3416

British Journal of Mathematical and Statistical Psychology (1990) **43**, 217–230 Printed in Great Britain

Non-linear redundancy analysis†

Eeke van der Burg‡

University of Utrecht, Faculty of Sciences, VOS, Methods Section, PO Box 80140, 3508 TC Utrecht, The Netherlands

Jan de Leeuw

Department of Psychology and Mathematics, UCLA, CA 90024, USA

A non-linear version of redundancy analysis is introduced. The technique is called REDUNDALS. It is implemented within the computer program for canonical correlation analysis called CANALS (Van der Burg & De Leeuw, 1983). The REDUNDALS algorithm is of an alternating least squares (ALS) type. The technique is defined as minimization of a squared distance between criterion variables and weighted predictor variables. The matrix of weights can be restricted to a specified rank. With the help of optimal scaling the variables are transformed non-linearly (cf. Young, 1981). An application of redundancy analysis is provided.

1. Introduction

In many situations data are available from different sources. Suppose the data are of the form: objects × variables, and let us suppose that each source corresponds with a *subset of variables*. In case two (sub)sets of variables are available, a possible technique to relate the sets to each other is *redundancy analysis* (RA). This technique predicts one set of variables from the other set by taking linear combinations of the variables of one set that maximize the explained variance by the other set. The variables of the first set are called *predictors* and the variables of the second set *criterion* variables. The name redundancy analysis dates back to Van den Wollenberg (1977). However, already long before 1977 several authors introduce the so-called Redundancy Index (Glahn, 1969; Horst, 1955; Rao, 1962; Stewart & Love, 1968). This index is a measure of the predictability of one set from the other. In *multivariate multiple regression* the Redundancy Index is computed. In the next section it will be shown that RA according to Van den Wollenberg (1977) is a way of decomposing the Redundancy Index. It will also be shown that Van den Wollenberg's redundancy analysis is similar to Reduced Rank Regression (Izenman, 1975; Davies & Tso, 1982).

A non-linear version of RA has been proposed by Israëls (1984). His technique makes it possible to incorporate qualitative variables by the use of 'dummies'. In this

†An earlier version of this paper has been published in Van der Burg, E. (1988). *Nonlinear Canonical Correlation and Some Related Techniques.* Leiden: DSWO Press.

‡Requests for reprints.

paper another version of non-linear RA is proposed. Van den Wollenberg's technique is extended with so-called *optimal scaling* (Young, 1981). The non-linear generalization is called REDUNDALS. It can handle a larger choice of measurement levels for each variable than the technique of Israëls (1984).

In the case of redundancy analysis, an asymmetric situation between the sets is presumed. In the case of a symmetric role between the sets *canonical correlation analysis* (CCA) is much more obvious. This technique is described in many multivariate analysis textbooks (e.g. Gnanadesikan, 1977, Chapter 3.3; Tatsuoka, 1988, Chapter 7). RA and CCA are highly related. Van der Burg & De Leeuw (1983) proposed a non-linear version of CCA, and called this technique CANALS. As the algorithm for non-linear redundancy analysis shows many correspondences with the algorithm for non-linear CCA (see Section 5), the computer program for non-linear RA, the REDUNDALS program, is embedded in the CANALS program.

2. Redundancy analysis

Suppose the data consist of observations for n objects on m variables, and assume that the m variables can be divided into m_1 criterion variables and m_2 predictors. In addition assume that each variable is standardized, i.e. it has zero mean and unit variance. Collect the criterion variables in the matrix $\mathbf{H}_1$ of dimensions $(n \times m_1)$ and the predictors in $\mathbf{H}_2$ $(n \times m_2)$. The Redundancy Index of Stewart & Love (1968) is obtained by a *multivariate multiple regression* of $\mathbf{h}_i$, the columns of $\mathbf{H}_1$ $(i=1,\ldots,m_1)$ on $\mathbf{H}_2$. Thus

$$\text{minimize} \sum_{i=1}^{m_1} (\mathbf{h}_i - \mathbf{H}_2\mathbf{b}_i)'(\mathbf{h}_i - \mathbf{H}_2\mathbf{b}_i)/nm_1 \text{ over } \mathbf{b}_1,\ldots,\mathbf{b}_{m_1}, \tag{1}$$

where the vector $\mathbf{b}_i$ (m_2 elements) consists of regression weights. The squared distance or *loss* in (1) is divided by nm_1 for the sake of comparing the various techniques. In (1) each criterion variable explains maximally the predictor set. The matrix formulation of (1) is:

$$\text{minimize tr}(\mathbf{H}_1 - \mathbf{H}_2\mathbf{B})'(\mathbf{H}_1 - \mathbf{H}_2\mathbf{B})/nm_1 \text{ over } \mathbf{B}. \tag{2}$$

This expression is minimized by

$$\mathbf{B} = (\mathbf{H}_2'\mathbf{H}_2)^{-1}\mathbf{H}_2'\mathbf{H}_1, \tag{3}$$

provided that $\mathbf{H}_2'\mathbf{H}_2$ is of full rank. Substitution of (3) in (2) gives the minimum:

$$\text{tr}(\mathbf{H}_1'\mathbf{H}_1 - \mathbf{H}_1'\mathbf{H}_2(\mathbf{H}_2'\mathbf{H}_2)^{-1}\mathbf{H}_2'\mathbf{H}_1)/nm_1 \tag{4}$$

