The Geometry of Biplot Scaling

J. C. Gower

Statistics Department

Walton Hall

The Open University

Milton Keynes, MK7 6AA, U. K.

j.c.gower@open.ac.uk

SUMMARY

A simple geometry allows the main properties of matrix approximations used in biplot displays to be developed. It establishes orthogonal components of an analysis of variance, from which different contributions to approximations may be assessed. Particular attention is paid to approximations that share the same singular vectors, in which case the solution space is a convex cone. Two and three-dimensional approximations are examined in detail and then the geometry is interpreted for different forms of the matrix being approximated.

Some key words: Biplots; Correspondence analysis; Matrix approximation; Scaling.

1. INTRODUCTION

Expressing any matrix X in terms of its singular value decomposition, $X = U\Sigma V'$, Eckart & Young (1936) showed that the matrix Y of rank r that minimises $\|X - Y\|^2$ is given by $Y = U\Sigma_r V'$, where Σ_r is zero apart from the first r dominant ordered (diagonal) values of Σ. In biplot displays, the rows and columns of Y are represented

by points whose coordinates are the rows of $U\Sigma_r^{\alpha}$ and $V\Sigma_r^{\beta}$. To represent Y by its inner-product we need $\gamma = \alpha + \beta = 1$ but, as is well known, non-unit values of γ are also used. We therefore consider the γ-family of rank r matrices of the form $X_r = U\Sigma_r^{\gamma}V'$, so that X_1 is a synonym for Y.

It has long been observed that the commonly used scalings α, β give displays that are minor variants of one another that have little effect on interpretation. Gabriel (2002) quantified and confirmed this observation. Here, I give a geometrical approach that I believe offers additional insights and which extends Gabriel's essentially two-dimensional results to three dimensions.

2. The Basic Geometry

Any $p \times q$ matrix X may be represented by a point X in a Euclidean space of pq dimensions, with coordinates x given by the concatenation of the columns of X. If y gives the coordinates of a point Y, representing Y, a rank r matrix, then, for any λ, λy also represents a rank-r matrix. Thus, rank-r matrices form a cone in this space; because the sum of two rank-r matrices is not generally of rank r, the cone is not convex. However, the cone property is sufficient to establish that, if y minimises $||x - y||^2$ over all rank r matrices, then y is given by the orthogonal projection of x on to the ray through y, thus establishing the orthogonal analysis of variance:

$$||x||^2 = ||y||^2 + ||x - y||^2. \qquad (1)$$

As Γ_r varies, the set of rank-r matrices $U\Gamma_r V'$, for fixed singular vectors, do form a convex cone. We term this the UV-cone; the γ-family is a subset where $\Gamma_r = \Sigma_r^{\gamma}$.

When $\gamma \neq 1$, any rank-r matrix defines a ray different from the ray through X_1. The geometry is shown in Fig. 1, where X is denoted by a point X, its optimal least-squares approximation X_1 by X_1 and a sub-optimal matrix X_γ of the γ-family by X_γ. The cone of all matrices of rank r, including the γ-family, is indicated by the shaded area. The distance between two matrices with the same singular vectors is given by $||U\Sigma V' - U\Gamma V'||^2 = ||\Sigma - \Gamma||^2$, and so depends only on the singular values. Hence, we may represent the coordinates of X_γ by $(\sigma_1^\gamma, \sigma_2^\gamma, ..., \sigma_r^\gamma, 0, 0, ..., 0)$. The convex cone is defined by the constraints on the singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ where the final constraint may be relaxed to $\sigma_r \geq 0$, in which case the cone also includes all matrices with rank less than r. In the important case where $r = 2$, we have $\sigma_1 \geq \sigma_2 \geq 0$ and the cone is planar, bounded by the rays $\sigma_2 = 0$ and $\sigma_1 = \sigma_2$; we label these rays by selecting their most simple points, (1,0) and (1,1). This represents a triangular wedge subtending an angle $\pi/4$ at the origin. No two rank-two matrix can subtend an angle that exceeds this angle. When $r = 3$ the wedge shape is bounded by rays (1,0,0), (1,1,0) and (1,1,1), and the maximum angle is given by $\cos(\theta) = 1/\sqrt{3}$ and in general by $\cos(\theta) = 1/\sqrt{r}$.

