PSYCHOMETRIKA—VOL. 55, NO. 3, 519–531
SEPTEMBER 1990

INTERPRETING CANONICAL CORRELATION ANALYSIS THROUGH BIPLOTS OF STRUCTURE CORRELATIONS AND WEIGHTS

CAJO J. F. TER BRAAK

AGRICULTURAL MATHEMATICS GROUP

RESEARCH INSTITUTE FOR NATURE MANAGEMENT

This paper extends the biplot technique to canonical correlation analysis and redundancy analysis. The plot of structure correlations is shown to be optimal for displaying the pairwise correlations between the variables of the one set and those of the second. The link between multivariate regression and canonical correlation analysis/redundancy analysis is exploited for producing an optimal biplot that displays a matrix of regression coefficients. This plot can be made from the canonical weights of the predictors and the structure correlations of the criterion variables. An example is used to show how the proposed biplots may be interpreted.

Key words: biplot, canonical correlation analysis, canonical weight, interbattery factor analysis, partial analysis, redundancy analysis, regression coefficient, reduced rank regression, structure correlations.

Introduction

Despite its elegant theory, canonical correlation analysis has yielded few useful applications (Kendall, 1975; Thompson, 1984). The reason for this are reviewed in depth by Thorndike and Weiss (1973) and Israëls (1987). One major reason is the difficulty of interpretation of the canonical variates. Should the interpretation be based on the canonical weights or on the structure correlations? (Structure correlations are the correlations of the original variables with the canonical variates.) Meredith (1964), and many textbooks thereafter, found the structure correlations "very enlightening", but more recently Rencher (1988) argued that they are "not useful" and "redundant" because they "merely reproduce univariate statistics". This paper abstains from interpreting the canonical variates. Instead, they are used as a means to produce graphical displays of the relationships between the two sets of variables. Plots of the structure correlations have been proposed (Caillez & Pagès, 1976; Israëls, 1987; van der Geer, 1986) and have been used now and then (e.g., van der Burg & de Leeuw, 1983), but it has not fully been explained how to read the plot and what its optimality properties are. By using a matrix approximation approach (Corsten, 1976; Rao, 1980) in conjunction with the biplot technique (Gabriel, 1971, 1982), we show that the plot yields (by way of scalar inner products) approximate values of the correlations between the variables of the one set and those of the other set, and that the approximation is best in a weighted least-squares sense. The optimality of the plot of structure correlations relates to the factorization of the between-set correlation matrix proposed earlier by McKeon (1966, p. 7) and Rao (1975, p. 585) and to interbattery factor analysis (Browne, 1979). In an example, we give explicit and easy-to-use rules for reading the plot.

The case where two sets of variables should be treated symmetrically is rare in

I am indebted to L. C. A. Corsten, K. R. Gabriel, A. Z. Israëls, J. M. F. ten Berge, H. A. L. Kiers, J. de Bree and A. A. M. Jansen for comments on the manuscript.

Requests for reprints should be sent to Cajo J. F. ter Braak, Agricultural Mathematics Group, Box 100, 6700 AC Wageningen, THE NETHERLANDS.

0033-3123/90/0900-89109$00.75/0

practice (Gifi, 1981; Israëls, 1987). Often, one set contains criterion variables to be predicted by the second. Multivariate regression is then an apt tool yielding a matrix of regression coefficients and a matrix of fitted values for the criterion variables. Rather than inspecting the full numerical output of such an analysis, one may want to display the main patterns in a graph. We show that a joint plot of the weights of the predictor variables and the structure correlations of criterion variables is useful for this; this plot displays the approximate values of the (standardized) regression coefficients by way of scalar innerproducts. The plot is, again, optimal in a weighted least-squares sense and can even be enriched to display the approximate significance of the regression coefficients. The fitted values for the criterion variables can also optimally be displayed by adding the canonical variate scores for the predictor set to the plot. This plot has its basis in the asymmetric interpretation of canonical correlation analysis as investigated under the name of reduced-rank regression by Anderson (1951, 1984), Izenman (1975), Tso (1981), and Velu, Reinsel and Wichern (1986). Redundancy analysis (van den Wollenberg, 1977) is thus not the only multivariate technique allowing an asymmetric treatment of the sets! We point out the distinction between these techniques and propose an intermediate technique that differs from the intermediate proposal by DeSarbo (1981).

With the graphical potential of canonical techniques fully exploited, they are likely to find more useful applications in the future than they have in the past. After deriving the optimality properties of structure correlations and weights for factoring matrices of correlations and of regression coefficients, we illustrate the use of the proposed plots. Thereafter, the results of the paper are extended to partial canonical correlation analysis and to redundancy analysis.

Theory

Let $\mathbf{X}$ and $\mathbf{Y}$ be real matrices of order $n \times p$ and $n \times q$, containing n observations of p predictor variables and q criterion variables, respectively. On assuming that each column of $\mathbf{X}$ and $\mathbf{Y}$ has been standardized to zero mean and unit sum of squares, the sample correlation matrices are $\mathbf{R}_{xx} = \mathbf{X}'\mathbf{X}$, $\mathbf{R}_{yy} = \mathbf{Y}'\mathbf{Y}$, and $\mathbf{R}_{yx} = \mathbf{Y}'\mathbf{X}$. Further, let $\|\mathbf{A}\|^2 = \text{trace}\,(\mathbf{A}'\mathbf{A})$ and $[\mathbf{A}]_r$ the matrix consisting of the first r columns of the matrix $\mathbf{A}$. For a square, symmetric and positive definite matrix $\mathbf{A}$, let $\mathbf{A}^{1/2}$ denote the symmetric matrix such that $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$.

