PSYCHOMETRIKA–VOL. 47, NO. 1
MARCH 1982

ON THE OPTIMALITY OF THE SIMULTANEOUS REDUNDANCY TRANSFORMATIONS

DAVID E. TYLER

OLD DOMINION UNIVERSITY

The objective of this paper is to introduce and motivate additional properties and interpretations for the redundancy variables. It is shown that these variables can be derived by application of certain invariance arguments and without reference to the index of redundancy. In addition, an optimality property for the variables is presented which is important whenever one restricts attention in a study to a subset of the redundancy variables. This optimality property pertains to the subset rather than to the individual variables.

Key words: Canonical correlation analysis, index of redundancy, principal components, redundancy transformations.

1. Introduction and Summary

Stewart and Love [1968] proposed an index to measure the degree to which one set of variables can predict another set of variables, or equivalently, how redundant one set is relative to another. Their index is commonly referred to as the "index of redundancy" and was orginially defined by the summation

$$R^2(\mathbf{Y}:\mathbf{X}) = \sum_{i=1}^{p} \frac{\rho_i^2 V_e(\mathbf{Y}:\mathbf{b}'_{(i)}\mathbf{Y})}{\mathrm{tr}(R_{YY})}, \tag{1}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are q-dimensional and p-dimensional multivariate responses respectively, ρ_i is the ith largest canonical correlation between $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{b}'_{(i)}\mathbf{Y}$ is the canonical variable for the $\mathbf{Y}$ set associated with ρ_i, and $V_e(\mathbf{b}'\mathbf{Y})$ is the amount of total variance of $\mathbf{Y}$ explained by the component $\mathbf{b}'\mathbf{Y}$. The joint variance-covariance matrix of $\mathbf{X}$ and $\mathbf{Y}$ is

$$R = \begin{pmatrix} R_{YY} & R_{YX} \\ R_{XY} & R_{XX} \end{pmatrix}. \tag{2}$$

Unless otherwise stated, R is assumed to be nonsingular. If the variables are standardized, then R represents the correlation matrix.

The redundancy index is asymmetric. That is, $R^2(\mathbf{Y}:\mathbf{X}) \neq R^2(\mathbf{X}:\mathbf{Y})$ except for very special cases. It distinguishes between the dependent variables ($\mathbf{Y}$) and the independent variables ($\mathbf{X}$). An important interpretation for the index of redundancy which was alluded to by Stewart and Love and formally proven by Gleason [1976] is that the index represents the proportion of the total variance in the $\mathbf{Y}$ set which is accounted for by the linear prediction of $\mathbf{Y}$ by $\mathbf{X}$. More specifically,

$$R^2(\mathbf{Y}:\mathbf{X}) = \frac{\mathrm{tr}(R_{YX} R_{XX}^{-1} R_{XY})}{\mathrm{tr}(R_{YY})}. \tag{3}$$

From this representation for the index of redundancy, it easily follows that the index is invariant under orthogonal transformations of the dependent variables and under non-

This paper is based in part on the author's doctoral dissertation, Department of Statistics, Princeton, University. Research was conducted under the supervision of Lawrence S. Mayer.

Requests for reprints should be sent to David E. Tyler, Department of Mathematical Sciences, Old Dominion University, Norfolk, Virginia 23508.

0033-3123/82/0300-4005

singular transformations of the independent variables. That is, if P is a $(p \times p)$ orthogonal matrix and A is a $(q \times q)$ nonsingular matrix, then

$$R^2(\mathbf{Y}: \mathbf{X}) = R^2(P'\mathbf{Y}: A'\mathbf{X}). \tag{4}$$

The index of redundancy, however, is not invariant under arbitrary nonsingular transformations of the dependent variables.

In practice, the index of redundancy is usually used as a summary index in conjunction with canonical correlation and variable analysis. It is argued, though, by Nicewander and Wood [1974, 1975] and by Cramer and Nicewander [1979] that the association of the index of redundancy with canonical correlation and variable analysis is somewhat artificial. Canonical correlation and variable analysis does not distinguish between dependent and independent variables whereas the index of redundancy does. In addition, the index of redundancy is only invariant under orthogonal transformations of the dependent variables, whereas the canonical correlations and variables are invariant under all nonsingular transformations of the dependent set of variables.