We wish to measure how close are the configurations X_1 and X_γ. Direct comparison raises problems, mainly because the orthogonal analysis of variance (1) is invalid when $\gamma \neq 1$. The residual sum-of-squares increases, i.e. $||X_\gamma - X||^2 \geq ||X_1 - X||^2$ with equality only when $\gamma = 1$, but the fitted sum-of-squares, $||X_\gamma||^2$, may be greater or less than $||X_1||^2$. However, rather than a direct comparison, we may be more interested in how the shape of the configuration X_γ compares with the shape of the configuration X_1. When we are comparing the shapes of two configurations, the simplest approach

is to find the rigid body translation, orthogonal transformation, i.e. rotation, and isotropic scaling of X_γ that matches X_1. When, as here, X_1 and X_γ share the same singular vectors, they are already optimally rotated. If, for the present, we assume that they are also optimally translated, that only leaves us to determine the isotropic scaling ρ that maximises $\|\rho X_\gamma - X_1\|^2$, which has the simple solution

$$\rho = \frac{\mathrm{tr}(X_1' X_\gamma)}{\| X_\gamma \|^2}.$$

Let us write $X_{\gamma^*} = \rho X_\gamma$ for the optimally scaled version of X_γ and label the corresponding point on the ray γ by X_{γ^*}, as is shown in Fig. 1.

To compare two approximations X_1 and X_{γ^*} needs some justification, when we are primarily concerned with how well each approximates X. The comparison may be justified as follows. In Fig. 1, the least-squares definitions imply that XX_1, denoted by the vector r, is orthogonal to OX_1 and XX_{γ^*}, denoted by the vector s, is orthogonal to OX_γ. Indeed, XX_1 is normal to the whole cone of rank-r approximations with the same singular vectors as those of X; otherwise there would be a rank r approximation with smaller residual vector than the minimum, $\|r\|^2$. Thus, $X_1X_{\gamma^*}$, denoted by the vector $\boldsymbol{t}$, is orthogonal to XX_1. Algebraically, these orthogonal relationships may be written:

$$\left.\begin{aligned} \| s \|^2 &= \| r \|^2 + \| t \|^2 \\ \| x \|^2 &= \| r \|^2 + \| x_1 \|^2 \\ \| x \|^2 &= \| s \|^2 + \| x_{\gamma^*} \|^2 \end{aligned}\right\}, \qquad (2)$$

from which it follows that

$$\|x_1\|^2 = \|t\|^2 + \|x_{\gamma^*}\|^2.$$

This establishes that t is orthogonal to x_γ, showing that not only is X_1 the best approximation to X but also that X_{γ^*} is the best approximation, on its particular ray, to both X_1 and X. This is, at least in part, a justification for confining attention to the cone that contains both approximations. It also follows from the relationships that

$$||x||^2 = ||x_{\gamma^*}||^2 + ||t||^2 + ||r||^2, \tag{3}$$

so that $||t||^2$ represents the increase in the residual sum-of-squares induced by fitting the sub-optimal X_{γ^*} to X rather than the optimal X_1. Equation (3) allows the effect of replacing X_1 by X_γ, and thence X_{γ^*}, to be assessed.

Since the relative size of the representations is immaterial, we are concerned mainly with the angle θ_γ. It is clear from Fig. 1 that $||x_{\gamma^*}||^2 = ||x_1||^2\cos^2(\theta_\gamma)$. The angle θ_γ may be derived directly from the coordinate representations of X_1 and X_γ as

$$\cos^2(\theta_\gamma) = \frac{(\sum_{i=1}^{r}\sigma_i^{\gamma+1})^2}{\sum_{i=1}^{r}\sigma_i^2\sum_{i=1}^{r}\sigma_i^{2\gamma}} = \frac{(1+\sum_{i=2}^{r}\psi_i^{\gamma+1})^2}{(1+\sum_{i=2}^{r}\psi_i^2)(1+\sum_{i=2}^{r}\psi_i^{2\gamma})}. \tag{4}$$

where $\psi_i = \sigma_i / \sigma_1$. This is the measure used by Gabriel (2002).

Just as X_γ is obtained from $\boldsymbol{X_1}$ by replacing each σ_i by σ_i^γ, so is X_1 obtained from X_γ by replacing each σ_i^γ by $(\sigma_i^\gamma)^{1/\gamma}$. This implies that, for ψ_2 and ψ_3,

$$\cos^2\theta_\gamma(\psi_2,\psi_3,...,\psi_r) = \cos^2\theta_{1/\gamma}(\psi_2^\gamma,\psi_3^\gamma,...,\psi_r^\gamma). \tag{5}$$

and that this extends to any number of ratios of singular values. This result, which may be readily verified algebraically from (4), is useful because it allows results found for $\gamma > 1$ to be transformed into corresponding results for $\gamma < 1$.