To show in which sense the structure correlations are optimal, we derive a "rank r weighted least-squares approximation" to $\mathbf{R}_{yx}$ of the form $\mathbf{BC}'$ with $\mathbf{B}$ and $\mathbf{C}$ matrices of order $q \times r$ and $p \times r$, respectively. For producing plots on the basis of the approximation, a convenient choice of r is 2. Whether $r = 2$ is adequate can be judged by the approximation error or, if the lower dimensionality hypothesis is credible, by Barlett's test of dimensionality (see the example section). In the approximation we take as weight matrices the inverses of $\mathbf{R}_{xx}$ and $\mathbf{R}_{yy}$; this choice makes the loss function independent of linear transformations of $\mathbf{X}$ and $\mathbf{Y}$. Thus, we seek the minimum over $\mathbf{B}$ and $\mathbf{C}$ of

$$\|\mathbf{R}_{yy}^{-1/2}(\mathbf{R}_{yx} - \mathbf{BC}')\mathbf{R}_{xx}^{-1/2}\|^2 = \|\mathbf{R}_{yy}^{-1/2}\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1/2} - (\mathbf{R}_{yy}^{-1/2}\mathbf{B})(\mathbf{R}_{xx}^{-1/2}\mathbf{C})'\|^2. \tag{1}$$

As follows from the Eckhart-Young theorem (Eckhart & Young, 1936; Greenacre, 1984) the minimum is obtained from the singular value decomposition

$$\mathbf{R}_{yy}^{-1/2}\,\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1/2} = \mathbf{P\Lambda Q}', \tag{2}$$

where $\mathbf{P}$ and $\mathbf{Q}$ are orthonormal matrices of order $q \times t$ and $p \times t$ with $t = \min(p, q)$ containing the singular vectors, and $\boldsymbol{\Lambda}$ a diagonal matrix with the singular values on the diagonal, arranged in decreasing order ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_t \geq 0$). The minimum of (1) is $\lambda_{r+1}^2 + \cdots + \lambda_t^2$ and is attained by setting

$$\mathbf{R}_{yy}^{-1/2}\mathbf{B} = [\mathbf{P}\boldsymbol{\Lambda}]_r \quad \text{and} \quad \mathbf{R}_{xx}^{-1/2}\mathbf{C} = [\mathbf{Q}]_r. \tag{3}$$

or, equivalently,

$$\mathbf{B} = [\mathbf{R}_{yy}^{1/2}\mathbf{P}\boldsymbol{\Lambda}]_r \quad \text{and} \quad \mathbf{C} = [\mathbf{R}_{xx}^{1/2}\mathbf{Q}]_r. \tag{4}$$

The singular value decomposition is one computational route for obtaining a canonical correlation analysis (Gittins, 1985, Horst, 1961); the singular values are the canonical correlations, $\mathbf{R}_{yy}^{-1/2}$ $\mathbf{P}$ and $\mathbf{R}_{xx}^{-1/2}$ $\mathbf{Q}$ contain the canonical weights, and $\mathbf{B}$ and $\mathbf{C}$, therefore, contain the correlations of the original variables with the canonical variates of the predictor set (Gittins, pp. 17–18, 38–39). In the terminology of canonical correlation analysis, $\mathbf{B}$ and $\mathbf{C}$ contain structure correlations; more precisely, $\mathbf{B}$ contains the interset correlations of the criterion variables and $\mathbf{C}$ the intraset correlations of the predictor variables (Gittins). The plot of structure correlations is thus obtained by plotting each row of $\mathbf{B}$ and of $\mathbf{C}$ as a vector in an r-dimensional Cartesian coordinate system (see Figure 1 for an example). The matrix product $\mathbf{BC}'$ is represented in the plot by the scalar inner products between the vectors for the criterion variables (the rows of $\mathbf{B}$) and the vectors for the predictor variables (the rows of $\mathbf{C}$). Such a plot is called a biplot (Gabriel, 1971; 1982); it displays the weighted least-squares approximation $\mathbf{BC}'$ to $\mathbf{R}_{yx}$. This was first noted by Haber and Gabriel (1976). The traditional rationale for the plot of structure correlations is based on the observation that the rows of $\mathbf{B}$ and $\mathbf{C}$ contain the coordinates of the original variables with respect to the canonical variates of the predictor variables (Caillez & Pagès, 1976). The rationale that the plot optimally displays $\mathbf{R}_{yx}$ according to the biplot rules, is much stronger, and more helpful for the interpretation. The example section provides easy-to-use rules for reading the biplot.

In (3) and (4), $\boldsymbol{\Lambda}$ can be moved from the equation for $\mathbf{B}$ to the equation for $\mathbf{C}$ without affecting the loss-function (1); then, $\mathbf{B}$ contains intraset correlations and $\mathbf{C}$ interset correlations. For the biplot one might therefore equally well use intraset correlations for the criterion variables and interset correlations for the predictor variables. This yields the same approximation to $\mathbf{R}_{yx}$. Alternatively, one can treat $\mathbf{B}$ and $\mathbf{C}$ symmetrically with respect to $\boldsymbol{\Lambda}$ and choose

$$\mathbf{B} = [\mathbf{R}_{yy}^{1/2}\mathbf{P}\boldsymbol{\Lambda}^{1/2}]_r \quad \text{and} \quad \mathbf{C} = [\mathbf{R}_{xx}^{1/2}\mathbf{Q}\boldsymbol{\Lambda}^{1/2}]_r. \tag{5}$$

This factorization of $\mathbf{R}_{xy}$ is the most natural in the common factor model (Rao, 1973, p. 585–586) and in inter-battery factor analysis (Browne, 1979), because $\mathbf{Y}$ and $\mathbf{X}$ are placed on equal footing in these models. Under the assumption of multinormality, the various factorizations (e.g., (4) and (5)) are, however, equivalent, because they yield the same value of the likelihood (Browne).

