

Graphical Representation of Asymmetric Matrices

By A. G. CONSTANTINE and J. C. GOWER

CSIRO, Adelaide, South Australia

*Rothamsted Experimental Station,
Harpenden, Herts, Britain*

[Received August 1977. Revised May 1978]

SUMMARY

When sample units are plotted against orthogonal principal axes in a components analysis, the concept of (Euclidean) distance plays a central point in interpretation. Variants of such diagrams are common throughout multivariate analysis. Because distance between a pair of points is independent of the order in which they are taken, displays of this kind are especially associated with symmetric matrices, but methods are also required for displaying asymmetric matrices. In this paper two methods for displaying asymmetric square matrices are presented and illustrated by examples. In the first method (multidimensional unfolding) the square matrix is regarded as part of an otherwise unknown symmetric matrix and the resulting diagram is interpreted using distances, much as with classical methods. In the second method the matrix is partitioned into its symmetric and skew-symmetric components. While the symmetric part is represented by some established distance-based method, the skew-symmetric part is represented by points whose relationships are interpreted in terms of areas of triangles and co-linearities.

Keywords: GRAPHICAL DISPLAY; ORDINATION; SCALING; UNFOLDING; CANONICAL ANALYSIS; ASYMMETRY; SYMMETRY; SKEW-SYMMETRY

1. INTRODUCTION

DISTANCE relationships are symmetric so distance gives a convenient basis for representing the values of symmetric matrices graphically. This idea is much exploited by classical multivariate methods. For example, in principal components analysis samples are exhibited as points plotted relative to orthogonal principal axes. In the full multivariate space similar samples are represented by neighbouring points and dissimilar samples by distant points. Here the relevant distance d_{ij} between the i th and j th samples is given by

$$d_{ij}^2 = \sum_{k=1}^p (x_{ik} - x_{jk})^2,$$

where x_{ik} represents the value observed for the k th variate on the i th sample. In canonical variate analysis $d_{ij} = D_{ij}$, the Mahalanobis distance. In multidimensional scaling and various forms of hierarchical cluster analysis d_{ij} is (a function of) an observed or calculated dissimilarity. In all these cases the data values d_{ij} form a symmetric matrix which is approximated and exhibited graphically as a two (or higher)-dimensional map giving rise to fitted distances d_{ij} . With hierarchical dendrograms the d_{ij} form a set of Euclidean distances satisfying the ultrametric inequality (Holman, 1972; Gower and Banfield, 1974). The type of approximation varies according to the method of analysis, but the symmetric property $d_{ij} = d_{ji}$ is always implicit in the resulting map, even when not a property of the data, i.e. when $d_{ij} \neq d_{ji}$.

The problem of representing general rectangular matrices has had some attention (see, for example, Coombs, 1964; Schonemann, 1970; Gabriel, 1971; Hill, 1974). Rectangular matrices cannot be symmetric, so direct distance interpretations are less attractive although, as we shall see below, are still sometimes available. Commonly, symmetric matrices are derived from a rectangular matrix \mathbf{A} giving either row by row (e.g. \mathbf{AA}') or column by column (e.g. $\mathbf{A}'\mathbf{A}$) comparisons, or both (see below). Methods available for rectangular matrices are applicable, and sometimes useful, but special problems arise, especially when the rows and columns are classified by the same entities, or variants of the same entities. Asymmetric relationships between pairs of items are not uncommon. A simple example might be the time it takes to walk from one place to another; if at different altitudes, this time will be less for the downhill journey than for the uphill journey. More serious examples can arise for example with confusion matrices (giving the number of times in a series of trials that item i is said to be the same as item j , taken in that order) or in diallel cross-experiments (giving the number of progeny or yield when a male of line i is crossed with a female of line j) or the number of people who live in place i and carry out some activity in place j , or the frequency with which journal i cites journal j and so on. Square matrices arising from these and similar examples of asymmetric relationships have their rows classified by a set of entities in one mode and their columns classified by the same entities but in a different mode.

Sometimes, as with the diallel cross-experiment, a specific model is available and it is then appropriate to fit that model, estimating its parameters in accordance with accepted statistical procedures. Usually no model is readily available, and what is wanted is some data-analytical tool that allows the data to be exhibited for preliminary investigation in digestible form, analogous to a scatter-plot. As is demonstrated below, such plots of the data may suggest parametric models that permit a more formal analysis. With matrices, the usual multivariate problem arises, that the potential scatter of the data occupies too many dimensions for direct representation, so the possibilities of approximate representation in few dimensions have to be explored. A special problem with square matrices with similarly classified rows and columns is that although it often seems reasonable to assume that a direct distance interpretation partially underlies the observed values, the lack of symmetry makes established methods unavailable. A distance interpretation is often so attractive that a matrix \mathbf{D} is not infrequently analysed as $\frac{1}{2}(\mathbf{D} + \mathbf{D}')$, but this ignores departures from symmetry that may be informative.

