how many times row-stimulus i was identified as column-stimulus j . A classical example are the Morse code signals studied by Rothkopf (1957). Confusion matrices are not unlike the input-output matrices of economics.

In psychology (and marketing) researchers also collect direct similarity judgments in various forms to map cognitive domains. Ekman's color similarity data is one of the prime examples (Ekman 1963), but many measures of similarity (rankings, ratings, ratio estimates) have been used.

1.5. Psychology/Political Science/Choice Theory. Another source of distance information are *preference data*. If a number of individuals indicate

their preferences for a number of objects, then many choice models use geometrical representations in which an individual prefers the object she is closer to. This leads to ordinal information about the distances between the individuals and the objects, e.g., between the politicians and the issues they vote for, or between the customers and the products they buy.

1.6. Biochemistry. Fairly recently, MDS has been applied in the conformation of molecular structures from nuclear resonance data. The pioneering work is Crippen (1977), and a more recent monograph is Crippen & Havel (1988). Recently, this work has become more important because MDS techniques are used to determine protein structure. Numerical analysts and mathematical programmers have become involved, and as a consequence there have been many new and exciting developments in MDS.

2. AN EXAMPLE

The previous section shows that it will be difficult to find an example that illustrates all aspects of MDS. We select one that can be used in quite a few of the techniques discussed below. It is taken from Coombs (1964, page 464). The data are cross-references between ten psychological journals. The journals are given in Table 1. The actual data are in Table 2. The basic

	Journal	Label
A	<i>American Journal of Psychology</i>	AJP
B	<i>Journal of Abnormal and Social Psychology</i>	JASP
C	<i>Journal of Applied Psychology</i>	JAP
D	<i>Journal of Comparative and Physiological Psychology</i>	JCPP
E	<i>Journal of Consulting Psychology</i>	JCP
F	<i>Journal of Educational Psychology</i>	JEP
G	<i>Journal of Experimental Psychology</i>	JExP
H	<i>Psychological Bulletin</i>	PB
I	<i>Psychological Review</i>	PR
J	<i>Psychometrika</i>	Pka

TABLE 1. Ten Psychology Journals

idea, of course, is that journals with many cross-references are similar.

3. TYPES OF MDS

There are two different forms of MDS, depending on how much information is available about the distances. In some of the applications reviewed above the dissimilarities are known numbers, equal to distances, except perhaps for measurement error. In other cases only the rank order of the dissimilarities is known, or only a subset of them is known.

	A	B	C	D	E	F	G	H	I	J
A	122	4	1	23	4	2	135	17	39	1
B	23	303	9	11	49	4	55	50	48	7
C	0	28	84	2	11	6	15	23	8	13
D	36	10	4	304	0	0	98	21	65	4
E	6	93	11	1	186	6	7	30	10	14
F	6	12	11	1	7	34	24	16	7	14
G	65	15	3	33	3	3	337	40	59	14
H	47	108	16	81	130	14	193	52	31	12
I	22	40	2	29	8	1	97	39	107	13
J	2	0	2	0	0	1	6	14	5	59

TABLE 2. References in Row-Journal to Column-Journal

3.1. Metric Scaling. In metric scaling the dissimilarities between all objects are known numbers, and they are approximated by distances. Thus objects are mapped into a metric space, distances are computed, and compared with the dissimilarities. Then objects are moved in such a way that the fit becomes better, until some loss function is minimized.

In geodesy and molecular genetics this is a reasonable procedure, because dissimilarities correspond rather directly with distances. In analyzing input-output tables, however, or confusion matrices, such tables are often clearly asymmetric and not likely to be directly translatable to distances. Such cases often require a model to correct for asymmetry and scale. The most common class of models (for counts in a square table) is $\mathbf{E}(f_{ij}) = \alpha_i \beta_j \exp\{-\phi(d_{ij}(X))\}$, where ϕ is some monotone transformation through the origin. For ϕ equal to the identity this is known as the choice model for recognition experiments in mathematical psychology (Luce 1963), and as a variation of the quasi-symmetry model in statistics (Haberman 1974). The negative exponential of the distance function was also used by Shepard (1957) in his early theory of recognition experiments.

