Biometrika (1971), 58, 3, p. 453
With 3 text-figures
Printed in Great Britain

The biplot graphic display of matrices with application to principal component analysis

By K. R. GABRIEL
The Hebrew University, Jerusalem

SUMMARY

Any matrix of rank two can be displayed as a biplot which consists of a vector for each row and a vector for each column, chosen so that any element of the matrix is exactly the inner product of the vectors corresponding to its row and to its column. If a matrix is of higher rank, one may display it approximately by a biplot of a matrix of rank two which approximates the original matrix. The biplot provides a useful tool of data analysis and allows the visual appraisal of the structure of large data matrices. It is especially revealing in principal component analysis, where the biplot can show inter-unit distances and indicate clustering of units as well as display variances and correlations of the variables.

1. Exact biplot of any rank two matrix

Any matrix may be represented by a vector for each row and another vector for each column, so chosen that the elements of the matrix are the inner products of the vectors representing the corresponding rows and columns. This is conceptually helpful in understanding properties of matrices. When the matrix is of rank 2 or 3, or can be closely approximated by a matrix of such rank, the vectors may be plotted or modelled and the matrix representation inspected physically. This is of obvious practical interest for the analysis of large matrices.

Any $n \times m$ matrix $\mathbf{Y}$ of rank r can be factorized as

$$\mathbf{Y} = \mathbf{G}\mathbf{H}' \qquad (1)$$

into a $n \times r$ matrix $\mathbf{G}$ and a $m \times r$ matrix $\mathbf{H}$, both necessarily of rank r (Rao, 1965*a*, 1*b*.2.3). This factorization is not unique. One way of factorizing $\mathbf{Y}$ is to choose the r columns of $\mathbf{G}$ as an orthonormal basis of the column space of $\mathbf{Y}$, and to compute $\mathbf{H}$ as $\mathbf{Y}'\mathbf{G}$.

Factorization (1) may be written as

$$y_{ij} = \mathbf{g}_i'\mathbf{h}_j \qquad (2)$$

for each i and j, where y_{ij} is the element in the ith row and jth column of $\mathbf{Y}$, $\mathbf{g}_i'$ is the ith row of $\mathbf{G}$ and $\mathbf{h}_j$ is the jth row of $\mathbf{H}$. In this form, the factorization assigns the vectors $\mathbf{g}_1, \ldots, \mathbf{g}_n$, one to each of the n rows of $\mathbf{Y}$ and the vectors $\mathbf{h}_1, \ldots, \mathbf{h}_m$, one to each column of $\mathbf{Y}$. Each of these vectors is of order r, and thus (2) gives a representation of $\mathbf{Y}$ by means of $n+m$ vectors in r-space. The vectors $\mathbf{g}_1, \ldots, \mathbf{g}_n$ may be considered as 'row effects' in that $\mathbf{g}_i = k\mathbf{g}_e$ means that row i is k times row e, and similarly the $\mathbf{h}_j$s may be considered as 'column effects'.

For a matrix of rank one, factorization (1) assigns scalar row effects $g_1, \ldots, g_n$ and column effects $h_1, \ldots, h_m$ and y_{ij} is simply the product $g_i h_j$. Such a matrix is therefore said to have a multiplicative structure. Contrast this with the additive structure $y_{ij} = \beta_i + \tau_j$ assumed for matrices of means in two-way analysis of variance.

In a matrix of rank two, the effects $\mathbf{g}_1, \ldots, \mathbf{g}_n$ and $\mathbf{h}_1, \ldots, \mathbf{h}_m$ are vectors of order two. These $n+m$ vectors may be plotted in the plane, giving a representation of the nm elements of $\mathbf{Y}$ by means of the inner products of the corresponding row effect and column effect vectors. Such a plot will be referred to as a *biplot* since it allows row effects and column effects to be plotted jointly. In the rest of this section only matrices $\mathbf{Y}$ of rank $r = 2$ will be considered.

The biplot represents a rank two matrix exactly, to the accuracy of plotting. This graphical representation is likely to be useful in allowing rapid visual appraisal of the structure of the matrix. An inner product of two vectors may be appraised visually by considering it as the product of the length of one of the vectors times the length of the other vector's projection onto it. This allows one to see easily which rows or columns are proportional to what other rows or columns (same directions), which entries are zero (right angles between row and column effects), etc.

To illustrate the biplot, Fig. 1 shows the graphic display of a 4×3 matrix in two different factorizations. The matrix with its alternative factorizations is

$$\begin{bmatrix} 2 & 2 & -4 \\ 2 & 1 & -3 \\ 0 & -\frac{3}{2} & \frac{3}{2} \\ -1 & -\frac{1}{2} & \frac{3}{2} \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 1 \\ 0 & -\frac{3}{2} \\ -1 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & -4 \\ 0 & -1 \\ -3 & 4\frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} -3 & -1 & 4 \\ -2 & -1 & 3 \end{bmatrix}.$$

Despite the considerable visual disparity of the two biplots it is readily seen that the 12 inner products of $\mathbf{g}$ vectors with $\mathbf{h}$ vectors are the same for both.

The disparity between the two biplots of the same matrix in Fig. 1 illustrates the non-uniqueness of factorization (1), which can be replaced by

$$\mathbf{Y} = (\mathbf{G}\mathbf{R}')(\mathbf{H}\mathbf{R}^{-1})' \tag{3}$$

for any nonsingular $\mathbf{R}$. To examine this nonuniqueness, consider the singular value decomposition of $\mathbf{R}'$,

$$\mathbf{R}' = \mathbf{V}'\mathbf{\Theta}\mathbf{W}, \tag{4}$$

where $\mathbf{V}$ and $\mathbf{W}$ are 2×2 orthonormal matrices and $\mathbf{\Theta} = \operatorname{diag}(\theta_1, \theta_2)$ and the transposed inverse is

$$\mathbf{R}^{-1} = \mathbf{V}'\mathbf{\Theta}^{-1}\mathbf{W} \tag{5}$$

(Good, 1969). Evidently transformations $\mathbf{G} \to \mathbf{G}\mathbf{R}'$ and $\mathbf{H} \to \mathbf{H}\mathbf{R}^{-1}$ each consist of a rotation of axes due to $\mathbf{V}'$, a stretching and possible reflexion along the resulting axes, and a further rotation of axes due to $\mathbf{W}$. Only the stretchings differ, the first transformation using factors θ_1 and θ_2, whereas the second uses the reciprocal factors $1/\theta_1$ and $1/\theta_2$. In the example of Fig. 1, the matrix is

$$\mathbf{R}' = \begin{bmatrix} -1 & 1 \\ 2 & -3 \end{bmatrix}$$

and one obtains

$$\mathbf{V}' = \begin{bmatrix} -0{\cdot}361 & -0{\cdot}932 \\ +0{\cdot}932 & -0{\cdot}361 \end{bmatrix},$$

$$\boldsymbol{\Theta} = \begin{bmatrix} 3{\cdot}8643 & 0 \\ 0 & 0{\cdot}2588 \end{bmatrix},$$

$$\mathbf{W} = \begin{bmatrix} 0{\cdot}576 & -0{\cdot}817 \\ 0{\cdot}817 & 0{\cdot}576 \end{bmatrix}.$$

