RR 92-08

SOME REMARKS ON ORTHOGONALITY IN THE ANALYSIS OF ASYMMETRY

John C. Gower

Berrie Zielman

Department of Data Theory

University of Leiden

SOME REMARKS ON ORTHOGONALITY IN THE ANALYSIS OF ASYMMETRY

J. C. Gower and B. Zielman
Department of Data Theory
University of Leiden, Netherlands

Summary.

Some basic results concerning skew-symmetric and orthogonal matrices and the relationships between them are brought together and their relevance for data-analysis is discussed. Given a square matrix **A**, particular emphasis is given to least-squares approaches, in which one seeks a general symmetric matrix **M** and some constrained matrix **C** which satisfy (i) **A** = **M** + **C** and (ii) **A** = **MC**. Typical constraints on **C** are low rank, orthogonality and low-rank departures from a unit matrix. The consequences for graphical representation are discussed

1. Introduction

Gower (1977) discussed a general approach to the analysis of a square asymmetric matrix **A** of order n, which was based on the well-known decomposition

$$\mathbf{A} = \mathbf{M} + \mathbf{N} \tag{1}$$

where $\mathbf{M} = \frac{1}{2}(\mathbf{A} + \mathbf{A}')$ is symmetric and $\mathbf{N} = \frac{1}{2}(\mathbf{A} - \mathbf{A}')$ is skew-symmetric. Writing $\|\mathbf{A}\|$ for the sum-of-squares of the matrix **A**, the decomposition (1) has the useful property that :

$$\|\mathbf{A}\| = \|\mathbf{M}\| + \|\mathbf{N}\| \tag{2}$$

showing that the total sum-of-squares may be partitioned into independent parts attributable to symmetry and skew-symmetry. It follows that, within the context of least-squares, symmetry and skew-symmetry may be analysed separately. One may be interested in approximations to **M** and **N** that are of some constrained form. The most simple constraints on approximation are those of reduced rank, which, in the context of least-squares, lead immediately to consideration of the Eckart-Young theorem (1936). The relationship between rank and geometrical dimensionality allows approximations to be represented by configurations of points and this is especially convenient when the rank is two, in which case the configurations are two-dimensional. The Eckart-Young Theorem requires the singular value decomposition of the matrix being approximated. Some of the basic canonical properties of skew-symmetric matrices are derived in an appendix. Here, the decompositions of **M** and **N** will be written:

$$\mathbf{M} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{K}\mathbf{U}' \qquad \text{and} \qquad \mathbf{N} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{J}\mathbf{V}' \tag{3}$$

where **U** and **V** are orthogonal matrices, and $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$ are non-negative matrices of singular values, assumed arranged in non-increasing order along the diagonal. The matrices **K** and **J** are essentially matrices of signs associated with the columns of **U** and **V**, respectively. **K** is merely a diagonal matrix of entries plus one or minus one, where the negative values are needed when **M** is not positive semi-definite. The eigenvalues of **M** are given by the diagonal values of $\boldsymbol{\Lambda}\mathbf{K}$; for any zero eigenvalue, the corresponding value in **K** is immaterial but is conventionally set to one. The form of **N** is a little more complicated, because the singular values of a skew-symmetric matrix occur in equal pairs and so $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \sigma_1, \sigma_2, \sigma_2, \ldots)$. Corresponding to each pair, the matrix **J** has a 2×2 skew-symmetric orthogonal diagonal block of the form :

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{4}$$

When the order of **N** is odd, then the last diagonal element of $\boldsymbol{\Sigma}$ must be zero and the corresponding value of **J** is conventionally set to one; the convention is adopted that for any zero pairs of singular values of **N**, the corresponding blocks of **J** will be assumed to be set to the form (4). The settings of **J** and **K** that correspond to zero singular values are essentially arbitrary but the convention adopted here ensures that $\mathbf{J}\mathbf{V}'$ and $\mathbf{K}\mathbf{U}'$ remain orthogonal matrices so that in both cases (3) gives singular value decompositions with full orthogonal matrices, even in cases of deficient rank.

From (3) it is clear that the singular value decompositions of symmetric and skew symmetric matrices share the property of having the same pre- and post-vectors, apart from possible changes of sign and, in the case of skew-symmetry, permutation. In both cases, the rows of the singular vector matrices, $\mathbf{U}\boldsymbol{\Lambda}^{1/2}$ and $\mathbf{V}\boldsymbol{\Sigma}^{1/2}$, may be plotted in as many dimensions as required and as justified by the adequacy of the retained singular values expressed as a proportion of the sum-of-squares of all the singular values. However, the forms of **K** and **J** have a fundamental impact on how such diagrams are to be interpreted. When **M** is positive semi-definite, the singular value and eigenvalue decompositions coincide and the familiar inner-product $\mathbf{u}_i\boldsymbol{\Lambda}\mathbf{u}_j'$, essentially the cosine rule of elementary trigonometry, validly interprets the interaction between the ith and jth rows of **U**. When **K** is not positive semi-definite the inner product is replaced by $\mathbf{u}_i\boldsymbol{\Lambda}\mathbf{K}\mathbf{u}_j'$ and if "negative" dimensions are included, one has to remember to subtract their contributions; such representations are non-Euclidean. With skew-symmetry, things are more complicated and, as explained by Gower (1977), interpretation is always non-Euclidean and the interaction between two rows of **V** is in terms of areas of triangles formed from the points with coordinates $\mathbf{v}_i$, $\mathbf{v}_j$ and the origin. Furthermore these areas have to be evaluated in sets of two-dimensional spaces that have been termed bimensions. The decreasing pairs of singular values impose a natural ordering on the bimensions in decreasing order of importance; a single bimension, i.e. a single two-dimensional space, may suffice for adequate approximation. When two dimensions are not sufficient, then the areas in two, or more, bimensions have to be summed.

Although the usual restriction on approximating **M** and **N** is that they should have low ranks, other restrictions on the approximations are permissible. For example, when **A** has a zero diagonal, so then does **M** which may then be approximated by some form of distance matrix, usually Euclidean, that is generated by points in a low-dimensional space. An approximation to **N** that is of special importance is when it can be written as the linear form $\mathbf{1n}' - \mathbf{n1}'$, which is a special case of a rank two skew-symmetric matrix that is often termed one-dimensional skew-symmetry because the points plotted from the first two columns of **V** are collinear. The least-squares estimate of **n** is the vector of row-means of **N**.

1.1. Notation

In the above, diagonal block matrices have already been required. Matrices of these forms are frequent in the following and it is convenient to represent them by a special notation. We shall write:

$$\mathbf{B} = \left\{ \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix}, p, q, [r] \right\}$$

to denote a sequence of 2×2 diagonal blocks, followed by a sequence of scalars p and then a series of scalars q. Any, but not all, of these block-types may be null. The value [r] denotes a scalar r that appears in the final position when n is odd. Thus, the block-diagonal associated with **N** is $\mathbf{\Sigma J} = \left\{ \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}, [0] \right\}$. The number of such blocks, including the unit blocks for scalars, is denoted by m. The ith block will be written $\mathbf{B}_i^*$ and as $\mathbf{B}_i$ when it is augmented by units along the diagonal. Thus:

$$\mathbf{B}_i^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_i & b_i & \cdots & 0 \\ 0 & 0 & \cdots & c_i & d_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \qquad \mathbf{B}_i = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_i & b_i & \cdots & 0 \\ 0 & 0 & \cdots & c_i & d_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}$$

where it is to be understood that a scalar p, q or r, may be written in place of the 2×2 block. The following results are immediate:

$$\mathbf{B} = \sum_{i=1}^{m} \mathbf{B}_i^* \qquad \text{and} \qquad \mathbf{B} = \prod_{i=1}^{m} \mathbf{B}_i = \sum_{i=1}^{m} \mathbf{B}_i - (m-1)\mathbf{I}. \qquad (5)$$

