PSYCHOMETRIKA—VOL. 47, NO. 2.
JUNE 1982

RESTRICTED MULTIDIMENSIONAL SCALING MODELS FOR ASYMMETRIC PROXIMITIES

DAVID G. WEEKS
WASHINGTON UNIVERSITY SCHOOL OF MEDICINE

AND

P. M. BENTLER

UNIVERSITY OF CALIFORNIA, LOS ANGELES

Restricted multidimensional scaling models [Bentler & Weeks, 1978] allowing constraints on parameters, are extended to the case of asymmetric data. Separate functions are used to model the symmetric and antisymmetric parts of the data. The approach is also extended to the case in which data are presumed to be linearly related to squared distances. Examples of several models are provided, using journal citation data. Possible extensions of the models are considered.

Key words: multidimensional scaling, asymmetric data.

Restricted Multidimensional Scaling Models for Asymmetric Proximities

Data that are both asymmetric and distance-like occur in several situations. Examples are confusions of one stimulus for another, migration rates, frequencies of journal citations, and initiation of contact. The magnitude of an observation h_{ij} is plausibly related to the similarity of or distance between entities i and j, but in general $h_{ij} \neq h_{ji}$. A number of multidimensional scaling (MDS) methods have been used to analyze such data. The simplest method is to ignore the asymmetry (e.g., by averaging responses h_{ij} and h_{ji} if a symmetric matrix is required) and then use any traditional MDS method. The problem here is that the asymmetry may embody important information. Models which attempt to represent the asymmetry include those of Chino [Note 1], Cunningham [1978], Constantine and Gower [1978], Gower [1977], Bishop, Fienberg, and Holland [1975, Chapter 8], Harshman [Note 2], Holman [1979], Hutchinson [Note 3], Tobler [1976, 1979], and Young [Note 4].

Any square nonsymmetric matrix can be additively decomposed into a symmetric and a skew-symmetric matrix, $\mathbf{P} = \mathbf{Q} + \mathbf{R}$, where $\mathbf{Q}$ is symmetric, $q_{ij} = q_{ji} = (p_{ij} + p_{ji})/2$, and $\mathbf{R}$ is skew-symmetric, $\mathbf{R} = \mathbf{P} - \mathbf{Q}$, $r_{ij} = -r_{ji}$. The model proposed here attempts to represent the symmetric and skew-symmetric parts of the data matrix with two separate functions:

$$h_{ij} = bd_{ij} + k + c_i - c_j + e_{ij} \tag{1}$$

where h_{ij} is an observation, d_{ij} is a euclidean distance, b and k are the parameters of a linear equation (b is particularly useful in the case of similarity data when it can be set to -1; k is the traditional additive constant), c_i, c_j represent the skew-symmetric component of the data, and e_{ij} is a random error component. In matrix terms, (1) may be written as

This research was supported in part by USPHS Grant 0A01070, P. M. Bentler, principal investigator, and NIMH Grant MH-24819, E. J. Anthony and J. Worland, principal investigators.

The authors wish to thank E. W. Holman and several anonymous reviewers for their valuable suggestions concerning this research.

Requests for reprints should be sent to David G. Weeks, Washington University School of Medicine, 369 North Taylor, St. Louis, Missouri 63108.

0033-3123/82/0600-3025$00.75/0

$\mathbf{H} = b\mathbf{D} + k(\mathbf{1}\mathbf{1}' - \mathbf{I}) + \mathbf{c}\mathbf{1}' - \mathbf{1}\mathbf{c}' + \mathbf{E}$. The distances are specified as

$$d_{ij} = [(\mathbf{a}_i - \mathbf{a}_j)\mathbf{S}(\mathbf{a}_i - \mathbf{a}_j)']^{1/2} \tag{2}$$

where $\mathbf{a}_i$ is a $(1 \times t)$ row vector of the projections of object i on t dimensions, and $\mathbf{S}$ is a $(t \times t)$ symmetric matrix representing the covariances of the (possibly oblique) dimensions. The parameters of the model consist of b, k, and each a_{ip}, c_i, s_{pq}. Any parameter Θ is subject to constraints of the forms

$$\begin{gathered} \Theta_u = \text{constant} \\ w_u \Theta_u = w_v \Theta_v . \end{gathered} \tag{3}$$