Thus, to pass from Fig. 1(*a*) to Fig. 1(*b*) the axes are first rotated through an angle of $-68{\cdot}8^\circ = \arcsin(-0{\cdot}932)$, then the $\mathbf{g}$ co-ordinates are reflected and stretched by 3·8643

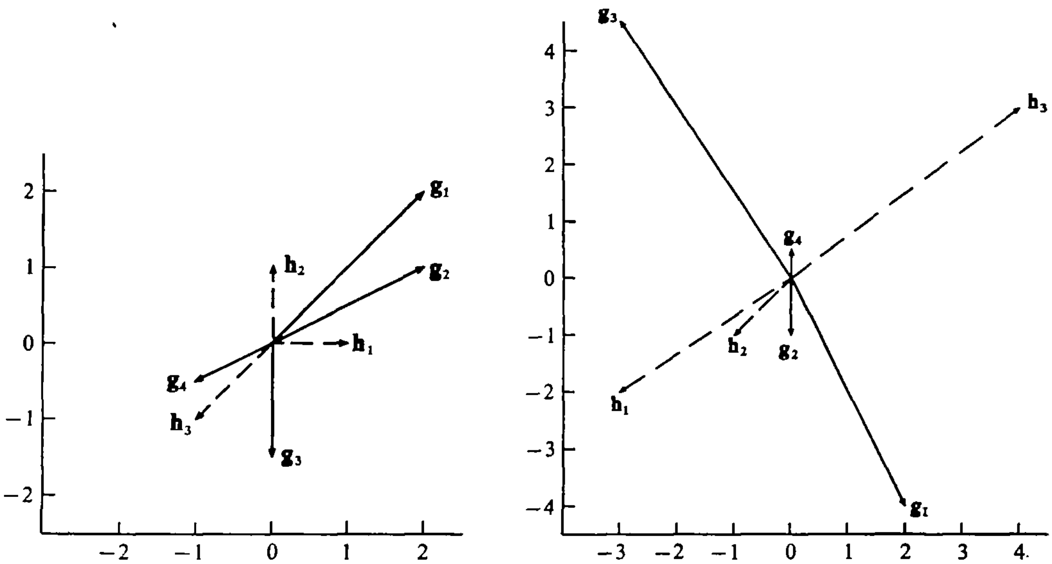

Fig. 1. Two biplots of the matrix $\begin{bmatrix} 2 & 2 & -4 \\ 2 & 1 & -3 \\ 0 & -\tfrac{3}{2} & \tfrac{3}{2} \\ -1 & -\tfrac{1}{2} & \tfrac{3}{2} \end{bmatrix}$.

and 0·2488 respectively, whereas the $\mathbf{h}$ co-ordinates are reflected and stretched by 1/3·8643 and 1/0·2488, respectively, and finally the axes are rotated again through an angle of $54{\cdot}8^\circ = \arcsin(0{\cdot}817)$. To see what happens, rotate biplot (*a*) of Fig. 1 through

$$68{\cdot}8^\circ + 180^\circ - 54{\cdot}8^\circ = 194^\circ$$

and note it to differ from biplot (*b*) only by the reciprocal stretching along two axes which are now at $-54{\cdot}8^\circ$ from the given axes, of biplot (*b*) and rotated biplot (*a*). The disparity between different factorizations (3) of $\mathbf{Y}$, and thus between the resulting biplots, as illustrated in Fig. 1, is such that relations, apart from collinearity, among the different $\mathbf{g}$ vectors, as well as among the $\mathbf{h}$ vectors, depend almost entirely on the particular factorization chosen.

To employ the biplot usefully for the inspection of relations between rows of the $\mathbf{Y}$ matrix and/or between its columns, one therefore has to impose a particular metric and make the resulting factorization and biplot unique. For example, if one wishes relations between rows of $\mathbf{Y}$ to be represented by corresponding relations of $\mathbf{g}$ vectors one may impose the requirement

$$\mathbf{H}'\mathbf{H} = \mathbf{I}_2, \tag{6}$$

which yields

$$\mathbf{YY}' = \mathbf{GG}', \tag{7}$$

so that, for any two rows $\mathbf{y}_i$ and $\mathbf{y}_e$ of $\mathbf{Y}$,

$$\mathbf{y}_i'\mathbf{y}_e = \mathbf{g}_i'\mathbf{g}_e, \tag{8}$$

$$\|\mathbf{y}_i\| = \|\mathbf{g}_i\|, \tag{9}$$

$$\cos(\mathbf{y}_i, \mathbf{y}_e) = \cos(\mathbf{g}_i, \mathbf{g}_e), \tag{10}$$

where $(\mathbf{x}, \mathbf{y})$ denotes the angle between vectors $\mathbf{x}$ and $\mathbf{y}$, and also

$$\|\mathbf{y}_i - \mathbf{y}_e\| = \|\mathbf{g}_i - \mathbf{g}_e\|. \tag{11}$$

Note that with this requirement (6),

$$\mathbf{Y}'(\mathbf{YY}')^{-}\mathbf{Y} = \mathbf{HH}', \tag{12}$$

for any conditional inverse $(\mathbf{YY}')^{-}$ of $\mathbf{YY}'$, and this is the matrix projecting onto the row space of $\mathbf{Y}$. The inner products of the $\mathbf{h}$ vectors are therefore those of the $\mathbf{Y}$ columns taken through any metric $(\mathbf{YY}')^{-}$, i.e.

$$\boldsymbol{\eta}_j'(\mathbf{YY}')^{-}\boldsymbol{\eta}_g = \mathbf{h}_j'\mathbf{h}_g, \tag{13}$$

where $\boldsymbol{\eta}_j$ and $\boldsymbol{\eta}_g$ indicate the jth and gth columns of $\mathbf{Y}$.