2. Skew-symmetry and Orthogonality

In the middle of the nineteenth century, Cayley showed that if $\mathbf{N}$ is skew then

$$\mathbf{Q} = (\mathbf{I} - \mathbf{N})(\mathbf{I} + \mathbf{N})^{-1}$$

is orthogonal. Also if $\mathbf{Q}$ is orthogonal, then

$$\mathbf{N} = (\mathbf{I} + \mathbf{Q})^{-1}(\mathbf{I} - \mathbf{Q})$$

is skew, provided the inverse exists. Because sign does not affect the orthogonality and skewness properties, the signs associated with both $\mathbf{Q}$ and $\mathbf{N}$ may be interchanged; the forms chosen here are mutually consisitent. These formulae are useful for expressing the n^2 dependent parameters of an orthogonal matrix in terms of the $\frac{1}{2}n(n-1)$ independent parameters of a skew-symmetric matrix and have found applications in applied mathematics and in integrations over the elements of orthogonal matrices. Substituting the singular value decomposition of $\mathbf{N}$ into Cayley's formula for $\mathbf{Q}$ gives

$$\mathbf{Q} = \mathbf{V}\mathbf{H}\mathbf{V}' \qquad (6)$$

where $\mathbf{H} = \left\{ \begin{pmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{pmatrix}, [1] \right\}$ with $\cos\theta_i = (1-\sigma_i^2)/(1+\sigma_i^2)$ and $\sin\theta_i = -2\sigma_i/(1+\sigma_i^2)$

and where the final [1] is necessary when $\mathbf{N}$ has odd order. Clearly $\mathbf{H}$ is itself an orthogonal matrix. Notice that if $\sigma_i > \sigma_j$ then $\cos\theta_i > \cos\theta_j$. The decomposition (6) may seem somewhat circular as it expresses the orthogonal matrix $\mathbf{Q}$ in terms of the product of three other orthogonal matrices. However one of these is a transpose of another and the third, $\mathbf{H}$, has a very special form. Indeed $\mathbf{H}$ may be interpreted in terms of matrices $\mathbf{H}_i$ giving rotations through an angle θ_i in the plane determined by the vectors $\mathbf{v}_{(2i-1)}$ and $\mathbf{v}_{(2i)}$. These planes of rotation have strong links with the bimensions of skew-symmetry. When $\sigma_i = 0$, $\mathbf{H}_i$ becomes a unit matrix and has no effect but as $\sigma_i \to \infty$, $\mathbf{H}_i$ has two values of -1 on its diagonal which correspond to two reflections rather than a single rotation. These two reflections are in the planes whose normals are $\mathbf{v}_{(2i-1)}$ and $\mathbf{v}_{(2i)}$, respectively. Of course, two reflections are equivalent to a rotation through π but (6) cannot accommodate this possibility for any finite σ_i. Thus, Cayley's formulae do not embrace all orthogonal matrices.

Cayley's formulae are deceptively simple. Doubling $\mathbf{N}$ induces a complicated change in $\mathbf{Q}$. This effect is easily parameterised by noting the generalization

$$\mathbf{Q} = (k\mathbf{I} - \mathbf{N})(k\mathbf{I} + \mathbf{N})^{-1}$$

$$\mathbf{N} = k(\mathbf{I} + \mathbf{Q})^{-1}(\mathbf{I} - \mathbf{Q})$$

where k is an arbitrary scalar. This form shows that the symmetry of Cayley's form is a consequence of choosing k = 1. Thus for given **N** a whole range may be obtained of related orthogonal matrices **Q.** Choosing different values of k leaves the bimensions **V** invariant (planes of rotation) but influences the effects of the scaling of **N** and the angles of rotation in **Q** . The latter now become $\cos\theta_i = (k^2 - \sigma_i^2)/(k^2 + \sigma_i^2)$ and $\sin\theta_i = -2k\sigma_i/(k^2 + \sigma_i^2)$. The question is open as to whether there are any advantageous choices, such as $k = \sigma_1$ or $k^2 = \|\mathbf{N}\|$. Note that the possible singularity of **I** + **Q** remains a difficulty.

Equation (6) is very close to a classical decomposition of any orthogonal matrix **Q**. The classical result (see Theorem 4 of the appendix) is that for all orthogonal **Q** there exist orthogonal matrices **W** and **H** such that

$$\mathbf{Q} = \mathbf{WHW'} \tag{7}$$

where $\mathbf{H} = \left\{\begin{pmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{pmatrix}, -1, 1, [1]\right\}$. The continued use of **H** in (7) is justified because it contains the **H** of (6) as a minor special case. Now, as well as the previous plane rotations, there are reflections in the planes normal to any vector that corresponds to -1 on the diagonal of **H**. Thus, using (5), (7) allows **Q** to be interpreted as a product of elementary orthogonal transformations:

$$\mathbf{Q} = \mathbf{WHW'} = \prod_{i=1}^{m}(\mathbf{WH_iW'}). \tag{8}$$

Also from (5), **Q** may be expressed as a sum:

$$\mathbf{Q} = \sum_{i=1}^{m}(\mathbf{WH_iW'}) - (m - 1)\mathbf{I}. \tag{9}$$

Selecting an arbitrary orthogonal matrix **R** then $\mathbf{Q} = \mathbf{RI(R'Q)}$, which is in the standard form of a singular value decomposition with all singular values equal to unity. Thus, the singular values of an orthogonal matrix are all unity and there is no question of finding *any* general best least-squares approximation to **Q** that is of lower rank, but the possibility of finding an approximation matrix that has some special form is not precluded. For example, there might be interest in determining **P**, the best plane rotation that approximates **Q**. This requires **P** to be chosen to maximise Trace(**QP'**). Indeed, from (7) it is clear that that if one of the elementary rotations $\mathbf{H}_k$ is selected, then $\mathbf{WHH}_k'\mathbf{W}'$ is the same as **Q** except that in the diagonal block corresponding to $\mathbf{H}_k$, the two diagonal values of $\cos\theta_k$ are replaced by units and the off-diagonal values by zeroes, so that $\text{Trace}(\mathbf{QWH}_k\mathbf{W}') = \text{Trace}(\mathbf{Q}) + 2(1 - \cos\theta_k)$. This trace is maximised by choosing the block k with the smallest value of $\cos\theta_k$, and hence the maximum angle of rotation. Thus, setting $\mathbf{P} = \mathbf{WH}_k'\mathbf{W}'$, gives the plane rotation, chosen from all $\mathbf{H}_j$ included in **H**, that best approximates **Q**. This is a special case of the general result, proved in §4, that $\mathbf{H}_k$ generates the best of all possible plane rotations.

3. Multiplicative Transformations to Symmetry

One of the ways of arriving at the additive decomposition outlined in §1, is to ask what unit rank matrix $\mathbf{ab}'$, when added to $\mathbf{A}$ gives the greatest symmetry. This requires the minimum of $\|(\mathbf{A} - \mathbf{ab}') - (\mathbf{A} - \mathbf{ab}')'\|$, so we have to minimise $\|\mathbf{N} - \frac{1}{2}(\mathbf{ab}' - \mathbf{ba}')\|$ which immediately leads to the singular-value decomposition of $\mathbf{N}$ and is strongly suggestive of its rank-2 bimension properties.

Similarly we could ask what unit-rank matrix would give the greatest symmetry when post multiplying $\mathbf{A}$. This requires $\mathbf{Aab}' = \mathbf{M}$, to be as symmetric as possible. When $\mathbf{A}$ is of full rank, this has the solution of choosing $\mathbf{M}$ to be an arbitrary unit-rank symmetric matrix $\mathbf{cc}'$ and setting $\mathbf{a} = \mathbf{A}^{-1}\mathbf{c}$ and $\mathbf{b} = \mathbf{c}$, so exact symmetry is trivially obtained. Further, if we try the approximation $\mathbf{A} = \mathbf{Mab}'$, this constrains $\mathbf{A}$ to be of unit rank, which is rarely acceptable. This approach must be modified in some way if it is to lead to anything useful. Clearly some constraint must be put on permissible post-multiplying matrices. Levin and Brown (1979) choose a pre-multiplying diagonal matrix $\mathbf{V}$ such that $\mathbf{VA}$ is as symmetric as is possible. However this constraint is not in itself sufficient to avoid trivial solutions, so $\mathbf{V}$ is further constrained either (i) so that $\mathbf{1}'\mathbf{V}\mathbf{1} = 1$, or (ii) so that $\|\mathbf{VA}\| = \|\mathbf{A}\|$. The motivation for (i) is that when $\mathbf{A}$ is assumed to be a conditional proximity matrix then the vector $\mathbf{V1}$ may be regarded as estimating probabilities for the row-items; note that this interpretation seems to require the additional constraint that the elements of $\mathbf{V1}$ be non-negative. The constraint (ii) preserves the size of $\mathbf{A}$ but note that if $\mathbf{VA} = \mathbf{M}$ then $\mathbf{A} = \mathbf{V}^{-1}\mathbf{M}$ and $\|\mathbf{A}\| \neq \|\mathbf{V}^{-1}\mathbf{M}\|$. The constraints have a major effect on the analytical form of the solution. As usual, the linear constraint (i) leads to a matrix inversion and the quadratic constraint (ii) leads to a symmetric eigenvalue problem. If $\mathbf{V}$ were a post-multiplying diagonal matrix, a second set of solutions would be obtained.