The model defined by (1)–(3) is equivalent to that of Bentler and Weeks [1978] except for the asymmetry induced by the c_i terms. It should also be noted that the form $c_i - c_j$ (in the unrestricted case) is equivalent to the one-dimensional case of Gower's [1977] decomposition of skew-symmetric matrices. One could also consider (1) as a particular metric form of Holman's [1979] model with the appropriate choice of symmetric, row, and column functions (d_{ij}, c_i, $-c_j$ respectively) which were specified by Holman as arbitrary monotonic functions. In some cases it may be more reasonable to assume the symmetric component of the data are more like squared distances than distances (e.g., Takane, Young, & de Leeuw, [1977]). In such cases (1) may be written as

$$h_{ij} = bd_{ij}^2 + k + c_i - c_j + e_{ij}. \tag{4}$$

Identification and Estimation

Some constraints must be imposed in order for the model to be identified. Since distances are invariant with respect to translation and rotation, $t(t + 1)/2$ constraints can be placed in the projection matrix $\mathbf{A}$. For example, all elements of A on or above the diagonal can be constrained to equal zero. This requirement holds in the case of $\mathbf{S} = \mathbf{I}$. If elements of $\mathbf{S}$ are free, additional constraints in $\mathbf{A}$ would be required. If, for example, only the upper triangle of $\mathbf{A}$ is constrained, $\mathbf{A}$ could be multiplied by some constant u and the parameter b could be divided by u [u^2 for (4)]. Thus either b or an additional element of $\mathbf{A}$ must be constrained. One constraint must also be placed on $\mathbf{c}$. In matrix form, the skew-symmetric component of the model is given by $\mathbf{c}\mathbf{1}' - \mathbf{1}\mathbf{c}'$ where $\mathbf{1}$ is a column vector of ones. Let $\mathbf{c}^* = \mathbf{c} + u\mathbf{1}$, where u is a scalar. Then $\mathbf{c}^*\mathbf{1}' - \mathbf{1}\mathbf{c}^{*\prime} = \mathbf{c}\mathbf{1} + u\mathbf{1}\mathbf{1}' - \mathbf{1}\mathbf{c}' = u\mathbf{1}\mathbf{1}' = \mathbf{c}\mathbf{1}' - \mathbf{1}\mathbf{c}'$. One constraint, e.g., $c_1 = 0$, requires $u = 0$.

We have employed the Gauss-Newton algorithm (sometimes preceded by several steepest descent steps) to obtain least-squares estimates of the parameters. A step in the Gauss-Newton algorithm is defined by

$$\Theta^{k+1} = \Theta^k + \alpha^k(\mathbf{F}^k\mathbf{F}^{k\prime})^{-1}\mathbf{F}^k \operatorname{Vec}(\mathbf{H} - \hat{\mathbf{H}}^k)$$

where the superscript indicates the iteration number, $\mathbf{F}^k = \partial\mathbf{H}/\partial\Theta$, the derivatives of the model, with respect to the parameters Θ evaluated at the current parameter estimates; rows of $\mathbf{F}$ are indexed by parameters, columns are indexed by elements of $\mathbf{H}$. $\hat{\mathbf{H}} = b\mathbf{D} + k(\mathbf{1}\mathbf{1}' - \mathbf{I}) + \mathbf{c}\mathbf{1}' - \mathbf{1}\mathbf{c}'$ for the distance model, and $\hat{\mathbf{H}} = b\mathbf{D}^{(2)} + k(\mathbf{1}\mathbf{1}' - \mathbf{I}) + \mathbf{c}\mathbf{1}' - \mathbf{1}\mathbf{c}'$ for the squared distance model (the exponent in parentheses denotes an elementwise operation). A stepsize parameter, $0 < \alpha < 1$, is included to ensure a nonincreasing error function $[tr(\mathbf{H} - \hat{\mathbf{H}})^2]$ at each iteration.