An alternative factorization is the one which reproduces inner products of $\mathbf{Y}$ columns by those of $\mathbf{h}$ vectors but does not do so for $\mathbf{Y}$ rows and $\mathbf{g}$ vectors. Here one would impose

$$\mathbf{G}'\mathbf{G} = \mathbf{I}_2 \tag{14}$$

instead of (6) and obtain

$$\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-}\mathbf{Y}' = \mathbf{GG}' \tag{15}$$

for any conditional inverse $(\mathbf{Y}'\mathbf{Y})^{-}$ of $\mathbf{Y}'\mathbf{Y}$, as well as the desired

$$\mathbf{Y}'\mathbf{Y} = \mathbf{HH}'. \tag{16}$$

In general, if a metric $\mathbf{M}$ is used for rows, that is, one requires

$$\mathbf{YMY}' = \mathbf{GG}', \tag{17}$$

one must choose $\mathbf{H}$ so as to satisfy

$$\mathbf{H}'\mathbf{MH} = \mathbf{I}_2, \tag{18}$$

and any conditional inverse $(\mathbf{YMY}')^{-}$ can serve as metric for the columns, giving

$$\mathbf{Y}'(\mathbf{YMY}')^{-}\mathbf{Y} = \mathbf{HH}'. \tag{19}$$

To prove (19), introduce (1) and (18) and make use of the fact that

$$\mathbf{G}'(\mathbf{GG}')^{-}\mathbf{G} = \mathbf{I}_2 \tag{20}$$

because $\mathbf{G}'(\mathbf{GG}')^{-}\mathbf{G}$ is the projection matrix onto the column space of $\mathbf{G}'$ which is readily seen to be the Euclidean space $\mathscr{R}_2$ whose projection matrix is $\mathbf{I}_2$.

Similarly, for any column metric $\mathbf{N}$ one must choose $\mathbf{G}$ so that

$$\mathbf{G}'\mathbf{NG} = \mathbf{I}_2. \tag{21}$$

Then

$$\mathbf{Y}'\mathbf{NY} = \mathbf{HH}' \tag{22}$$

as well as

$$\mathbf{Y}(\mathbf{Y}'\mathbf{NY})^{-}\mathbf{Y}' = \mathbf{GG}' \tag{23}$$

for any conditional inverse $(\mathbf{Y}'\mathbf{NY})^{-}$.

In conclusion, the biplot can be made unique, apart from rotations and reflexions, operations which do not change the relations between the vectors, by introducing the requirement of a particular metric for either row or column comparisons.

2. Approximate biplot of any matrix

Matrices of ranks higher than two cannot be represented exactly by a biplot. However, if a matrix $\mathbf{Y}$ can be satisfactorily approximated by a rank two matrix $\mathbf{Y}_{(2)}$, the biplot of $\mathbf{Y}_{(2)}$ may allow useful approximate visual inspection of $\mathbf{Y}$ itself. In such a case, the inner products of the plotted row and column effects will be approximations to the elements of $\mathbf{Y}$.

To approximate any rectangular $n \times m$ matrix $\mathbf{Y}$ of rank r by a $n \times m$ matrix of lower rank, one may use the singular value decomposition (Eckart & Young, 1939; Good, 1969; Golub & Reinsch, 1970). This is

$$\mathbf{Y} = \sum_{\alpha=1}^{r} \lambda_\alpha \mathbf{p}_\alpha \mathbf{q}'_\alpha, \tag{24}$$

where, for each $\alpha = 1, \ldots, r$, the singular value λ_α, singular column $\mathbf{p}_\alpha$ and singular row $\mathbf{q}'_\alpha$ are chosen to satisfy

$$\mathbf{p}'_\alpha \mathbf{Y} = \lambda_\alpha \mathbf{q}'_\alpha, \tag{25}$$

$$\mathbf{Y}\mathbf{q}_\alpha = \lambda_\alpha \mathbf{p}_\alpha, \tag{26}$$

$$\mathbf{Y}\mathbf{Y}'\mathbf{p}_\alpha = \lambda_\alpha^2 \mathbf{p}_\alpha, \tag{27}$$

$$\mathbf{Y}'\mathbf{Y}\mathbf{q}_\alpha = \lambda_\alpha^2 \mathbf{q}_\alpha, \tag{28}$$

$$\lambda_1 \geqslant \ldots \geqslant \lambda_r > 0, \tag{29}$$

$$\mathbf{p}'_\alpha \mathbf{p}_\epsilon = \mathbf{q}'_\alpha \mathbf{q}_\epsilon = \delta_{\alpha,\epsilon}, \tag{30}$$

$\delta_{\alpha,\epsilon}$ being Kronecker's delta. Any solution of a pair of equations (25) and (26), (25) and (27) or (26) and (28) will satisfy the remaining two equations.

The method of least squares then provides

$$\mathbf{Y}_{(s)} = \sum_{\alpha=1}^{s} \lambda_\alpha \mathbf{p}_\alpha \mathbf{q}'_\alpha \tag{31}$$

as the rank s approximation to $\mathbf{Y}$ (Householder & Young, 1938), i.e. the $n \times m$ matrix $\mathbf{M}$ of rank s which minimizes

$$\|\mathbf{Y} - \mathbf{M}\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} (y_{ij} - m_{ij})^2. \tag{32}$$

Moreover, the extent of lack of fit is measured by

$$\|\mathbf{Y} - \mathbf{Y}_{(s)}\|^2 = \lambda_{s+1}^2 + \ldots + \lambda_r^2. \tag{33}$$

Because

$$\|\mathbf{Y}\|^2 = \lambda_1^2 + \ldots + \lambda_r^2, \tag{34}$$

an absolute measure of goodness of fit can be defined as

$$\rho_s^{(2)} = 1 - \|\mathbf{Y} - \mathbf{Y}_{(s)}\|^2 / \|\mathbf{Y}\|^2 = \sum_{\alpha=1}^{s} \lambda_\alpha^2 \Big/ \sum_{\alpha=1}^{r} \lambda_\alpha^2. \tag{35}$$

Of particular importance in interpreting this least squares criterion is the fact that it is equivalent to the criterion of least squares on the differences between all rows as well as between all columns. From Rao (1965*b*)

$$2n\|\mathbf{Y} - \mathbf{M}\|^2 = \sum_{i+e=1}^{n} \|(\mathbf{y}_i - \mathbf{y}_e) - (\mathbf{m}_i - \mathbf{m}_e)\|^2, \tag{36}$$

if

$$\mathbf{1}'\mathbf{Y} = \mathbf{1}'\mathbf{M} = \mathbf{0}', \tag{37}$$

that is, if the columns of $\mathbf{Y}$ and of $\mathbf{M}$ sum to zero.

It follows that $\mathbf{Y}_{(s)}$ is the rank s matrix whose row differences best approximate the row differences of the matrix $\mathbf{Y}$, and they do so with goodness of fit $\rho_s^{(2)}$. The same applies for columns.

The approximate biplot of $\mathbf{Y}$ is then the exact biplot of

$$\mathbf{Y}_{(2)} = [\mathbf{p}_1, \mathbf{p}_2]\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}\begin{bmatrix}\mathbf{q}_1'\\ \mathbf{q}_2'\end{bmatrix} \tag{38}$$

and its goodness of fit is measured by

$$\rho_2^{(2)} = (\lambda_1^2+\lambda_2^2)\Big/\sum_{\alpha=1}^{r}\lambda_\alpha^2. \tag{39}$$

If $\rho_2^{(2)}$ is near to one, such a biplot will give a good approximation of $\mathbf{Y}$.