Denoting $\mathbf{R}_{11}$ for $\mathbf{H}_1'\mathbf{H}_1/n$ and $\mathbf{R}_{12}$ and $\mathbf{R}_{22}$ for $\mathbf{H}_1'\mathbf{H}_2/n$ and $\mathbf{H}_2'\mathbf{H}_2/n$ respectively, expression (4) is equivalent to

$$1-\mathrm{tr}(\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21})/m_1. \tag{5}$$

The expression $\mathrm{tr}(\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21})/m_1$ is equal to the Redundancy Index of Stewart & Love (1968). Thus minimizing (1) or (2) corresponds to computing the Redundancy Index.

If we reformulate (2) with rank restructions on $\mathbf{B}$ we are able to see the relationship between Reduced Rank Regression, redundancy analysis in the sense of Van den Wollenberg (1977) and multivariate multiple regression. Suppose matrix $\mathbf{B}$ is restricted to be of rank r with $1 \leqq r \leqq \min(m_1, m_2)$. Then $\mathbf{B}$ can always be written as a product of matrices, i.e. $\mathbf{B}=\mathbf{V}\mathbf{W}'$ with $\mathbf{V}$ $(m_2 \times r)$ and $\mathbf{W}$ $(m_1 \times r)$. As there are many solutions for $\mathbf{V}$ and $\mathbf{W}$, normalization constraints are necessary to determine the decomposition. Suppose that matrix $\mathbf{V}$ is restricted to $\mathbf{V}'\mathbf{R}_{22}\mathbf{V}=\mathbf{I}$, which means that $\mathbf{H}_2\mathbf{V}$ is a basis for the space spanned by the columns of $\mathbf{H}_2$. Then expression (2) can be written in terms of $\mathbf{V}$ and $\mathbf{W}$ as follows:

$$\min \mathrm{tr}(\mathbf{H}_1-\mathbf{H}_2\mathbf{V}\mathbf{W}')'(\mathbf{H}_1-\mathbf{H}_2\mathbf{V}\mathbf{W}')/nm_1 \text{ over } \mathbf{V} \text{ and } \mathbf{W} \tag{6}$$

$$\text{subject to the condition that } \mathbf{V}'\mathbf{R}_{22}\mathbf{V}=\mathbf{I}.$$

When r is equal to $\min(m_1, m_2)$ (6) is similar to multivariate multiple regression (2). When r is small (6) corresponds to Reduced Rank Regression (Izenman, 1975).

Some computational work shows that the columns of $\mathbf{V}$ in (6) correspond to the right eigenvectors of the matrix $\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{R}_{12}$. Denote the diagonal matrix of ordered eigenvalues by $\mathbf{\Phi}$. Then

$$\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{R}_{12}\mathbf{V}=\mathbf{V}\mathbf{\Phi} \quad \text{with} \quad \mathbf{V}'\mathbf{R}_{22}\mathbf{V}=\mathbf{I}. \tag{7}$$

Solving (7) for the first r eigenvectors corresponds to redundancy analysis of Van den Wollenberg (1977). He searches for r (normalized) weight vectors that optimize the explained variance between criterion variables and weighted predictors. These weight vectors are similar to the first r columns of matrix $\mathbf{V}$. If all weights are solved for, the sum of eigenvalues/m_1 is equal to the Redundancy Index (Israëls, 1984). From (6) and (7) we see that Van den Wollenberg's redundancy analysis is a special case of multivariate multiple regression, namely the case in which rank-r restrictions are imposed on $\mathbf{B}$. However, we can also argue that multivariate multiple regression is a special case of Van den Wollenberg's redundancy analysis, as Van den Wollenberg has the choice of r, that is how many eigenvectors will be computed.

In Section 4 non-linear transformations will be introduced in the form of optimal scaling. The combination of formulation (6) and optimal scaling is called REDUNDALS. The next section gives an overview of authors dealing with forms of redundancy analysis.

3. Historical overview

As mentioned already in the introduction the name redundancy analysis dates back to Van den Wollenberg (1977). Although he was the first one to name the technique, it actually stems from an earlier period. De Leeuw (1986) discusses the history of RA. We briefly summarize it. Horst (1955), Rao (1962), Stewart & Love (1968) and Glahn (1969) all propose the *Redundancy Index*. Rao (1964) and Robert & Escoufier (1976) discuss techniques for decomposing this Redundancy Index into uncorrelated components. Fortier (1966) proposes 'simultaneous linear predictions' which is equivalent with RA (cf. Ten Berge, 1985). Izenman (1975) and Davies & Tso (1982) also treat RA, but under the name Reduced Rank Regression. So far the discussion of De Leeuw (1986). Johansson (1981) proposes several forms of RA, which vary with orthogonality constraints, and DeSarbo (1981) discusses a technique which is a mixture between CCA and RA. Van de Geer (1984) places various types of RA in a larger framework of k-sets CCA. Israëls (1986) treats RA with various normalizations and rotations. Meulman (1986, Chapter 5.2.1) discusses a version of RA which can be shown to be a generalization of Van den Wollenberg's RA. However Meulman uses a completely different approach, formulating RA in terms of distances between objects or individuals.

