ON CANONICAL CORRELATION AND REDUNDANCY

Ke, Hui-xin
Department of Information Systems, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University

Asano, Choichiro
Department of Information Systems, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University

https://doi.org/10.15017/17703

出版情報：九州大学大学院総合理工学報告. 9 (2), pp.263-274, 1988-01-01. 九州大学大学院総合理工学研究科

ON CANONICAL CORRELATION AND REDUNDANCY

Hui-xin Ke* and Chooichiro Asano**

(Received September 28, 1987)

Relationships between the original definition of redundancy index first proposed by Stewart and Love [1] and the modified definition developed by Van Den Wollenberg [2] are presented, thereby supporting Van Den Wollenberg's method [2] on redundancy analysis and Desarbo's method [3] on canonical/redundancy analysis and showing that redundancy index has both theoretical and practical use in meeting the arguments between Nicewander and Wood [4], [5], and Miller [6].

The relative efficiencies of prediction of one set on the other using both canonical correlations and redundancies are given so that similar points and differences between canonical correlations and redundancies are shown.

A summary about two definitions of redundancy, canonical analysis and redundancy analysis is also presented.

1. Introduction

Suppose there are two sets of variables

$$\mathbf{X}^{(1)}=(X_1^{(1)}, X_2^{(1)}, \cdots, X_{p_1}^{(1)})' \quad \text{and} \quad \mathbf{X}^{(2)}=(X_1^{(2)}, X_2^{(2)}, \cdots, X_{p_2}^{(2)})'. \tag{1}$$

Assume that

$$\left.\begin{array}{l} E[\mathbf{X}^{(1)}]=\mathbf{0}, \quad E[\mathbf{X}^{(2)}]=\mathbf{0}, \quad k=\min(p_1, p_2), \\ V[X_i^{(1)}]=V[X_j^{(2)}]=1, \quad i=1, 2, \cdots, p_1; \quad j=1, 2, \cdots, p_2. \end{array}\right\} \tag{2}$$

The correlation matrix of $\mathbf{X}=(\mathbf{X}^{(1)\prime}, \mathbf{X}^{(2)\prime})'$ is given as

$$\mathbf{R}=E(\mathbf{XX}')=\begin{pmatrix}\mathbf{R}_{11} & \mathbf{R}_{12}\\ \mathbf{R}_{21} & \mathbf{R}_{22}\end{pmatrix}, \text{ where } \mathbf{R}_{ij}=E[\mathbf{X}^{(i)}\mathbf{X}^{(j)\prime}], \quad i, j=1, 2. \tag{3}$$

In canonical correlation analysis, k pairs of canonical variables (u_i, v_i)

$$\left.\begin{array}{l} \mathbf{U}=(u_1 \cdots u_i \cdots u_k)'=\boldsymbol{\alpha}'\mathbf{X}^{(1)}=(\boldsymbol{\alpha}^{(1)}, \cdots, \boldsymbol{\alpha}^{(i)}, \cdots, \boldsymbol{\alpha}^{(k)})'\mathbf{X}^{(1)} \\ \mathbf{V}=(v_1 \cdots v_i \cdots v_k)'=\boldsymbol{\beta}'\mathbf{X}^{(2)}=(\boldsymbol{\beta}^{(1)}, \cdots, \boldsymbol{\beta}^{(i)}, \cdots, \boldsymbol{\beta}^{(k)})'\mathbf{X}^{(2)} \end{array}\right\} \tag{4}$$

are derived so as to maximize the correlations between every pair of $u_i=\boldsymbol{\alpha}^{(i)\prime}\mathbf{X}^{(1)}$ and $v_i=\boldsymbol{\beta}^{(i)\prime}\mathbf{X}^{(2)}$:

*Department of Information Systems, Visiting Scholar
**Department of Information Systems

$$\varphi_i = E(u_i\, v_i) = \boldsymbol{\alpha}^{(i)\prime}\mathbf{R}_{12}\boldsymbol{\beta}^{(i)} \tag{5}$$

under the constraints:

$$\left.\begin{aligned}
&V(u_i) = E(u_i^2) = E(\boldsymbol{\alpha}^{(i)\prime}\mathbf{X}^{(1)}\mathbf{X}^{(1)\prime}\boldsymbol{\alpha}^{(i)}) = \boldsymbol{\alpha}^{(i)\prime}\mathbf{R}_{11}\boldsymbol{\alpha}^{(i)} = 1,\\
&V(v_i) = E(v_i^2) = E(\boldsymbol{\beta}^{(i)\prime}\mathbf{X}^{(2)}\mathbf{X}^{(2)\prime}\boldsymbol{\beta}^{(i)}) = \boldsymbol{\beta}^{(i)\prime}\mathbf{R}_{22}\boldsymbol{\beta}^{(i)} = 1,\\
&E(u_{i+1}u_i) = E(u_{i+1}v_i) = E(v_{i+1}v_i) = E(v_{i+1}u_i) = 0, \quad i = 1, 2, \cdots, k-1.
\end{aligned}\right\} \tag{6}$$