Another possibility, that is explored in the following, is based on the polar decomposition of a matrix $\mathbf{A} = \mathbf{M}_1\mathbf{Q} = \mathbf{Q}\mathbf{M}_2$, where $\mathbf{M}_1$ and $\mathbf{M}_2$ are symmetric and $\mathbf{Q}$ is orthogonal. The polar decomposition is a classical result (see e.g. Gantmacher, 1959) which follows easily from the singular value decomposition $\mathbf{A} = \mathbf{S\Gamma T}'$ by writing $\mathbf{A} = (\mathbf{S\Gamma S}')\mathbf{ST}' = \mathbf{ST}'(\mathbf{T\Gamma T}')$. Thus $\mathbf{M}_1 = \mathbf{S\Gamma S}'$, $\mathbf{M}_2 = \mathbf{T\Gamma T}'$, both positive semi-definite, and $\mathbf{Q} = \mathbf{ST}'$. An interesting feature of the polar decomposition is that the orthogonal matrix $\mathbf{Q}$ is unchanged whether it occurs as a pre-multiplier or as a postmultiplier. Note that $\|\mathbf{A}\| = \|\mathbf{M}_1\| = \|\mathbf{M}_2\|$ so, like (1), the transformation preserves the total sum-of-squares of **A**.

Another way of arriving at the polar decomposition is to ask what orthogonal matrix **Q**, possibly constrained in some way, makes **AQ'** as symmetric as possible. Thus, in a variant of the orthogonal Procrustes problem, it is required to match **AQ'** to **QA'**. This requires the minimisation of $\|\mathbf{AQ}' - \mathbf{QA}'\|$, which is the same as the maximisation of Trace($\mathbf{AQ}'\mathbf{AQ}'$). Substituting the singular value decomposition of **A** and permuting under the trace operator gives Trace($\mathbf{AQ}'\mathbf{AQ}'$) = Trace($\mathbf{\Gamma T}'\mathbf{Q}'\mathbf{S\Gamma T}'\mathbf{Q}'\mathbf{S}$) = Trace($\mathbf{\Gamma R \Gamma R}$), where $\mathbf{R} = \mathbf{T}'\mathbf{Q}'\mathbf{S}$ is orthogonal. It follows that provided **Q** is not constrained, the maximum of $\sum_{i=1}^{n}\gamma_i^2$ occurs when $\mathbf{R} = \mathbf{I}$, that is $\mathbf{Q} = \mathbf{ST}'$, as before. When **Q** is constrained to a special class of orthogonal matrices $\mathbf{Q_c}$, say, different solutions will apply. The constrained form that minimises $\|\mathbf{AQ}_c' - \mathbf{Q}_c\mathbf{A}'\|$ also minimises $\|\mathbf{Q}_c'\mathbf{A} - \mathbf{A}'\mathbf{Q}_c\|$, showing that the property that the pre- and post- multiplying orthogonal matrices are the same, remains true even for constrained orthogonal solutions. However, this minimisation problem does not seem to be straightforward and in the following, we address the more simple problem of minimising $\|\mathbf{Q} - \mathbf{Q}_c\|$ where $\mathbf{Q} = \mathbf{ST}'$. Yet another approach is to consider the Procrustes problem of matching **S** to **T**, that is finding **Q** which minimises $\|\mathbf{T}' - \mathbf{S}'\mathbf{Q}\|$, which yields an exact fit, again with $\mathbf{Q} = \mathbf{ST}'$.

Clearly, there is a sense in which **Q** embodies departures from symmetry in **A**. When **A** is itself symmetric then, from (3), $\mathbf{Q} = \mathbf{UKU}'$ which becomes a regular unit matrix when **A** is positive semi-definite; when **A** is skew-symmetric then, from the singular value decomposition of **N**, $\mathbf{Q} = \mathbf{VJV}'$, which may be regarded as a unit skew-orthogonal matrix. Thus **Q** is worth considering as an encapsulation of departures from symmetry. Just as **A** - **N** is always symmetric, so is **AQ'** and just as approximations to **N** are useful, so might be approximations to **Q**.

4. The Approximation of an Orthogonal Matrix

As we have seen, all orthogonal matrices have full rank and only unit singular values, so there is no question of approximating **Q** by an orthogonal matrix of lower rank, but as was discussed at the end of §2, **Q** might be approximated by a plane rotation. A rotation through an angle θ in the plane of vectors **p** and **q** may be written as

$$\|(\mathbf{I} - \mathbf{Q}) - \omega_k \mathbf{W}_k \mathbf{L}_k^* \mathbf{W}_k'\| = \| \mathbf{Q} - \mathbf{W}(\mathbf{I} - \mathbf{H}_k^*)\mathbf{W}'\|,$$

where $\mathbf{H}_k = \mathbf{I} - \mathbf{H}_k^*$ is a rotation in the plane defined by $\mathbf{w}_{(2k-1)}$ and $\mathbf{w}_{(2k)}$ and is a component of the canonical decomposition (7). Notice the link between planes of rotation and the bimensions associated with skew-symmetry. Thus, (10) defines the plane rotation $\mathbf{H}_k$ which minimises $\|\mathbf{Q} - \mathbf{W}\mathbf{H}_k\mathbf{W}'\|$. The best unrestricted rank-two approximation to $\mathbf{I} - \mathbf{Q}$ has been shown to generate a plane rotation, so that better cannot be done by considering general non-orthogonal linear transformations in the plane (see Constantine and Gower, 1978, for a discussion of plane transformations including elementary orthogonal transformations). Thus the best plane rotation deriving from the canonical decomposition (7) of $\mathbf{Q}$, as described in §2, is proved to be the best possible of all plane rotations.

Now, consider the case where the target matrix for $\mathbf{Q}$ is $\mathbf{WKW}'$, which implies that the target for $\mathbf{H}$ is $\mathbf{K}$ rather than $\mathbf{I}$. Clearly any values of plus or minus one on the diagonal of $\mathbf{H}$ should be transferred to the same position in $\mathbf{K}$. By so-doing, all reflections are excluded from further consideration. When no 2×2 blocks remain there is nothing further to do. Thus we may assume that there is at least one 2×2 block and we must determine whether to match its diagonals to plus one or to minus one. The former case is as discussed above and leads to a singular value of $2\sin\frac{1}{2}\theta_i$. The latter case gives

$$\mathbf{H}_i^* = \begin{pmatrix} -1-\cos\theta_i & -\sin\theta_i \\ \sin\theta_i & -1-\cos\theta_i \end{pmatrix} = 2\cos\tfrac{1}{2}\theta_i \begin{pmatrix} \cos\frac{1}{2}\theta_i & -\sin\frac{1}{2}\theta_i \\ \sin\frac{1}{2}\theta_i & \cos\frac{1}{2}\theta_i \end{pmatrix}$$

and consideration of the singular value $2\cos\frac{1}{2}\theta_i$. Thus for the best rank-two solution, $\mathbf{H}_i^*$ must be chosen to maximise $\omega_i = \max(\cos\frac{1}{2}\theta_i, \sin\frac{1}{2}\theta_i)$ and the corresponding values of $\mathbf{K}$ set to minus or plus one, according to whether the choice is the cosine or the sine. This process should be done for all 2×2 blocks and $\mathbf{H}_k^*$ set to the block that has the greatest singular value. Then $\mathbf{H}_k$ is formed from $\mathbf{K}$ by inserting $\mathbf{H}_k^*$ in the corresponding position and leaving the remainder of $\mathbf{K}$ unchanged. The approximation $\mathbf{W}\mathbf{H}_k\mathbf{W}'$ minimises $\|\mathbf{W}\mathbf{H}_k\mathbf{W}' - \mathbf{Q}\|$ but now incorporates reflections and is no longer a simple plane rotation as it is when $\mathbf{K} = \mathbf{I}$.

5. The Geometrical Interpretation of the Polar Decomposition

The polar decomposition and its approximation, as described in §3 has an algebraic validity in terms of providing a parsimonious model for describing departures from symmetry. As has been shown, it also has a geometrical interpretation in terms of plane rotations. To be really useful, this geometrical interpretation should have a characterisation that allows data to be exhibited in an attractive way that aids interpretation. In analogy with the additive decomposition outlined in §1, in which skew-symmetry is represented by areas generated from n points in one or more bimensions, it is possible to represent each of the plane rotations of the canonical decomposition of **Q** in two dimensions, which we shall continue to refer to as bimensions. Writing the vectors $\mathbf{w}_{(2k-1)}$ and $\mathbf{w}_{(2k)}$ as $\mathbf{w}_1$ and $\mathbf{w}_2$ and the rotation angle as θ, then **I** - **Q** is approximated by $\omega_k \mathbf{W}_k \mathbf{L}_k^* \mathbf{W}_k'$ and its i,j th element is given by:

$$q_{ij} = \cos\theta\,(w_{1i}w_{2i} + w_{1j}w_{2j}) + \sin\theta\,(w_{1i}w_{2j} - w_{2j}w_{1j}). \qquad (11)$$

Thus if we plot the points $P_i(w_{1i}, w_{2i})$, i = 1,2,...,n then (11) gives the formula for reconstructing q_{ij}. At first sight this looks as if it requires a complicated composition of linear combinations of inner-products, for the term in cosθ, and areas, for the term in sinθ, but closer examination shows that this is not so.