In choosing, as in (1), factors $\mathbf{G}$ and $\mathbf{H}$ of $\mathbf{Y}_{(2)}$ for biplotting, one may use the factorization provided by the singular decomposition (38). Writing

$$\mathbf{p}_\alpha' = (p_{\alpha 1}, \ldots, p_{\alpha n}), \quad \mathbf{q}_\alpha' = (q_{\alpha 1}, \ldots, q_{\alpha m}),$$

one obtains

$$\mathbf{Y}_{(2)} = \begin{bmatrix}p_{11} & p_{21}\\ \vdots & \vdots\\ p_{1n} & p_{2n}\end{bmatrix}\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}\begin{bmatrix}q_{11}\cdots q_{1m}\\ q_{21}\cdots q_{2m}\end{bmatrix}. \tag{40}$$

One choice of $\mathbf{G}$ and $\mathbf{H}$ would be in terms of rows

$$\begin{aligned}\mathbf{g}_i' &= (\sqrt{\lambda_1}p_{1i}, \sqrt{\lambda_2}p_{2i}) \quad (i = 1, \ldots, n),\\ \mathbf{h}_j' &= (\sqrt{\lambda_1}q_{1j}, \sqrt{\lambda_2}q_{2j}) \quad (j = 1, \ldots, m).\end{aligned} \tag{41}$$

Other choices of $\mathbf{G}$ and $\mathbf{H}$ are obtained by defining

$$\begin{aligned}\mathbf{g}_i' &= (p_{1i}, p_{2i}) \quad (i = 1, \ldots, n),\\ \mathbf{h}_j' &= (\lambda_1 q_{1j}, \lambda_2 q_{2j}) \quad (j = 1, \ldots, m),\end{aligned} \tag{42}$$

which satisfies requirement (14), or as

$$\begin{aligned}\mathbf{g}_i' &= (\lambda_1 p_{1i}, \lambda_2 p_{2i}) \quad (i = 1, \ldots, n),\\ \mathbf{h}_j' &= (q_{1j}, q_{2j}) \quad (j = 1, \ldots, m),\end{aligned} \tag{43}$$

which satisfies requirement (6).

As an illustration consider the data $\mathbf{Y}$ in Table 1 showing percentages of households having various facilities and appliances in East Jerusalem Arab areas, by quarters of the town. I am obliged to Israel Sauerbrun for bringing this example to my attention. The average percentages in each quarter indicate the standard of living of that area and the average percentage of each facility or appliance its over-all prevalence. With a multiplicative model, such averages are fitted by least squares as the first singular component $\mathbf{Y}_{(1)}$. To study the differential prevalence of different facilities and appliances in the different quarters, this first component was subtracted out, leaving $\mathbf{Y}-\mathbf{Y}_{(1)}$ in Table 2. Note that

$$\mathbf{Y}-\mathbf{Y}_{(1)} = \sum_{\alpha=2}^{8}\lambda_\alpha^2\mathbf{p}_\alpha\mathbf{q}_\alpha'$$

is the singular decomposition of that residual matrix.

Table 1. *Facilities and equipment in East Jerusalem in* 1967, *by subquarter, from Israel* (1968)

Percentage of households possessing:	Old city quarters				Modern		Other		Rural
	Christian	Armenian	Jewish	Moslem	Amer. Colony Sh. Jarah	Shaafat Bet-Hanina	A-Tur Isawiye	Silwan Abu-Tor	Sur-Bahar Bet-Safafa
Toilet	98·2	97·2	97·3	96·9	97·6	94·4	90·2	94·0	70·5
Kitchen	78·8	81·0	65·6	73·3	91·4	88·7	82·2	84·2	55·1
Bath	14·4	17·6	6·0	9·6	56·2	69·5	31·8	19·5	10·7
Electricity	86·2	82·1	54·5	74·7	87·2	80·4	68·6	65·5	26·1
Water*	32·9	30·3	21·1	26·9	80·1	74·3	46·3	36·2	9·8
Radio†	73·0	70·4	53·0	60·5	81·2	78·0	67·9	64·8	57·1
TV set	4·6	6·0	1·5	3·4	12·7	23·0	5·6	2·7	1·3
Refrigerator‡	29·2	26·3	4·3	10·5	52·8	49·7	21·7	9·5	1·2

* In dwelling. † Or transistor radio. ‡ Electric.

Table 2. *Residual matrix after subtracting out multiplicative least squares fit (first singular component) from Table* 1, *with next two singular values, rows and columns*

	Christian	Armenian	Jewish	Moslem	American	Shaafat	A-Tur	Silwan	Sur-Bahar	$\mathbf{p}_2$	$\mathbf{p}_3$
Toilet	1·60	2·17	21·16	10·62	−17·12	−16·72	−2·44	4·81	12·64	0·394	0·185
Kitchen	−3·11	0·43	1·04	0·15	−5·87	−5·52	3·65	8·58	6·05	0·107	0·213
Bath	−15·49	−11·80	−17·56	−17·09	20·71	35·12	3·14	−8·09	−7·20	−0·575	0·371
Electricity	11·71	8·83	−4·21	8·18	−1·25	−5·28	−2·83	−3·27	−18·51	0·023	−0·768
Water	−11·66	−13·53	−14·02	−12·90	27·19	23·04	3·57	−4·94	−16·89	−0·525	0·059
Radio	2·30	0·85	−2·73	−2·65	−2·76	−3·33	0·10	−0·47	14·76	0·071	0·244
TV set	−3·27	−1·74	−4·70	−3·63	3·36	13·95	−1·94	−4·56	−3·41	−0·173	0·042
Refrigerator	2·78	0·31	−16·53	−13·10	21·42	19·31	−3·64	−14·89	−14·62	−0·437	−0·358
$\mathbf{q}_2'$	0·171	0·172	0·381	0·307	−0·495	−0·574	−0·027	0·195	0·297	$\lambda_2 = 88·35$	
$\mathbf{q}_3'$	−0·486	−0·340	0·151	−0·223	−0·070	0·209	0·152	0·207	0·679	$\lambda_3 = 33·67$	

The residual matrix $\mathbf{Y}-\mathbf{Y}_{(1)}$ corresponds to interaction residuals after fitting an additive model in two-way analysis of variance. Thus, for example, the large positive values for toilets and for radios in Sur-Baher and Bet-Safafa do not indicate a higher prevalence of toilets and radios in that rural area than in the Eastern city as a whole; see Table 1. It indicates that, relative to the general paucity of facilities and appliances in that area, toilets and radios are not as rare as other items.

Along with the residual matrix $\mathbf{Y}-\mathbf{Y}_{(1)}$, Table 2 also shows its first two singular values λ_2 and λ_3, columns $\mathbf{p}_2$ and $\mathbf{p}_3$ and rows $\mathbf{q}_2'$ and $\mathbf{q}_3'$. The goodness of fit of the second and third singular components to the residual matrix $\mathbf{Y}-\mathbf{Y}_{(1)}$ is

$$(\lambda_2^2+\lambda_3^2)\Big/\sum_{\alpha=2}^{8}\lambda_\alpha^2 = 0{\cdot}937,$$

so that the matrix of rank 2 should give a very close approximation to the residuals under consideration. The biplot of Fig. 2 has been constructed from these values by means of factorization (43). This factorization was considered appropriate since in the resulting biplot the relations between appliances would be approximated by the relations between the corresponding $\mathbf{g}$ vectors; see (6) to (11). Similarly distances between $\mathbf{h}$ vectors would approximate standardized statistical distances between subquarters; see (12) and (13).