As we noted above, in systematic zoology and ecology, the basic data matrix is often a matrix in which n objects are measured on p variables. The first step in the analysis is to convert this into a $n \times n$ matrix of similarities or dissimilarities. Which measure of (dis)similarity is chosen depends on the types of variables in the problem. If they are numerical Euclidean distances or Mahalanobis distances can be used, but if they are binary other dissimilarity measures come to mind (Gower & Legendre 1986). In any case, the result is a matrix which can be used as input in a metric MDS procedure.

3.2. Nonmetric Scaling. In various situations, in particular in psychology, only the rank order of the dissimilarities is known. This is either because only ordinal information is collected (for instance by using paired or triadic comparisons) or because while the assumption is natural that the function relating dissimilarities and distances is monotonic, the choice of a specific functional form is not.

There are other cases in which there is incomplete information. For example, observations may be available on only a subset of the distances, either by design or by certain natural restrictions on what is observable. Such cases lead to a *distance completion problem*, where the configuration is constructed from a subset of the distances, and at the same time the other (missing) distances are estimated. Such distance completion problems (assuming that the observed distances are measured without error) are currently solved with mathematical programming methods (Alfakih, Khandani & Wolkowicz 1998).

3.3. Three-way Scaling. In *three-way scaling* information is available on dissimilarities between n objects on m occasions, or for m subjects. Two easy ways of dealing with the occasions is to perform either a separate MDS for each subject or to perform a single MDS for the average occasion. Three-way MDS constitutes a strategy between these two extremes.

This technique requires computation of m MDS solutions, but they are required to be related to each other. For instance, one can impose the restriction that the configurations are the same, but the transformations relating dissimilarities and distances are different. Or one could require that the projections on the dimensions are linearly related to each other in the sense that $d_{ij}(X_k) = d_{ij}(XW_k)$, where W_k is a diagonal matrix characterizing occasion k . A very readable introduction to three-way scaling is Arabie, Carroll & DeSarbo (1987).

3.4. Unfolding. In *multidimensional unfolding* information is available only about off-diagonal dissimilarities, either metric or nonmetric. This means dealing with two different sets of objects, for instance individuals and stimuli or members of congress and political issues, and dissimilarities between members of the first set and members of the second set, and not on the within-set dissimilarities. This typically happens with preference and choice data, in which how individuals like candies, or candidates like issues is known, but not how the individuals like other individuals, and so on.

In many cases, the information in unfolding is also only ordinal. Moreover, it is *conditional*, which means that while it is known that a politician prefers one issue over another, it is not known if a politician's preference for an issue is stronger than another politician's preference for another issue. Thus the ordinal information is only within rows of the off-diagonal matrix.

This makes unfolding data, especially nonmetric unfolding data, extremely sparse.

3.5. Restricted MDS. In many cases it makes sense to impose restrictions on the representation of the objects in MDS. The design of a study may be such that the objects are naturally on a rectangular grid, for instance, or on a circle or ellipse. Often, incorporating such prior information leads to a more readily interpretable and more stable MDS solution.

As we have seen above, some of the more common applications of restricted MDS are to three-way scaling.

4. EXISTENCE THEOREM

The basic existence theorem in Euclidean MDS, in matrix form, is due to Schoenberg (1937). A more modern version was presented in the book by Torgerson (1958).

We give a simple version here. Suppose E is a non-negative, hollow¹, symmetric matrix of order n , and suppose $J_n = I_n - \frac{1}{n}e_n e_n'$ is the *centering* operator. Here I_n is the identity, and e_n is a vector with all elements equal to one. Then E is a matrix of squared Euclidean distances between n points in \mathcal{R}^p if and only if $-\frac{1}{2}J_n E J_n$ is positive semi-definite of rank less than or equal to p .

This theorem has been extended to the classical non-Euclidean geometries, for instance by Blumenthal (1953). It can also be used to show that any non-negative, hollow, symmetric E can be imbedded *nonmetrically* in $n - 2$ dimensions.