The necessary derivatives for the distance model (1) are given by

$$
\begin{aligned}
&\frac{\partial \hat{h}_{ij}}{\partial b} = d_{ij} \\
&\frac{\partial \hat{h}_{ij}}{\partial a_{iq}} = b \sum_{p}^{t} s_{qp} \frac{(a_{ip} - a_{jp})}{d_{ij}} \\
&\frac{\partial \hat{h}_{ij}}{\partial k} = 1 \\
&\frac{\partial \hat{h}_{ij}}{\partial s_{pq}} = \frac{b(a_{ip} - a_{jp})(a_{iq} - a_{jq})}{2d_{ij}} \\
&\frac{\partial \hat{h}_{ij}}{\partial c_i} = 1, \qquad \frac{\partial h_{ji}}{\partial c_i} = -1.
\end{aligned}
\tag{5}
$$

For the squared distance model (4) the derivatives are given by

$$
\begin{aligned}
&\frac{\partial \hat{h}_{ij}}{\partial b} = d_{ij}^2 \\
&\frac{\partial \hat{h}_{ij}}{\partial a_{iq}} = 2b \sum_{p}^{t} s_{qp}(a_{ip} - a_{jp}) \\
&\frac{\partial \hat{h}_{ij}}{\partial k} = 1 \\
&\frac{\partial \hat{h}_{ij}}{\partial s_{pq}} = b(a_{ip} - a_{jp})(a_{iq} - a_{jq}) \\
&\frac{\partial \hat{h}_{ij}}{\partial c_i} = 1, \qquad \frac{\partial h_{ji}}{\partial c_i} = -1.
\end{aligned}
\tag{6}
$$

Note that in (5), derivatives with respect to elements of $\mathbf{A}$ and $\mathbf{S}$ are discontinuous at $d_{ij} = 0$, but in (6) they are continuous. In some cases this implies a distinct computational advantage for the squared distance model. The parameter vector Θ refers to only free, nonredundant parameters—some subset of the possible parameters b, a_{iq}, k, s_{pq}, and c_i. Computation of derivatives allowing for constraints is identical to the procedure in Bentler and Weeks [1978, pp. 141–142].

Examples

Data are the number of citations of ten psychological journals to each other in 1960 [Coombs, 1964, p. 464], for eight of these journals in 1964 [Coombs, Dawes, & Tversky, 1969, p. 73], and citations between 12 journals in 1979. An observation h_{ij} represents the number of citations in journal i to journal j. The journals represented are listed in Table 1, along with abbreviations to be used subsequently. The 1979 data are presented in Table 2. These data have the advantage of being familiar to most people involved with MDS, yet with the inclusion of the 1979 data, offer the potential of something new as well.

Prior to the analysis, data were converted to proportions of citations given. This served to eliminate any effects due to differences in number of pages or number of articles per year. The first step was to determine the appropriate dimensionality for each data set independently. Solutions in one, two, and three dimensions were obtained. These solutions

Table 1

Journals Represented in Citation Data Sets

JOURNAL	ABBREVIATIONS	YEARS		
American Journal of Psychology	AJP	60	64	79
Journal of Abnormal and Social Psychology	JASP	60	64	
Journal of Abnormal Psychology	JABN			79
Journal of Personality and Social Psychology	JPSP			79
Journal of Applied Psychology	JAPPL	60	64	79
Journal of Comparative and Physiological Psychology	JCPP	60	64	79
Journal of Consulting Psychology	JCP	60	64	
Journal of Consulting and Clinical Psychology	JCCP			79
Journal of Educational Psychology	JEDP	60	64	79
Journal of Experimental Psychology	JEP	60	64	79[a]
Psychometrika	PKA	60	64	79
Psychological Bulletin	PB	60		79
Psychological Review	PR	60		79
Multivariate Behavioral Research	MBR			79