Israëls (1984, 1987, Chapter 9) and Meulman (1986, Chapter 5.2.1) describe non-linear forms of redundancy analysis. In the next section another non-linear version of redundancy analysis is proposed, which is more general than the version of Israëls (1984, 1987, Chapter 9). However this author does not assume standardized variables for the criterion set, which means that our results must be corrected for variances to be comparable.

4. Optimal scaling

In many ways *non-linear transformations* can be implemented in redundancy analysis. To do so Israëls (1984) employed dummies for variables measured on a nominal measurement level. Meulman (1986, Chapter 5.2.1) uses monotone regression in her version of non-linear RA. Monotone regression is a form of *optimal scaling* (cf. Young, 1981). This means that the transformations (scaling parameters) minimize the loss, and at the same time *measurement restrictions* are maintained. We also use optimal scaling. The non-linear transformations treated in this article are *nominal* and *ordinal* (a definition will follow). In addition, of course, *linear* or *numerical* transformations are dealt with. 'Dummy transformations', as employed by Israëls (1984), are not discussed, however they can always be obtained by simply coding variables as dummies, and, in addition, by treating these dummies numerically. Another way to obtain these 'dummy transformations' is by using *copies* of a variable within the corresponding set, and by treating these copies as nominal. This gives a multiple nominal (or dummy) transformation (see Gifi, 1990, Chapter 4.4). Using copies instead of dummies has the advantage that one may choose both the dimensionality of the transformation and the measurement level of each copy separately. More information about copies can be found in De Leeuw (1984) and Van der Burg (1988, Chapter 4.6).

The nominal, ordinal and numerical transformations employed in this article agree with the transformations used by Van der Burg & De Leeuw (1983) in their version of non-linear CCA (CANALS). Together these three transformations form the optimal scaling. Our definition of optimal scaling corresponds to the definition of Young (1981). We mentioned already that optimal scaling refers to the fact that variables are optimally scaled in the sense of the model. This means that the data matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ are replaced by parameter matrices $\mathbf{Q}_1$ $(n \times m_1)$ and $\mathbf{Q}_2$ $(n \times m_2)$ such that they optimize the model, i.e. minimize the original loss, but at the same time satisfy the measurement restrictions. The original loss was formulated in (2). If the parameter matrix $\mathbf{Q}_1$ is substituted for $\mathbf{H}_1$ and $\mathbf{Q}_2$ for $\mathbf{H}_2$, and a rank restriction on $\mathbf{B}$ is added, this expression can be rewritten as follows. Denote the set of possible transformations for the ith variable, i.e., ith column of $[\mathbf{H}_1, \mathbf{H}_2]$, by C_i and use the notation $\mathbf{q}_i$ for the ith column of $[\mathbf{Q}_1, \mathbf{Q}_2]$. Non-linear redundancy analysis or REDUNDALS is

$$\text{minimize tr}\,(\mathbf{Q}_1 - \mathbf{Q}_2\mathbf{B})'(\mathbf{Q}_1 - \mathbf{Q}_2\mathbf{B})/nm_1 \tag{8}$$

over $\mathbf{Q}_1$, $\mathbf{Q}_2$ and $\mathbf{B}$ subject to the conditions that

$$\mathbf{q}_i \in C_i \ (i = 1, \ldots, m),$$

$$\mathbf{B} = \mathbf{V}\mathbf{W}' \quad \text{and} \quad \mathbf{V}'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{V} = n\mathbf{I}.$$

The sets of possible transformations are determined by *tie* and *normalization restrictions* for nominal variables, and, in addition, by *monotone* constraints for ordinal variables or by *linear* constraints for numerical variables (cf. De Leeuw, 1977). Tie restrictions imply that ties in the data correspond to ties in the transformation. Normalization restrictions result in standardized transformations (i.e. zero mean and unit variance). The monotone transformations discussed here correspond to the secondary approach of Kruskal & Shephard (1974). Finally linear transformations are equal to the variable itself, as standardization of the columns of the data matrix was assumed. A more extensive discussion of optimal scaling restrictions can be found in Young, De Leeuw & Takane (1976) and Young (1981).

In (8) the condition $\mathbf{V}'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{V} = n\mathbf{I}$ is necessary to determine $\mathbf{V}$ and $\mathbf{W}$. However, if (8) is solved for $\mathbf{V}$ and $\mathbf{W}$ without restrictions on $\mathbf{V}$ the same loss is reached. Then the corresponding solutions of $\mathbf{V}$ and $\mathbf{W}$ can always be rotated (without changing the loss) such that $\mathbf{V}'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{V} = n\mathbf{I}$ is satisfied.