An alternative to canonical analysis was suggested by van den Wallenberg [1977], which he refers to as redundancy analysis. In redundancy analysis one successively extracts uncorrelated linear combinations of the independent variables which maximizes $R^2(\mathbf{X}: \mathbf{w}'\mathbf{X})$. This leads to the linear combinations $\mathbf{w}_1'\mathbf{X}, \mathbf{w}_2' X, \ldots, \mathbf{w}_r'\mathbf{X}$ which correspond to the eigenvectors

$$R_{XX}^{-1} R_{XY} R_{YX} \mathbf{w}_i = \lambda_i \mathbf{w}_i, \tag{5}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ are the nonzero eigenvalues of $R_{XX}^{-1} R_{XY} R_{YX}$, and

$$r = \text{rank}(R_{XX}^{-1} R_{XY} R_{YX}) = \text{rank}(R_{XY}) \leq \min(p, q). \tag{6}$$

By normalizing $\mathbf{w}_i$ such that $\mathbf{w}_i' R_{XX} \mathbf{w}_i = 1$, we then have $R^2(\mathbf{Y}: \mathbf{w}_i'\mathbf{X}) = \lambda_i$.

In van den Wallenberg's paper, it is suggested that $\mathbf{Y}$ be transformed in a manner similar to $\mathbf{X}$, that is, that the eigenvectors of $R_{YY}^{-1} R_{YX} R_{XY}$ be used. A disadvantage to this approach is that the transformation for $\mathbf{Y}$ is not related to the transformation for $\mathbf{X}$, whereas in canonical analysis the transformations for the two sets of variables are naturally related.

In view of this argument, Johansson [1981] suggested alternative transformations for the $\mathbf{Y}$ set which are naturally associated with the transformation for the $\mathbf{X}$ set. One of these approaches successively extracts linear combinations of the dependent variables $\mathbf{v}_1'\mathbf{Y}, \mathbf{v}_2'\mathbf{Y}, \ldots, \mathbf{v}_r'\mathbf{Y}$ such that the absolute value of $\mathbf{v}_i' R_{YX}\mathbf{w}_i$ is maximized subject to the constraints $\mathbf{v}_i'\mathbf{v}_i = 1$ and $\mathbf{v}_i'\mathbf{v}_j = 0$ for $i \neq j$. The vector $\mathbf{w}_i$ is defined in (5). The resulting solution is

$$\mathbf{v}_i = \eta_i^{-1} R_{YX} \mathbf{w}_i \tag{7}$$

where $\eta_i = (\mathbf{w}_i' R_{XY} R_{YX} \mathbf{w}_i)^{1/2}$, see Johansson [1981], section 4. Although not stated by Johansson, it easily follows that $\eta_i^2 = \lambda_i$ where λ_i, is defined as in (5).

The objective of this paper is to present a unified treatment of redundancy analysis and to introduce and motivate additional properties and interpretations for the eigenvectors $\{\mathbf{w}_i, \mathbf{v}_i\}$ and eigenvalues $\{\lambda_i\}$, which are hereafter referred to as the simultaneous redundancy transformations, and the redundancy roots respectively. In section 2, it is shown that the redundancy transformations can be derived by application of certain invariance arguments and without reference to the index of redundancy. This approach emphasizes their use in general for studies where a distinction is made between the dependent and independent variables and where only invariance under orthogonal transformations of the dependent variables is desirable, rather than simply for studies where the index of redundancy is used as a summary index. In section 3, additional optimality properties are given for the simultaneous redundancy transformations with respect to the index of redundancy. These optimality properties are important whenever one restricts attention in a study to a reduced set of redundancy variables. The relationships between the properties given in section 2 and 3 with the properties given by van den Wollenberg and Johansson are discussed in section 4. An example illustrating the use of the simultaneous redundancy transformations is given in section 5.

2. *The Redundancy Transformations*

Many scalar-valued indices which are strictly functions of the canonical correlations have been proposed to measure the relationship between two multivariate responses. These indices have the property that they are invariant under any nonsingular transformation of either set of responses. This property is not always a desirable property. For example, the concept of total variance of a set of variables is not invariant under all nonsingular transformations of the variables, only under orthogonal transformations. The ability of **X** to predict a linear combination of **Y** which accounts for a large proportion of the total variance of **Y** may be of more interest than the ability of **X** to predict a linear combination of **Y** which accounts for a small proportion of the total variance of **Y**. This distinction cannot be considered in an index which is strictly a function of the canonical correlations. This argument is similar to the argument given by Stewart and Love in motivating the definition of the index of redundancy.

Rather than simply considering a summary index, it is natural to address the following more general problem. In studies where there is a distinction between dependent and independent variables, how can one parsimoniously represent the interrelationship between the two sets of variables if only orthogonal transformations of the dependent variables are to be tolerated? For such studies, a canonical analysis would not be appropriate. In Theorem 1 of this section, it is shown that it is possible for R to be transformed into a relatively simple form by applying an orthogonal transformation to the dependent variables and a nonsingular transformation to the independent variables. It is also shown that the simultaneous redundancy transformations are the only transformations which can transform R into this simpler form. For brevity, the Kronecker delta, δ_{ij}, is used in the proof of the theorem. It is defined by $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$.