When the relationship between the two sets of variables is asymmetric, we can obtain more insight into the relationships by performing a multivariate regression of $\mathbf{Y}$ on $\mathbf{X}$ using the model

$$\mathbf{Y} = \mathbf{XM}' + \mathbf{E}, \tag{6}$$

where $\mathbf{M}$ is a $q \times p$ matrix of regression coefficients, and $\mathbf{E}$ a $n \times q$ matrix of random errors. The usual least-squares estimator for $\mathbf{M}$ is

$$\hat{\mathbf{M}} = \mathbf{R}_{yx}\,\mathbf{R}_{xx}^{-1}. \tag{7}$$

One may grasp the main features of $\hat{\mathbf{M}}$ more easily from a plot than from the numbers themselves. Such a plot can be obtained from an approximation to $\hat{\mathbf{M}}$ of the form $\mathbf{B}_0\mathbf{C}_0'$, with $\mathbf{B}_0$ and $\mathbf{C}_0$ matrices of order $q \times r$ and $p \times r$, respectively. Some elements of $\hat{\mathbf{M}}$ are more precise than other elements, as indicated by the standard error of estimate of the regression coefficients. Therefore, we use a weighted least-squares approximation. As weight matrix we choose the inverse of the covariance matrix of the estimated regression coefficients (to allow not only for the standard errors of the estimates but also for the covariances among them). The estimated covariance matrix of the i-th row and j-th row of $\hat{\mathbf{M}}$ is $(n - p - 1)^{-1}s_{ij}\mathbf{R}_{xx}^{-1}$ (Anderson, 1984, p. 291), with s_{ij} the (i, j)-th element of the residual sum of squares and products matrix of $\mathbf{Y}$ with respect to $\mathbf{X}$:

$$\mathbf{S}_e = \mathbf{R}_{yy} - \mathbf{R}_{yx}\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}. \tag{8}$$

The loss function is thus proportional to

$$\|\mathbf{S}_e^{-1/2}(\hat{\mathbf{M}} - \mathbf{B}_0\mathbf{C}_0')\mathbf{R}_{xx}^{1/2}\|^2 = \|\mathbf{S}_e^{-1/2}\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1/2} - (\mathbf{S}_e^{-1/2}\mathbf{B}_0)(\mathbf{R}_{xx}^{1/2}\mathbf{C}_0)'\|^2. \tag{9}$$

As before, the minimum follows from a singular value decomposition, now of

$$\mathbf{S}_e^{-1/2}\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1/2} = \mathbf{G}\boldsymbol{\Delta}\mathbf{H}', \tag{10}$$

where $\mathbf{G}$ and $\mathbf{H}$ are orthonormal matrices of order $q \times t$ and $p \times t$ containing the singular vectors, and $\boldsymbol{\Delta}$ a diagonal matrix with the singular values in decreasing order on the diagonal ($\delta_1 \geq \delta_2 \cdots \geq \delta_t \geq 0$). The minimum of (9) is therefore $\delta_{r+1}^2 + \cdots + \delta_t^2$, and is attained by setting

$$\mathbf{B}_0 = [\mathbf{S}_e^{1/2}\mathbf{G}\boldsymbol{\Delta}]_r \quad \text{and} \quad \mathbf{C}_0 = [\mathbf{R}_{xx}^{-1/2}\mathbf{H}]_r. \tag{11}$$

But, the singular value decompositions of (2) and (10) are closely related (Anderson, 1984, pp. 506–507; Bock, 1975, p. 391), in particular

$$\delta_i^2 = \frac{\lambda_i^2}{1 - \lambda_i^2}\,[i = 1, \ldots, t], \quad \mathbf{H} = \mathbf{Q}, \quad \text{and} \quad \mathbf{S}_e^{-1/2}\mathbf{G}\boldsymbol{\Delta}^{-1} = \mathbf{R}_{yy}^{-1/2}\mathbf{P}\boldsymbol{\Lambda}^{-1}. \tag{12}$$

From (11) and (12) we obtain the expressions

$$\mathbf{B}_0 = [\mathbf{R}_{yy}^{1/2}\mathbf{P}\boldsymbol{\Lambda}]_r \quad \text{and} \quad \mathbf{C}_0 = [\mathbf{R}_{xx}^{-1/2}\mathbf{Q}]_r, \tag{13}$$

that is, $\mathbf{B}_0$ equals $\mathbf{B}$ in (4) and contains the interset correlations of the criterion variables, whereas $\mathbf{C}_0$ contains the canonical weights of the predictor variables (Gittins, 1985, p. 18). In conclusion, these interset correlations and these weights minimize the loss-function (9). By plotting each row of $\mathbf{B}_0$ and of $\mathbf{C}_0$ as a vector in an r-dimensional Cartesian coordinate system, we therefore obtain a biplot that approximates $\hat{\mathbf{M}}$ in a weighted least-squares sense.

The biplot approximation of $\hat{\mathbf{M}}$ can also be derived from reduced rank regression. Izenman (1975), Brillinger (1981), Davies and Tso (1982), and Velu, Reinsel, and Wichern (1986) considered a loss-function of the form $\|(\mathbf{Y} - \mathbf{X}\mathbf{C}_1\mathbf{B}_1')\boldsymbol{\Gamma}^{1/2}\|^2$ with $\boldsymbol{\Gamma}$ a given weight matrix. This is also the loss-function of a fixed factor score model with linear restrictions (de Leeuw, Mooijaart, & van der Leeden, 1985). If $\boldsymbol{\Gamma} = \mathbf{S}_e^{-1}$ (i.e., weight inverse with residual variance), the minimum is attained by setting $\mathbf{B}_1 = \mathbf{B}_0$ and $\mathbf{C}_1 = \mathbf{C}_0$. This weighted least-squares solution is also the maximum likelihood solution when the errors are normally distributed with unknown error covariance matrix (see