According to Anderson [7], the solutions come from the following eigenstructure equations:

$$\left.\begin{aligned}
&\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\boldsymbol{\alpha} = \boldsymbol{\alpha}\Lambda_1^2 \quad \text{or} \quad \mathbf{H}_1\boldsymbol{\alpha} = \boldsymbol{\alpha}\Lambda_1^2,\\
&\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\boldsymbol{\beta} = \boldsymbol{\beta}\Lambda_2^2 \quad \text{or} \quad \mathbf{H}_2\boldsymbol{\beta} = \boldsymbol{\beta}\Lambda_2^2,
\end{aligned}\right\} \tag{7}$$

where Λ_i^2 $(i=1, 2)$ are diagonal matrices, whose diagonal elements are eigenvalues of $\mathbf{H}_1$ and $\mathbf{H}_2$ respectively arranged in decreased order, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are matrices with corresponding eigenvectors as their column vectors. It can be shown that k diagonal elements of Λ_1^2 and Λ_2^2 are equal. They are $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_k^2 \geq 0$.

So in canoncal correlation analysis, it is usually to draw canonical variables or canonical factors $u_1, \cdots, u_k$ and $v_1, \cdots, v_k$, which have the highest correlations at the respective stages and show the common characteristics of the two sets.

In dealing with the relationships between two sets of variables, such canonical correlations present some interpretive difficulties. Stewart and Love [1] pointed out that: '… canonical correlations cannot be interpreted as correlations between sets of variables. It is important to note that a relatively strong canonical correlation may obtain between two linear functions (composits), even though these linear functions may not extract significant portions of variance from their respective batteries'. In their article, a nonsymmetric index of redundancy was proposed, which may be represented as the intersection of two sets of variables, that is, the proportion of one set which is in the intersection.

Suppose u_i and v_i are the ith canonical variables of $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$, $i=1, 2, \cdots, k$, as described above:

$$\left.\begin{aligned}
&\mathbf{U} = \boldsymbol{\alpha}'\mathbf{X}^{(1)}, \quad \mathbf{V} = \boldsymbol{\beta}'\mathbf{X}^{(2)},\\
&E(\mathbf{U}) = E(\mathbf{V}) = \mathbf{0}, \quad V(u_i) = V(v_i) = 1, \quad E(\mathbf{U}\mathbf{V}') = \Lambda.
\end{aligned}\right\} \tag{8}$$

Now we use canonical factors to present original variables in sets of $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ as follows,

$$\mathbf{X}^{(1)} = \mathbf{F}\mathbf{U} + \boldsymbol{\varepsilon}, \quad \mathbf{X}^{(2)} = \mathbf{F}^*\mathbf{V} + \boldsymbol{\varepsilon}^*, \tag{9}$$

where $\mathbf{F}$ and $\mathbf{F}^*$ are factor loading matrices of $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ on $\mathbf{U}$ and $\mathbf{V}$, and $\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}^*$ are error vectors with $E[\boldsymbol{\varepsilon}]=E[\boldsymbol{\varepsilon}^*]=\mathbf{0}$. Stewart and Love's definition of redundancy $R(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ of $\mathbf{X}^{(2)}$-variables given $\mathbf{X}^{(1)}$-variables is an index of proportion of variance of $\mathbf{X}^{(2)}$ predictable from $\mathbf{X}^{(1)}$.

$$R(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=\sum_{i=1}^{k}\lambda_i^2 VX_i^{(2)}=\sum_{i=1}^{k}\lambda_i^2\left(\sum_{j=1}^{p2}f_{ji}^2/p_2\right),\quad k=\min(p_1,p_2), \tag{10}$$

where f_{ji} is the element of the jth row and the ith column in $\mathbf{F}^*$, the loading of $X_j^{(2)}$ on v_i (the correlation between $X_j^{(2)}$ and v_i). So $\sum_{j=1}^{p2}f_{ji}^2$ means the variance of the set of $\mathbf{X}^{(2)}$ extracted by canonical factor v_i, and $VX_i^{(2)}=\sum_{j=1}^{p2}f_{ji}^2/p_2$ is the proportion of the variance. Multiplied by λ_i^2, the squared canonical correlation between u_i and v_i, $\lambda_i^2VX_i^{(2)}$ shows the variance proportion predictable from u_i. Therefore $\sum_{i=1}^{k}\lambda_i^2VX_i^{(2)}$ may be interpreted as the proportion of variance proportion of one set of $\mathbf{X}^{(2)}$ explained by the other set of $\mathbf{X}^{(1)}$. The redundancy of the $\mathbf{X}^{(1)}$-variables given the $\mathbf{X}^{(2)}$-variables $R(\mathbf{X}^{(1)}/\mathbf{X}^{(2)})$ is completely anologous to $R(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$.