Note that the loss of (8) together with the condition $\mathbf{B} = \mathbf{V}\mathbf{W}'$ can be written as

$$tr\{(\mathbf{Q}_1 - \mathbf{Q}_2\underline{\mathbf{B}})'(\mathbf{Q}_1 - \mathbf{Q}_2\underline{\mathbf{B}}) + (\mathbf{Q}_2\underline{\mathbf{B}} - \mathbf{Q}_2\mathbf{V}\mathbf{W}')'(\mathbf{Q}_2\underline{\mathbf{B}} - \mathbf{Q}_2\mathbf{V}\mathbf{W}')\}/nm_1 \tag{9}$$

$$\text{with } \underline{\mathbf{B}} = (\mathbf{Q}_2'\mathbf{Q}_2)^{-1}\mathbf{Q}_2'\mathbf{Q}_1.$$

This shows that, if $\mathbf{B}$ is computed without a rank restriction (which corresponds to $\underline{\mathbf{B}}$), $\mathbf{V}$ and $\mathbf{W}$ can be computed from minimizing the second part of the loss in (9). This is used is the following section.

5. REDUNDALS algorithm

The algorithm for non-linear redundancy analysis follows easily from (8) and (9). Using an alternating least squares method results in solving the parameters in the following order (δ is a very small number)

(a) initialize $\mathbf{Q}_1$, $\mathbf{Q}_2$
(b) compute $\mathbf{B}$ (unrestricted)
(c) decompose $\mathbf{B}$
(d) compute $\mathbf{Q}_2$
(e) compute $\mathbf{Q}_1$
(f) if $(\text{loss}_{\text{previous}} - \text{loss}_{\text{present}}) \leqq \delta$ go to (*b*)
(g) end

If one set of parameters is updated the remaining ones are kept at a constant level. As non-linear RA can be viewed as a special case of non-linear CCA the solutions for the various parameters can be found in Van der Burg & De Leeuw (1983). These authors formulate non-linear CCA as follows. Define $\mathbf{A}$ $(m_1 \times p)$ and $\mathbf{B}$ $(m_2 \times p)$ as the weight matrices for the first set and the second set respectively (this new definition of $\mathbf{B}$ does not interfere with the earlier one). The p corresponds to the number of dimensions or the number of canonical variates. Then non-linear CCA is, according to Van der Burg & De Leeuw (1983),

$$\text{minimize tr}(\mathbf{Q}_1\mathbf{A} - \mathbf{Q}_2\mathbf{B})'(\mathbf{Q}_1\mathbf{A} - \mathbf{Q}_2\mathbf{B})/np \tag{10}$$

over $\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{A}$ and $\mathbf{B}$, subject to the conditions that

$$\mathbf{A}'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{A} = n\mathbf{I} \text{ or } \mathbf{B}'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{B} = n\mathbf{I} \text{ and}$$

$$\mathbf{q}_i \in C_i \; (i = 1, \ldots, m).$$

This technique is called CANALS. In comparing this definition with the definition of non-linear RA in (8) we see that, from a CCA point of view,

(a) the number of dimensions p is fixed to m_1,
(b) matrix $\mathbf{A}$ is equal to the identity matrix,
(c) no normalizations are used,
(d) $\mathbf{B}$ is decomposed.

The CANALS algorithm is based on one normalization, either of $\mathbf{A}$ or of $\mathbf{B}$. In addition the CANALS algorithm uses transfer of normalization in the iterative process (i.e. weight matrices $\mathbf{A}$ and $\mathbf{B}$ are rescaled such that the matrix to be updated is *not* normalized). This rescaling involves a transformation of each weight matrix that does not change the loss (but does change the normalization). Details can be found in Van der Burg & De Leeuw (1983). The sequence of solving the parameters in the CANALS program is the following:

(1) initialize $\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{A}$
(2) rescale $\mathbf{A}$ such that $\mathbf{A}'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{A} = n\mathbf{I}$
(3) compute $\mathbf{B}$ (unrestricted)
(4) compute $\mathbf{Q}_2$

(5) rescale **A** and **B** such that $\mathbf{B}'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{B} = n\mathbf{I}$
(6) compute **A** (unrestricted)
(7) compute $\mathbf{Q}_1$
(8) rescale **A** and **B** such that $\mathbf{A}'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{A} = n\mathbf{I}$
(9) if $(\text{loss}_{\text{previous}} - \text{loss}_{\text{present}}) \leqq \delta$ go to 3
(10) rescale **A** and **B** such that both **A** and **B** are normalized
(11) end

Again the remaining parameters are supposed to be at a constant level when one set of parameters is updated. As the REDUNDALS solutions are similar to the CANALS solutions (as long as **I** is substituted for **A**, p is taken as m_1 and no decomposition of **B** is used), we see that REDUNDALS corresponds to steps 1, 3, 4, 7, 9, and 11 of CANALS. Therefore it is easy to combine the two algorithms. The REDUNDALS program is simply embedded in the CANALS program by only employing the equivalent steps and by skipping the other ones. A difference between the CANALS and the REDUNDALS program is the fact that in the case of REDUNDALS, matrix **A** is initialized on **I** and in the case of CANALS, **A** starts with random values. In addition the final solution for REDUNDALS is not rotated while the CANALS solution is (for CCA rotated weights give a similar loss). Finally the weights **B** are not decomposed for CANALS. They can be decomposed for REDUNDALS, depending on the value of r, i.e. if $r < \min(m_1, m_2)$. If r is equal to $\min(m_1, m_2)$ no rank restriction is applied. In that case another option is possible, namely a rank restriction on **B** afterwards. The latter solution corresponds to multivariate analysis of variance with optimal scaling combined with Van den Wollenberg's redundancy analysis.