Theorem 1. There exists an orthogonal matrix V and a nonsingular matrix W such that

$$\begin{bmatrix} V' & 0 \\ 0 & W' \end{bmatrix} R \begin{bmatrix} V & 0 \\ 0 & W \end{bmatrix} = \begin{bmatrix} V'R_{YY}V & D \\ D' & I \end{bmatrix} \tag{8}$$

where for $p \geq q$, $D = [\Delta\colon 0]$, and for $p > q$, $D' = [\Delta\colon 0]$, and where Δ is a diagonal matrix of order min(p, q) with diagonal entries $\lambda_1^{1/2} \geq \lambda_2^{1/2} \geq \cdots \geq \lambda_{\min[p,q]}^{1/2} \geq 0$. Furthermore, any such matrices V and W must satisfy the relationships

$$R_{YX}R_{XX}^{-1}R_{XY}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \tag{9}$$

$$R_{XX}^{-1}R_{XY}R_{YX}\mathbf{w}_i = \lambda_i \mathbf{w}_i, \tag{10}$$

and

$$\lambda_i^{1/2}\mathbf{w}_i = R_{XX}^{-1} R_{XY}\mathbf{v}_i \tag{11}$$

where $\mathbf{v}_i$ and $\mathbf{w}_i$ are the ith columns of V and W respectively, and with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ being the nonzero roots of $R_{YX}R_{XX}^{-1}R_{XY}$ and $\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_{\max[p,q]} = 0$.

Proof. Since $R_{YX}R_{XX}^{-1}R_{XY}$ is symmetric and positive semidefinite its eigenvalues are nonnegative. Also, if $\{v_i\}$ is defined by (9), then they can be chosen such that $\mathbf{v}_i'\mathbf{v}_j = \delta_{ij}$. By construction, V is an orthogonal matrix. If $\mathbf{w}_i$ is defined by (11) for $i = 1, 2, \ldots, r$ then it follows that $\mathbf{w}_i' R_{XX}\mathbf{w}_j = \delta_{ij}$ for $i, j = 1, 2, \ldots, r$. It should be noted that multiplying both sides of (11) by $R_{XX}^{-1} R_{XY} R_{YX}$ gives (10) for $i = 1, 2, \ldots, r$. For i $= r + 1, r + 2, \ldots, q$, let $\mathbf{w}_i$ be chosen so that (10) also holds, that is $R_{XX}^{-1} R_{XY} R_{YX}\mathbf{w}_i = 0$, and so that $\mathbf{w}_i' R_{XX}\mathbf{w}_j = \delta_{ij}$ for $i, j = r + 1, r+2, \ldots, q$. Hence, by construction $W'R_{XX}W = I$. Finally, for $i \le r$, $\mathbf{v}_j' R_{YX}\mathbf{w}_i = \lambda_i^{-1/2}\mathbf{v}_j' R_{YX}R_{XX}^{-1}R_{XY}\mathbf{v}_i = \lambda_i^{1/2}\delta_{ij}$, and for $i > r$, $\mathbf{v}_j' R_{YX}\mathbf{w}_i = 0$, and so $V'R_{YX}W = D$. Thus, it has been shown that there exists an orthogonal matrix V and a nonsingular matrix W, namely those whose columns satisfy (9), (10) and (11), such that statement (8) holds.

To complete the proof, it must be shown that the only choices for the orthogonal and nonsingular matrices V and W for which (8) holds are those satisfying (9), (10), and (11). Note that $V'R_{YX}W = D$ and $W'R_{XX}W = I$ implies

$$R_{YX} = VDW^{-1} \qquad \text{and} \qquad R_{XX}^{-1} = WW'. \tag{12}$$

Thus, $R_{YX}R_{XX}^{-1}R_{XY} = VDW^{-1}WW'(W')^{-1}DV' = VDD'V' = V\Lambda_1 V'$, where $\Lambda_1 =$ diagonal$(\lambda_1, \lambda_2, \ldots, \lambda_p)$. This implies that $\mathbf{v}_i$ and λ_i, $i = 1, 2, \ldots, p$ satisfy $R_{YX}R_{XX}^{-1}R_{XY}\mathbf{v}_i = \lambda_i\mathbf{v}_i$. Likewise, $R_{XX}^{-1}R_{XY}R_{YX} = WW'(W')^{-1}D'V'VDW^{-1} = W\Lambda_2 W^{-1}$, where $\Lambda_2 =$ diagonal$(\lambda_1, \lambda_2, \ldots, \lambda_q)$, implies that $\mathbf{w}_i$ and λ_i, $i = 1, 2, \ldots, p$ satisfy $R_{XX}^{-1}R_{XY}R_{YX}\mathbf{w}_i = \lambda_i\mathbf{w}_i$. Finally, $R_{XX}^{-1}R_{XY}\mathbf{v}_i = WW'(W')^{-1}D'\mathbf{v}_i = \lambda_i^{1/2}\mathbf{w}_i$. This completes the proof. □