Note that $\Sigma^{\gamma} \to \sigma_1^{\gamma}(1,0,0,\ldots)$ as $\gamma \to \infty$ and $\Sigma^{\gamma} \to \sigma_1^{\gamma}(1,1,1,\ldots)$ as $\gamma \to 0$. When Σ_1 is close to either of these two extremes, then by choosing γ sufficiently large, or small, Σ_1^{γ} will tend to the other extreme. Thus, the extreme angle with $\cos^2(\theta_\gamma) = 1/r$ may always be achieved within the γ-family giving a global minimum for any rank r approximation to X that shares the same singular vectors.

3. Two- and Three-dimensional Approximations

Figure 2 shows the UV-cone for $r = 2$ with the attainable extreme value of $\cos^2(\theta) = ½$, or $\theta = \frac{\pi}{4}$. The ray OX_1, best least-squares fit, has gradient $\psi = \psi_2 = \sigma_2/\sigma_1$, the ratio of the singular values relative to the (1,0) ray. The ray OX_γ has gradient ψ^{γ}, where $\psi^{\gamma} < \psi$ if $\gamma > 1$ and $\psi^{\gamma} > \psi$ if $0 < \gamma < 1$. The angle θ_γ increases with $\gamma \oplus 1$ until it is maximal on one of the boundaries. For example, when $\psi = ½$, a fairly typical value, the minimal possible value of $\cos^2(\theta_\gamma)$ is 0.8, corresponding to $\psi = 0$; on the other boundary, $\psi = 1$, we have $\cos^2(\theta_\gamma) = 0.9$. Thus, with values of ψ found in practice, fits much better than the global minimum of ½ will be found, whatever value is assigned to γ. When γ is restricted to conventional values, $½ \le \gamma \le 2$ say, even better fits will be found; for example for $\psi = ½$, minimal values of $\cos^2(\theta_\gamma)$ are 0.9529 for $\gamma = 2$ and 0.9771 for $\gamma = ½$. Indeed, when $\gamma = 2$ the global minimal value of $\cos^2(\theta_\gamma) = 0.9510$

occurs for $\psi = 0.435$. This ray is shown as a dotted line in Fig. 2. From (5) with $r = 2$, we know that the same global minimum applies to $\gamma = ½$ for $\psi = 0.435^2 = 0.190$.

In a sense, the value of 0.951 subsumes the whole of Fig. 1 of Gabriel (2002). It shows that for $½ \leq \gamma \leq 2$, which includes all values of practical importance, $\cos^2(\theta_\gamma) > 0.9510$ so the contour surface is essentially flat; a similar remark applies to Fig. 3, introduced below. Thus, biplot scalings within this range of γ are all acceptable as judged by the criterion (4), whatever may be the actual singular values.

Similar arguments apply to the three dimensional case. In three dimensions, it suffices to consider a two-dimensional cross-section of the cone, where the optimal ray is represented in the plane $\psi_1 = 1$ by the point $X_1(1, \psi_2, \psi_3)$ and X_γ by $(1, \psi_2^\gamma, \psi_3^\gamma)$. We may consider the locus of X_γ in this plane. As in the two-dimensional case, $X_\gamma \rightarrow (1,0,0)$ as $\gamma \rightarrow \infty$ and $\gamma \rightarrow (1,1,1)$ as $\gamma \rightarrow 0$. It is not immediately obvious, as it was in the two-dimensional case, that $\cos(\theta_\gamma)$ decreases monotonically with γ. However, if we consider the alternative intersection-plane that passes through X_1 and is orthogonal to OX_1, it is immediate that the circular cones with fixed θ_γ intersect this plane in circles centred at X_1. It follows that the locus of X_γ intersects these circles at points corresponding to increasing values of both γ and θ_γ. The same must remain true in planes oblique to OX_1, such as the plane $(1, \psi_2, \psi_3)$. Thus, as in the two-dimensional case, we may consider the extreme cases $\gamma = 2$ and $\gamma = ½$, confident that better fits, i.e. smaller θ_γ, must occur for intermediate values of γ.