In the following two sections we illustrate two graphical methods for retaining information on asymmetry: multidimensional unfolding and canonical analysis of skew-symmetry.

The two methods complement one another. In the examples examined below canonical analysis is the more informative, but this is not always so and a contrary example is given by Gower (1977). An important difference between the two methods is that unfolding attempts to display the whole of the given matrix, whereas canonical analysis displays only the skew-symmetric part, leaving the symmetric part for separate analysis.

2. THE USE OF MULTIDIMENSIONAL UNFOLDING TO ANALYSE ASYMMETRY

Multidimensional unfolding is a variant of multidimensional scaling, or ordination, methods. Specifically given a matrix $\mathbf{D}(p \times q)$ of distances, it is required to find matrices $\mathbf{X}(p \times r)$ and $\mathbf{Y}(q \times r)$ giving coordinates of two sets of points in some specified number (r) of dimensions that reproduce \mathbf{D} , or approximate \mathbf{D} optimally. Thus the problem is one of drawing maps in r dimensions, given only a $p \times q$ rectangular portion of a complete $(p+q) \times (p+q)$ distance matrix.

Schonemann (1970) has suggested a least-squares solution to the problem of fitting the distances generated by \mathbf{X}, \mathbf{Y} to the given distances \mathbf{D} . However, all the results discussed below have been obtained by "non-metric methods" for which there are several widely available computer programs. Such programs seek to fit in a Euclidean space a set of points generating

distances d_{ij}^* that try to reproduce the rank orderings of the elements d_{ij} of \mathbf{D} (which need not be a distance-matrix). The goodness of fit is defined as some function that measures the monotonicity of the relationship between d_{ij} and d_{ij}^* and which is invariant to monotonic transformations of the d_{ij} . The form of this function differs from program to program. We used the program MDSCAL where the goodness of fit function is termed *stress* and in a modified residual sum of squares for the monotonic regression of d_{ij} plotted against d_{ij}^* (Kruskal, 1964). In unfolding, the program operates as if the whole symmetric distance matrix were available, but computes stress by operating only on the pq pairs (i, j) actually present in the data.

When \mathbf{D} is square, with its rows and columns similarly classified, multidimensional unfolding will give p points representing the row items and p points representing the column items. Interpretation of the data lies mainly in investigating the distances *between* the two sets, and there is only a secondary interest in the within-set distances. Because row and column items are represented by separate points, both row-column and column-row distances are portrayed. Thus unfolding attempts to display the whole matrix \mathbf{D} .

As a first example of multidimensional unfolding consider Table 1 which gives the distances between eight English towns. The table is not symmetric because a simple skew-symmetric matrix of form $\mathbf{1x}' - \mathbf{x1}'$ has been added to the true symmetric distances. Thus a distance d_{ij} becomes $d_{ij} + x_i - x_j$, where $x_i - x_j$ is taken to be about 10 per cent of the distance that town j is west of town i , so the contamination is worst for towns which differ most in longitude.

TABLE 1

Distance-matrix contaminated by skew-symmetric matrix of form $\mathbf{x1} - \mathbf{1x}'$

	K	M	N	O	P	S	T	Y
Kendal	0	70	230	206	419	270	270	80
Manchester	74	0	172	138	354	202	205	60
Norwich	258	196	0	147	412	198	262	186
Oxford	216	146	131	0	264	65	115	172
Penzance	401	332	366	236	0	203	131	382
Southampton	280	210	180	65	231	0	93	235
Taunton	270	201	234	103	149	81	0	251
York	92	68	170	172	412	237	263	0

Fig. 1 is a two-dimensional unfolding of Table 1 in which the zero diagonals have been ignored. The stress of 4.65 per cent indicates an excellent fit. Each town is represented twice, once as a row-entity and once as a column-entity. Kendal, Manchester, York, Norwich and Oxford form near coincident pairs but moving westwards from Southampton through Taunton to Penzance the pairs increasingly diverge. This divergence accommodates the asymmetry of Table 1 with, for example, Penzance-Norwich being closer than Norwich-Penzance. If the zero diagonal had been included then all towns would have occurred in nearly coincident pairs. This is because the requirement of zero distance between a town and itself is a strong constraint that dilutes, but does not totally obscure, asymmetrical representation.