Figure 1 shows the two points P_i and P_j subtending an angle α_{ij} at the origin O. Writing r_i for the length OP_i, then (10) becomes:

$$q_{ij} = \omega r_i r_j \sin\tfrac{1}{2}\theta\,\cos\alpha_{ij} + \omega r_i r_j \cos\tfrac{1}{2}\theta\,\sin\alpha_{ij} = \omega r_i r_j \sin(\alpha_{ij} + \tfrac{1}{2}\theta),$$

and similarly $\quad q_{ji} = \omega r_i r_j \sin(\alpha_{ij} - \tfrac{1}{2}\theta),$

where ω is the dominant singular value from the decomposition of (10). The matrix (**I** - **Q**) is asymmetric and this is represented by noting that the angle $\tfrac{1}{2}\theta$ is to be added or subtracted, depending on the order of the suffices in q_{ij}. Thus, as with skew-symmetry, the representation is interpreted in terms of areas, as is shown in Figure 1. Corresponding to every pont P_i is a point P_i^* obtained by rotating OP_i through an angle $\tfrac{1}{2}\theta$ in, say, a clockwise direction. The element q_{ij} is then approximated by the area of the triangle $OP_jP_i^*$ and q_{ji} by $OP_iP_j^*$. Alternatively, an anti-clockwise convention could be adopted, in which P_j^* transforms to P_j^{**} and then q_{ij} is approximated by the area $OP_iP_j^{**}$. In practice, there is no need to plot any of the transformed points because the effects of a simple rotation on area are easily assessed by eye.

Apart from the singular value ω_k, the component $\mathbf{H}_k^*$ of **I** - **Q** represents a rotation through an angle $\tfrac{1}{2}\theta$ in the exhibited bimension. Being a rigid body transformation this ensures that rotated configurations are the same as the unrotated configurations. Thus straight lines transform into straight lines, circles into circles and so on; moreover, distances remain unaltered by transformation. All points P_k^* predicting the same value of q_{ij} for fixed P_j must lie on a line through P_i^* parrallel to OP_j. Because of the rigid body property, corresponding to the locus of P_k^* is another line which is the locus of P_k. Furthermore, the two lines must be inclined at the angle of rotation. or, equivalently, because of the parallelism, the locus of P_k makes this angle with OP_j as is shown in Figure 1.

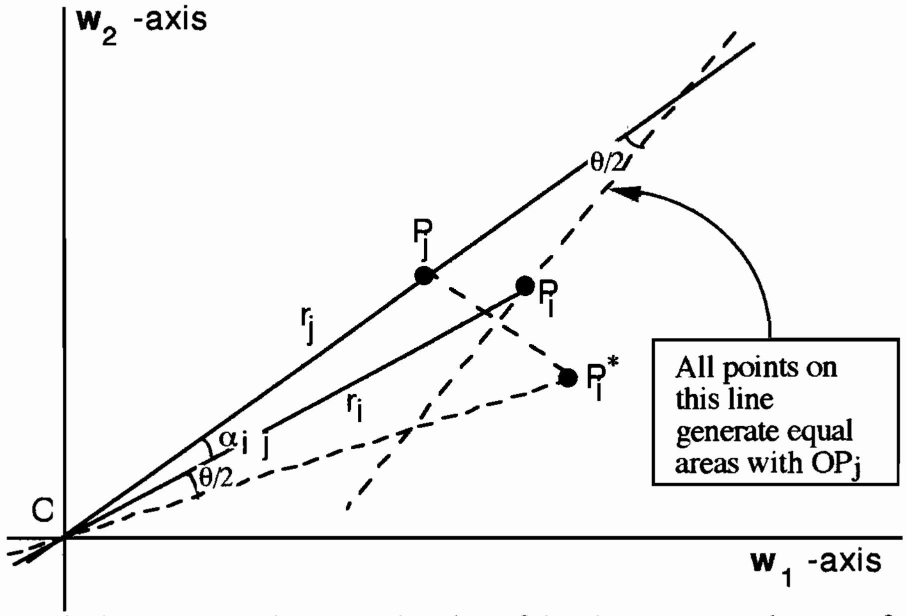

Figure 1. This demonstrates the approximation of the element q_{ij} as the area of triangle $OP_jP_i^*$ in the plane of the vectors $\mathbf{w}_1$ and $\mathbf{w}_2$. OP_i has to be rotated clockwise through the angle $\frac{1}{2}\theta$ to position OP_i^*. Alternatively, rotate OP_j anticlockwise through $\frac{1}{2}\theta$ to a postion OP_j^* and take the triangle $OP_iP_j^*$. To represent q_{ji} requires either OP_i^* to be obtained by rotating OP_i anticlockwise through $\frac{1}{2}\theta$ or by rotating OP_j clockwise through $\frac{1}{2}\theta$ to a postion OP_j^*.

The values of **Q** have to be related to those of **A** by multiplication. This may not be easy, but it is hoped that the examples of section 6 will show that useful information can be conveyed by representations based on the geometry shown in Figure 1. When more than one bimension is used there is a need to interpret two or more products and this would appear to raise serious difficulties. However, the singular value decomposition of **I** - **Q** permits the effects of the products to be viewed additively. Specifically, with b bimensions:

$$\mathbf{A} = \mathbf{MQ} = \mathbf{M} - \mathbf{M}(\mathbf{I} - \mathbf{Q}) = \mathbf{M} - \mathbf{MW\Omega LW'} \sim \mathbf{M} - \mathbf{M}\sum_{k=1}^{b} \omega_k \mathbf{WL_kW'}.$$

This implies that multiplicative departures from symmetry may be viewed and interpreted as the sum of separate simple multiplicative terms $\mathbf{W}_k\mathbf{L}_k\mathbf{W}_k'$.

A problem with the above representation seems to be the difficulty of combining the results with the symmetric part, **M**. Indeed, considering only the case **K** = **I** of §4 we have:

$$\mathbf{A} = \mathbf{MQ} = \mathbf{M} - \mathbf{M}(\mathbf{I} - \mathbf{Q})$$

where $\mathbf{I} - \mathbf{Q} = \mathbf{W}(\mathbf{I} - \mathbf{H})\mathbf{W}'$ and $\mathbf{I} - \mathbf{H}$ is to be approximated by $\mathbf{I} - \mathbf{H}_k$ and hence $\mathbf{H}$ is approximated by $\mathbf{H}_k = \mathbf{I} - \mathbf{H}_k^*$. Thus **A** is approximated by $\mathbf{M} - \mathbf{MW}_k\mathbf{H}_k^*\mathbf{W}_k'$ and departures from symmetry are confined to $\mathbf{MW}_k\mathbf{H}_k^*\mathbf{W}_k'$ where $\mathbf{H}_k^*$ is a matrix of rank-two:

$$2\sin\tfrac{1}{2}\theta(\mathbf{w}_1, \mathbf{w}_2)\begin{pmatrix} \sin\tfrac{1}{2}\theta & -\cos\tfrac{1}{2}\theta \\ \cos\tfrac{1}{2}\theta & \sin\tfrac{1}{2}\theta \end{pmatrix}\begin{pmatrix} \mathbf{w}_1' \\ \mathbf{w}_2' \end{pmatrix}.$$

We may plot two sets of points **X**, **Y** in two dimensions. The first set has coordinates given by the rows of $\mathbf{X} = \mathbf{MW}_k\mathbf{H}_k^*$ and the second set by the rows of $\mathbf{Y} = \mathbf{W}_k$. Alternatively, we could choose $\mathbf{X} = \mathbf{MW}_k$ and $\mathbf{Y} = \mathbf{W}_k\mathbf{H}_k^{*\prime}$ to determine the two sets or indeed, any other partition of $\mathbf{H}_k^*$ between the two sets. Whichever form is adopted, interpretation is through the usual inner product, taking the ith point from one set and the jth point from the other set. Inner products between pairs of points from the same set do not seem to yield anything useful.