Figure 3 shows the contours of constant $\cos^2(\theta_\gamma)$ for $\gamma = 2$ and for different values of ψ_2 and ψ_3. There are three minima: (i) $\cos^2(\theta_2) = 0.917$ when $\psi_2 = \psi_3 = 0.399$; (ii)

$\cos^2(\theta_2) = 0.951$ when $\psi_2 = 0.435$ and $\psi_3 = 0$; and (iii) $\cos^2(\theta_2) = 0.973$ when $\psi_2 = 1$ and $\psi_3 = 0.460$. These minima all correspond to pathological cases: (i) the second and third singular values are equal; (ii) X_1 has rank two; and (iii) the first two singular values are equal. The global minimum occurs for case (i) but is, nevertheless, encouragingly large. Case (ii) is the two-dimensional case already discussed; the line $\psi_3 = 0$ of Fig. 3 corresponds to the vertical dotted line of Fig. 2. Despite the busy-looking contours in Fig. 3, the surface is essentially flat. The sharp corners in the contours at the minima on $\psi_2 = 1$ and $\psi_3 = 0$ occur not because of any discontinuity in the gradients. If the contours are continued into the analytically acceptable regions beyond the convex cone, that correspond to inadmissible singular values, the sharp corners are seen as the nodes of looped contours.

4. DISCUSSION

4.1 Preamble

The above depends entirely on γ and not on the individual values of α and β. The usual choices are (i) $\alpha = 1$, $\beta = 0$, (ii) $\alpha = 0$, $\beta = 1$, (iii) $\alpha = ½$, $\beta = ½$ or (iv) $\alpha = 1$, $\beta = 1$. Which one to choose depends on what substantive meaning is to be assigned to X. There are three major considerations: is X a multivariate data-matrix with p columns referring to different variables or does it refer to a two-way table of a single variable classified by two factors with p and q levels; are the entries in X quantitative measurements, counts or categorical; and are we more interested in approximating X itself or in approximating a derived symmetric matrix, such as $X'X$ or XX', or a distance matrix of some kind, or a similarity matrix? These topics are covered in the literature; for a review of most of the issues, see Gower and Hand (1996). In this section, we examine the effect of the different forms of X on our geometry.

4.2 X is a data matrix

The classical case is of a quantitative data matrix X, centred to have zero means and possibly normalised to have unit sums-of-squares. Then we choose $\alpha = 1$, $\beta = 0$ to obtain the usual representation $(U\Sigma)V'$ of the samples and variables in principal components analysis. We may compute the singular value decomposition via the spectral decomposition of XX' or $X'X$, but this is a computational convenience and does not necessarily signify an interest in approximating these matrix products. However, these matrices may have substantive interest and replace X itself as the primary matrix. Thus, $XX' = U\Sigma^2U'$, so that $U\Sigma$ also generates XX' and hence the distances between the rows of X. Furthermore, $X'X = V\Sigma^2V'$, so that $V\Sigma$ generates the correlation matrix, provided X has been centred and normalised. The squared distance between the ith and jth columns of X is $2(1-\rho_{ij})$, where ρ_{ij} is the correlation between the ith and jth variables (Hills, 1969). This representation needs $\alpha = 0$, $\beta = 1$, though we could also choose $\alpha = \beta = 1$, in which case the induced inner-product is $U\Sigma^2V'$ which might be regarded as our basic matrix. The Eckart-Young approximation now has singular values Σ^2 and hence contributes fourth powers to the fit, so the fits to X and to $X'X$ are assessed differently. It now becomes ambivalent as to whether we are assigning weights $\alpha = \beta = 1$ to Σ or weights $\alpha = \beta = ½$ to Σ^2. This does not affect the geometry of §§2 and 3 because the only constraint required on X is that it shares its singular vectors with those of X_γ. Indeed, the analysis of §3 shows that X will be well approximated even when our primary objective is to approximate $X'X$ and XX'.