Tso, 1981, for fixed **X**; Israëls, 1987, p. 271, for random **X**). The choice $\mathbf{\Gamma} = \mathbf{R}_{yy}^{-1}$ gives the same solution (Haber & Gabriel, 1976; Velu, Reinsel, & Wichern, 1986). The reduced-rank fit for **Y** is displayed in a biplot of the rows of $\mathbf{XC}_0$ and the rows of $\mathbf{B}_0$, that is, a plot of both the canonical variate scores of the predictor set ($\mathbf{V} = \mathbf{XC}_0$) and the interset correlations of the criterion variables. Yet another formulation is that **V** and $\mathbf{B}_0$ minimize

$$\|(\hat{\mathbf{Y}} - \mathbf{VB}_0')\mathbf{S}_e^{-1/2}\|^2, \tag{14}$$

over all rank r fits, where $\hat{\mathbf{Y}}$ is the ordinary least-squares fit of the regression of **Y** on **X** (Davies & Tso). It follows (Gabriel, 1971) that $\mathbf{B}_0$ is also the solution for the minimization of

$$\|\mathbf{S}_e^{-1/2}(\hat{\mathbf{Y}}'\hat{\mathbf{Y}} - \mathbf{B}_0\mathbf{B}_0')\mathbf{S}_e^{-1/2}\|^2. \tag{15}$$

Inner products of the rows of $\mathbf{B}_0$ therefore approximate, in a weighted least-squares sense, the covariances among the fitted criterion variables. What they represent exactly are the covariances among the reduced-rank fits of the criterion variables, because $(\mathbf{XC}_0\mathbf{B}_0')'(\mathbf{XC}_0\mathbf{B}_0') = \mathbf{B}_0\mathbf{B}_0'$.

The biplot for $\hat{\mathbf{M}}$ can be enriched in a simple way to show the approximate significance of the regression coefficients as judged by the usual t-ratio (estimate/estimated standard error). The matrix of t-ratios T, say, and its approximation are

$$\mathbf{T} = \mathbf{D}^{-1/2}\hat{\mathbf{M}}\mathbf{F}^{-1/2} \approx (\mathbf{D}^{-1/2}\mathbf{B}_0)(\mathbf{F}^{-1/2}\mathbf{C}_0)', \tag{16}$$

where **D** and **F** are diagonal matrices of orders q and p containing the diagonal elements of $(n - p - 1)^{-1}\mathbf{S}_e$ and of $\mathbf{R}_{xx}^{-1}$, respectively. Note that **F** contains the variance inflation factors for the predictor variables (e.g., Montgomery & Peck, 1982). A biplot for **T** is therefore obtained from the biplot for $\hat{\mathbf{M}}$ by changing the lengths of vectors in the latter. The vector for any predictor variable must be divided by the square root of its variance inflation factor (a number ≥ 1) and will therefore be reduced in length. The approximate t-ratios are then obtained by taking the scalar inner products between vectors of the two sets. Geometrically, a scalar inner product can be obtained by projecting a vector for a predictor on a vector for a criterion variable and by multiplying the length of the projection with the length of the vector for the criterion variable on to which it is projected. Rather than changing the lengths of the vectors for the criterion variables, it is convenient to mark the position on each vector where a projection point would precisely yield the critical t-ratio. The coordinates of the mark for the i-th criterion variable are $(t_c(n - p - 1)^{-1/2}s_{ii}^{1/2}\|\mathbf{b}_i\|^{-2})\mathbf{b}_i$, where t_c is the critical t-ratio and $\mathbf{b}_i'$ the i-th row of $\mathbf{B}_0$. An example of such a plot is in Figure 2.

Greenacre (1984, p. 349) proposed yet another biplot, namely of canonical weights only. This biplot yields a weighted least-squares approximation of $\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1}$, the weights involved in the approximation being $\mathbf{R}_{yy}$ and $\mathbf{R}_{xx}$. Because the quantities being approximated are difficult to interpret, this type of plot cannot be recommended.

The optimality properties of the proposed biplots can be generalized beyond the Euclidean matrix norm to norms that are invariant under unitary transformations (Rao, 1979, 1980; Sabatier, Jan, & Escoufier, 1984). Velu, Reinsel, and Wichern (1986) derived the asymptotic covariances of the elements of $\mathbf{B}_0$ and $\mathbf{C}_0$.

Example of Proposed Biplots

Adelman, Geier, and Morris (1969) applied canonical correlation analysis to study "the simultaneous relationships among instruments and goals in the process of national

TABLE 1

Correlation Matrix between Goals (y) and Instruments (x) in National Development (a: $\mathbf{R}_{yx}$, b: $\mathbf{R}_{yy}$ below and $\mathbf{R}_{xx}$ above Diagonal)

a	x_1	x_2	x_3	x_4	x_5	x_6
y_1	0.45	0.40	0.35	0.57	0.63	0.52
y_2	0.77	0.61	0.29	0.87	0.65	0.80
y_3	0.86	0.54	0.27	0.80	0.67	0.82
y_4	0.69	0.36	0.19	0.72	0.46	0.73
y_5	0.16	0.38	0.51	0.16	0.21	0.18
y_6	0.78	0.60	0.32	0.89	0.65	0.77

b	x_1	x_2	x_3	x_4	x_5	x_6
y_1	-	0.45	0.14	0.73	0.60	0.80
y_2	0.66	-	0.51	0.60	0.52	0.46
y_3	0.50	0.80	-	0.32	0.40	0.26
y_4	0.40	0.64	0.79	-	0.68	0.78
y_5	0.18	0.21	0.27	0.16	-	0.56
y_6	0.55	0.86	0.82	0.72	0.16	-

development and modernization''. Table 1 (extracted from Adelman & Morris 1967, Table A.1) shows the correlation coefficients between 6 goals and 6 instruments in their ''full'' sample of 74 developing noncommunist nations. The goals and instruments are those analysed by Adelman et al. for their ''low sample'':

goals: y_1 = rate of growth of real per capita GNP: 1950/51-1963/64;
y_2 = extent of dualism;
y_3 = extent of social mobility;
y_4 = degree of national integration and sense of national unity;
y_5 = extent of political stability;
y_6 = level of modernization of techniques in agriculture.
instruments: x_1 = extent of literacy;
x_2 = degree of administrative efficiency;
x_3 = extent of leadership commitment to economic development;
x_4 = level of adequacy of physical overhead capital;
x_5 = gross investment rate;
x_6 = rate of improvement in human resources.