6. Application

For illustration of REDUNDALS an example is taken which was also used to demonstrate the CANALS technique (Van der Burg & De Leeuw, 1983). A more detailed description of this example can be found in the latter article. The data are from a Parliamentary Survey carried out in 1972. Among other things, the Dutch members of parliament (MPs) gave their opinion on seven issues, and their preference votes for the political parties of which only the four larger parties interest us, The opinions were measured on a nine-point scale of which the lowest and the highest category were described (Table 1). The preference votes were recorded as a table of rank orders. As there were 14 parties, the preference votes take values 2 (highest preference) to 15 (lowest preference) (Table 1).

We wondered whether party preferences could predict the opinions on the issues. The idea behind it is that MPs have their traditional, party sympathies and take over the official party viewpoints. In Dutch politics a rather strong party discipline exists, so that many parties carry a clear image.

Several types of analyses are used. In the first place an ordinal REDUNDALS analysis (i.e. REDUNDALS with ordinal measurement restrictions for all variables) without a rank restriction on **B** is applied. Next rank-2 and rank-1 restrictions are added after the iteration process. Furthermore an ordinal solution with a rank-2

Table 1. Dutch Parliament. The issues and party preferences and the meaning of the lowest and the highest category

DEV:	development aid
(1)	the government should spend more money on aid to developing countries
(9)	the government should spend less money on aid to developing countries
ABO:	abortion
(1)	the goovernment should prohibit abortion completely
(9)	a woman has the right to decide for herself about absortion
LAW:	law and order
(1)	the government takes too strong action against public disturbances
(9)	the government should take stronger action against public disturbances
INC:	income differences
(1)	income differences should remain as they are
(9)	income differences should become much less
PAR:	participation
(1)	only management should decide important matters in industry
(9)	workers must also have participation in decision important for industry
TAX:	taxation
(1)	taxes should be increased for general welfare
(9)	taxes should be decreased so that people can decide for themselves how to spend their money
DEF:	defence
(1)	the government should insist on shrinking Western armies
(9)	the government should insist on maintaining strong Western armies
PvdA:	socialists
(2)	highest preference, (15) lowest preference
ARP:	christian democrats (protestants)
(2)	highest preference, (15) lowest preference
KVP:	christian democrats (catholics)
(2)	highest preference, (15) lowest preference
VVD:	conservatives
(2)	highest preference, (15) lowest preference

decomposition during the iteration process is computed, and finally a numerical REDUNDALS analysis is applied. Apart from the fitting of missing values (see discussion), the latter analysis is comparable with Van den Wollenberg's RA.

Results of the ordinal analysis without rank restriction are shown in Table 2. In the first place the multiple correlation coefficients (MC-4) are rather high. Thus the preference votes are good predictors for the opinions on the issues. In Table 2 also the correlations between the preferences and the issues (both transformed monotonically) are given. We do not interpret the weights as they do not give a clear idea of the relations between the variables due to multicollinearity (cf. Gnanadesikan, 1977, p. 22). Table 2 shows that the preference votes for PvdA and VVD are more strongly correlated with the issues than are the preference votes for ARP and KVP. The highest correlations are between PvdA and LAW, and VVD and INC. This means that the amount of sympathy for the socialists (PvdA) goes together with ideas

Table 2. Dutch Parliament, ordinal solution with rank restriction afterwards. Multiple correlations (MC-4 no rank restriction, MC-2 rank-2 restrictions, MC-1 rank-1 restriction) and correlations between issues (columns) and preference votes (rows)

	DEV	ABO	LAW	INC	PAR	TAX	DEF
MC-4	0.662	0.782	0.786	0.804	0.767	0.805	0.765
MC-2	0.661	0.776	0.783	0.788	0.760	0.801	0.758
MC-1	0.653	0.283	0.745	0.770	0.715	0.786	0.741
PvdA	0.591	−0.430	0.719	−0.624	−0.632	0.716	0.705
ARP	0.000	0.542	−0.336	−0.022	−0.152	0.031	−0.294
KVP	−0.001	0.637	−0.274	0.051	−0.100	0.033	−0.182
VVD	−0.537	−0.212	−0.418	0.722	0.667	−0.654	−0.450

about law and order: more sympathy corresponds to 'too strong action', and antipathy to 'stronger action'. MPs with great sympathy for the VVD agree with the proposition 'income differences should remain', and MPs with antipathy agree that 'income differences should become much less'. Law and order is a hot topic for the PvdA (but also TAX and DEF) and income differences for the VVD. Both parties take up clear positions on these issues (cf. Van der Burg & De Leeuw, 1983). All the other opinions are also correlated strongly with the preference votes for these two parties, and nearly always in different directions (except for ABO). The socialist party (PvdA) and the conservative party (VVD) are antagonists in Dutch politics, as in most countries. Apparently the profile of those parties is clear to many members of parliament.