Inspection of Fig. 2 shows the Old City quarters to be opposite a cluster of the modern quarters. The one rural area is roughly orthogonal to both clusters, somewhat nearer to the Old City than to the modern quarters, whereas the other quarters are less prominent in a similar direction, with the poorer Silwan and Abu-Tor areas closer to the poorest of the Old City quarters, and the richer A-Tur and Isawiye slightly in the direction of the modern quarters.

The modern quarters appear to have a particularly high prevalence of baths, water inside the dwelling and refrigerators, whereas the poorer quarters have relatively high prevalences of toilets and electricity. Evidently the last two items were pretty generally available in all urban sections and thus are not indicative of better living conditions, whereas the former three items were much more available in better off homes.

It is interesting to note that electricity is noticeably rarer in the rural area than in all urban quarters, whereas radios, presumably battery operated, and kitchens are the items which least reflect the low general level of the rural area.

In the present example, attention was focused on residuals from a multiplicative fit so that the second and third components were biplotted. In other instances it might be more interesting to biplot the first two components and study the data matrix itself.

3. Principal component biplot

A $n \times m$ matrix $\mathbf{Y}$ of observations of n units on m variables is considered, in which the mean of each variable has been subtracted out, i.e. (37) is satisfied. Then

$$\mathbf{S} = \frac{1}{n}\mathbf{Y}'\mathbf{Y} \tag{44}$$

is the corresponding m-variate estimated variance-covariance matrix. A standardized measure of the distance between the ith and eth units is given by

$$d_{i,e}^2 = (\mathbf{y}_i-\mathbf{y}_e)'\,\mathbf{S}^{-1}(\mathbf{y}_i-\mathbf{y}_e) \tag{45}$$

(Seal, 1964, pp. 126–7).

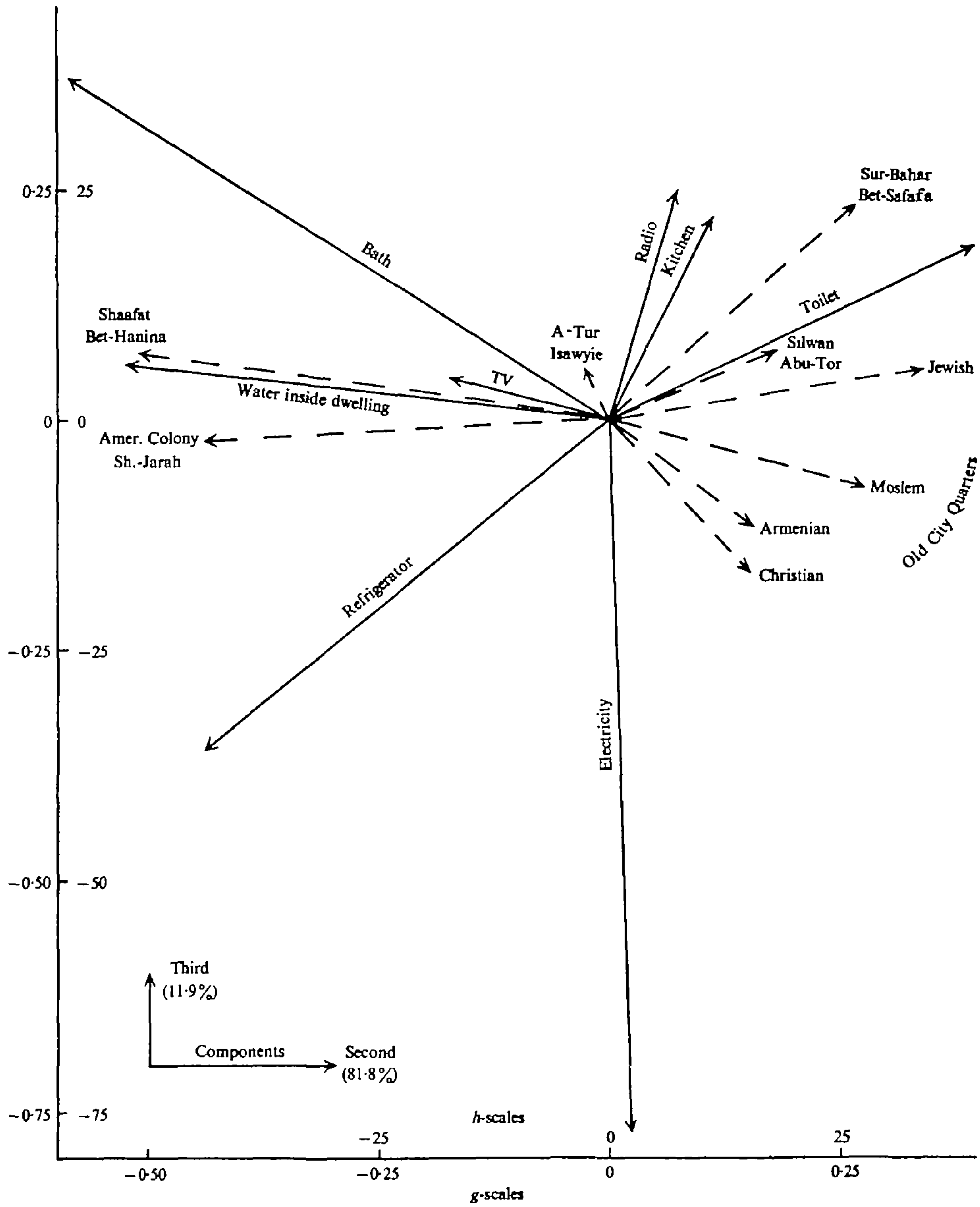

Fig. 2. Differential prevalence of facilities and equipment in East Jerusalem households (Arab) in 1967.

Principal component analysis consists of singular decomposition of such a matrix **Y** (Whittle, 1952). Note that (28) becomes

$$n\mathbf{S}\mathbf{q}_\alpha = \lambda_\alpha^2 \mathbf{q}_\alpha, \tag{46}$$

the usual form of the equations for principal component analysis, except that the factor n is often omitted and λ_α^2/n computed instead of λ_α^2.

In view of (25), the singular rows $\mathbf{q}'_\alpha$ are seen to be weighted sums of the actual n rows. Similarly, by (26), the canonical columns $\mathbf{p}_\alpha$ are weighted sums of the actual m columns.