The transformations V' and W' when simultaneously applied to $\mathbf{Y}$ and $\mathbf{X}$ respectively, are the simultaneous redundancy transformations. Note that multiplying both sides of (7) by $R_{YX}R_{XX}^{-1}R_{XY}$ gives (9). This shows that the definition of $\mathbf{v}_i$ given in Theorem 1 is consistent with Johansson's definition and that $\eta_i^2 = \lambda_i$. Equations (9) and (11) are easier to use for calculations than (5) and (7) since they involve the eigenvalues and eigenvectors of a symmetric matrix rather than a nonsymmetric matrix.

By examining (9), it is interesting to note that the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ correspond to the principal component vectors for $\hat{\mathbf{Y}}$, where $\hat{\mathbf{Y}}$ is the linear regression of $\mathbf{Y}$ on $\mathbf{X}$. That is,

$$\hat{\mathbf{Y}} = \boldsymbol{\mu}_Y + R_{YX}R_X^{-1}(\mathbf{X} - \boldsymbol{\mu}_X), \tag{13}$$

where $\boldsymbol{\mu}_Y$ and $\boldsymbol{\mu}_X$ are the expected values of $\mathbf{Y}$ and $\mathbf{X}$ respectively, which without loss of generality are hereafter taken to be zero. The variance-covariance matrix for $\hat{\mathbf{Y}}$ is $R_{\hat{Y}} = R_{YX}R_{XX}^{-1}R_{XY}$. By examining (11), note that the variable $\lambda_i^{1/2}\mathbf{w}_i'\mathbf{x}$ is simply the linear regression of $\mathbf{v}_i'\mathbf{Y}$ on X. Thus, (9) and (11) provide easily understood interpretations for the redundancy variables.

Equation (8) gives the resulting joint variance-covariance matrix when the simultaneous redundancy transformations are applied to the variables $\mathbf{Y}$ and $\mathbf{X}$. As an exploratory technique, the redundancy transformations do not simplify the joint variance-covariance matrix to the extent that the canonical transformations do. This is to be expected since less information on the joint variance-covariance matrix is lost when only considering orthogonal transformations of $\mathbf{Y}$. In particular, the index of redundancy is preserved. That is, by applying statement (4) with $P = V$ and $A = W$, we obtain

$$R^2(\mathbf{Y}: \mathbf{X}) = R^2(V'\mathbf{Y}: W'\mathbf{X}) = \frac{\sum_{i=1}^{r}\lambda_i}{\text{tr}(R_{YY})}. \tag{14}$$

3. *The Optimality of the Redundancy Transformations*

When the index of redundancy is used in conjunction with canonical correlation and variable analysis, the quantity

$$R^2(\mathbf{Y}: \mathbf{a}'_{(i)}\mathbf{X}) = \frac{\rho_i^2 V_e(\mathbf{Y}: \mathbf{b}'_{(i)}\mathbf{Y})}{tr(R_{YY})}, \tag{15}$$

where $\mathbf{a}'_{(i)}\mathbf{X}$ is the canonical variable for the $\mathbf{X}$ set associated with ρ_i, is usually used as an aid in determining which canonical variables deserve interpretation and further attention rather than simply using the canonical correlations themselves. As noted by van den Wollenberg, the practice of using the value of (15) for each of the canonical variables to reduce, in essence, the dimensionality of the two sets of multivariate responses is not an optimal procedure. The canonical variables are extracted because they best explain the intercorrelations between the sets of responses. They are not necessarily the best linear combinations to consider when attempting to account for the overall size of the index of redundancy. It is shown in this section that the redundancy variables are best suited for this purpose. Before doing so, it is first necessary to extend the concept of the contribution made by a canonical variable to the overall size of the index of redundancy, given by (15), to the contribution made by any set of linear combinations of the dependent or of the independent variables to the overall size of the index.