Table 2 displays the results of a canonical correlation analysis for these data. The canonical correlations are 0.96, 0.59, 0.51, 0.38, 0.29, and 0.17. As judged by Bartlett's (1938) test of dimensionality (Gittins, 1985, p. 61), the first two canonical correlations

TABLE 2

Canonical Correlation Analysis of Table 1*: Canonical Weights and Correlations with the Canonical Variates of the Instruments

	weights 1	2	3	correlations 4	5	6
y_1	0.01	0.37	-0.26	0.58	0.22	-0.09
y_2	0.35	-0.48	0.38	0.90	-0.00	0.05
y_3	0.31	0.20	-0.96	0.89	0.03	-0.14
y_4	0.08	-0.39	0.26	0.76	-0.13	0.01
y_5	0.00	0.45	0.26	0.21	0.45	0.15
y_6	0.30	0.26	0.46	0.91	0.01	0.09
x_1	0.32	0.06	-0.45	0.89	-0.06	-0.29
x_2	0.08	0.29	0.25	0.63	0.44	0.31
x_3	0.02	0.72	0.43	0.32	0.81	0.37
x_4	0.49	-0.77	1.16	0.95	-0.07	0.20
x_5	0.04	0.56	-1.02	0.72	0.42	-0.37
x_6	0.18	-0.16	-0.30	0.89	-0.07	-0.10

* Using 5 digits in the Correlation Matrix

are nonzero at the 1%-significance level. The third canonical correlation is just significant at the 5%-level.

Figure 1 is a plot of the correlations of the original variables of both sets with the first two canonical variates of the instruments (see Table 2; Columns 4 and 5). This plot is a biplot approximation of the correlations between goals and instruments; so the innerproduct of a goal-arrow (y) and an instrument-arrow (x) is an approximation of their sample correlation coefficient. Although this type of joint plot is not new at all (Caillez & Pagès, 1976; Gifi, 1981; van der Geer, 1986; Israëls, 1987; among others), it seems appropriate to remember some useful rules for interpreting the biplot (Gabriel, 1971, 1982). The inner product is by definition equal to the product of the lengths of the corresponding arrows and the cosine of the angle between them. Thus, the correlation as displayed in the biplot is positive if the angle is sharp, negative if the angle is obtuse, and zero if the arrows are perpendicular. Alternatively, the inner product is derived by projecting an x-vector onto a y-vector and multiplying the lengths of the y-vector and the projected x-vector (the result is multiplied by -1 if the y-vector and the projected x-vector point in opposite directions). One can therefore obtain the approximate order of the correlations of x-vectors with a particular y-vector from the order of the projection points of these x-vectors onto this y-vector. One can of course equally well interchange x and y in the above rule.

Using these rules, we see from Figure 1 that in the approximation used, the goals y_2, y_3, and y_6 are highly correlated with the instruments x_1, x_4, and x_6 (correlations between 0.76 and 0.89 in Table 1), somewhat less with x_2 and x_5 (correlations between 0.53 and 0.65 in Table 1) and least with x_3 (correlations between 0.26 and 0.32 in Table 1). Goal y_4 shows about the same correlation pattern, but at a lower level of correlation. Goal y_5 points in a quite different direction and therefore shows another pattern; the

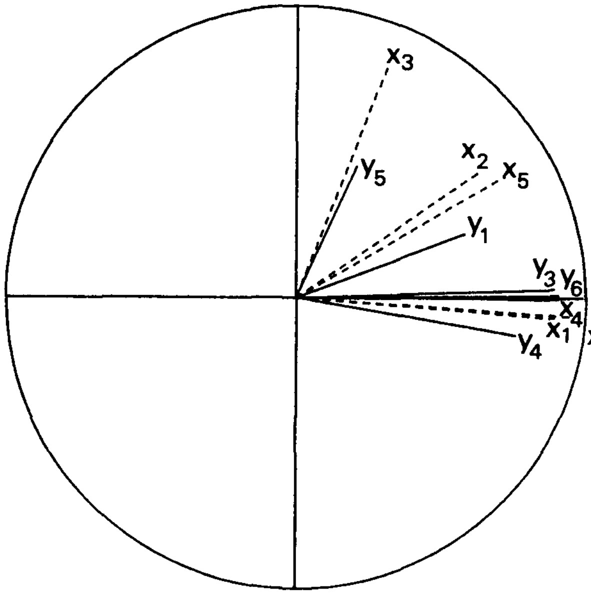

FIGURE 1
Biplot of correlations between goals and instruments (Table 1a) based on canonical correlation analysis (Table 2). Plotted are the interest correlations of the goals (solid lines) and the intraset correlations of the instruments (dashed lines); for explanation, see example Section. The circle has radius 1.