5. LOSS FUNCTIONS

5.1. Least Squares on the Distances. The most straightforward loss function to measure fit between dissimilarities and distances is STRESS, defined by

$$(1) \quad \text{STRESS}(X) \triangleq \sum_{i=1}^n \sum_{j=1}^n w_{ij} (\delta_{ij} - d_{ij}(X))^2.$$

Obviously this formulation applies to metric scaling only. In the case of nonmetric scaling the major breakthrough in a proper mathematical formulation of the problem was Kruskal (1964). For this case, STRESS is defined as,

$$(2) \quad \text{STRESS}(X, \hat{D}) \triangleq \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (\hat{d}_{ij} - d_{ij}(X))^2}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (d_{ij}(X) - \bar{d}(X))^2},$$

¹i.e. zero-diagonal

and this function is minimized over both X and \hat{D} , where \hat{D} satisfies the constraints imposed by the data. In nonmetric MDS the \hat{D} are called *disparities*, and are required to be monotonic with the dissimilarities. Finding the optimal \hat{D} is an *isotonic regression problem*. In the case of distance completion problems (with or without measurement) error, the \hat{d}_{ij} must be equal to the observed distances if these are observed, and they are free otherwise.

One particular property of the STRESS loss function is that it is not differentiable for configurations in which two points coincide (and a distance is zero). It is shown by de Leeuw (1984) that at a local minimum of STRESS pairs of points with positive dissimilarities cannot coincide.

5.2. Least Squares on the Squared Distances. A second loss function, which has been used a great deal, is SSTRESS, defined by

$$(3) \quad \text{SSTRESS}(X) \triangleq \sum_{i=1}^n \sum_{j=1}^n w_{ij} (\delta_{ij}^2 - d_{ij}^2(X))^2.$$

Clearly, this loss function is a (fourth-order) multivariate polynomial in the coordinates. There are no problems with smoothness, but often a large number of local optima results.

Of course a nonmetric version of the SSTRESS problem can be confronted, using the same type of approach used for STRESS.

5.3. Least Squares on the Inner products. The existence theorem discussed above suggest a third way to measure loss. Now the function is known as STRAIN, and it is defined, in matrix notation, as

$$(4) \quad \text{STRAIN}(X) \triangleq \text{tr}\{J(\Delta^{(2)} - D^{(2)}(X))J(\Delta^{(2)} - D^{(2)}(X))\},$$

where $D^{(2)}(X)$ and $\Delta^{(2)}$ are the matrices of squared distances and dissimilarities, and where J is the centering operator. Since $JD^{(2)}(X)J = -2XX'$ this means that $-\frac{1}{2}J\Delta^{(2)}J$ is approximated by a positive semi-definite matrix of rank r , which is a standard eigenvalue-eigenvector computation.

Again, nonmetric versions of minimizing STRAIN are straightforward to formulate (although less straightforward to implement).

6. ALGORITHMS

6.1. STRESS. The original algorithms (Kruskal 1964) for minimizing STRESS use gradient methods with elaborate step-size procedures. In de Leeuw (1977) the *majorization method* was introduced. It leads to a globally convergent algorithm with a linear convergence rate, which is not bothered by the nonexistence of derivatives at places where points coincide. The majorization method can be seen as a gradient method with a constant step-size, which uses convex analysis methods to prove convergence.

More recently, faster linearly or superlinearly convergent methods have been tried successfully (Glunt, Hayden & Rayden 1993, Kearsley, Tapia & Trosset 1998).

One of the key advantages of the majorization method is that it extends easily to restricted MDS problems (de Leeuw & Heiser 1980*b*). Each sub-problem in the sequence is a least squares projection problem on the set of configurations satisfying the constraints, which is usually easy to solve.

6.2. **SSTRESS.** Algorithms for minimizing SSTRESS were developed initially by Takane, Young & de Leeuw (1984). They applied cyclic coordinate descend, i.e. one coordinate was changed at the time, and cycles through the coordinates were alternated with isotonic regressions in the nonmetric case. More efficient alternating least squares algorithms were developed later by De Leeuw, Takane, and Browne (cf. Browne (1987)), and superlinear and quadratic methods were proposed by W. Glunt & Liu (1991) and Kearsley et al. (1998).