A natural extension for an arbitrary set of linear combinations of the independent variables is the proportion of the total variance of the dependent variables which can be explained by its linear regression on these linear combinations only. That is,

$$R^2(\mathbf{Y}: A'\mathbf{X}) = \frac{tr[R_{YX} A(A'R_{XX}A)^{-1}A'R_{XY}]}{tr(R_{YY})}, \tag{16}$$

where A is a $(q \times k)$ matrix, can be considered the contribution made by the set of linear combinations $A'\mathbf{X}$ to the overall size of the index $R^2(\mathbf{Y}: \mathbf{X})$. If $A'R_{XX}A$ is singular, say $\text{rank}(A'R_{XX}A) = t < k$, then the definition of the index of redundancy can be logically extended by defining

$$R^2(\mathbf{Y}: A'\mathbf{X}) = R^2(\mathbf{Y}: C'A'\mathbf{X}), \tag{17}$$

where C is a $(k \times t)$ matrix such that $C'A'R_{XX}AC$ is nonsingular. This definition does not depend upon the particular choice of C, and also represent the proportion of the total variance of $\mathbf{Y}$ which can be explained by the linear combinations $A'\mathbf{X}$.

Thus defined, the contributions to the index made by uncorrelated linear combinations of the independent variables are additive. If $A = [A_1 : A_2]$ with $A'_1 R_{XX} A_2 = 0$, then

$$R^2(\mathbf{Y}: A'\mathbf{X}) = R^2(\mathbf{Y}: A'_1\mathbf{X}) + R^2(\mathbf{Y}: A'_2\mathbf{X}). \tag{18}$$

In particular, if A_0 is a $(q \times k)$ matrix with $\text{rank}(A_0) = k$ and whose columns are a subset of the canonical vectors, say $\{\mathbf{a}_{(i)}, i \in I\}$, then we have the desired result

$$R^2(\mathbf{Y}: A'_0\mathbf{X}) = \frac{\sum_{i \in I} \rho_i^2 V_e(\mathbf{Y}: \mathbf{b}'_{(i)}\mathbf{Y})}{tr(R_{YY})}. \tag{19}$$

In addition, it should be noted that the index of redundancy can be decomposed over any complete set of uncorrelated linear combinations of the independent variables. That is, if $\mathbf{a}_1$,

$\mathbf{a}_2, \ldots, \mathbf{a}_q$ is any set of nonzero vectors such that $\mathbf{a}_i' R_{XX} \mathbf{a}_j = 0$ for $i \neq j$, then

$$R^2(\mathbf{Y}: \mathbf{X}) = \sum_{i=1}^{q} R^2(\mathbf{Y}: \mathbf{a}_i'\mathbf{X}). \tag{20}$$

By applying (15), one observes that (20) is a generalized version of the summation given in the definition of the index of redundancy.

For an arbitrary set of linear combinations of the dependent variables, a suitable extension is not obvious. One extension proposed by Miller and Farr [1971] for any linear combination $\mathbf{b}'\mathbf{Y}$ is the product $R^2(\mathbf{Y}: \mathbf{b}'\mathbf{Y})R^2(\mathbf{b}'\mathbf{Y}: \mathbf{X})$. This product is the proportion of the total variance of $\mathbf{Y}$ which can be explained by the linear combination $\mathbf{b}'\mathbf{Y}$ times the proportion of the variance of $\mathbf{b}'\mathbf{Y}$ which can be explained by $\mathbf{X}$. If $\mathbf{b}'\mathbf{Y}$ is taken to be the ith canonical variable $\mathbf{b}_{(i)}'\mathbf{Y}$, then this product is the same as (15), since $R^2(\mathbf{Y}: b_{(i)}'\mathbf{Y}) = V_e(\mathbf{Y}: \mathbf{b}_{(i)}'\mathbf{Y})/\text{tr}(R_{YY})$ and $R^2(\mathbf{b}_{(i)}'\mathbf{Y}: \mathbf{X}) = \rho_i^2$. Unfortunately, it has recently been shown that it is possible for the product $R^2(\mathbf{Y}: \mathbf{b}'\mathbf{Y})R^2(\mathbf{b}'\mathbf{Y}: \mathbf{X})$ to be greater than the index of redundancy itself, see Tyler [1982]. Thus an alternative generalization is needed.

To motivate an alternative generalization, we note that

$$R^2(\mathbf{Y}: B_0'\hat{\mathbf{Y}}) = \frac{\sum_{i \in I} \rho_i^2 V_e(\mathbf{Y}: \mathbf{b}_{(i)}'\mathbf{Y})}{tr(R_{YY})}, \tag{21}$$

where B_o is a $(p \times k)$ matrix with $\text{rank}(B_0) = k$ and whose columns are the canonical vectors $\{\mathbf{b}_{(i)}, i \in I\}$. It is therefore proposed that $R^2(\mathbf{Y}: B'\hat{\mathbf{Y}})$ be considered as a generalization of (15) for the contribution made by the set of linear combinations $B'\mathbf{Y}$ to the overall size of the index of redundancy, where B is any matrix of order $(p \times k)$. This quantity represents the proportion of the total variance of $\mathbf{Y}$ which can be accounted for by the linear regression of $B'\mathbf{Y}$ on $\mathbf{X}$.