The singular decomposition (24) shows that matrix **Y** can be factorized as

$$\mathbf{Y} = (\mathbf{p}_1, \ldots, \mathbf{p}_r)(\lambda_1 \mathbf{q}_1, \ldots, \lambda_r \mathbf{q}_r)'. \tag{47}$$

This factorization has, in view of (30), the following properties:

$$(\mathbf{p}_1, \ldots, \mathbf{p}_r)'(\mathbf{p}_1, \ldots, \mathbf{p}_r) = \mathbf{I}_r, \tag{48}$$

$$(\mathbf{p}_1, \ldots, \mathbf{p}_r)(\mathbf{p}_1, \ldots, \mathbf{p}_r)' = \frac{1}{n}\mathbf{Y}\mathbf{S}^{-1}\mathbf{Y}', \tag{49}$$

$$(\lambda_1 \mathbf{q}_1, \ldots, \lambda_r \mathbf{q}_r)'(\lambda_1 \mathbf{q}_1, \ldots, \lambda_r \mathbf{q}_r) = \operatorname{diag}(\lambda_1, \ldots, \lambda_r), \tag{50}$$

$$(\lambda_1 \mathbf{q}_1, \ldots, \lambda_r \mathbf{q}_r)(\lambda_1 \mathbf{q}_1, \ldots, \lambda_r \mathbf{q}_r)' = n\mathbf{S}. \tag{51}$$

Now consider the rank two approximation $\mathbf{Y}_{(2)}$ of (38) and, for the purpose of biplotting, choose

$$\mathbf{G} = (\mathbf{p}_1, \mathbf{p}_2)\sqrt{n}, \qquad \mathbf{H} = \frac{1}{\sqrt{n}}(\lambda_1 \mathbf{q}_1, \lambda_2 \mathbf{q}_2), \tag{52}$$

which, apart from a constant factor, consists of introducing requirement (14). Write $\sim$ for 'is approximated by means of a least squares fit of rank two'. Then (47) to (51) yield

$$\mathbf{Y} \sim \mathbf{G}\mathbf{H}', \tag{53}$$

$$\mathbf{Y}\mathbf{S}^{-1}\mathbf{Y}' \sim \mathbf{G}\mathbf{G}', \tag{54}$$

$$\mathbf{S} \sim \mathbf{H}\mathbf{H}', \tag{55}$$

(54) and (55) corresponding to (15) and (16).

Any approximate biplot of **Y**, or exact biplot of $\mathbf{Y}_{(2)}$, allows the following approximations: the individual observations

$$y_{ij} \sim \mathbf{g}'_i \mathbf{h}_j, \tag{56}$$

the ith and eth units' difference on variable j

$$y_{ij} - y_{ej} \sim (\mathbf{g}_i - \mathbf{g}_e)'\mathbf{h}_j, \tag{57}$$

the ith unit's difference between variables j and g

$$y_{ij} - y_{ig} \sim \mathbf{g}'_i(\mathbf{h}_j - \mathbf{h}_g), \tag{58}$$

the ith and eth units' interaction with variables j and g

$$y_{ij} - y_{ej} - y_{ig} + y_{eg} \sim (\mathbf{g}_i - \mathbf{g}_e)'(\mathbf{h}_j - \mathbf{h}_g). \tag{59}$$

All these follow from (53).

The biplot of **Y** with particular choice (52) of **G** and **H** allows additional approximations. From (54), one approximates the standardized distance (45) between the ith and eth units by means of

$$d_{i,e} \sim \|\mathbf{g}_i - \mathbf{g}_e\|. \tag{60}$$

Also, from (55), approximations of covariances, variances and correlations of the m variables are given by

$$s_{j,g} \sim \mathbf{h}_j'\mathbf{h}_g, \tag{61}$$

$$s_j^2 \sim \|\mathbf{h}_j\|^2, \tag{62}$$

$$r_{j,g} \sim \cos(\mathbf{h}_j, \mathbf{h}_g), \tag{63}$$

$s_{j,g}$ being the j, gth element of matrix S and $r_{j,g} = s_{j,g}/\sqrt{(s_{j,g}s_{g,g})}$. The expression

$$\frac{1}{n}\sum_{i=1}^{n}(y_{ij}-y_{ig})^2 \sim \|\mathbf{h}_j - \mathbf{h}_g\|^2 \tag{64}$$

gives an approximation to the average squared difference between variables.

This particular choice of approximate biplot for **Y** therefore not only allows one to view the individual observations and their differences, but further permits one to scan the standardized differences between units and to inspect the variances, covariances and correlations of the variables. This is likely to provide a most useful graphical aid in interpreting multivariate matrices of observations, provided, of course, that these can be adequately approximated at rank two.

The elements of **Y** are biplotted with goodness of fit

$$\rho_2^{(2)} = (\lambda_1^2+\lambda_2^2)\Big/\sum_{\alpha=1}^{r}\lambda_\alpha^2, \tag{65}$$

as was pointed out in §2. The elements of **S**, however, are biplotted, (61) and (62), with even better fit

$$\rho_2^{(4)} = (\lambda_1^4+\lambda_2^4)\Big/\sum_{\alpha=1}^{r}\lambda_\alpha^4, \tag{66}$$

as will be shown. On the other hand, the standardized distances $d_{i,e}$ of (45) are approximated only to the extent of

$$\rho_2^{(0)} = (\lambda_1^0+\lambda_2^0)\Big/\sum_{\alpha=1}^{r}\lambda_\alpha^0 = 2/r \tag{67}$$

on the biplot, as is shown next. Therefore, whereas the matrix elements themselves and the variances, covariances and correlations may often be excellently represented in the biplot, the standardized distances are not well represented. In fact the biplot distances must be regarded as distances standardized in the plane of best fit, rather than as approximations to standardized distances in the entire r space, which latter cannot be approximated any better on a plane. Such a standardized planar distance may indeed be a more attractive measure than the wholly standardized distance which gives equal weight to all dimensions; see also Rao (1952, § 9*c*).

To consider the approximation of distances $d_{i,e}$ consider the canonical decomposition

$$\mathbf{Y}\mathbf{S}^{-\frac{1}{2}} = [\mathbf{p}_1, \ldots, \mathbf{p}_r]\begin{bmatrix}\sqrt{n} & & 0\\ & \ddots & \\ 0 & & \sqrt{n}\end{bmatrix}\begin{bmatrix}\mathbf{q}_1'\\ \vdots\\ \mathbf{q}_r'\end{bmatrix}, \tag{68}$$

which is readily checked from (47) and (51). Now, noting (37), use (36) to write

$$\frac{1}{2}\sum_{i\neq e=1}^{n}\left\|(\mathbf{y}_i-\mathbf{y}_e)'\,\mathbf{S}^{-\frac{1}{2}} - (\mathbf{g}_i-\mathbf{g}_e)'\begin{bmatrix}\mathbf{q}_1'\\ \mathbf{q}_2'\end{bmatrix}\right\|^2 = n^2(r-2), \tag{69}$$

$\mathbf{g}$'s being rows of $\mathbf{G}$ of (52a). Also

$$\frac{1}{2}\sum_{i\neq e=1}^{n} d_{i,e}^2 = \frac{1}{2}\sum_{i\neq e=1}^{n} \|(\mathbf{y}_i-\mathbf{y}_e)'\,\mathbf{S}^{-\frac{1}{2}}\|^2 = n^2 r, \tag{70}$$

so that (67) is established as a measure of goodness of fit. Also, in view of the least squares argument in §2, it is clear that no other vectors of order 2 can approximate the $\mathbf{y}_i'\mathbf{S}^{-\frac{1}{2}}$ differences better than the $\mathbf{g}_i$'s do.