The nice geometrical properties of plane rotations seem to find no place in the interpretation. However, we could seek a solution in which $\mathbf{X} = \mathbf{YH}_k^*$ to approximate **A** by $\mathbf{YH}_k^*\mathbf{Y}'$. This brings us into the realm of DEDICOM models (Harshman, *et al.* 1982) with the constraint that the central matrix is to be a 2 × 2 orthogonal rotation matrix. Only the points **Y** need be plotted. If θ is the angle of rotation and P_i and P_j are the two points given by the ith and jth rows of **Y**, which subtend an angle α_{ij} at the origin O, then a_{ij} is estimated by $r_ir_j\cos(\alpha_{ij} - \theta)$ and a_{ji} is estimated by $r_ir_j\cos(\alpha_{ij} + \theta)$. This representation is shown in Figure 2 and is similar to Figure 1 but with inner-products replacing areas. Thus, as usual, interpretation is in terms of the inner-product but we have to remember to add or subtract θ, depending on the order of the suffices. All points P_k that contribute the same a_{ij} lie on the projection of P_j onto OP_i. This representation might be regarded as an angular analogue to that given by J. B. Kruskal's slide-vector model (see Carroll and Wish,1974 and De Leeuw and Heiser, 1982). A difference from the representation of **Q**, discussed above, is that the orthogonal matrix **W** derived from the canonical decomposition is replaced by a generally non-orthogonal matrix **Y** chosen to minimise $\|\mathbf{A} - \mathbf{YH}_k^*\mathbf{Y}'\|$.

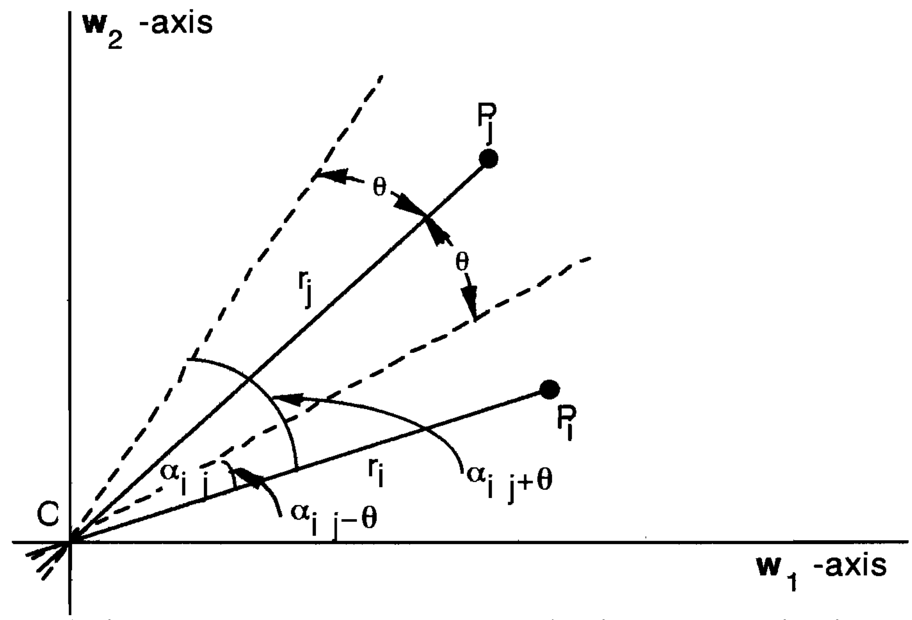

Figure 2. This demonstrates the approximation of the elements a_{ij} and a_{ji} in terms of an inner-product supplemented by a rotation through an angle θ in the plane of the vectors $\mathbf{w}_1$ and $\mathbf{w}_2$.

The emphasis throughout has been on the approximation of the orthogonal part **Q**. In so doing the symmetric part may be undervalued. For example $\mathbf{MW}_k$ represents a projection of **M** onto the plane determined by $\mathbf{W}_k$ and unless these happen to span the space of the first two eigenvectors of **M**, much will be lost. More emphasis could be put on approximating the symmetric part, for example by $\mathbf{U\Sigma}_r\mathbf{U}'$, where $\mathbf{\Sigma}_r$ indicates an r-dimensional approximation obtained by setting the n-r smallest singular values to zero. With r = 2, the coordinates given by the rows of $\mathbf{X} = \mathbf{U\Sigma}_r^{1/2}$ give a familiar two-dimensional inner product representation of **M**. Combining this with **Q**, gives a second set of coordinates $\mathbf{Y} = \mathbf{Q}'\mathbf{U\Sigma}_r^{1/2}$. When **A** is symmetric, then **X** and **Y** are the same thing. However, the equality of inner products for two pairs of points does not depend on this property, so one would be expressing asymmetry in terms of pairs of points that differed in position and in innerproduct.

In analogy with the additive model where **M** and **N** are analysed independently and secondarily an attempt may be made to combine the two parts, perhaps the best thing is to analyse the matrices **M** and **Q** of the polar decomposition independently and not worry too much about models that share the same parameters for describing symmetry and departures from symmetry. In practice, very often different mechanisms govern the two aspects.

6. Examples

This first example uses a set of data concerned with the tendency of customers to switch brands when purchasing a new product. Table 1 shows a subset of the complete data. The brands have been coded to preserve anonymity.

	A	B	C	D	E	F
A	163	19	39	40	9	14
B	17	136	4	10	3	5
C	23	23	2928	728	164	195
D	37	33	896	4861	334	308
E	5	3	134	148	526	50
F	10	8	121	146	53	696

Table 1. Brand-switching data. Numbers of customers changing brands. The rows pertain to previous brands and the columns the successor brands.

Figures 3(a,b) show the analysis of $\mathbf{N}$ given by the additive model (1) and Figures 3(c,d) show the analysis of $\mathbf{Q}$ given by the multiplicative model $\mathbf{A} = \mathbf{MQ}$, and using the rank two approximation to $\mathbf{I} - \mathbf{Q}$ as given by (11). Thus Figures 3(c,d) are to be interpreted as in Figure 1. There is a striking similarity between Figures 3(a,b) and 3(c,d); why should this be so? A clue is given by considering the exact case when $\mathbf{A} = \mathbf{M} - \mathbf{M}(\mathbf{I} - \mathbf{Q})$ where $\mathbf{I} - \mathbf{Q}$ is of rank two. Thus $\mathbf{N} = \mathbf{A} - \mathbf{A}' = (\mathbf{I} - \mathbf{Q})'\mathbf{M} - \mathbf{M}(\mathbf{I} - \mathbf{Q})$ is at most of rank four. This argument shows that sometimes the multiplicative model can display asymmetry in fewer dimensions than are required for the additive model. Thus, in the exact case, the single bimension of the multiplicative model must be embedded in, at most, two bimensions of the additive model. There seems to be no reason why the same bimensions of the additive and multiplicative models should correspond but they should span the same space. Indeed, in a second example, discussed below, although similar bimensions occur for both models, the first (second) bimension of the additive model corresponds to the second (first) of the multiplicative model. With both of the examples discussed here, the multiplicative and additive models seem to require the same number of bimensions. This is not a general property because it is trivial to construct examples where the additive model requitres the fewer bimensions and where the multiplicative model requires the fewer bimensions; there is no reason why real data should be constrained to either form.