As defined above, the contributions to the index of redundancy made by uncorrelated linear combinations of the dependent variables are not necessarily additive. The contributions made by the linear combinations of $\mathbf{Y}$ whose linear regression on $\mathbf{X}$ are uncorrelated, however, are additive. That is, if $B = [B_1 : B_2]$ with $B_1' R_{YX} R_{XX}^{-1} R_{XY} B_2 = 0$, then

$$R^2(\mathbf{Y}: B'\hat{\mathbf{Y}}) = R^2(\mathbf{Y}: B_1'\hat{\mathbf{Y}}) + R^2(\mathbf{Y}: B_2'\hat{\mathbf{Y}}). \tag{22}$$

In particular, if $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p$ are any set of vectors such that $\mathbf{b}_i' R_{YX} R_{XX}^{-1} R_{XY} \mathbf{b}_j = 0$ for $i \neq j$, then

$$R^2(\mathbf{Y}: \mathbf{X}) = \sum_{i=1}^{p} R^2(\mathbf{Y}: \mathbf{b}'\hat{\mathbf{Y}}). \tag{23}$$

Since $R^2(\mathbf{Y}: \mathbf{b}_{(i)}'\hat{\mathbf{Y}}) = R^2(\mathbf{Y}: \mathbf{a}_{(i)}'\mathbf{X})$, it can be observed by application of (15) that (23) is a generalization of the summation defining the index of redundancy.

In view of these extensions of (15), an important optimality property of the redundancy transformations is given in the next theorem. This theorem states that of all sets of k pairs of linear combinations of $\mathbf{X}$ and $\mathbf{Y}$, the redundancy variables associated with the k largest redundancy roots best account for the overall size of the index of redundancy.

Theorem 2. Let $\mathbf{v}_i$ and $\mathbf{w}_i$ be defined as in Theorem 1, let V_R be a $(q \times k)$ matrix with columns $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, and let W_k be a $(p \times k)$ matrix with columns $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$.

(i) For any $(q \times k)$ matrix A, $R^2(\mathbf{Y}: A'X) \leq R^2(\mathbf{Y}: V_k'\mathbf{X})$.
(ii) For any $(p \times k)$ matrix B, $R^2(\mathbf{Y}: B'\hat{\mathbf{Y}}) \leq R^2(\mathbf{Y}: W_k'\hat{\mathbf{Y}})$.

Proof. Part (i) follows from the results of Rao [1964], section 8. In that paper, it is shown that the quantity $\mathrm{tr}[R_{YY} - R_{YX}A'(AR_{XX}A')^{-1}A'R_{XY}]$ is minimized over all A of order $(k \times q)$ with $\mathrm{rank}(A) = k$ by choosing $A = V_k$. For all such A, the inequality in part (i) holds by noting the relationship between the above form and (16). The inequality easily extends to any A of order $(k \times q)$ with $\mathrm{rank}(A) \leq k$ by application of (17).

To prove part (ii), we note that for all B of rank less than or equal to k, $R^2(\mathbf{Y}: B'\hat{\mathbf{Y}})$ is maximized by choosing B such that $B'R_{YX}R_{XX}^{-\frac{1}{2}} = MV_k'$, where M has full rank. This follows from part (i). For $k \leq r$, it then follows from (12) that $W_k'R_{YX}R_{XX}^{-\frac{1}{2}} = W_k'WDV^{-1}VV' = D_kV_k'$ where $D_k = \mathrm{diagonal}(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_k^{1/2})$. If $k > r$, part (ii) is immediate, since $R^2(\mathbf{Y}: W_k'\hat{\mathbf{Y}}) = R^2(\mathbf{Y}:\mathbf{X})$. □

If only the redundancy variables associated with the k largest redundancy roots are retained for further study, then the contribution of this reduced set of variables to the overall size of the index of redundancy is given by

$$R^2(\mathbf{Y}: W_k'\mathbf{X}) = R^2(\mathbf{Y}: V_k'\hat{\mathbf{Y}}) = \frac{\sum_{i=1}^{k}\lambda_i}{\mathrm{tr}(R_{YY})}. \tag{24}$$

By Theorem 2, we know that this contribution cannot be improved upon for any other set of k linear combinations of the two sets of variables.