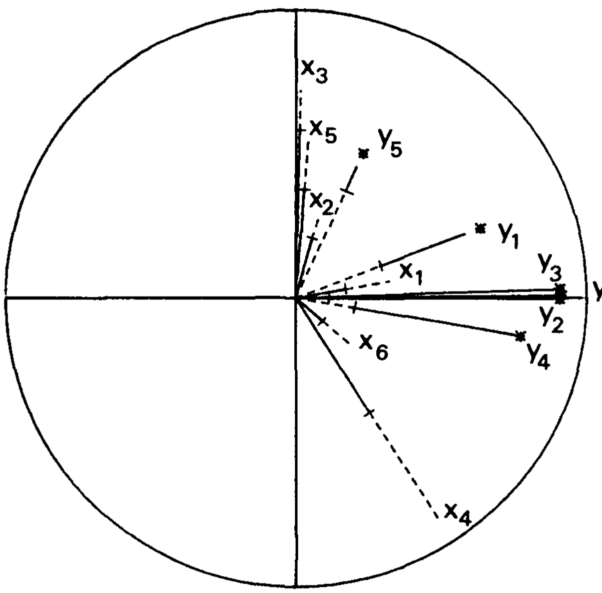

FIGURE 2
Biplot of the coefficients and associated t-ratios of the multivariate regression of goals on instruments (Table 3) based on canonical correlation analysis (Table 2). Plotted are the interset correlations of the goals (lines ending solid) and the canonical weight for the instruments (lines ending dashed). The length of a line for a goal is equal to its multiple correlation in the displayed Rank-2-regression model. The multiple correlation in the full rank model is indicated by the distance of the star to the origin. The positions where the lines change from solid in dashed (mark) are important for inferring t-ratios. For explanation, see example section. The circle has radius 1.

highest correlation is with x_3. Because y_5 is closest to the origin of the plot, the correlations with y_5 tend to be small. Goal y_1 is intermediate between y_3 and y_5; of all goals, y_1 has the least differentiation in its correlations with the instruments. The approximation of correlations by Figure 1 is quite good; from the 36 correlations in $\mathbf{R}_{yx}$, only 2 correlations differ by more than 0.07 from their value displayed in the biplot.

Because Figure 1 is a plot of correlations of the original variables with the canonical variates of the instruments, the squared lengths of arrows are Rank 2 communalities: intraset communalities for the instruments and interset communalities for the goals (Gittins, 1985, p. 43).

In Figure 2 the canonical weights of the instruments are plotted together with the interset correlations of the goals. Figure 2 is thus a biplot approximation of the regression coefficients of the multiple regression of each goal on the instruments (Table 3). Each arrow is subdivided in a solid part and a dashed part. For instruments (x-vectors), the solid part indicates the reduced vectors for the approximation of t-ratios (see the theory Section). If a reduced x-vector is projected onto a particular y-vector and the projection point falls in the dashed part of the y-vector (or its mirror-image on the other side of the origin), then the t-ratio as displayed in the biplot is less than 2 (in absolute value) and the hypothesis that the corresponding regression coefficient is equal to zero cannot be rejected at the 5%-significance level. Note that this graphical test is not exact–even if the assumptions of a t-test hold true–because the biplot displays the observed t-ratios with some error.

For example, by projecting the solid part of the x-vectors on y_5, we see that the projection point of x_3 is the only one falling outside the dashed part of y_5 and its mirror image. The observed t-ratio is 3.6 (Table 3). Notice that the projection point of x_4 falls in the mirror image; its t-ratio is -1.3 in Table 3. Similarly we derive from Figure 2 that x_4 has a significant effect ($p < 0.05$) on y_4, y_2, y_6, and y_3. This agrees with Table 3, except that the t-ratio for y_3 is only 1.8. The major discrepancy between Figure 2 and Table 3 is the effect of x_5 on y_1, which is not well displayed in the biplot. The third canonical variate may help in detecting discrepancies; the magnitude of the weights of x_4 and x_5 on this variate suggests that the biplot displays their effect the worst.

The length of a y-vector in Figure 2 is its multiple correlation in the Rank-2-regression model. The usual (full rank) multiple correlation, indicated by a star in Figure 2, is only slightly higher. From Figure 2 we can thus see that the goals y_2, y_3, and y_6 can be explained best by the instruments ($R \approx 0.9$), the goals y_1 and y_4 less ($R \approx 0.6$–0.7), and goal y_5 least ($R \approx 0.5$). The first canonical dimension accounts for 57% of the variance in the goals, the second one for 4%, and the third one for only 1%. In full space, the instruments account for 64% of the variance in the goals; so only slightly more than the 61% displayed in Figure 2.

If the original observations would have been available, one could have plotted in Figure 2 the canonical variate scores of the nations as well, to obtain a biplot of the fitted values (see (14)).

On comparing Figure 1 and Figure 2, one should notice that Figure 1 displays marginal "effects" (pairwise correlations) and Figure 2 conditional effects (regression coefficients). The comparison of these effects is easy because the y-vectors take the same position in both Figures, whereas the x-vectors differ. From the comparison we see that the instruments x_1 and x_6 have about the same marginal effect as x_4 but a much lower conditional effect. The change in position of x_2 and x_5 is most striking. These instruments have a high correlation ($r > 0.5$) with y_2, y_3, and y_6, but a non-significant conditional effect.

TABLE 3

Multivariate Regression of Goals (y) on Instruments (x)

	x_1	x_2	x_3	x_4	x_5	x_6	R^{**}
y_1	-0.13*	-0.03	0.09	0.17	0.44	0.23	0.68
	(-0.77)	(-0.23)	(0.78)	(0.99)	(3.30)	(1.33)	
y_2	0.17	0.13	-0.02	0.50	0.03	0.20	0.91
	(1.75)	(1.79)	(-0.30)	(5.05)	(0.39)	(2.01)	
y_3	0.46	0.05	0.01	0.18	0.12	0.21	0.91
	(4.85)	(0.74)	(0.21)	(1.82)	(1.63)	(2.06)	
y_4	0.24	-0.11	0.04	0.44	-0.11	0.30	0.78
	(1.70)	(-1.07)	(0.43)	(3.01)	(-1.00)	(1.99)	
y_5	0.17	0.22	0.46	-0.25	-0.04	0.04	0.55
	(0.89)	(1.53)	(3.61)	(-1.27)	(-0.25)	(0.22)	
y_6	0.27	0.07	0.05	0.61	-0.01	0.04	0.91
	(2.91)	(1.03)	(0.84)	(6.35)	(-0.15)	(0.38)	
VIF***	3.44	1.96	1.55	3.77	2.18	3.82	