Strictly, the above argument concerns the goodness of fit of the differences $(\mathbf{y}_i-\mathbf{y}_e)'\,\mathbf{S}^{-\frac{1}{2}}$ by differences

$$(\mathbf{g}_i-\mathbf{g}_e)'\begin{bmatrix}\mathbf{q}_1'\\ \mathbf{q}_2'\end{bmatrix}$$

rather than that of their lengths $d_{i,e}$ by the lengths $\|\mathbf{g}_i-\mathbf{g}_e\|$.

Next, to consider the goodness of fit of the variances and covariances note that the biplot of $\mathbf{Y}$ with factors (52) gives the same plot of vectors $\mathbf{h}$ for variables as the corresponding biplot of the variance-covariance matrix $\mathbf{S}$. To see this, note from (47), (24) and (30) that

$$\mathbf{S} = \frac{1}{n}\sum_{\alpha=1}^{r}\lambda_\alpha^2\,\mathbf{q}_\alpha\mathbf{q}_\alpha' \tag{71}$$

and this is readily checked to be the singular decomposition of $\mathbf{S}$. Since $\mathbf{S}$ is symmetric, this is also its spectral decomposition (Good, 1969). It follows that

$$\mathbf{S}_{(2)} = \frac{1}{\sqrt{n}}(\lambda_1\mathbf{q}_1,\lambda_2\mathbf{q}_2)\frac{1}{\sqrt{n}}\begin{bmatrix}\lambda_1\mathbf{q}_1'\\ \lambda_2\mathbf{q}_2'\end{bmatrix}, \tag{72}$$

so that the biplot of $\mathbf{S}$ may be reduced to the plot of a single set of vectors $\mathbf{h}_1,\ldots,\mathbf{h}_m$ which are the rows of matrix $(1/\sqrt{n})\,(\lambda_1\mathbf{q}_1,\lambda_2\mathbf{q}_2)$. But this is exactly the choice of $\mathbf{H}$ in (52), proving the equivalence of the $\mathbf{h}$'s in the biplots of $\mathbf{Y}$ and of $\mathbf{S}$. The measure of goodness of fit (39) for the plot of $\mathbf{S}$ becomes $\rho_2^{(4)}$ of (66) because in (71) λ_α^2's play the role of the singular values whereas in (24) λ_α's played that role.

The plot of vectors $\mathbf{h}$ for variables, based on the decomposition of $\mathbf{S}$, is not novel. Hills (1969) points out that for standardized data, i.e. each column standardized to have unit variance, the inter-variable squared distance (64) provides the approximation

$$2(1-r_{jg}) \sim \|\mathbf{h}_j-\mathbf{h}_g\|^2. \tag{73}$$

The biplot of vectors for units jointly with vectors for variables, and the particular choice (52) of factors for principal component analysis are apparently novel. It is interesting to note, however, that Bennett (1956) was aware of the possibility of a similar plot.

An alternative biplot of $\mathbf{Y}_{(2)}$ may be obtained by choosing

$$\begin{aligned}\mathbf{G} &= (\lambda_1\mathbf{p}_1,\lambda_2\mathbf{p}_2),\\ \mathbf{H} &= (\mathbf{q}_1,\mathbf{q}_2),\end{aligned} \tag{74}$$

which is equivalent to introducing requirement (6), so that properties (7)–(13) hold. This may be of interest when the quantities in the different columns of $\mathbf{Y}$ are of a similar nature and it is preferred to compare rows of $\mathbf{Y}$ by giving all their elements the same weight, and not weights inverse to the variance-covariance matrix. In other words, factorization (74) is appropriate if we prefer to approximate the simple distance

$$\|\mathbf{y}_i-\mathbf{y}_e\| \sim \|\mathbf{g}_i-\mathbf{g}_e\|, \tag{75}$$

as in (11), instead of the standardized distance

$$d_{i,e} \sim \|\mathbf{g}_i - \mathbf{g}_e\|, \tag{76}$$

as in (39) and (53). This would, however, invalidate approximations (61)–(64) to the variance, covariances and correlations, and introduce instead something like (12).

As noted in §1, different biplots may be obtained with different metrics. Thus, for example, $\mathbf{N} = (1/n)\,\mathbf{I}_n$ and/or $\mathbf{M} = \mathbf{S}^{-1}$ has given choice (52) corresponding to (14), whereas $\mathbf{M} = \mathbf{I}_m$ gives choice (74) corresponding to (6). Another choice commonly used for standardization is $\mathbf{M}^{-1} = \mathrm{diag}\,(s_{1,1}, \ldots, s_{m,m})$.

Table 3. *New variables $X_{i,j}$ determined from $Z_{i,j}$*

i	$X_{i,1}$	$X_{i,2}$	$X_{i,3}$	$X_{i,4}$
1–10	$Z_{i,1}+10$	$Z_{i,1}+Z_{i,2}$	$-Z_{i,1}-Z_{i,2}-Z_{i,3}$	$Z_{i,1}+\ldots+Z_{i,4}$
11–20	$Z_{i,1}$	$Z_{i,1}+Z_{i,2}$	$-Z_{i,1}-Z_{i,2}-Z_{i,3}$	$Z_{i,1}+\ldots+Z_{i,4}$
21–30	$Z_{i,1}$	$Z_{i,1}+Z_{i,2}$	$-Z_{i,1}-Z_{i,2}-Z_{i,3}$	$Z_{i,1}+\ldots+Z_{i,4}+10$

To illustrate the principal component biplot, first type, with choice of factors (52), an artificial 4-variate example has been constructed. I am obliged to Mrs Irith Hocherman for constructing and analyzing this example. One hundred and twenty independent $N(0,1)$ variables

$$Z_{i,j} \quad (i = 1, \ldots, 30; j = 1, \ldots, 4)$$

were generated and four new variables $X_{i,j}$ were computed; see Table 3.

The first two components were found to provide a goodness of fit of

$$\rho_2^{(2)} = (\lambda_1^2 + \lambda_2^2) \Big/ \sum_{\alpha=1}^{4} \lambda_\alpha^2 = 0{\cdot}9425$$

to the 30×4 matrix of deviations from the 4-variate means. The biplot is shown in Fig. 3 with vectors $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{h}_3$ and $\mathbf{h}_4$ labelled as such and the end-points of vectors $\mathbf{g}_1, \ldots, \mathbf{g}_{30}$ indicated by the unit indices $1, \ldots, 30$; the $\mathbf{g}$ vectors have not here been plotted as lines.