The subject abortion needs some extra explanation. As the VVD originally is a liberal party, there still exists some liberal ideas. Especially with regard to abortion several VVD-members kept liberal thoughts (cf. Van der Burg & De Leeuw, 1983), so that socialist MPs and these conservative MPs agree on this subject. One has to know the historical background to understand such behaviour. Apparently sympathy (or antipathy) for the VVD comes from both MPs against abortion and MPs pro abortion, as the correlation between ABO and the VVD preference is not very high.

The christian democratic parties (KVP and ARP) appear to be less clear (or extreme) than the socialists or conservatives. Sympathy for the KVP includes a strong position against free abortion, but other issues hardly correlate with a KVP preference. A similar thing holds for the ARP. This agrees with the fact that the christian democrats form a middle party (they combined after 1972). Sometimes they cooperate with left, sometimes with right. Both socialists and conservatives need christian democrats to have a majority in the parliament. Therefore it is clear that MPs from both left and right have sympathy for the KVP or ARP, while having completely different ideas on the issues (except for ABO).

Addition of a rank-2 restriction on the weight matrix gives multiple correlation coefficients (MC-2) which are very similar to the multiple correlations (MC-4) already discussed (Table 2). This means that the rank of **B** was already nearly two. However, if a rank-1 restriction is added, the multiple correlations (MC-1) differ from the MC-2 values (Table 2). Then the issue ABO is hardly explained. In fact the scores on the issues (except for ABO) correlate with the left–right contrast in politics, whereas ABO

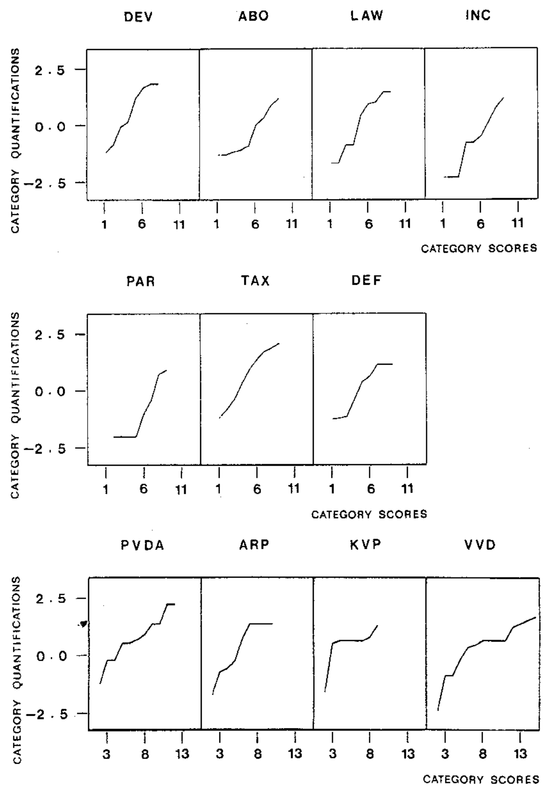

Figure 1. Dutch Parliament. Montone transformations of the variables. Original scores (horizontal) against category quantifications (vertical).

represents the pro–con abortion contrast. Apparently the left–right aspect is represented much better by the preferences for issues other than the ideas on abortion.

The difference between seven separate (non-linear) multiple correlations and a REDUNDALS analysis (without rank restriction) lies mainly in the fact that only one transformation is obtained for all the analyses together. This is a great advantage as the results have to be interpreted for each set of transformations separately. This problem is avoided by using one set of transformations. As all variables had a natural ordering in the categories, the data were treated at an ordinal measurement level. The monotone transformations are given in Fig. 1. The original scores (horizontal) are plotted against the transformed values, the so called category quantifications (vertical). The most striking transformation is the one belonging to the KVP preference. We see that the lowest score is separated from the rest. This means that one has either a very large sympathy (i.e. one is a member of the KVP party) or not. The nots are not distinguished from each other. Therefore it is clear that the KVP preference hardly correlates with the opinions on the issues. The ties accentuate once more the middle position of the KVP.

Table 3. Dutch Parliament, ordinal solution with rank-2 restriction in iteration process. Multiple correlations (MC) and correlations between issues (columns) and preference votes (rows)

	DEV	ABO	LAW	INC	PAR	TAX	DEF
MC	0.664	0.779	0.786	0.791	0.758	0.802	0.759
PvdA	0.590	−0.424	0.721	−0.640	−0.631	0.711	0.704
ARP	−0.017	0.555	−0.332	0.014	−0.102	0.026	−0.281
KVP	−0.011	0.651	−0.296	0.069	−0.083	0.023	−0.242
VVD	−0.543	−0.208	−0.427	0.708	0.667	−0.665	−0.448

The transformation of the ARP preference also shows ties, but only for categories regarding great antipathy. Thus strong or very strong aversion to the ARP does not matter in this analysis. The issue INC shows some ties in the lower categories and then ascends rather steeply. This means that the differentiation is in the fact of how much income differences should decrease. The amount of decrease corresponds to the PvdA and VVD preference (in different directions). The remaining transformations look normal, i.e. they do not contain many ties, nor very big jumps.