* Standardized regression coefficient and, between brackets, the associated t-ratio

** Multiple correlation coefficient for the goals

*** Variance inflation factor for the instruments

Partial Canonical Correlation Analysis

In partial canonical correlation analysis, one wants to study the relationships between criterion and predictor variables while taking account for their correlations with variables of a third set, say **Z** (Röhr, 1987; Roy & Whittlesey, 1952; Timm & Carlson, 1976). The technique reduces to the usual canonical correlation analysis applied to partial correlations and nothing new arises: the biplot based on (4) approximates partial correlations and the biplot based on (13) approximates "partial" regression coefficients. These partial regression coefficients are essentially the regression coefficients corresponding to the predictor variables in the full model (compare (6)):

$$\mathbf{Y}=\mathbf{X}\mathbf{M}' + \mathbf{Z}\mathbf{N}' + \mathbf{E}. \tag{17}$$

A complication is that the analysis on partial correlations linearly rescales the criterion and predictor variables in such a way that their length is 1 after the adjustment for **Z**. We will assume this scaling in (17). The rescaling affects the definition of the regression coefficients in an obvious way but does not affect the *t*-ratios in (16) because the latter are scale-invariant. (Alternatively, the analysis specified by (2), (7) through (13) can be applied to partial covariances rather than correlations). Loss function (9) is appropriate indeed for the partial analysis because with the redefinitions, the covariance matrix of the *i*-th and *j*-th row of $\hat{\mathbf{M}}$ in (17) is as specified above (9). The analysis of Model (17) with partial canonical correlation analysis is a form of reduced-rank regression with concomitant variables (Davies & Tso, 1982).

Redundancy Analysis

The results of this paper carry over to redundancy analysis (van den Wollenberg, 1977) and the equivalent reduced-rank regression (Davies & Tso, 1982; Israëls, 1984, 1987) by redefining $\mathbf{R}_{yy}$ in (1) and $\mathbf{S}_e$ in (9) to be a $q \times q$ identity matrix. This change reflects the fact that in redundancy analysis, no adjustment is made for correlations among criterion variables nor for differences in error variance. In redundancy analysis, the elements of $\mathbf{B}$ ((4)) and of $\mathbf{B}_0$ ((10)) are not only correlations but also canonical coefficients. Partial redundancy analysis is discussed briefly by Davies and Tso (1982) under the name of reduced rank regression with concomitant variables.

Discussion

This paper uses canonical weights and structure correlations to construct low-dimensional views of the relationships between two sets of variables. These views, in the form of biplots, display familiar statistics: correlations between pairs of variables and regression coefficients. The canonical variates are simply a means to construct these views, and are not needed for interpretation. This makes the interpretation of canonical correlation output easier. In contrast to Rencher (1988), we find canonical correlation analysis useful *because* it reproduces univariate statistics! The gain lies in the dimension reduction that makes the biplot possible.

The question of whether to use structure correlations or weights for interpretation is now back at the choice between pairwise correlations between variables and regression coefficients to study relationships. Multivariate regression, whence regression coefficients, is useful when one set of variables is to be predicted from the other set; that is, when the role of the two sets of variables is asymmetric. A regression coefficient estimates the change in the criterion variable for one unit of change in the corresponding predictor variable when the other predictor variables are held constant. It measures a conditional effect. By contrast, a sample correlation coefficient measures marginal association. It depends in general on how other variables covary in the sample and may therefore be spurious. Of course, one must also be careful in interpreting regression coefficients, when (a) the predictor variables are intrinsically related, so that predictor variables cannot be held constant when the predictor under consideration is varied; (b) the sample is such that the predictor variables are almost multicollinear (the "bouncing beta" problem). In the first case one should consider (latent) path models (LISREL; Jöreskog & Sörbom, 1983; Saris & Stronkhorst, 1984). The second case is indicated by the variance inflation factors (Table 3) which are displayed in the biplot by the squared relative lengths of the full vector and its solid part for each predictor (Figure 2).

Following Tso (1981) and Davies and Tso (1982), we showed that both canonical correlation analysis and redundancy analysis can analyze asymmetric relationships. Redundancy analysis is based on an unweighted least-squares criterion, whereas canonical correlation analysis uses a weighted criterion. The choice between them should therefore be based on whether one wants to weight criterion variables equally or with statistically chosen optimal weights. An interesting intermediate case is to set $\mathbf{R}_{yy}$ equal to a diagonal matrix with the same diagonal as $\mathbf{S}_e$ ((8)). This weights criterion variables in dependence of how well they can be predicted without the reduced-rank restriction. Similar considerations apply to the choice between a weighted (Browne, 1979) and an unweighted (Tucker, 1958) approximation of $\mathbf{R}_{yx}$. Related discussion is provided by Wold (1982) in his choice between the modes A, B, and C in PLS.

Timm and Carlson (1976) proposed part canonical correlation as an alternative for partial canonical correlation analysis, for the case where the third set of variables (**Z**) influences only one of the other two sets, say **Y**, and not the other, say **X**. This is the case in the asymmetric model (17). Yet, under normality of errors, the maximum likelihood solution of (17) with a rank-restriction on **M** is obtained by partial (*not* part) canonical correlation analysis. The proof runs along the same lines as in Tso (1981).

The proposed biplots can also be used in non-linear canonical correlation analysis (van der Burg & de Leeuw, 1983) by basing the plots on the optimally scaled variables. Gower and Harding (1988) propose a modification of the classical biplot that has some data analytic appeal. Instead of a vector, they suggest drawing for each variable, a line segment through the origin, the end points of which indicate the minimum and maximum value of the variable in the data. The segments warn the user against inferring values from the plot that are outside the observed range. This modification can be useful as well in the biplots we propose, although a plot like our Figure 2 would quickly become too crowded. As this paper shows, the strength of canonical correlation analysis and related multivariate techniques lies in their ability to turn tables into biplots.