6.3. **STRAIN.** Minimizing STRAIN was, and is, the preferred algorithm in metric MDS. It is also used as the starting point in iterative nonmetric algorithms. Recently, more general algorithms for minimizing STRAIN in nonmetric and distance completion scaling have been proposed by Trosset (1998*b*) and Trosset (1998*a*).

7. ANALYSIS OF THE EXAMPLE

7.1. **Initial Transformation.** In the journal reference example we suppose $\mathbf{E}(f_{ij}) = \alpha_i \beta_j \exp\{-\phi(d_{ij}(X))\}$. In principle this model can be tested by contingency table techniques, but we shall not go that route. We merely use the model to transform the frequencies to estimated distances, using

$$-\log \sqrt{\frac{\tilde{f}_{ij} \tilde{f}_{ji}}{\tilde{f}_{ii} \tilde{f}_{jj}}} \approx \phi(d_{ij}(X)),$$

where $\tilde{f}_{ij} = f_{ij} + \frac{1}{2}$. This transformed matrix is given in Table 3.

7.2. **Metric Analysis.** In the first analysis we suppose the numbers in Table 3 are approximate distances, i.e. we suppose ϕ is the identity. We then minimize STRAIN, using metric MDS, i.e. we find the dominant eigenvalues and corresponding eigenvectors of the doubly-centered squared distance matrix. This results in the following two-dimensional configurations. In the second analysis we iteratively minimize metric STRESS, using the majorization algorithm. The solutions are given in Figure 1. In both figures we see the same grouping of journals, with Pka as an outlier, the journals central to the discipline, such as AJP, JExp, PB, and PR, in the middle, and more

0.00	2.93	4.77	1.89	3.33	2.78	0.77	1.02	1.35	3.79
2.93	0.00	2.28	3.32	1.25	2.61	2.39	0.53	1.41	4.24
4.77	2.28	0.00	3.87	2.39	1.83	3.13	1.22	3.03	2.50
1.89	3.32	3.87	0.00	5.62	4.77	1.72	1.11	1.41	4.50
3.33	1.25	2.39	5.62	0.00	2.44	3.89	0.45	2.71	3.67
2.78	2.61	1.83	4.77	2.44	0.00	2.46	1.01	2.90	2.27
0.77	2.39	3.13	1.72	3.89	2.46	0.00	0.41	0.92	2.68
1.02	0.53	1.22	1.11	0.45	1.01	0.41	0.00	0.76	1.42
1.35	1.41	3.03	1.41	2.71	2.90	0.92	0.76	0.00	2.23
3.79	4.24	2.50	4.50	3.67	2.27	2.68	1.42	2.23	0.00

TABLE 3. Transformed Journal Reference Data

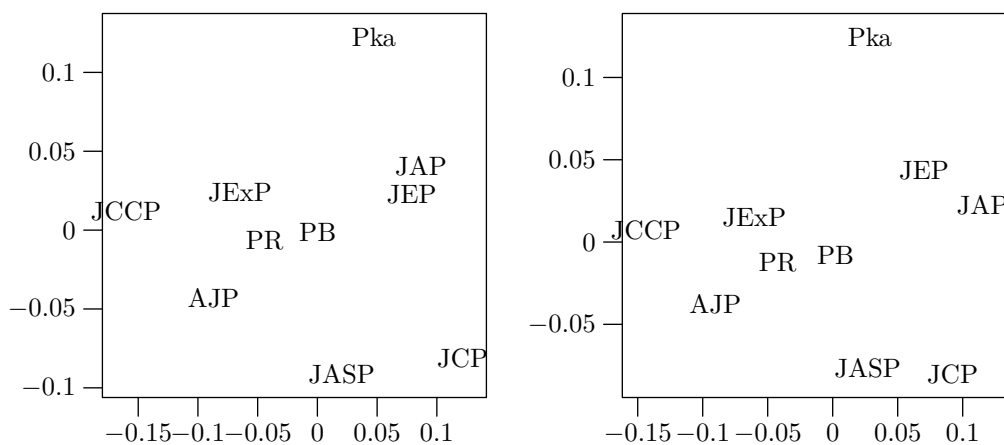


FIGURE 1. Metric Analysis (STRAIN left, STRESS right)

specialized journals generally in the periphery. For comparison purposes the STRESS of the first solution is 0.0687, that of the second solution is 0.0539. Finding the second solution takes about 30 iterations.