Van Den Wollenberg [2] rewrote the definition of redundancy and basing on it developed a method called redundandcy analysis. Also basing on Van Den Wollenberg's definition, Desarbo [3] proposed a method called canonical/redundancy factoring analysis. If we write Stewart and Love's redundancy as

$$R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=\sum_{i=1}^{k}\lambda_i^2 f_{x^{(2)}v_i}{}'f_{x^{(2)}v_i}/p_2, \tag{11}$$

where $f_{x^{(2)}v_i}$ is the ith column vector of loading matrix $\mathbf{F}^*$ and its components are correlations between variables in the set of $\mathbf{X}^{(2)}$ and the canonical factor v_i of the same set. Whereas Van Den Wollenberg's definition is given by use of vector $f_{x^{(2)}u_i}$ whose components are correlations between variables in the set of $\mathbf{X}^{(2)}$ and the canonical factor u_i of the other set of $\mathbf{X}^{(1)}$. It can be written as

$$R_2(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=\sum_{i=1}^{k}f_{x^{(2)}u_i}{}'f_{x^{(2)}u_i}/p_2. \tag{12}$$

It is clear that when u_i and v_i $(i=1, 2, \cdots, k;\ k=\min(p_1, p_2))$ are pairs of canonical variables of sets $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$, the two definitions $R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ and $R_2(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ are equal according to the derivation of the canonical correlation [7].

$$\left.\begin{aligned}\psi_i=&\ \boldsymbol{\alpha}^{(i)'}\mathbf{R}_{12}\boldsymbol{\beta}^{(i)}-0.5\lambda_i(\boldsymbol{\alpha}^{(i)'}\mathbf{R}_{11}\boldsymbol{\alpha}^{(i)}-1)-0.5\nu_i(\boldsymbol{\beta}^{(i)'}\mathbf{R}_{22}\boldsymbol{\beta}^{(i)}-1)\\&+\sum_{j=1}^{i-1}\mu_j\boldsymbol{\alpha}^{(i)'}\mathbf{R}_{11}\boldsymbol{\alpha}^{(j)}+\sum_{j=1}^{i-1}\theta_j\boldsymbol{\beta}^{(i)'}\mathbf{R}_{22}\boldsymbol{\beta}^{(j)},\end{aligned}\right\} \tag{13}$$

$$\frac{\partial\psi_i}{\partial\boldsymbol{\beta}^{(i)}}=\mathbf{0}\ \Longrightarrow\ \mathbf{R}_{21}\boldsymbol{\alpha}^{(i)}-\lambda_i\mathbf{R}_{22}\boldsymbol{\beta}^{(i)}=\mathbf{0},\quad(\text{note: }\lambda_i=\nu_i,\ \mu_j=\theta_j=0), \tag{14}$$

so

$$R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}) = \sum_{i=1}^{k} \lambda_i^2 f_{x^{(2)}v_i}' f_{x^{(2)}v_i}/p_2 \tag{15}$$

$$= \sum_{i=1}^{k} \lambda_i \boldsymbol{\beta}^{(i)'} \mathbf{R}_{22}\mathbf{R}_{22} \boldsymbol{\beta}^{(i)} \lambda_i/p_2 \tag{16}$$

$$= \sum_{i=1}^{k} \boldsymbol{\alpha}^{(i)'} \mathbf{R}_{12}\mathbf{R}_{21} \boldsymbol{\alpha}^{(i)}/p_2 \tag{17}$$

$$= \sum_{i=1}^{k} f_{x^{(2)}u_i}' f_{x^{(2)}u_i}/p_2$$

$$= R_2(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}). \tag{18}$$

If we use R_2^a and R_2^c to represent the redundancies when u_i and v_i are any component variables of $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$ and canonical variables of $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$ respectively, we have $R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}) = R_2^c(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$.

In Van Den Wollenberg's redundancy analysis and Desarbo's canonical/redundancy analysis, $R_2^a(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ is maximized (in the order $f_{x^{(2)}u_i}' f_{x^{(2)}u_i} \geq f_{x^{(2)}u_{i+1}}' f_{x^{(2)}u_{i+1}}$, $i=1, \cdots, k-1$) associated with other constraints and conditions so as to seek the component coefficient vectors $\boldsymbol{\alpha}^{(i)}$ and $\boldsymbol{\beta}^{(i)}$, and redundancy variables $\boldsymbol{\alpha}^{(i)'}\mathbf{X}^{(1)}$ and $\boldsymbol{\beta}^{(i)'}\mathbf{X}^{(2)}$ are obtained as the result. We represent the maximum of $R_2^a(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ as $R_2^r(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$.