[a]JEP: General in 1979.

were obtained using the squared distance model (4), with constraints $b = -1$ (because the data are similarity-like), $\mathbf{S} = \mathbf{I}$, one element of $\mathbf{c}$ set to zero, and an upper triangle of $\mathbf{A}$ set to zero to fix the origin and orientation. These constraints were necessary for identification of the model. The fit as indexed by the correlation of model to data, for each year and one to three dimensions is shown in Table 3. These values suggest a reasonably good fit, but they do not clearly indicate a best choice for dimensionality. However, the third dimension in the three-dimensional solution had no obvious interpretation, and for that reason a two-dimensional solution was chosen as best for the present purpose.

Results for the 1960 data are plotted in Figure 1. There seems to be a hard-soft or clinical-experimental dimension roughly parallel to the line between JCPP and JCP. A dimension orthogonal to that serves to separate PKA from the rest. Such might reasonably be considered a quantitative or psychometric dimension.

The goodness of the fit or of the interpretability of the dimensional solution, of course, says nothing about the skew-symmetric component of the model. A model equivalent to the two-dimensional model, but with $\mathbf{c}$ set to zero, was fit to each of the three data sets. Fit for these models are also given in Table 3. The correlations are much lower, clearly indicating that the vector $\mathbf{c}$ plays an important role in accounting for these data. The values of $\mathbf{c}$ are plotted in Figure 2. In this particular application a high value of $\mathbf{c}$ indicates a journal that receives proportionately fewer citations than it gives. So for all three years, JEDP, JAPPL, and AJP are in this position. In the two earlier years, JEP was the most "over-cited" journal, but it did not hold this position in 1979. The most likely reason for this change is the change in the definition of the journal—in 1979, only JEP-General was counted. An-

Table 2

1979 Citation Data. Rows represent journals giving citation; columns represent citations received.

	AJP	JABN	JPSP	JAPPL	JCPP	JEDP	JCCP	JEP	PKA	PB	PR	MBR
1. AJP	31	10	10	1	36	4	1	119	2	14	36	0
2. JABN	7	235	55	0	13	4	65	25	3	50	31	0
3. JPSP	16	54	969	28	15	21	89	62	16	149	141	16
4. JAPPL	3	2	30	310	0	8	5	7	6	71	14	0
5. JCPP	4	0	2	0	386	0	2	13	1	22	35	1
6. JEDP[a]	1	7	61	10	2	100	6	5	4	18	9	2
7. JCCP	0	105	55	7	3	10	331	3	19	89	22	8
8. JEP	9	20	16	0	32	6	1	120	2	18	46	0
9. PKA	2	0	0	0	0	6	0	6	152	31	7	10
10. PB	23	46	124	117	138	7	86	84	62	186	90	7
11. PR	9	2	21	6	3	0	0	51	30	32	104	2
12. MBR	0	7	14	4	0	0	24	3	95	46	2	56

[a]Numbers 1-3 only.

other possibility is that within the journals represented, JEP no longer holds the same position of dominance or centrality that it once did.

The third step was to conduct a longitudinal restricted analysis. It was hypothesized that those journals represented in the two later data sets as well as in 1960 would not move from their 1960 positions, but that the relative size of the dimensions would vary. Examination of the values of **c** for the three years indicated that stability over time was not to be expected. The specific constraints for this model were: for 1964, the location of all eight journals was set equal to their locations in 1960, (all of **A** fixed) but the diagonal of **S** was free. As before, b and one element of **c** were fixed, and k was free. This specification was repeated for the 1979 data, except that the locations of those journals not present in 1960 were left free. Fit of these models is also shown in Table 3. These numbers should be

Table 3

Correlations of Predicted Data to Data for Various Models

MODEL	DIMENSIONS	60	64	79
Unconstrained	1	.6504	.6546	.5292
	2	.7408	.7498	.6521
	3	.8123	.8387	.7363
No c	2	.5651	.5567	.5397
Constrained	2		.7126	.6396