After reducing a multivariate response to a smaller set of linear combinations of the response, it is customary in practice to consider linear transformations of the reduced set of linear combinations. These linear transformations are usually made to facilitate the interpretation of the reduced set. So, it is important to note that the optimality property for $W_k'\mathbf{X}$ and $V_k'\mathbf{Y}$ given in Theorem 2 still holds if either is transformed by a nonsingular linear transformation. However, in view of the discussion in section 2, only orthogonal transformations of $V_k'\mathbf{Y}$ would be appropriate.

4. Discussion

As stated in the introduction, the importance of the results of section 2, is that a derivation is given for the redundancy transformations without reference to the index of redundancy. An analogy to such a derivation can be found in principal components analysis. It is well known that the principal component variables can be derived by successively maximizing the variances of uncorrelated linear combinations of the original set of variables under the constraint that the linear combination has a sum of squared weights equal to one. Alternatively, the principal component variables can also be derived by successively maximizing the proportion of total variance explained by uncorrelated linear combinations. Whether variance, variance explained or some other concept is a useful criterion can always be debated. Thus, principal components analysis is often viewed as simply an orthogonal, hence "nondistorting", transformation to uncorrelated variables. In addition to the two optimality properties stated above, the principal components transformation can then be shown to have many other optimality properties, see Okamoto [1969].

The importance of the optimality properties given in section 3 is that they pertain to a set of redundancy variables. The properties for the redundancy variables given by van den Wollenberg and Johansson pertain to the redundancy variables when extracted successively. A suitably defined optimality property for a set of variables does not follow from the optimality of the individual variables without proof.

It should be noted that the redundancy transformation for the dependent set, that is $W'\mathbf{X}$, was first proposed by Rao [1964], section 8, as an alternative to the canonical transformation. He derives the transformation in the context of optimizing "predictive efficiency", a concept which is equivalent to redundancy. He refers to the transformation W' as the principal components transformation for the instrumental variable **X** with respect to the variable **Y**. This transformation also arises in reduced-rank regression problems, for example see Izenman [1976].

Before concluding the discussion, it should also be noted that Johansson also proposed an alternative analysis for the dependent set of variables (**Y**). This alternative analysis consists of successively extracting linear combinations of the dependent variables, $\mathbf{u}_1'\mathbf{Y}$, $\mathbf{u}_2'\mathbf{Y}, \ldots, \mathbf{u}_r'\mathbf{Y}$, such that the absolute value of $\mathbf{u}_1' R_{YX} \mathbf{w}_i$ is maximized subject to the constraints that $\mathbf{u}_i' R_{YY} \mathbf{u}_i = 1$ and $\mathbf{u}_i' R_{YY} \mathbf{u}_j = 0$ for $i \neq j$, see Johansson [1981], section 3. A disadvantage to this procedure is that the solution does not give an orthogonal transformation of the dependent variables, and thus does not preserve the index of redundancy. In addition, it is not obvious what optimality properties a set of such transformed variables possess. For these reasons, this approach was not considered in this paper.

5. *An Example*

To illustrate the use of the simultaneous redundancy transformations, the following example is taken from section 4.15 of Timm [1975]. As reported by Timm, this example is based on a study by Dr. William D. Rohwer of the University of California at Berkeley. Further background information for this study is given in sections 4.3 and 4.7 of Timm's book.

In the study two sets of measurements are recorded for 37 students. One set of measurements consisting of basic tasks scores for "named action" (NA) and "sentence still" (SS) are considered to be independent variables. Another set of measurements consisting of scores for a standard achievement test (SAT), the Peabody Picture vocabulary test (PEA) and the Ravin Progression Matrices Test (RAV) are considered to be dependent variables. The joint sample correlation matrix for **Y** = (SAT PEA RAV)′ and **X** = (NA SS)′ is

$$R = \left[\begin{array}{ccc|cc} 1.0000 & & & & \\ .3703 & 1.0000 & & & \\ .2114 & .3548 & 1.0000 & & \\ \hline .2617 & .6720 & .3390 & 1.0000 & \\ .3341 & .5876 & .3404 & .7951 & 1.0000 \end{array}\right]$$

where the upper triangular entries are obtained by the symmetry of R.

Timm uses this example to illustrate the use of canonical correlation analysis. However, since there is a distinction between the dependent and independent variables, a redundancy analysis may be more appropriate. In addition, after extracting canonical variables, Timm calculates the individual redundancies for each of the canonical variates using (15) since "they better summarize the overlap between the two sets of variables than the squares of the canonical correlations". This further suggests that redundancy analysis should be considered in order to avoid the inconsistencies inherent in simultaneously using canonical variates and redundancy indices.