4.3 Approximating symmetric matrices

In the above situations, we are approximating symmetric matrices, with zero diagonals for distance matrices and unit diagonals for correlation and similarity matrices. When using the implicit Eckart-Young approximation we note that the

diagonals get half the weight of the symmetric off-diagonal terms; it might be more appropriate to give zero weight to the diagonal terms. Bailey & Gower (1990) investigate some of the algebraic consequences of using differential weightings. Diagonals are excluded when we use most forms of metric multidimensional scaling, and special handling of the diagonal is one of the motivations for the factor analysis of a correlation matrix. By ignoring the diagonal, we can find better fits to the substantive parts of the matrix, but the approximations are no longer part of the γ-family of the UV-cone, so the criterion (4) is invalid and the geometry discussed earlier is invalid. We could continue to calculate the angle between X, or $X'X$, and the approximation, but the geometry and the algebra both become much more complicated. In practice, the approximations found by different methods of multidimensional scaling are usually similar, so it might be expected that small angles would be found as with the γ-family, but this has not been formally established. The handling of diagonal terms for categorical variables is discussed below.

4.4 Approximating a two-way table of quantitative values

Let us now turn to the case where X is a quantitative two-way table. If we fit a biadditive model, the main effects are estimated in the usual way as the row and column means of X. Then the residual matrix is $Z = (I - P)X(I - Q)$, where $I - P$ and $I - Q$ are row and column centring matrices that eliminate main effects. Thus, Z estimates multiplicative interaction terms which can be biplotted with $\alpha = ½$, $\beta = ½$. In this case the rows and columns of Z have similar status and there is no case for differential weighting.

4.5 Correspondence analysis

Now consider the case where the variables are categorical. First we consider a two-way contingency table Y with p rows and q columns. This is the set-up for

correspondence analysis. Rather than analyse Y directly, we consider $X = R^{-½} Y C^{-½}$, where R and C are the row and column totals of Y expressed as diagonal matrices. The singular value decomposition of X has a dominant unit singular value with corresponding unnormalised vectors $R^{½} 1$ and $C^{½} 1$**.** These may be removed from X to give

$$R^{-½} Y C^{-½} - \frac{R^{½} 1 1' C^{½}}{y_{..}}, \qquad (6)$$

where $y_{..}$ is the grand total of Y used as a normaliser. Apart from the removed first term, this matrix has the same singular value decomposition as X. The elements of (6) are proportional to the terms in Pearson's chi-squared for the independence of rows and columns and may be exhibited in the usual way as a biplot. Rows and columns are of similar status so we would use $\alpha = \beta = ½$. The distances between the rows and between the columns of (6) seem to be of little interest. Correspondence analysis has a principal interest in chi-squared distance. These are the distances between the rows of $R^{-1} Y C^{-½}$ and between the columns of $R^{-½} Y C^{-1}$, and could be represented by any desired method of multidimensional scaling. They can also be derived as distances between points whose coordinates are given by the rows of $A = R^{-½} U \Sigma_r$ and $B = C^{-½} V \Sigma_r$. These are not ordinary least-squares solutions but may be expressed in terms a weighted least-squares analysis. So far as our geometry is concerned, the main issue is that A and B are not members of the γ-family and the geometry developed above does not apply. When R and C are approximately proportional to unit matrices, we are close to the γ-family and so the inner product is close to being proportional to $U\Sigma^2 V'$ and will give a good approximation to (6), now with $\alpha = \beta = 1$. However, if the row and column totals of Y are disparate the approximation could be poor.

4.6 Multiple correspondence analysis

Multiple correspondence analysis can be developed in several ways, one of which is as the simple correspondence analysis of an indicator matrix $G = (G_1, G_2,\ldots, G_p)$. Here G_k is an indicator matrix for the kth categorical variable, zero apart from a single unit in each row, recording the occurrence of a category for the corresponding case. Consequently the row sums of G are constant, equal to p, and the correspondence analysis depends only on the column sums which give the frequencies of every category. We write this as $1'G = 1'L$ where $L = \text{diag}(L_1, L_2,\ldots, L_p)$ are the category frequencies written as diagonal matrices. Correspondence analysis requires the singular value decomposition of $GL^{-½}$. This can be a large matrix and it is common to simplify the calculations by requiring the spectral decomposition of the normalised Burt matrix $B = L^{-½}G'GL^{-½}$. The Burt matrix has interesting structure, with unit diagonal blocks and symmetrically placed off-diagonal blocks of the contingency tables, scaled as for ordinary correspondence analysis, for all pairs of categorical variables. Thus, an analysis of B can be regarded as a generalisation of ordinary correspondence analysis that gives a simultaneous analysis of all the two-way contingency tables. When $p = 2$, $L_1 = R$ and $L_2 = C$ and we may ask how the analysis of B relates to a simple correspondence analysis of $L_1^{-½}G_1'G_2L_2^{-½}$ which corresponds precisely with $X = R^{-½}YC^{-½}$ of the previous paragraph. A crucial difference is that B contains the uninteresting unit diagonal blocks. It turns out (Gower & Hand, 1996, §10.2) that B recovers the singular vectors U and V but replaces the singular values Σ by $½(I + \Sigma)$. This approximation is not part of the γ-family but does belong to the convex UV-cone of rank-r matrices with singular vectors U and V. Corresponding to (4) we may evaluate