The above example may seem unrealistic but it exemplifies the simplest form of departure from asymmetry and one that not infrequently occurs. Gower (1977) has discussed (i) a model in which flight times between pairs of cities are modified by the effects of a prevailing wind and (ii) a matrix of first passage times, both of which exhibit this simple type of departure from symmetry. A further example is discussed below.

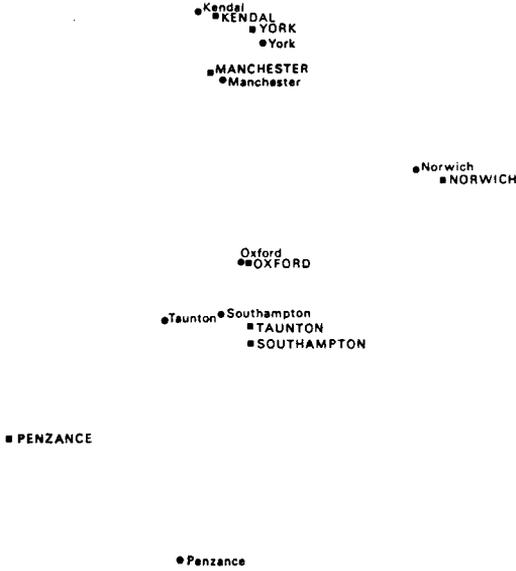


FIG. 1. Unfolding of contaminated distance-matrix of Table 1, obtained by MDSCAL (stress = 4.65 per cent). ■ YORK, column-name; ● York, row-name.



FIG. 2. Unfolding of response reduction caused by cross-adaptation of seven odours (see Table 2(a)), obtained by MDSCAL (stress = 21.19 per cent). ■ CITRAL, test stimulus; ● Citral, adapting stimulus.

As an example with real data, consider Table 2 reproduced from Köster (1971).

TABLE 2
Amounts of response reduction caused by cross-adaptation (Köster, 1971)

Table 2(a)

<i>Adapting stimuli</i>	<i>Test stimuli</i>						
	1	2	3	4	5	6	7
1. Citral	—	12.0	4.1	10.3	16.0	20.0	11.4
2. Cyclopentanone	9.5	—	15.7	9.6	8.5	14.4	11.8
3. Benzyl acetate	2.4	11.0	—	8.3	21.1	17.4	15.4
4. Safrole	13.3	6.9	3.1	—	4.9	18.5	9.7
5. Butyl acetate	10.2	5.1	19.1	6.6	—	17.7	12.5
6. <i>m</i> -Xylene	6.6	5.6	7.2	8.6	19.1	—	19.5
7. Methyl salicylate	5.5	4.5	12.4	4.3	18.5	11.6	—

Table 2(b)

<i>Adapting stimuli</i>	<i>Test stimuli</i>						
	1	2	3	4	5	6	7
1. Dioxan	—	14.2	8.1	8.3	23.7	17.5	28.6
2. Cyclopentanone	5.5	—	17.9	8.9	6.0	14.6	14.3
3. Cyclohexanone	4.1	17.7	—	4.5	10.6	22.3	19.7
4. <i>a</i> -Thujone	7.1	7.2	1.9	—	13.7	13.9	20.4
5. <i>m</i> -Xylene	13.1	4.4	9.2	9.3	—	11.6	10.4
6. iso-Propanol	3.7	12.5	10.1	15.6	11.4	—	13.8
7. Cyclopentanol	8.5	4.5	8.2	0.5	7.4	11.9	—

This table gives a measure of the effect of one odour on the ability of a subject to detect a second odour. In the absence of the initial odour (termed the adapting stimulus) some proportion, p , of the subjects would be able to detect the second odour (test stimulus). p is not necessarily unity because not all people can detect all odours. However, if the adapting stimulus is strong then this will have a paralysing or delaying effect and after a specified time (30 s for these data) only a proportion p' of subjects can detect the test stimulus. Tables 2(a) and 2(b) give the differences $p - p'$ multiplied by 100. Gross asymmetry can arise when a weak adapting stimulus is followed by a strong test stimulus, in which case p and p' are likely to be similar in value, whereas when taken in the reverse order p' may be greatly reduced from p . Such results can be seen in Table 2 (e.g. Table 2(b), for stimuli 4 and 7). When the adapting and test stimuli are the same, the differences $p - p'$ (not shown in Table 2) tend to be big. This is because it is generally easier to detect a change in odour than to detect the continuance of the same odour. This fact suggests that one should not necessarily seek, or expect, an unfolding in which like pairs of points are necessarily close.