7.3. Nonmetric STRESS Analysis. Next, we minimize nonmetric STRESS on the same data (using only their rank order). The solution is in Figure 2. On the left we see the transformation relating the data in Table 3 to the optimally transformed data, a monotone step function. Again we see basically the same configuration of journals, with the same groupings. The nonmetric solution has a (normalized) STRESS of 0.0195, and again finding it takes about 30 iterations of the majorization method. The optimal transformation does not seem to deviate systematically from linearity.

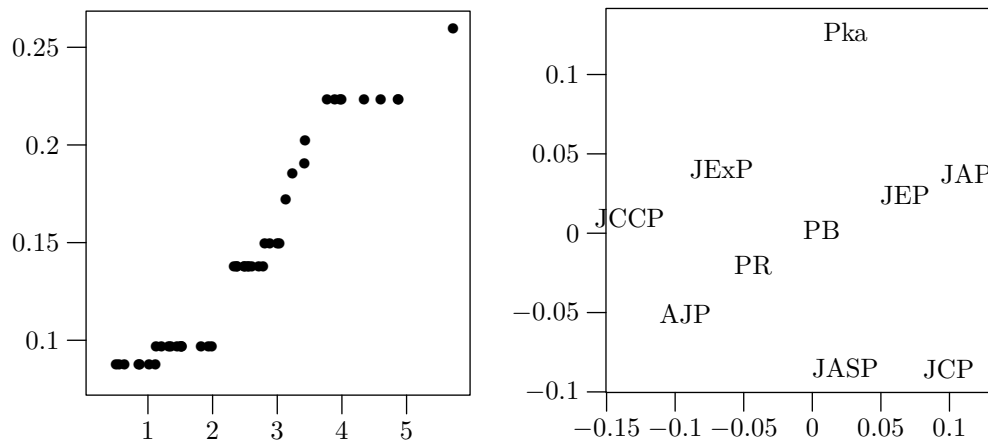


FIGURE 2. Nonmetric Analysis (Transformation left, Solution right)

8. FURTHER READING

Until recently, the classical MDS reference was the little book by Kruskal & Wish (1978). It is clearly written, but very elementary. A more elaborate practical introduction is by Coxon (1982), which has a useful companion volume (Davies & Coxon 1982) with many of the classical MDS papers. Some additional early intermediate-level books, written from the psychometric point of view, are Davison (1983) and Young (1987).

More recently, more modern and advanced books have appeared. The most complete treatment is no doubt Borg & Groenen (1997), while Cox & Cox (1994) is another good introduction especially aimed at statisticians.

REFERENCES

- Alfakih, A., Khandani, A. & Wolkowicz, H. (1998), 'Solving Euclidean distance matrix completion problems via semidefinite programming', *Computational Optimization and Applications* **12**, 13–30.
- Arabie, P., Carroll, J. & DeSarbo, W. (1987), *Three-way scaling and clustering*, Sage Publication.
- Blumenthal, L. (1953), *Distance Geometry*, Oxford University Press.
- Borg, I. & Groenen, P. (1997), *Modern Multidimensional Scaling*, Springer-Verlag.
- Boyden, A. (1931), 'Precipitin tests as a basis for a comparative phylogeny', *Proceedings of the Society for Experimental Biology and Medicine* **29**, 955–957.
- Browne, M. (1987), 'The Young-Householder algorithm and the least squares multidimensional scaling of squared distances', *Journal of Classification* **4**, 175–190.
- Coombs, C. (1964), *A Theory of Data*, Wiley.
- Cox, T. & Cox, M. (1994), *Multidimensional Scaling*, Chapman & Hall.
- Coxon, A. (1982), *The User's Guide to Multidimensional Scaling*, Heinemann.