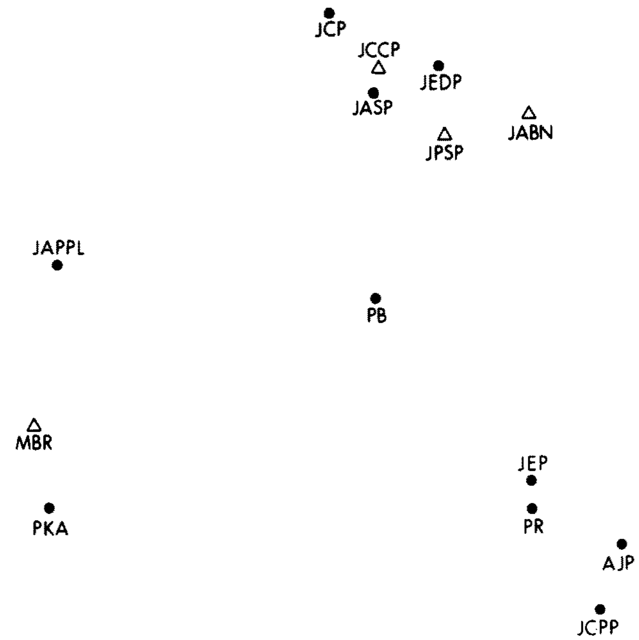

FIGURE 1
Location of journals in two dimensions. Dot indicates journals in 1960 and 1964. Triangle indicates journals included only in 1979.

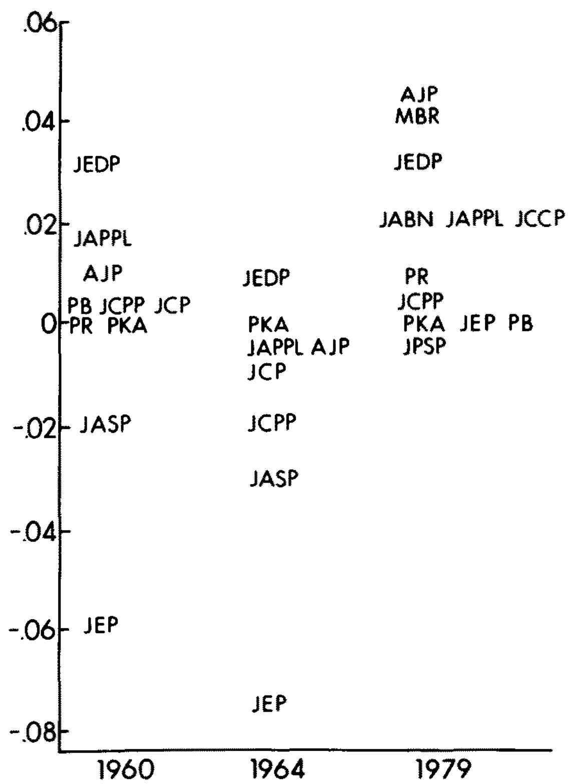

FIGURE 2
Values of c for journals for all three years.

compared to the correlations for the unconstrained model in two dimensions. As can be seen, the decline in fit due to the restrictions is very slight. It can be concluded that the constrained models are almost as good as the unconstrained models. The constrained models are much more parsimonious, using 10 and 22 parameters for 1964 and 1979 respectively, as opposed to 21 and 33 parameters for the unconstrained models. The values of the diagonal elements of S were .86 and .92 for the 1964 data, and .84 and 1.10 for 1979. The journals unique to the 1979 data set are also plotted in Figure 1. Their locations appear reasonable—JCCP, JPSP, and JABN appear near the clinical-social cluster of previous years; MBR appears near PKA but offset a bit towards JAPPL.

Discussion

The model proposed here represents one simple way to conceptualize and evaluate the process of data generation for asymmetric proximities. In the completely unconstrained case, the model can be conceived as one that provides a formal mathematical basis for the common procedure of symmetrizing asymmetric data prior to multidimensional scaling. In fact, it can be shown that the parameter estimates for the asymmetric and symmetric components of the model are separable in this case, and the skew-symmetric scale values can be estimated by the differences between the row and column sums of the proximities. In more complicated constrained models, however, where functional relations between parameters of the model are specified, this simple estimation method is not generally available, and a formal estimation process such as the one proposed here must be adopted.