Fig. 2 gives a two-dimensional multidimensional unfolding of Table 2(a). The two-dimensional fit with stress of 21.2 is not especially good. The figure suggests that cyclopentanone, benzyl acetate, methyl salicylate and butyl acetate have similar effects as adapting stimuli. The large asymmetry in response of citral, benzyl acetate and safrole with *m*-xylene comes out clearly. The representation is a faithful one but gives no special insight into interpreting the data. Perhaps a better chemical understanding of these seven substances would help.

Similarly the unfolding of Table 2(b) is illustrated in Fig. 3. The fit is good with a stress of 11.4. From Fig. 3 we see that the pairs of points for dioxan, cyclopentanone, *a*-thujone and cyclohexanone form mirror-image sets for test and adapting stimuli and hence their intra-distances are roughly symmetric. The remaining odours (*m*-xylene, iso-propanol, cyclopentanol) are represented by less regularly disposed pairs of co-ordinates but, on examination, their intra-distances are again seen to be roughly symmetric. The inter-distances between the two sets are asymmetric. This is a fair representation of Table 2(b). The longest distances between dioxan and cyclopentanol and between dioxan and *m*-xylene are clear, together with the relatively small distances of their inversions. However, the large distance between cyclohexanone and iso-propanol is less well represented. It seems to us that the graphical display of the unfolding has not revealed the structure of Table 2(b) clearly although the main features of the table seem to have been preserved.

Unfolding can give a very informative graphical display of asymmetry (see Gower, 1977, for an example). However, the above examples have been a little disappointing in that

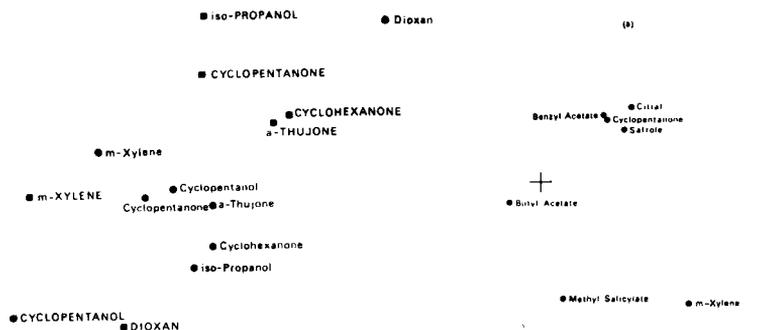


Fig. 3. Unfolding of response reduction caused by cross-adaptation of seven odours (see Table 2(b)) obtained by MDSCAL (stress = 11.37 per cent). ■ DIOXAN, test stimulus; ● Dioxan, adapting stimulus.

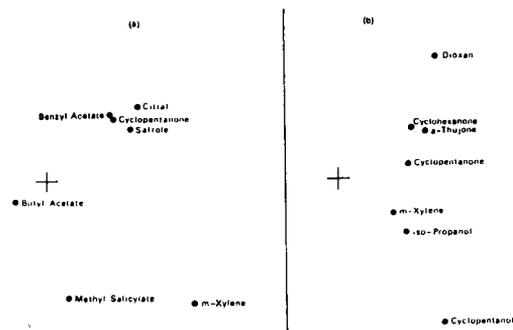


Fig. 4. Canonical analysis of the skew-symmetric parts of Tables 2(a) and 2(b).

although the data are fitted reasonably well, little insight is given into the nature of the asymmetry, even when of known simple form as in Table 1. Clearly the method is better at exhibiting some forms of asymmetry than others. A mathematically good graphical approximation to data need not be a useful one, and other methods of display are worth considering.