In calculating the redundancy roots and variables, the symmetric matrix $R_{YX} R_{XX}^{-1} R_{XY}$ is first calculated and then its eigenvalues and normalized eigenvectors are computed. This gives $\lambda_1 > \lambda_2 > \lambda_3 = 0$ and the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. The values of $\mathbf{w}_1$ and $\mathbf{w}_2$ are then computed using the relationships $\mathbf{w}_1 = \lambda_1^{-1/2} R_{XX}^{-1} R_{XY} \mathbf{v}_1$ and $\mathbf{w}_2 = \lambda_2^{-1/2} R_{XX}^{-1} R_{XY} \mathbf{v}_2$. The results are $\lambda_1 = 0.6735$, $\lambda_2 = 0.0261$,

$$V = \begin{bmatrix} -.3710 & -.8528 & .3676 \\ -.8203 & .4865 & .3008 \\ -.4353 & -.1900 & -.8800 \end{bmatrix}$$

and

$$W = \begin{bmatrix} -\ .6503 & .2982 \\ -2.0421 & -1.5991 \end{bmatrix}.$$

The index of redundancy is $R^2(\mathbf{Y}: \mathbf{X}) = \mathrm{tr}(R_{YX} R_{XX}^{-1} R_{XY})/3 = (\lambda_1 + \lambda_2)/3 = .2332$, and the covariance matrix for the redundancy variables of the dependent set is

$$V' R_{YY} V = \begin{bmatrix} 1.5471 & & \\ .2657 & .6957 & \\ .0919 & -.0570 & .7573 \end{bmatrix}$$

From this analysis, one can note that the "redundancy" of the **Y** set given the **X** set is principally due to the predictive ability of the variate .65(NA) + 2.04(SS) for the variate .37(SAT) + .82(PEA) + .44(RAV). For this particular example, the results of the redundancy analysis are similar to the results of the canonical variables analysis, see section 4.15 of Timm [1975]. This is to be expected whenever the canonical variate for **Y** associated with the largest canonical correlation explains a high percentage of the total variation within the **Y** set.

The results of a redundancy analysis would differ greatly from the results of a canonical analysis whenever the dependent set of variables contains a highly predictable component which only accounts for a small percentage of the total variation of the dependent set. Such a component could be interpreted as a highly predictable "noise" factor, and not a factor for which the dependent set of variables were selected to reflect. Canonical analysis would place emphasis on such components, whereas redundancy analysis would not.

In conclusion, the simultaneous redundancy transformations should prove to be a viable exploratory procedure. Their use is recommended in studies where the structure within the dependent set of variables up to an orthogonal transformation is considered important, and whenever one does not postulate a more refined model for the structure within and between the two sets of variables.

REFERENCES

Cramer, E. M. & Nicewander, W. A. Some symmetric, invariant measures of multivariate association. *Psychometrika*, 1979, *44*, 43–54.

Gleason, T. C. On redundancy in canonical analysis. *Psychological Bulletin*, 1976, 83, 1004–1006.

Izenman, A. J. Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, 1976, *5*, 248–264.

Johansson, J. K. An extension of Wollenberg's redundancy analysis. *Psychometrika*, 1981, *46*, 93–103.

Miller, J. K. & Farr, S. D. Bi-multivariate redundancy: a comprehensive measure of interbattery relationship. *Multivariate Behavior Research*, 1971, *6*, 313–324.

Nicewander, W. A. & Wood, D. A. Comments on "A general canonical correlation index." *Psychological Bulletin*, 1974, *81*, 92–94.

Nicewander, W. A. & Wood, D. A. On the mathematical bases of the general canonical correlation index: rejoinder to Miller. *Psychological Bulletin*, 1975, *82*, 210–212.

Okamoto, M. Optimality of principal components. In P. R. Krishnaiah (ed.), *Multivariate Analysis* II. New York: Academic Press, 1969, pp. 673–685.

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

Stewart, D. and Love, W. A general canonical correlation index. *Psychological Bulletin*, 1968, *70*, 160–163.

Timm, N. H. *Multivariate Analysis with Applications in Education and Psychology*. Monterey, California: Brooks/Cole Publishing Company, 1975.

Tyler, D. E. A counterexample to Miller and Farr's algorithm for the index of redundancy. To appear, *Multivariate Behavior Research*, 1982.

van den Wollenberg, A. L. Redundancy analysis: an alternative to canonical correlation analysis. *Psychometrika*, 1977, *42*, 207–219.

Manuscript received 7/10/81
Final version received 12/21/81