The results of an ordinal REDUNDALS analysis with a rank-2 restriction in the iteration process are given in Table 3. These results are comparable with the multivariate multiple regression solution (Table 2). In fact they are so much the same that the results do not have to be discussed separately. The category quantifications of both solutions are very similar too. Therefore no transformation plots of this solution are shown.

For the sake of comparing REDUNDALS with the existing techniques for RA a numerical REDUNDALS analysis is applied. The results are shown in Table 4. Again the multiple correlation coefficients (MC-4) are high. They are rather similar to the MC-4 values of the ordinal solutions. However there is a difference. For the ordinal solutions the transformations of the variables ARP and KVP were the more non-linear ones. Thus for these variables the numerical solution differs most from the ordinal solutions. This implies that the criterion variable ABO, which depends mainly on those two preferences, is predicted worse in the numerical solution than in the ordinal solutions. For the other criterion variables the difference with the ordinal solutions is not too large as these two preferences do not explain much variance. When the rank of the weight matrix is reduced, the same pattern occurs as in the ordinal case. The MC-2 and MC-4 values hardly differ, where the MC-1 values deviate mainly in the multiple correlation coefficient for ABO. In the numerical case it does not matter whether the rank restriction is implied during the iteration process or afterwards, provided that no missing data occur.

7. Discussion

The non-linear redundancy analysis presented in this article corresponds both to Van den Wollenberg's redundancy analysis (reduced rank regression) with optimal scaling

Table 4. Dutch Parliament, numerical solutions with rank restriction. Multiple correlations (MC-4 no rank restriction, MC-2 rank-2 restriction, MC-1 rank-1 restriction) and correlations between issues (columns) and preference votes (rows)

	DEV	ABO	LAW	INC	PAR	TAX	DEF
MC-4	0.610	0.668	0.721	0.728	0.687	0.769	0.698
MC-2	0.607	0.664	0.720	0.714	0.674	0.768	0.692
MC-1	0.600	0.296	0.696	0.694	0.635	0.755	0.671
PvdA	0.568	−0.386	0.675	−0.592	−0.602	0.699	0.626
ARP	0.028	0.517	−0.220	−0.127	−0.156	0.095	−0.211
KVP	−0.013	0.541	−0.290	0.075	−0.123	−0.034	−0.295
VVD	−0.523	−0.024	−0.515	0.680	0.589	−0.677	−0.527

and to multivariate multiple regression with optimal scaling. The technique maximizes the Redundancy Index of Stewart & Love (1968), and decomposes this index into orthogonal components. The algorithm is realized in the computer program REDUNDALS which can handle several types of non-linear transformations. This is a great advantage over other redundancy analysis programs. In addition only one transformation for each variable is obtained for all the multiple regressions together, which makes interpretation more simple than in case of separate analyses.

As the REDUNDALS program is implemented with a program for non-linear canonical correlation analysis (CANALS) the approach to missing data is the same. This means that missing observations are quantified such that the model is fitted optimally. Only one quantification for each missing value is computed instead of as many as there are criterion variables. Thus even in case of (incomplete) data with a numerical measurement level, one may prefer the REDUNDALS program over multivariate multiple regression, Van den Wollenberg's redundancy analysis or reduced rank regression with listwise or pairwise deletion.

The REDUNDALS program is written in Fortran. As CANALS and REDUNDALS are combined, the program can only be obtained together with the CANALS program.† Another computer program which can also perform multivariate multiple regression together with non-linear transformations is TRANSREG (Kuhfeld, Young & Kent, 1987). This program is implemented as a SAS procedure.

The difference between the results for CANALS and REDUNDALS is that CANALS finds direction(s) in both sets of variables (subspaces), that correlate maximally, independent of how much variance is explained, while REDUNDALS explains as much variance as possible in every criterion direction. This means that results are hardly comparable unless one of the canonical variates correlates strongly with one or more criterion variables. However they should never contradict each other.

In the case of the Dutch Parliamentary data the REDUNDALS results are mostly

†The present version of the CANALS/REDUNDALS program is only for personal use. It is meant to make a user's version. Please contact Eeke van der Burg if you are interested in the program.

comparable with the numerical CANALS analysis (Van der Burg & De Leeuw, 1983). The first dimension of this analysis is dominated by all the issues (except for ABO and DEV) and the VVD and PvdA preference, and the second dimension by ABO and the KVP and ARP preference. Of course the transformations differ, but we have seen from Fig. 1 that no large deviations exist from linearity (except perhaps for the KVP and the ARP transformations). The first CANALS dimension corresponds to the left–right contrast and the second dimension to the con–pro-abortion contrast. Indeed REDUNDALS shows a similar pattern both in the rank-4 (no restriction) as in the rank-2 solutions.

Although one expects that non-linear techniques give results that differ with results from a linear analysis, this does not happen in general. Many times only a different accent in interpretation is found. However, it is a result in its own right to know that the 'old' techniques were not too bad at all. We can only find this by using non-linear techniques.