3. CANONICAL ANALYSIS OF SKEW-SYMMETRY

Any square matrix \mathbf{D} may be expressed in the form $\mathbf{D} = \mathbf{M} + \mathbf{N}$, where \mathbf{M} is symmetric and \mathbf{N} skew-symmetric. This decomposition gives an orthogonal breakdown of sums of squares:

$$\sum_{i,j} d_{ij}^2 = \sum_{i,j} m_{ij}^2 + \sum_{i,j} n_{ij}^2 \quad (1)$$

which suggests the separate analysis of \mathbf{M} and \mathbf{N} . The symmetric part may be analysed to give a map, but what can be done with the skew-symmetric part? The canonical analysis of a skew-symmetric matrix \mathbf{N} rests on its canonical decomposition into a sum of elementary rank 2 skew-symmetric matrices. Specifically, we have

$$\mathbf{N} = \sum_1^{[p/2]} \lambda_i (\mathbf{u}_i \mathbf{v}_i' - \mathbf{v}_i \mathbf{u}_i'), \quad (2)$$

where all the vectors \mathbf{u}_i and \mathbf{v}_i may be taken to be the columns of an orthogonal matrix, and the values of λ_i are positive and may be taken in decreasing order. When p is odd $[p/2]$ is the greatest integer less than $p/2$. This is the special form taken by the singular value decomposition for skew-symmetric matrices, and hence, by the Eckart-Young theorem (1936), r terms of (2) give the best least-squares fit of rank $2r$ to the skew-symmetric matrix \mathbf{N} .

A consequence is that given \mathbf{N} , any singular value decomposition program (e.g. Golub and Reinsch, 1970) can be used to calculate $\lambda_i, \mathbf{u}_i, \mathbf{v}_i$. Failing that, an eigenvalue program operating on the symmetric matrix $\mathbf{N}\mathbf{N}' = -\mathbf{N}^2$ will give pairs of equal roots λ_i^2 with corresponding pairs of vectors \mathbf{u}_i and \mathbf{v}_i . Thus calculation presents no problem.

In the following we shall be concerned only with the case $r = 1$, so we have

$$\mathbf{N} \doteq \lambda(\mathbf{u}\mathbf{v}' - \mathbf{v}\mathbf{u}'), \quad (3)$$

where we have been able to drop the suffices. By plotting the vectors \mathbf{u}, \mathbf{v} as p points $P_i(u_i, v_i)$ relative to two orthogonal axes, just as in principal components analysis, we obtain a two-dimensional configuration. However, geometrical interpretation of this configuration needs some care.

Suppose $P_1(u_1, v_1)$ and $P_2(u_2, v_2)$ are two points in this representation, then the term n_{12} fitted by (3) is $(u_1 v_2 - v_1 u_2)$ which is twice the area of the triangle subtended at the origin, O , by P_1 and P_2 . The skew-symmetry in the representation arises because the areas of the triangles $OP_1 P_2$ and $OP_2 P_1$ are equal but opposite in sign. Thus the configuration of points has to be interpreted not in terms of distances but in terms of areas. Pairs of points lying on a line through the origin as well as coincident pairs of points will give zero areas. Triangles on the same base which includes a vertex at the origin and with the other vertex on a line parallel to the base are equal in area, so that parallel lines play an important part in interpretation.

The case when (3) is in the form

$$\mathbf{N} \doteq \lambda(\mathbf{w}\mathbf{1}' - \mathbf{1}\mathbf{w}') \quad (4)$$

is of special importance as the set of coordinates corresponding to the vector $\mathbf{1}$ is constant. Hence the set of all points in the two-dimensional configuration are co-linear, so that the areas of all triangles are proportional to the lengths of their bases. Thus

$$n_{ij} \doteq \lambda(w_i - w_j). \quad (5)$$

This linear model is the simplest form of skew-symmetry.

Let us now examine the data of Section 2 in the light of their canonical analyses. Because we know how Table 1 was constructed, there is little point in presenting its canonical analysis. The symmetric part of the matrix will give an accurate map of the eight towns. The skew-symmetric part will have rank 2 and can be represented as (4), or its equivalent (5), where the vector w exactly recovers the vector x representing the deliberate contamination. A minor difficulty is that the form (3), and hence also (4), is not unique, for u and v may be replaced by $v \sin \theta + u \cos \theta$ and $v \cos \theta - u \sin \theta$. When plotted, the two solutions differ only in rotation, and if the plot is linear we know that the solution is of type (4). To obtain the form (4), axes must be rotated so that one of them is parallel to and the other is at right angles to the plotted line.