References

Adelman, I., & Morris, C. T. (1967). *Society, politics & economic development*. Baltimore: John Hopkins Press.

Adelman, I., Geier, M., & Morris, C. T. (1969). Instruments and goals in economic development. *American Economic Review—Papers and Proceedings*, *59*, 409–426.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, *22*, 327–351.

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (2nd. ed.). New York: Wiley.

Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society*, *34*, 33–40.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.

Brillinger, D. R. (1981). *Time series: Data analysis and theory*. San Francisco: Holden-Day.

Browne, M. W. (1979). The maximum-likelihood solution in inter-battery factor analysis. *British Journal of Mathematical and Statistical Psychology*, *32*, 75–86.

Cailliez, F., & Pagès, J. P. (1976). *Introduction à l'analyse des données* [Introduction to data analysis]. Paris: SMASH.

Corsten, L. C. A. (1976). Matrix approximation, a key to application of multivariate methods. *Proceedings of the 9th International Biometric Conference* (pp. 61–77). Raleigh: The Biometric Society.

Davies, P. T., & Tso, M.K-S. (1982). Procedures for reduced-rank regression. *Applied Statistics*, *31*, 244–255.

de Leeuw, J., Mooijaart, A., & van der Leeden, R. (1985). *Fixed factor score models with linear restrictions* (Technical Report RR-85-06). Leiden: Department of Datatheory.

DeSarbo, W. S. (1981). Canonical/redundancy factoring analysis. *Psychometrika*, *46*, 307–329.

Eckhart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, *1*, 211–218.

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, *58*, 453–467.

Gabriel, K. R. (1982). Biplot. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Vol. 1, pp. 263–271). New York: Wiley.

Gifi, A. (1981). *Nonlinear multivariate analysis*. Leiden: Department of Datatheory.

Gittins, R. (1985). *Canonical analysis. A review with applications in ecology*. Berlin: Springer-Verlag.

Gower, J. C., & Harding, S. (1988). Nonlinear biplots. *Biometrika*, *75*, 445–455

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. New York: Wiley.

Haber, M., & Gabriel, K. R. (1976). *Weighted least squares approximation of matrices and its application to canonical correlations and biplot display* (Technical Report).Rochester: University of Rochester, Department of Statistics.

Horst, P. (1961). Relations among *m* sets of measures. *Psychometrika*, *26*, 129–150.

Israëls, A. Z. (1984). Redundancy analysis for qualitative variables. *Psychometrika, 49*, 331–346.
Israëls, A. Z. (1987). *Eigenvalue techniques for qualitative data*. Leiden: DSWO Press.
Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis, 5*, 248–264.
Jöreskog, K. G., & Sörbom, D. (1983). *LISREL VI Users Guide*. Uppsala: Department of Statistics.
Kendall, M. (1975). *Multivariate analysis*. London: Griffin.
McKeon, J. J. (1966). Canonical analysis: Some relations between canonical correlation, factor analysis, discriminant function analysis, and scaling theory. *Psychometrika Monograph No. 13*.
Meredith, W. (1964). Canonical correlations with fallible data. *Psychometrika, 29*, 55–65.
Montgomery, D. C., & Peck, E. A. (1982). *Introduction to linear regression analysis*. New York: Wiley.
Rao, C. R. (1975). *Linear statistical inference and its applications*. New York: Wiley.
Rao, C. R. (1979). Separation theorems for singular values of matrices and their applications in multivariate analysis. *Journal of Multivariate Analysis, 9*, 362–377.
Rao, C. R., (1980). Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In P. R. Krishnaiah (Ed.), *Multivariate Analysis V* (pp. 3–22). Amsterdam: North-Holland.
Rencher, A. C. (1988). On the use of correlations to interpret canonical functions. *Biometrika, 75*, 363–365.
Röhr, M. (1987). *Kanonische Korrelationsanalyse* [Canonical correlation analysis]. Berlin: Akademie-Verlag.
Roy, S. N., & Whittlesey, J. (1952). Canonical partial correlations. *Annals of Mathematical Statistics, 23*, 642.
Sabatier, R., Jan, Y., & Escoufier, Y. (1984). Approximations d'applications lineaires et analyse en composantes principales [Approximations of linear operators and principal components analysis]. In E. Diday et al. (Eds), *Data analysis and informatics III* (pp. 569–580). Amsterdam: North Holland.
Saris, W., & Stronkhorst, H. (1984). *Causal modelling in nonexperimental research*. Amsterdam: Sociometric Research Foundation.
Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretation*. Beverly Hills: Sage.
Thorndike, R. M., & Weiss, D. J. (1973). A study of the stability of canonical correlations and canonical components. *Educational Psychological Measurement, 33*, 123–134.
Timm, N. H., & Carlson, J. E. (1976). Part and bipartial canonical correlation analysis. *Psychometrika, 41*, 159–176.
Tso, M. K-S. (1981). Reduced-rank regression and canonical analysis. *Journal of the Royal Statistical Society, Series B, 43*, 183–189.
Tucker, L. R., (1958). An inter-battery method of factor analysis. *Psychometrika, 23*, 111–136.
van den Wollenberg, A. L. (1977). Redundancy analysis. An alternative for canonical correlation analysis. *Psychometrika, 42*, 207–219.
van der Burg, E., & de Leeuw, J. (1983). Nonlinear canonical correlation. *British Journal of Mathematical and Statistical Psychology, 36*, 54–80.
van der Geer, J. P. (1986). *Introduction to linear multivariate data analysis, Vol. 1*. Leiden: DSWO Press.
Velu, R. P., Reinsel, G. C., & Wichern, D. W. (1986). Reduced rank models for multiple time series. *Biometrika, 73*, 105–118.
Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation, Vol. 2* (pp. 1–54). Amsterdam: North-Holland.

Manuscript received 3/24/89
Final version received 10/12/89