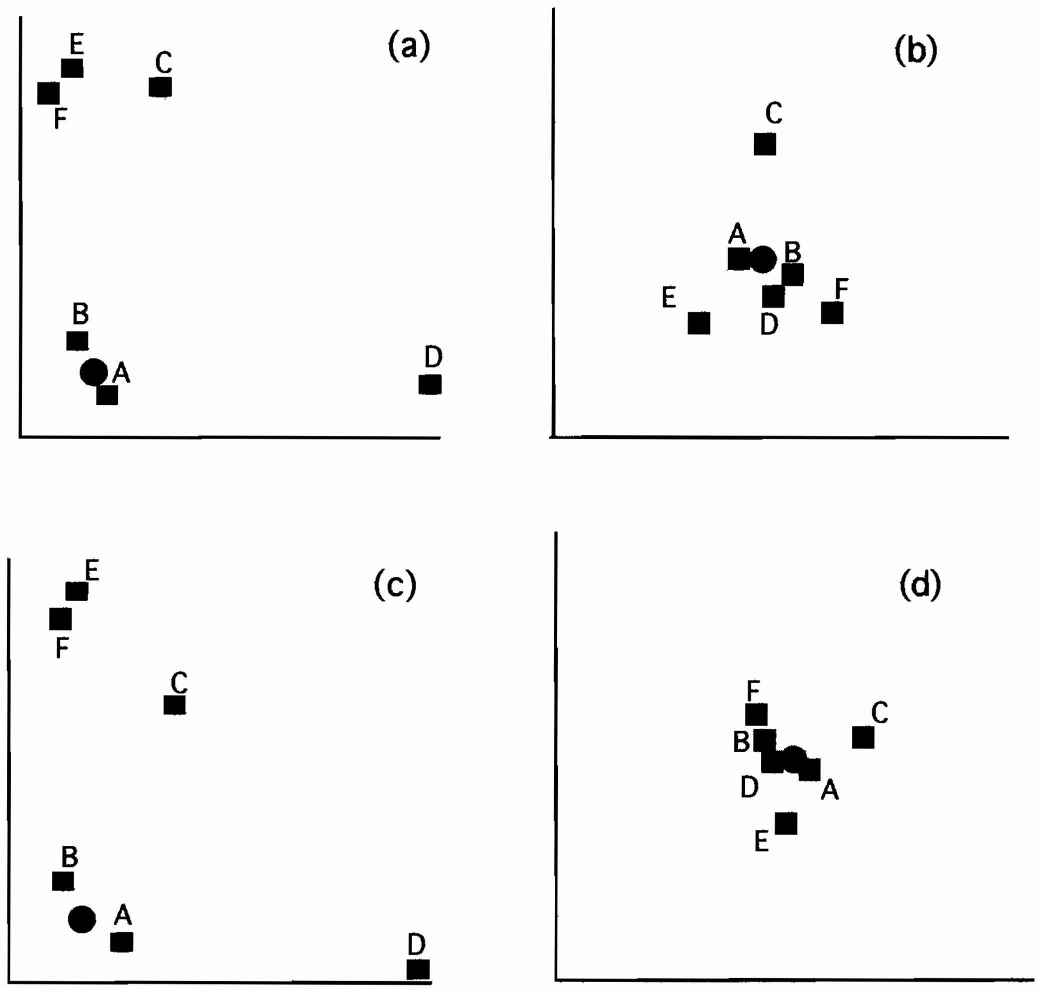

Figure 3.Switching between six brands A,B,C,D,E and F. The first bimension of the additive analysis is given in (a) and the second bimension in (b). The first bimension of the multiplicative analysis is given in (c) and the second bimension in (d). In each case the origin is marked by a black disc. The scales of (a) and (b) are comparable as are the scales of (c) and (d).

The singular values $2\sin\frac{1}{2}\theta_i$ of **I** - **Q** are .049, .012 and.005. Even the largest of these corresponds to a very small angle of rotation. The reason for this is likely to be governed by the dominance of the diagonal and perhaps a better model to fit would be **A** = **D** + **MQ**, where **D** is diagonal chosen to minimise the rank of the symmetric matrix **M** or, equivalently, minimise the rank of **A** - **D**; note that with this variant, **N** continues to have the form discussed in the previous paragraph although the values of **M** and **Q** will differ. Nevertheless, even without these adjustments, the structure for the two bimensions given in Figure 3(b) is closely related to that of the additive model, as has already been discussed. In the multiplicative model, although each bimension combines multiplicatively with **M**, equation (9) shows that the contributions of the different bimensions to total asymmetry may be considered additively. The singular values for the additive model are 309.1, 35.2 and 1.8 so that with this data the additive model is better explained by one bimension than is the multiplicative model. Because of the dominance of the diagonal and the consequent close relationship between the additive and multiplicative models, interpretation is as for the additive case, showing that of those customers who switch, the tendency is to substitute brands C, E or F for brand D. The second bimension suggests a secondary cyclic tendency to switch from D,B,F to C to E, and then back to D,B,F. These remarks are in terms of absolute numbers and take no account of relative market share.

	British	Irish	Scand	German	Italian	Polish	Jewish I	Jewish II
British	549	284	64	90	8	9	2	8
Irish	84	4455	31	52	7	5	2	1
Scandinavian	23	64	1140	43	2	14	1	3
German	64	130	75	2480	17	93	9	9
Italian	27	59	20	63	10001	41	19	48
Polish	2	18	5	34	0	3096	0	1
Jewish I	6	5	3	8	0	0	4840	1440
Jewish II	4	8	2	4	2	3	1906	17752

Table 2. Numbers of marriages between first generation immigrants to New- York - rows correspond to men and columns to women.

Table 2 is extracted from data reported by Pagnini and Morgan (1990). This table is dominated by the diagonal even more so than is Table 1. and a detailed analysis should certainly take account of the widely different row and column totals but here we are concerned only with representing the table as it is. Figures 4(a,b) show the additive analysis (singular values 468.9, 238.1, 67.9 and 10.5) and Figures 4(c,d) the multiplicatiave analysis (singular values .049, .021, .011 and .002). The clear propensity for the two Jewish groups to marry among themselves is shown in the first bimension of the additive and the second bimension of the multiplicative model. This propensity manifests itself in the the way that the non-Jewish groups cluster at the origin ensuring zero area triangles in both interpretations. The remaining bimensions agree apart from a reflection in a line through the origin at approximately 45 degrees. The British, Irish, Scandinavians and Germans are the only groups generating significant areas. In the additive case, the collinearity represents linear skew-symmetry ranking marriage preferences in the given order. For the same reasons as in the first example, the multiplicative model adds nothing to the additive one.

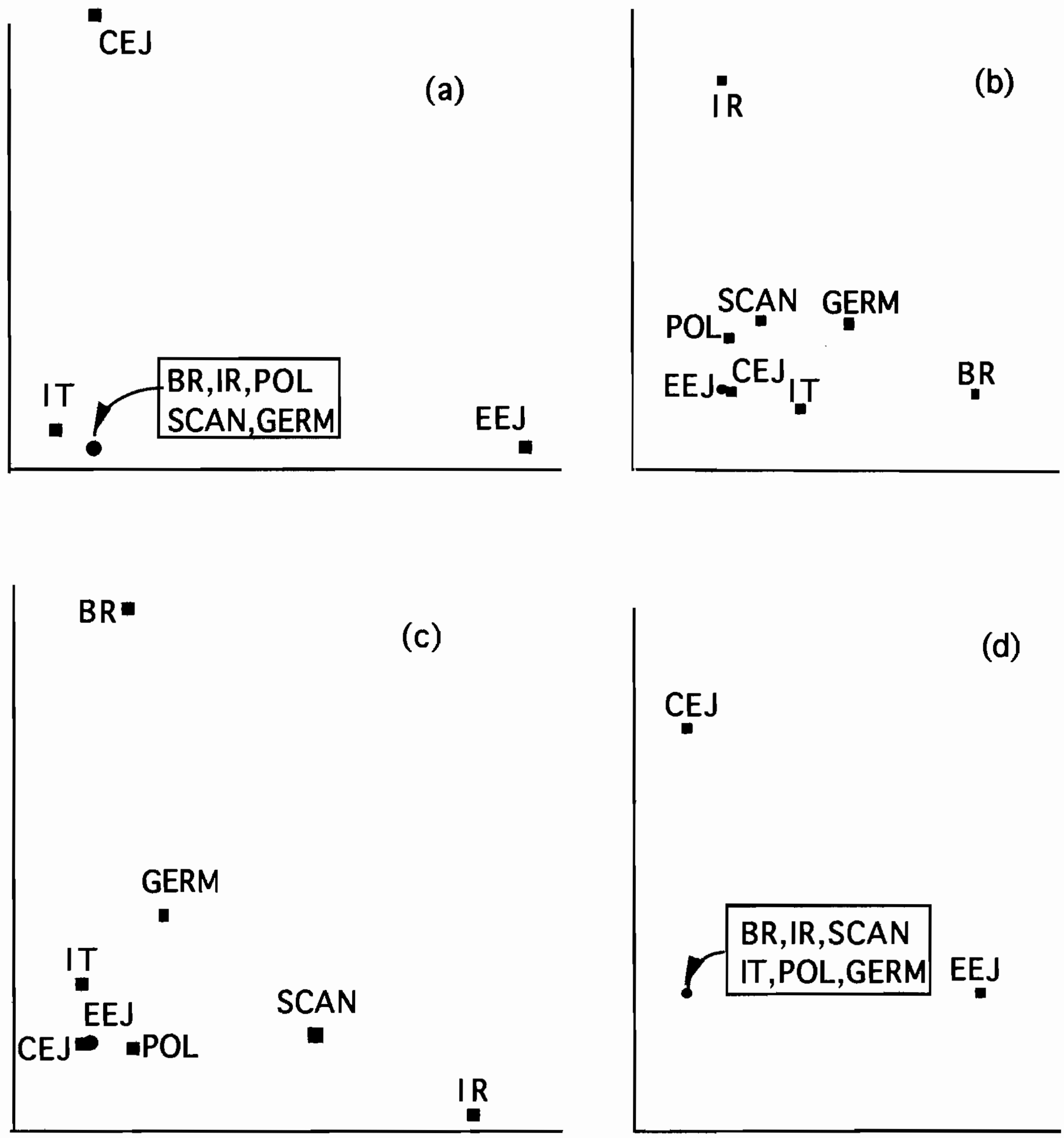

Figure 4. Geometrical representation of the asymmetries of Table2. The first two bimensions of the additive analysis are given in (a) and (b) and the first two bimensions of the multiplicative analysis in (c) and (d). The scales of (a) and (b) are comparable as are the scales of (c) and (d).