The canonical analysis of the skew-symmetric parts of Table 2(a) and 2(b) is shown in Fig. 4. The first two dimensions account for 76.6 per cent and 87.0 per cent of the total sum of squares of the skew-symmetric elements of Tables 2(a) and 2(b), respectively, so in both cases the fit is quite good. From Fig. 4(a) we see immediately that (citral, cyclopentanone, benzyl acetate, safrole and butyl acetate) are represented by points that, together with the origins, are nearly co-linear. This correctly indicates the general symmetry of Table 2(a) for these substances. Further, the gross asymmetry of *m*-xylene and methyl salicylate both with each other and with all the other substances, but less so for butyl acetate, is revealed. This ordering of the substances in Table 2(a) has been chosen to exhibit this interpretation and it is clear that the description is a good one. The two-dimensional approximation cannot avoid some inaccuracies, of which perhaps the worst are an exaggerated asymmetry between *m*-xylene and butyl acetate and an exaggerated symmetry between citral and butyl acetate. The unfolding previously discussed attempts to portray the size of the symmetric relationships as well as the departures from symmetry, and so, not surprisingly, it has failed to reveal the latter as well as has the canonical analysis.

Fig. 4(b) shows a near linear representation for the seven substances of Table 2(b). Hence the skew-symmetry can be approximated by the formulae (4) and (5). Although the vector w may be read directly from Fig. 4(b) as projections on a line fitted to the seven points, it is easier and better to obtain least-squares estimates, which are merely the row (or column) averages of the skew-symmetric matrix. These give $w = (4.2, 0.5, 1.7, 1.2, -1.1, -1.8, -4.7)$. These values of w order the substances as in Fig. 4(b) and it can be checked that inserting them into (5) approximates the skew-symmetric part of Table 2(b), especially for the higher values.

4. CONCLUSION

The above has concentrated on graphical displays and has demonstrated two ways of representing asymmetric square matrices; Gower (1977) outlines some others. Before using these methods the possibility of transforming the elements of any or all of the matrices D , M and N should be considered. For example, if it is thought that the asymmetric matrix D may be formed from a symmetric matrix S by multiplying the *i*th row by a constant c_i then we have $d_{ij} = c_i s_{ij}$ and transforming to logs gives $n_{ij} = \log d_{ij} - \log d_{ji} = \log c_i - \log c_j$, a form that is easily picked up by the canonical analysis of N . Similar transformations will often suggest themselves and it might sometimes be advantageous to combine the elements of these matrices in different ways, e.g. by operating on the skew matrix with elements n_{ij}/m_{ij} .

Although the canonical analysis separates the symmetric and skew-symmetric parts of D , their graphical displays can sometimes be combined in informative ways. This is especially true when N has form (5) because the values w_i may be superimposed on a map produced by symmetric scaling or ordination, and interpreted as an extra (usually third) dimension, to which contours can sometimes be fitted. Gower (1977) discusses such an example and other more complicated ways in which asymmetry may be combined with symmetry. A particularly simple, but useful, example of contouring is discussed by Banfield and Gower (1978).

A closer look is needed at various forms and models of asymmetry and how these are represented by unfolding and by canonical analysis. Finally, it would be nice to know something about the distributions of the singular values of skew-symmetric matrices arising from plausible null models.

REFERENCES

- BANFIELD, C. F. and GOWER, J. C. (1974). Goodness-of-fit criteria for hierarchical classification and their empirical distributions. *Proc. 8th Int. Biometric Conf.* (L. C. A. Corsten and T. Postelnicu, eds), pp. 347–361.
- (1978). A note on the graphical representation of multivariate binary data. *Biometrics* (in press).
- COOMBS, G. H. (1964). *A Theory of Data*. New York: Wiley.
- ECKART, C. and YOUNG, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, **1**, 211–218.
- GABRIEL, K. R. (1971). The biplot graphic display of matrices with applications to principal components. *Biometrika*, **58**, 453–467.
- GOLUB, G. H. and REINSCH, C. (1970). Handbook series linear algebra: Singular value decomposition and least squares solutions. *Numer. Math.*, **14**, 403–420.
- GOWER, J. C. (1977). The analysis of asymmetry and orthogonality. In *Recent Developments in Statistics*, (J. Barra *et al.* eds). Amsterdam: North Holland, p.p. 109–123.
- HILL, M. O. (1974). Correspondence analysis: A neglected multivariate method. *Appl. Statist.*, **23**, 340–354.
- HOLMAN, E. W. (1972). The relation between hierarchical and Euclidean models for psychological distances. *Psychometrika*, **37**, 417–423.
- KÖSTER, E. P. (1971). *Adaptation and Cross-adaptation in Olfaction*. Thesis, published by the University of Utrecht.
- KRUSKAL, J. B. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, **29**, 115–129.
- SCHÖNEMANN, P. H. (1970). On metric multidimension unfolding. *Psychometrika*, **35**, 349–366.