$$\cos^2(\theta) = \frac{(\sum_{i=2}^{r} \sigma_i(\sigma_i + 1))^2}{\sum_{i=2}^{r}(1+\sigma_i)^2 \sum_{i=2}^{r}(\sigma_i^2)}. \qquad (7)$$

Now, $\cos^2(\theta)$ depends on the absolute values of the singular values and, unlike (4), equation (7) cannot be expressed in terms of the ratios of singular values $\psi_i = \sigma_i / \sigma_2$.

For p = 2, consideration of the angle between the rays (σ_1, σ_2) and ½(1+σ_1, 1+σ_2) in Fig. 2 shows that, especially for the larger singular values that are of primary interest, the maximal angle that may be attained while remaining in the cone is more restricted than previously. A detailed investigation of (7) gives similar results to (4) but with simpler contours, i.e. arcs of circles passing through the origin, than those of Fig. 3. When p = 2, equation (7) expresses the difference between ignoring and not ignoring the uninteresting information on the diagonal blocks. Joint correspondence analysis (Greenacre, 1988) gives the methodology for ignoring diagonal blocks for general values of p. In a non-iterative form of joint correspondence analysis, all blocks retain their same singular vectors and the common singular values change. Then the two forms of each block to be compared belong to the same UV-cone and the geometry given above shows that they would have similar configurations. However, the fitted and residual parts are not now orthogonal, leading to some interpretational difficulties. An iterative form of joint correspondence analysis leads to a fully acceptable least-squares solution with orthogonal components but achieves this through allowing the singular vectors to change from block to comparable block. Then comparable blocks do not belong to the same UV-cone and our geometry is not available. When p = 2, both forms of joint correspondence analysis recover ordinary correspondence analysis. In general, joint correspondence analysis improves fits to the substantive part of the Burt matrix, and expressions may be found for the fitted and

residual sums-of squares and angles may continue to be evaluated. However, at present, when $p > 2$, there seems to be no way of investigating how these quantities vary with joint correspondence analysis approximations.

Multiple correspondence analysis offers a good example of the ambivalence between whether one is more interested in fitting X or $X'X$**,** in this case between $GL^{-½}$ or B. In part, this ambivalence is fuelled by the computational convenience of working in terms of the spectral decomposition of a symmetric matrix rather than the singular value decomposition of a rectangular matrix. A direct approximation of $GL^{-½}$ or even of G itself may be of interest, because the distances between rows may be interpreted in terms of dissimilarity coefficients. Alternatively, the categories may be quantified and then G is replaced by a quantitative matrix X, subsequently represented by a variety of biplot displays. Indeed, $L^{-½}$ may be regarded as giving one set of quantifications. Confusingly, a multiple correspondence analysis of $GL^{-½}$ leads to a further set of quantifications; see Homogeneity Analysis in Gifi (1990, chapter 3). For computational convenience, we may proceed via the Burt matrix; then the diagonal blocks appear as an essential part of the computation. However, as we have seen, there are substantive reasons for an interest in approximating a correlation matrix or a Burt matrix, with its constituent contingency tables and then improved fits can be found by excluding the uninformative diagonal blocks.

4.7 Interpreting biplots

Once we have our biplot, it has to be interpreted. Distance comparisons are easily made by eye. Inner-products, which are a major interpretative tool, are less easy to assess by eye because they involve the product of two lengths with the cosine of an angle. It is true, as pointed out by Gabriel (2002), that projections of a set of points on to an axis give the correct orderings of the inner products; this is because one of the

two lengths is constant, and is, of course, a basic property of Cartesian axes. However, two difficulties remain: projections allow us to compare inner products in one column of *X* but not across two columns; and there are differences in the degrees of approximation of different axes, resulting in the same unit difference being represented by different lengths. That projections onto different axes are on different scales should not be ignored. A solution is to provide each axis with its own scale, as proposed by Gower & Hand (1996, §§2.3 and 2.6). The axes then behave very much like familiar coordinate axes. Gabriel (2002) may have some justification in saying that this "proliferation of axes and scales then clutters up the display". Nevertheless, it reflects reality, and what one needs is a convenient way of incorporating scale information, not of ignoring it. One would discourage the publication of graphs and charts that did not have appropriate scale information; the situation is no different for biplots.