These examples are informative for drawing attention to the link between the additive and multiplicative models when the data is dominated by its diagonal. Note that the size of the diagonal does not affect **N** in the additive analysis. It is simple to construct examples in which the multiplicative model is more parsimonius, in terms of numbers of parameters, than is the additive model. Clearly, to demonstrate any advantage for the multiplicative model, we need appropriate data or to explore models which handle the diagonal separately. Models of the form **A** = **D** + **MQ** could be fitted directly, which would require the development of special algorithms, or first the diagonal matrix **D** could be fitted by requiring **A** - **D** to have some nominated rank and then, conditional on this solution, fitting **MQ** to **A** - **D**.

7. Conclusion

In the above, some results on the approximation of orthogonal matrices have been derived and how these might be used to model asymmetry in a square matrix has been discussed. An exposition has been given of the links between orthogonality and skew-symmetry. It has been shown that when the diagonal dominates the data then the model based on additive skew symmetry and the multiplicative model based on orthogonal plane rotations give similar results; the two models are likely to differ most when the diagonal is not dominant or when the diagonal is modelled independently but this aspect requires further investigation. Forms of graphical display have been suggested that are of potential use. My initial impressions are that we should not be too hopeful about their utility but novel graphical methods are notorious for being difficult to assimilate, so what is now needed is for people to gain some experience in their application.

References

Carroll, J. D. and Wish, M. (1974). Multidimensional perception models and measurement methods. In: E. C. Carterette and M. P. Friedman (Eds.), *Handbook of Perception*. Academic Press: New York.

Constantine, A. G. and Gower, J. C. (1978). Some properties and applications of simple orthogonal matrices. *Journal of the Institute of Mathematics and its Applications,* **21**, 445-454.

De Leeuw, J. and Heiser, W. J. (1982). Theory of multidimensional scaling. In: P. R. Krishnaiah and L. N. Kanal (Eds.), *Handbook of Statistics,* Vol 2. North Holland: Amsterdam.

Eckart, V. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika,* **1**, 211-218.

Gantmacher, F. R. (1959). *The Theory of Matrices* (two volumes). K. A. Hirsch (Trans.), Chelsea Publishing Company: New York.

Gower, J. C. (1977). The analysis of asymmetry and orthogonality. In: J. R. Barra, F. Brodeau, G. Romer and B. van Cutsem (Eds.), *Recent Developments in Statistics.* North Holland: Amsterdam.

Harshman, R. A., Green, P. E., Wind, Y. and Lundy, M. E., (1982). A model for the analysis of asymmetric data in marketing research. *Marketing Science*, **1**, 204-242.

Levin, J. and Brown, M. (1979). Scaling a conditional proximity matrix to symmetry. *Psychometrika*, **44**, 239-243.

Pagnini, D. L. and Morgan, S. P. (1990) Intermarriage and social distance among U.S. immigrants at the turn of the century. *American Journal of Sociology*, **96**, 405-432.

APPENDIX

PROPERTIES OF SKEW-SYMMETRIC AND ORTHOGONAL MATRICES

For convenience, elementary proofs are given of some of the spectral properties of skew-symmetric and orthogonal matrices. With the possible exception of the singular value decompositions of theorems 2 and 4(i), these results are well-known but are not readily found in modern algebra text-books.

Theorem 1

If $\mathbf{N}$ is a real skew-symmetric matrix, then its eigenvalues are imaginary and occur in conjugate pairs $+i\sigma$ and $-i\sigma$ corresponding to eigenvectors $\mathbf{x} + i\mathbf{y}$ and $\mathbf{x} - i\mathbf{y}$, respectively where $\mathbf{x}'\mathbf{x} = \mathbf{y}'\mathbf{y}$ and $\mathbf{x}'\mathbf{y} = 0$. When the order of $\mathbf{N}$ is odd, then there is an additional zero eigenvalue.

Proof

Because $\mathbf{N}$ is not symmetric, it may have complex eigenvalues and eigenvectors. Let $\rho + i\sigma$ be an eigenvalue corresponding to an eigenvector $\mathbf{x} + i\mathbf{y}$. Then:

$$\mathbf{N}(\mathbf{x} + i\mathbf{y}) = (\rho + i\sigma)(\mathbf{x} + i\mathbf{y})$$

and equating real and imaginary parts gives:

$$\left.\begin{aligned} \mathbf{N}\mathbf{x} &= \rho\mathbf{x} - \sigma\mathbf{y} \\ \mathbf{N}\mathbf{y} &= \sigma\mathbf{x} + \rho\mathbf{y}. \end{aligned}\right\} \qquad \text{(A1)}$$

Pre-multiplying by $\mathbf{x}'$ and $\mathbf{y}'$ and adding gives:

$$\mathbf{x}'\mathbf{N}\mathbf{x} + \mathbf{y}'\mathbf{N}\mathbf{y} = \rho(\mathbf{x}'\mathbf{x} + \mathbf{y}'\mathbf{y}).$$

But $\mathbf{x}'\mathbf{N}\mathbf{x} = \mathbf{y}'\mathbf{N}\mathbf{y} = 0$ and so $\rho = 0$, showing that non-zero eigenvalues are imaginary and $\mathbf{x}'\mathbf{y} = 0$. Premultiplying by $\mathbf{y}'$ and $\mathbf{x}'$ shows that $\sigma\mathbf{x}'\mathbf{x} = \sigma\mathbf{y}'\mathbf{y}$.
Setting $\rho = 0$ in (A1), gives:

$$\left.\begin{aligned} \mathbf{N}\mathbf{x} &= -\sigma\mathbf{y} \\ \mathbf{N}\mathbf{y} &= \sigma\mathbf{x}. \end{aligned}\right\} \qquad \text{(A2)}$$

and hence if $i\sigma$ is an eigenvalue satisfying $\mathbf{N}(\mathbf{x} + i\mathbf{y}) = i\sigma(\mathbf{x} + i\mathbf{y})$, then we also have that $\mathbf{N}(\mathbf{x} - i\mathbf{y}) = -i\sigma(\mathbf{x} - i\mathbf{y})$, which establishes the main result. When the order of $\mathbf{N}$ is odd, there is an extra eigenvalue ν, say, which is not one of a pair. Because the imaginary pairs cancel with one another, the sum of all the eigenvalues must be ν. Hence $\nu = \text{Trace } \mathbf{N} = 0$.

Theorem 2

The singular value decomposition of a real skew-symmetric matrix **N** has the form $\mathbf{U\Sigma JU}'$, where **U** is orthogonal and $\mathbf{\Sigma}$ and **J** are defined in §1.

Proof

Assume that **N** has a general singular value decomposition $\mathbf{N} = \mathbf{USV}'$. Then **U** and **V** are the eigenvectors of the symmetric matrices $\mathbf{NN}' = \mathbf{US}^2\mathbf{U}'$ and $\mathbf{N}'\mathbf{N} = \mathbf{VS}^2\mathbf{V}'$, respectively. But because **N** is skew-symmetric, $\mathbf{NN}' = \mathbf{N}'\mathbf{N} = -\mathbf{N}^2$, and hence **U** and **V** comprise the same column-vectors, but not necessarily in the same order, because $\mathbf{S}^2$ contains pairs of repeated eigenvalues σ^2.

From (A2) we have that $-\mathbf{N}^2\mathbf{x} = \sigma\mathbf{Ny} = \sigma^2\mathbf{x}$ and also $-\mathbf{N}^2\mathbf{y} = -\sigma\mathbf{Nx} = \sigma^2\mathbf{y}$, so that **x** and **y** are eigenvectors of $-\mathbf{N}^2$ and hence are singular vectors of **N**. Equation (A2) gives the relationship of this pair of singular vectors, showing that:

$$\mathbf{N}(\mathbf{x},\mathbf{y}) = (\mathbf{x},\mathbf{y})\begin{pmatrix} 0 & \sigma \\ -\sigma & 0 \end{pmatrix}$$

so that we may write $\mathbf{V} = \mathbf{UJ}'$ with $\mathbf{S} = \mathbf{\Sigma} = (\sigma_1,\sigma_1,\sigma_2,\sigma_2,\ldots)$. When the order of **N** is odd there is a zero singular value and **J** has to be augmented by a single unit in its final diagonal position.