Some of the features of the data that can be seen in this biplot are the following. The standard deviations of variables X_1 and X_4 are much larger than those of variables X_2 and X_3; this is evident from the lengths of the $\mathbf{h}$ vectors (55), exactly as one would expect from the factor 10 added to one third of the observations on those variables. From the angles between the $\mathbf{h}$ vectors one concludes by (54) that X_2 is positively correlated with X_1 and X_4 and negatively with X_3. All other correlations are slightly negative, as one would expect from the construction of the variables.

Inspection of the $\mathbf{g}$ vectors clearly shows these to fall into three quite distinct clusters, corresponding to the three types of units constructed. Units 1–10 are seen by (56) to have large positive deviations on X_1, no noticeable deviations on X_2 and X_3 and somewhat lesser negative deviations on X_4. Units 11–20 have altogether small deviations, negative on X_1 and X_4. Finally, units 21–30 have noticeable negative deviations on X_1 and quite sizeable positive ones on X_4. The average distance between units 1–10 and units 21–30 is about the same as that of either of these sets and units 11–20. This also agrees with the construction of these observations.

Another use of the biplot is in looking for linear combinations of variables with certain characteristics. Thus, the linear combination which maximizes the mean difference between units 11–20 versus the rest of the units is roughly $X_1 + X_4$, whose $\mathbf{h}$ vector is simply $\mathbf{h}_1 + \mathbf{h}_4$.

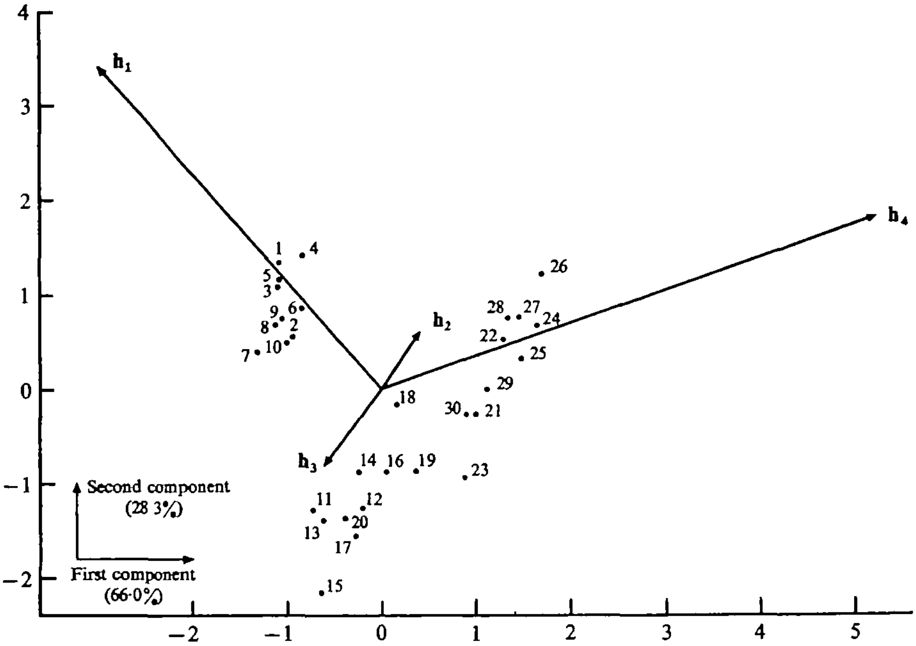

Fig. 3. Artificial 4-variate example of principal component analysis with three types of units ($\mathbf{h}_j$ represents variable X_j; number i, giving vector $\mathbf{g}_i$, represents unit i). $X_{i,1} = Z_1 + 10\delta_{i,1-10}$, $X_{i,2} = Z_1 + Z_2$, $X_{i,3} = -(Z_1 + Z_2 + Z_3)$, $X_{i,4} = Z_1 + \ldots + Z_4 + 10\delta_{i,21-30}$.

4. Extension to three dimensions and methods of computation

Factorization (1) for matrices of rank three can be represented by a bimodel consisting of spokes $\mathbf{g}_1, \ldots, \mathbf{g}_n$, $\mathbf{h}_1, \ldots, \mathbf{h}_m$ from a common origin. The interpretation of such a three dimensional model is analogous to that of the two dimensional biplot. For matrices of rank 3 or more, it would provide a better approximation than the biplot and might be worth constructing if λ_3 is large enough relative to the other roots.

Any program for principal component analysis may be used to obtain the $\mathbf{q}$ vectors as well as the λ roots from (46). The $\mathbf{p}$ vectors can then be calculated from (25) and the coordinates for plotting are available.

A special program CANDEC which carries out the singular decomposition for various types of input matrices is available from the author. This program is written in FORTRAN IV and has been run with a large variety of data on a CDC 6400 computer.

For efficient methods of computing the singular decomposition, especially for the smaller roots, see Golub & Reinsch (1970).

The development of the ideas underlying this paper and its formulation owe much to the critical insight of Dan Bardu and L. C. A. Corsten with whom this work was discussed in detail. I am also obliged to J. Putter and W. J. Hall for their helpful comments on an earlier version of this paper.

This research was supported by a Grant from the U.S. National Center for Health Statistics.

REFERENCES

BENNETT, J. F. (1956). Determination of the number of independent parameters of a score matrix from the examination of rank orders. *Psychometrika* **21**, 383–93.

ECKART, C. & YOUNG, G. (1939). A principal axis transformation for non-Hermitian matrices. *Am. Math. Soc. Bull.* **45**, 118–21.

GOLUB, G. H. & REINSCH, C. H. (1970). Singular value decomposition and least squares solution. *Numer. Math.* **14**, 403–20.

GOOD, I. J. (1969). Some applications of the singular decomposition of a matrix. *Technometrics* **11**, 823–31.

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

HOUSEHOLDER, A. S. & YOUNG, G. (1938). Matrix approximation and latent roots. *Am. Math. Monthly* **45**, 165–71.

ISRAEL, Central Bureau of Statistics (1968). *Statistical Abstract*, no. 19, Jerusalem, Government Press.

RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York: Wiley.

RAO, C. R. (1965*a*). *Linear Statistical Inference and Its Applications*. New York: Wiley.

RAO, C. R. (1965*b*). The use and interpretation of principal component analysis in applied research. *Sankhyā* A **26**, 329–58.

SEAL, H. L. (1964). *Multivariate Statistical Analysis for Biologists*. London: Methuen.

WHITTLE, P. (1952). On principal components and least square methods of factor analysis. *Skand. Aktuar.* **35**, 233–9.

[*Received December* 1970. *Revised June* 1971]

Some key words: Graphical representation of data matrix; Principal components; Cluster analysis; Singular value decomposition.