ACKNOWLEDGEMENT

I thank Dr. Karen Vines for preparing the contour plot.

REFERENCES

BAILEY, R.A. & GOWER, J. C. (1990). Approximating a symmetric matrix. *Psychometrika* **55**, 665-75.

ECKART, C. & YOUNG, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika* **1**, 211-8.

GABRIEL, K. R. (2002). Goodness of fit of biplots and correspondence analysis. *Biometrika* **89**, 423-36.

GIFI, A. (1990). *Nonlinear Multivariate Analysis*. Chichester: John Wiley and Sons.

GOWER, J. C. & HAND, D. J. (1996). *Biplots*. London: Chapman and Hall.

GOWER, J. C. (1990). Three-Dimensional Biplots. *Biometrika* **77**, 773-83.

GREENACRE, M. J. (1988). Correspondence analysis of multivariate categorical data by weighted least squares. *Biometrika* **75**, 457-67.

HILLS, M. (1969). On looking at large correlation matrices. *Biometrika* **56**, 249-53.

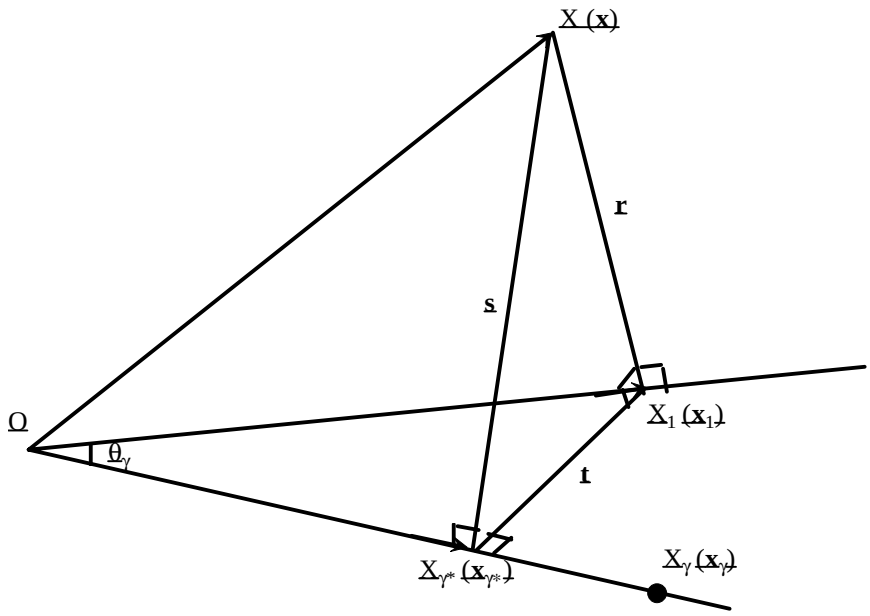

Figure 1 X is the point representing the original matrix and X_1 is its best least-squares rank-*r* approximation. The shaded area represents the cone of all rank-*r* matrices with the same singular vectors as **X**. X_γ is a sub-optimal matrix in the cone and X_{γ^*} is the nearest point to X on the ray γ.