Theorem 3

If **Q** is an orthogonal matrix, then its eigenvalues are +1, -1 or occur in conjugate pairs $e^{i\theta}$, $e^{-i\theta}$ with corresponding eigenvectors $\mathbf{x} + i\mathbf{y}$, $\mathbf{x} - i\mathbf{y}$, where $\mathbf{x}'\mathbf{x} = \mathbf{y}'\mathbf{y}$ and $\mathbf{x}'\mathbf{y} = 0$.

Proof

Replacing **N** by **Q** in (A1) gives:

$$\left.\begin{aligned} \mathbf{Qx} &= \rho\mathbf{x} - \sigma\mathbf{y} \\ \mathbf{Qy} &= \sigma\mathbf{x} + \rho\mathbf{y}. \end{aligned}\right\} \qquad \text{(A3)}$$

Because of the orthogonality:

$$\left.\begin{aligned} \mathbf{x}'\mathbf{Q}'\mathbf{Qx} &= \mathbf{x}'\mathbf{x} = \rho^2\mathbf{x}'\mathbf{x} + \sigma^2\mathbf{y}'\mathbf{y} - 2\rho\sigma\mathbf{x}'\mathbf{y} \\ \mathbf{y}'\mathbf{Q}'\mathbf{Qy} &= \mathbf{y}'\mathbf{y} = \sigma^2\mathbf{x}'\mathbf{x} + \rho^2\mathbf{y}'\mathbf{y} + 2\rho\sigma\mathbf{x}'\mathbf{y} \\ \mathbf{x}'\mathbf{Q}'\mathbf{Qy} &= \mathbf{x}'\mathbf{y} = (\rho^2 - \sigma^2)\mathbf{x}'\mathbf{y} + \rho\sigma(\mathbf{x}'\mathbf{x} - \mathbf{y}'\mathbf{y}). \end{aligned}\right\} \qquad \text{(A4)}$$

Adding the first two rows of (A4) gives:

$$\rho^2 + \sigma^2 = 1$$

and hence we may write $\rho = \cos\theta$, $\sigma = \sin\theta$. This gives eigenvalues $\rho + i\sigma = e^{i\theta}$ which for $\theta = 0$ and $\theta = \pi$ gives the real values of 1 and -1, respectively. For other values of θ,

$\rho - i\sigma = e^{-i\theta}$ is a conjugate eigenvalue in the usual way. Substituting for ρ and σ into (A4) shows that, provided $\sigma \neq 0$ (i.e. not in the real case), $\mathbf{x'y} = 0$ and $\mathbf{x'x} = \mathbf{y'y}$ giving orthogonality conditions on the real and imaginary components of the latent vectors.

Theorem 4

If **Q** is orthogonal, then:

(i) $\mathbf{I} - \mathbf{Q}$ has a singular value decomposition $\mathbf{W\Omega LW'}$, where:

$$\mathbf{\Omega} = \text{diag}(\omega_1, \omega_1, \omega_2, \omega_2, \ldots, 2, 2, \ldots, 2, 0, 0, \ldots, 0), \quad \omega_i = 2\sin\tfrac{1}{2}\theta_i \text{ and}$$

$$\mathbf{L} = \left\{ \begin{pmatrix} \sin\frac{1}{2}\theta_i & -\cos\frac{1}{2}\theta_i \\ \cos\frac{1}{2}\theta_i & \sin\frac{1}{2}\theta_i \end{pmatrix}, 1 \right\}.$$

(ii) **Q** has a canonical decomposition $\mathbf{Q} = \mathbf{WHW'}$, where:

$$\mathbf{H} = \left\{ \begin{pmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{pmatrix}, -1, 1, [1] \right\}.$$

Proof

<u>Part (i).</u>

Suppose the singular value decomposition of $\mathbf{I} - \mathbf{Q}$ is given by $\mathbf{I} - \mathbf{Q} = \mathbf{U\Theta V'}$. We have:

$(\mathbf{I} - \mathbf{Q'})(\mathbf{I} - \mathbf{Q}) = (\mathbf{I} - \mathbf{Q})(\mathbf{I} - \mathbf{Q'}) = 2\mathbf{I} - \mathbf{Q} - \mathbf{Q'} = \mathbf{U\Theta^2U'} = \mathbf{V\Theta^2V'}$. From theorem 3, the real eigenvalues of **Q** satisfy (a) when $\theta = 0$, $(\mathbf{I} - \mathbf{Q})\mathbf{x} = 0$ and (b) when $\theta = \pi$, $(\mathbf{I} - \mathbf{Q})\mathbf{x} = 2\mathbf{x}$. Thus vectors **x** arising from the real eigenvalues of **Q** give corresponding columns of both **U** and **V**, and hence of **W**, with singular values 0 and 2 in $\mathbf{\Omega}$ and corresponding unit diagonal values in **L**.

For other values of θ:

$$(\mathbf{I} - \mathbf{Q})\mathbf{x} = (1 - \cos\theta)\mathbf{x} + \sin\theta\mathbf{y}$$

$$(\mathbf{I} - \mathbf{Q})\mathbf{y} = -\sin\theta\mathbf{x} + (1 - \cos\theta)\mathbf{y}$$

giving

$$(\mathbf{I} - \mathbf{Q})(\mathbf{x},\mathbf{y}) = 2\sin\tfrac{1}{2}\theta(\mathbf{x},\mathbf{y})\begin{pmatrix} \sin\frac{1}{2}\theta & -\cos\frac{1}{2}\theta \\ \cos\frac{1}{2}\theta & \sin\frac{1}{2}\theta \end{pmatrix}.$$

Thus $\omega = 2\sin\frac{1}{2}\theta$ is a repeated singular value of $\mathbf{I} - \mathbf{Q}$ with $(\mathbf{x},\mathbf{y})$ two columns of **U** and $\begin{pmatrix} \sin\frac{1}{2}\theta & -\cos\frac{1}{2}\theta \\ \cos\frac{1}{2}\theta & \sin\frac{1}{2}\theta \end{pmatrix}\begin{pmatrix} \mathbf{x'} \\ \mathbf{y'} \end{pmatrix}$ two rows of $\mathbf{V'}$. Assigning $(\mathbf{x},\mathbf{y})$ to **W** and $\begin{pmatrix} \sin\frac{1}{2}\theta & -\cos\frac{1}{2}\theta \\ \cos\frac{1}{2}\theta & \sin\frac{1}{2}\theta \end{pmatrix}$ to **L** gives the form of part (i) of the theorem.

If required, the vectors $(\mathbf{x},\mathbf{y})$ that contribute to $\mathbf{U}$ (and $\mathbf{V}$) may be conveniently evaluated without recourse to complex arithmetic as follows. From theorem 3:

$$\mathbf{Q}(\mathbf{x} + i\mathbf{y}) = e^{i\theta}(\mathbf{x} + i\mathbf{y})$$

$$\mathbf{Q}'(\mathbf{x} + i\mathbf{y}) = e^{-i\theta}(\mathbf{x} + i\mathbf{y})$$

and therefore

$$(\mathbf{Q} + \mathbf{Q}')(\mathbf{x} + i\mathbf{y}) = 2\cos\theta(\mathbf{x} + i\mathbf{y}).$$

Thus $\mathbf{x}$ and $\mathbf{y}$ may be obtained as a pair of orthogonal eigenvectors of the real symmetric matrix $\mathbf{Q} + \mathbf{Q}'$, corresponding to a repeated eigenvalue $2\cos\theta$. Similarly, when $\theta = 0$, $(\mathbf{Q} + \mathbf{Q}')\mathbf{x} = 2\mathbf{x}$ and when $\theta = \pi$, $(\mathbf{Q} + \mathbf{Q}')\mathbf{x} = -2\mathbf{x}$.

Part (ii).

Write the result of part (i) in the form:

$$\mathbf{Q} = \mathbf{I} - \mathbf{W\Omega L W}'$$

$$= \mathbf{W}(\mathbf{I} - \mathbf{\Omega L})\mathbf{W}'$$

where it is easy to show that $\mathbf{I} - \mathbf{\Omega L} = \mathbf{H}$ has the form given in the statement of part (ii) of the theorem.