The analyses were presented primarily for the purpose of illustrating possible uses of the model. Even still, the analyses have revealed more of the structure of the journal citation data. Something like the hard-soft dimension found by Holman [1979] was found here, but in the present analyses, *Psychometrika* appears at the hard end, rather than the soft end. This seems a more reasonable placement. Holman also reported that the results were very similar for both the two older data sets. Our analyses allow a more precise statement: except for a scale factor, the location of journals on the dimension are virtually identical for all three years; the ordering on the skew-symmetric scale (**c**) is only moderately consistent from year to year.

There are several fairly straightforward generalizations of the models (1) and (4) which might be worthwhile to investigate. Following Gower [1977], one could decompose the skew-symmetric part of the data into several components. Several generalizations to individual differences models [cf. Bloxom, 1978] are possible. For example,

$$h_{ijy} = b_y[(\mathbf{a}_i - \mathbf{a}_j)\mathbf{S}_y(\mathbf{a}_i - \mathbf{a}_j)']^{\gamma/2} + k_y + c_{iy} - c_{jy} + e_{ijy}$$

assumes all subjects (y) share a common euclidean space, but the space may differ across individuals in the scale of dimensions and their degree of obliqueness ($\mathbf{S}_y$, k_y); and the skew-symmetric component may be different for each subject (c_{iy}, c_{jy}). Or a slightly more constrained version would write the skew-symmetric component as $b_y^*(c_i - c_j)$, a common scale but with individual weights.

REFERENCE NOTES

1. Chino, N. *A unified geometrical interpretation of the MDS techniques for the analysis of asymmetry and related techniques.* Paper presented at the meeting of the Psychometric Society, Iowa City, May, 1980.
2. Harshman, R. S. *Models for analysis of symmetrical relationships among N objects or stimuli.* Paper presented at the meeting of the Psychometric Society and the Society for Mathematical Psychology, McMaster University, Hamilton, Ontario, August, 1978.
3. Hutchinson, J. W. *Network representations of asymmetric data.* Paper presented at the Berkeley-Stanford joint meeting in Biological and Cognitive Psychology, Stanford, May, 1980.
4. Young, F. W. *An asymmetric euclidean model for multi-process asymmetric data.* Proceedings of the U.S.–Japan Seminar, University of California, San Diego, August, 1975.

REFERENCES

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

Bishop, Y. M. M., Fienberg, S. W., & Holland, P. W. *Discrete multivariate analysis: Theory and practice*. Cambridge: MIT Press, 1975.

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

Constantine, A. G., & Gower, J. C. Graphical representation of asymmetric matrices. *Applied Statistics*, 1978, *27*, 297–304.

Coombs, C. H. *A theory of data*. New York: Wiley, 1964.

Coombs, C. H., Dawes, R. M., & Tversky, A. *Mathematical psychology: An elementary introduction*. Englewood Cliffs, N.J.: Prentice-Hall, 1969.

Cunningham, J. P. Free trees and bidirectional trees as representations of psychological distance. *Journal of Mathematical Psychology*, 1978, *17*, 165–188.

Gower, J. C. The analysis of asymmetry and orthogonality. In J. R. Barra, F. Brodeau, G. Romier, & B. Van Clutsem (Eds.), *Recent developments in statistics*. Amsterdam: North-Holland, 1977.

Holman, E. W. Monotonic models for asymmetric proximities. *Journal of Mathematical Psychology*, 1979, *20*, 1–15.

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

Tobler, W. Spatial interaction patterns. *Journal of Environmental Systems*, 1976, *6*, 271–301.

Tobler, W. R. Estimation of attractivities from interactions. *Environment and Planning A*, 1979, *11*, 121–127.

Manuscript received 9/29/80
Final version received 2/8/82