In this paper nothing is said about significance of results. One way to check this is by using randomization techniques based on permutations (Edgington, 1980; Freedman & Lane, 1983). De Leeuw & Van der Burg (1986) made a small study on significance of results from non-linear canonical correlation analysis. The same method can be applied to REDUNDALS results. Another way to check stability is by using resampling techniques such as the jackknife and bootstrap (Efron, 1982). In Gifi (1990) many examples are found. Van der Burg & De Leeuw (1988) discuss estimation of confidence intervals for non-linear CCA with the help of jackknife and bootstrap. Both the permutation method as well as the randomization method will guard us against chance capitalization, and reveal outliers. Therefore, a stability study on non-linear redundancy analysis would be a valuable supplement to this paper.

References

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

De Leeuw, J. (1977). Normalized cone regression. University of Leiden, Department of Data Theory. Mimeo.

De Leeuw, J. (1984). The Gifi-system of nonlinear multivariate analysis. In E. Diday, M. Jambu, L. Lebart, J. Pages & R. Tomassone (Eds), *Data Analysis and Informatics III*, pp. 415–424. Amsterdam: North Holland.

De Leeuw, J. (1986). On the history of redundancy analysis. University of Leiden, Department of Data Theory.

De Leeuw, J. & Van der Burg, E. (1986). The permutational limit distribution of generalized canonical correlations. In E. Diday, Y. Escoufier, L. Lebart, J. P. Pagès, Y. Schektman & R. Tomassone (Eds), *Data Analysis and Informatics*, vol. IV, pp. 509–521. Amsterdam: North-Holland.

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

Edgington, E. S. (1980). *Randomization Tests.* New York: Marcel Dekker.

Fortier, J. J. (1966). Simultaneous linear prediction. *Psychometrika*, **31,** 369–381.

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans.* CBMS-NSF Regional Conference Series in Applied Mathematics, no. 38. Philadelphia: SIAM.

Freedman, D. A. & Lane, D. (1983). Significance testing in a nonstochastic setting. In P. J. Bickel, K. Doksum & J. L. Hodges Jr (Eds), *A Festschrift for Erich L. Lehmann.* Belmont: Wadworth.

Gifi, A. (1990). *Nonlinear Multivariate Analysis.* Chichester: Wiley.

Glahn, H. R. (1969). Some relationships derived from canonical correlation theory. *Econometrica*, **37**, 252–256.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations.* New York: Wiley.

Horst, P. (1955). A technique for the development of a multiple absolute prediction battery. *Psychological Monographs,* **69** (5), 1–22.

Israëls, A. Z. (1984). Redundancy analysis for qualitative variables. *Psychometrika*, **49,** 331–346.

Israëls, A. Z. (1986). Interpretation of redundancy analysis: rotated vs. unrotated solutions. *Applied Stochastic Models and Data Analysis*, **2,** 121–130.

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.

Johansson, J. K. (1981). An extension of Wollenberg's redundancy analysis. *Psychometrika*, **46,** 93–104.

Kruskal, J. B. & Shephard, R. N. (1974). A nonmetric variety of linear factor analysis. *Psychometrika*, **39,** 123–157.

Kuhfeld, W. F., Young, F. W. & Kent, D. P. (1987). New developments in psychometric and market research procedures. In *SUGI Proceedings*, **12,** 1101–1106. Cary, NC: SAS Institute.

Meulman, J. (1986). *A Distance Approach to Nonlinear Multivariate Analysis.* Leiden: DSWO Press.

Rao, C. R. (1962). Use of discriminant and allied functions in multivariate analysis. *Sankya A*, **24,** 149–154.

Rao, C. R. (1964). The use and interpretation of principal components analysis in applied research. *Sankya A*, **26,** 329–358.

Robert, P. & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: The RV-coefficient. *Applied Statistics*, **25,** 257–265.

Stewart, D. K. & Love, W. A. (1968). A general canonical correlation index. *Psychological Bulletin*, **70,** 160–163.

Tatsuoka, M. M. (1988). *Multivariate Analysis: Techniques for Educational and Psychological Research.* New York: Macmillan.

Ten Berge, J. M. F. (1985). On the relationship between Fortier's simultaneous linear prediction and Van den Wollenberg's redundancy analysis. *Psychometrika*, **50,** 121–122.

Van de Geer, J. P. (1984). Linear relations between *k* sets of variables. *Psychometrika*, **49,** 79–94.

Van den Wollenberg, A. L. (1977). Redundancy analysis: An alternative for canonical correlation analysis. *Psychometrika*, **42,** 207–219.

Van der Burg, E. (1988). *Nonlinear Canonical Correlation and Some Related Techniques.* Leiden: DSWO Press.

Van der Burg, E. & De Leeuw, J. (1983). Non-linear canonical correlation. *British Journal of Mathematical and Statistical Psychology*, **36,** 54–80.

Van der Burg, E. & De Leeuw, J. (1988). Use of the multinomial jackknife and bootstrap in generalized nonlinear canonical correlation analysis. *Applied Stochastic Models and Data Analysis*, **4**, 159–172.

Young, F. (1981). Qualitative analysis of qualitative data. *Psychometrika*, **46,** 247–388.

Young, F., De Leeuw, J. & Takane, Y. (1976). Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika*, **41,** 505–529.

Received 14 December 1987; revised version 10 March 1989