- Crippen, G. (1977), 'A novel approach to calculation of conformation: Distance geometry', *Journal of Computational Physics* **24**, 96–107.
- Crippen, G. & Havel, T. (1988), *Distance Geometry and Molecular Conformation*, Wiley.
- Davies, P. & Coxon, A. (1982), *Key Texts in Multidimensional Scaling*, Heinemann.
- Davison, M. (1983), *Multidimensional Scaling*, Wiley.
- de Leeuw, J. (1977), Applications of convex analysis to multidimensional scaling, in J. Barra, F. Brodeau, G. Romier & B. van Cutsem, eds, 'Recent developments in statistics', North Holland Publishing Company, Amsterdam, The Netherlands, pp. 133–145.
- de Leeuw, J. (1984), 'Differentiability of Kruskal's Stress at a local minimum', *Psychometrika* **49**, 111–113.
- de Leeuw, J. & Heiser, W. (1980a), Theory of multidimensional scaling, in P. Krishnaiah, ed., 'Handbook of Statistics, volume II', North Holland Publishing Company, Amsterdam, The Netherlands.
- de Leeuw, J. & Heiser, W. J. (1980b), Multidimensional scaling with restrictions on the configuration, in P. Krishnaiah, ed., 'Multivariate Analysis, volume V', North Holland Publishing Company, Amsterdam, The Netherlands, pp. 501–522.
- Ekman, G. (1963), 'Direct method for multidimensional ratio scaling', *Psychometrika* **23**, 33–41.
- Fisher, R. (1922), 'The systematic location of genes by means of cross-over ratios', *American Naturalist* **56**, 406–411.
- Glunt, W., Hayden, T. & Rayden, M. (1993), 'Molecular conformations from distance matrices', *Journal of Computational Chemistry* **14**, 114–120.
- Gower, J. & Legendre, P. (1986), 'Metric and Euclidean properties of dissimilarity coefficients', *Journal of Classification* **3**, 5–48.
- Haberman, S. (1974), *The Analysis of Frequency Data*, University of Chicago Press.
- Kearsley, A., Tapia, R. & Trosset, M. (1998), 'The solution of the metric STRESS and SSTRESS problems in multidimensional scaling using Newton's method', *Computational Statistics* **13**, 369–396.
- Kruskal, J. (1964), 'Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis', *Psychometrika* **29**, 1–27.
- Kruskal, J. & Wish, M. (1978), *Multidimensional Scaling*, Sage Publications.
- Luce, R. (1963), Detection and recognition, in 'Handbook of mathematical Psychology, I', Wiley.
- Rothkopf, E. (1957), 'A measure of stimulus similarity and errors in some paired-associate learning tasks', *Journal of Experimental Psychology* **53**, 94–101.
- Schoenberg, I. (1937), 'Remarks on Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espaces distanciés vectoriellement applicable sur l'espace de Hilbert"', *Annals of Mathematics* pp. 724–732.
- Shepard, R. (1957), 'Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space', *Psychometrika* **22**, 325–345.
- Shepard, R. (1962), 'The analysis of proximities: multidimensional scaling with an unknown distance function', *Psychometrika* **27**, 125–140, 219–246.
- Takane, Y., Young, F. & de Leeuw, J. (1984), 'Nonmetric individual differences in multidimensional scaling: An alternating least squares method with optimal scaling features', *Psychometrika* **42**, 7–67.
- Torgerson, W. (1958), *Theory and Methods of Scaling*, Wiley.
- Trosset, M. (1998a), 'Applications of multidimensional scaling to molecular conformation', *Computing Science and Statistics* **29**, 148–152.

- Trosset, M. (1998*b*), 'A new formulation of the nonmetric strain problem in multidimensional scaling', *Journal of Classification* **15**, 15–35.
- W. Glunt, T. H. & Liu, W.-M. (1991), 'The embedding problem for predistance matrices', *Bulletin of Mathematical Biology* **53**, 769–796.
- Young, F. (1987), *Multidimensional Scaling*, Earlbaum.

UNIVERSITY OF CALIFORNIA, LOS ANGELES