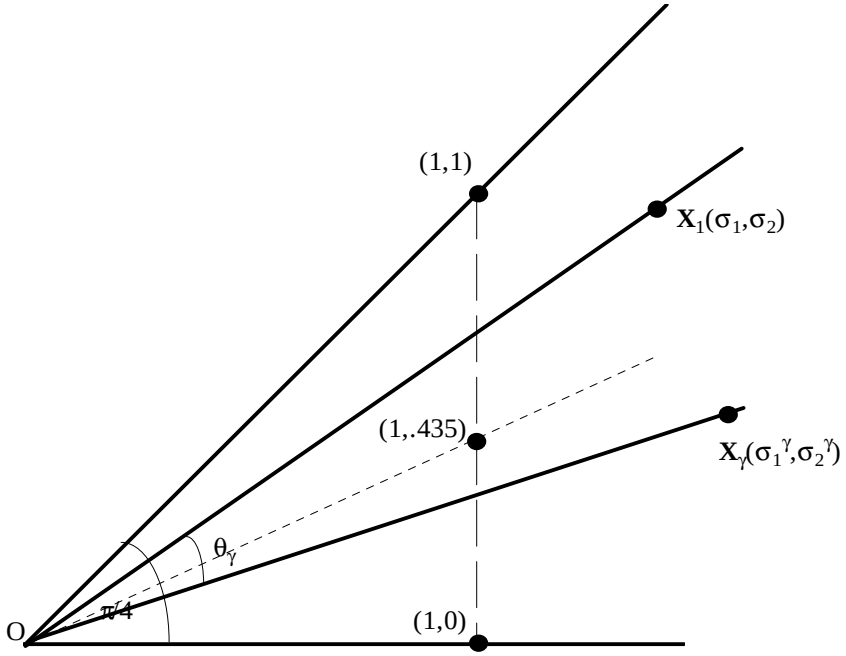

Figure 2. The cone for $r = 2$. As previously, X_1 is the optimal fit to X and X_γ is sub-optimal. The diagonal dotted line refers to the ray with the worst possible value of $\cos^2(\theta_\gamma) = 0.951$ for the restriction $\gamma = 2$. The ψ-coordinates appear on the vertical dotted line.

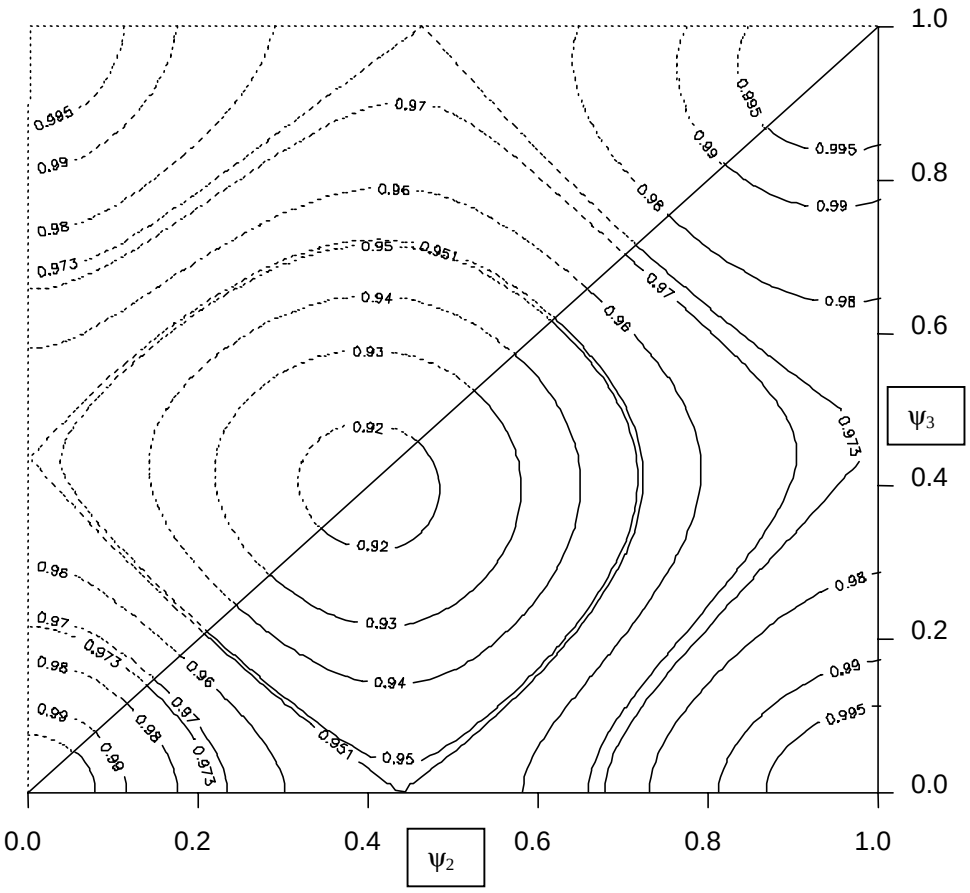

Figure 3. Contours of constant $\cos^2(\theta_\gamma)$ for $\gamma = 2$ and for different values of ψ_2 and ψ_3. The dotted half of the diagram is outside the convex cone of acceptable solutions and is shown only for convenience. A corresponding diagram for $\gamma = ½$ may be obtained by squaring the scale values of ψ_2 and ψ_3.