But the component variables obtained according to the maximization described above, that is redundancy variables, are not necessarily canonical variables. Therefore the questions arise that if the value $R_2^r(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ given by the redundancy variables is still equal to the original $R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ given by the canonical variables. In Van Den Wollenberg's redundancy analysis, the redundancies furnished by the main factors are of the importance. Here we also want to know what information the total redundancies furnished by all factors u_i, that is the redundancy index $R_2(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$, will tell us and if it is possible to further the meaning of redundancy index and redundancy analysis both in mathematical sense and in practical sense.

2. The relationships

Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in (4) be any $p_1 \times k$ and $p_2 \times k$ matrices to form new component variables of sets $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ respectively.

$$\left.\begin{array}{l} \mathbf{U} = \boldsymbol{\alpha}'\mathbf{X}^{(1)},\ \mathbf{V} = \boldsymbol{\beta}'\mathbf{X}^{(2)},\ E(\mathbf{U}) = E(\mathbf{V}) = \mathbf{0}, \\ V(u_i) = V(v_i) = 1, \quad i=1, 2, \cdots, k, \quad k = \min(p_1, p_2), \\ E(u_i v_i) = \lambda_i, \quad E(\mathbf{UU}') = E(\mathbf{VV}') = \mathbf{I}. \end{array}\right\} \tag{19}$$

Now consider the regression of the original variables $\mathbf{X}^{(2)}$ of one set on the component variables $\mathbf{U}$ (or $\mathbf{X}^{(1)}$) of other set. That is

$$\mathbf{X}^{(2)} = \mathbf{BU} + \boldsymbol{\varepsilon}, \tag{20}$$

where $\mathbf{B}$ is the regression matrix of vector $\mathbf{X}^{(2)}$ on vector $\mathbf{U}$ and ε is the error vector. From the regression theory we know

$$\mathbf{B}=E(\mathbf{X}^{(2)}\mathbf{U}')\ [E(\mathbf{U}\mathbf{U}')]^{-1}=E(\mathbf{X}^{(2)}\mathbf{X}^{(1)\prime})\,\alpha=\mathbf{R}_{21}\alpha. \tag{21}$$

So the regression variables $\tilde{\mathbf{X}}^{(2)}$ predicted from another set component variables $\mathbf{U}$ or original variables $\mathbf{X}^{(1)}$ may be presented as

$$\tilde{\mathbf{X}}^{(2)}=\mathbf{B}\mathbf{U}=\mathbf{R}_{21}\alpha\mathbf{U}=\mathbf{R}_{21}\alpha\alpha'\mathbf{X}^{(1)}. \tag{22}$$

The problem becomes to decide α so as to maximize the information in $\mathbf{U}$ (or $\mathbf{X}^{(1)}$) about $\mathbf{X}^{(2)}$ according to some criterion of adequacy. Let

$$\mathrm{RISK}=E[(\mathbf{X}^{(2)}-\tilde{\mathbf{X}}^{(2)})'(\mathbf{X}^{(2)}-\tilde{\mathbf{X}}^{(2)})]. \tag{23}$$

Then the risk may be chosen as the criterion of adequacy. When it is minimized, the best fit of $\tilde{\mathbf{X}}^{(2)}$ to $\mathbf{X}^{(2)}$ furnished by $\mathbf{U}$ will be obtained. By use of (22), we have

$$\left.\begin{aligned}\mathrm{RISK}^{a}&=E[(\mathbf{X}^{(2)}-\tilde{\mathbf{X}}^{(2)})'(\mathbf{X}^{(2)}-\tilde{\mathbf{X}}^{(2)})]=\mathrm{tr}[E(\mathbf{X}^{(2)}-\tilde{\mathbf{X}}^{(2)})\ (\mathbf{X}^{(2)}-\tilde{\mathbf{X}}^{(2)})']\\ &=\mathrm{tr}(\mathbf{R}_{22}-\mathbf{R}_{21}\alpha\alpha'\mathbf{R}_{12}-\mathbf{R}_{21}\alpha\alpha'\mathbf{R}_{12}+\mathbf{R}_{21}\alpha\alpha'\mathbf{R}_{11}\alpha\alpha'\mathbf{R}_{12}).\end{aligned}\right\} \tag{24}$$

Note that $\alpha'\mathbf{R}_{11}\alpha=\mathbf{I}$, we have

$$\left.\begin{aligned}\mathrm{RISK}^{a}&=\mathrm{tr}(\mathbf{R}_{22})-\mathrm{tr}(\mathbf{R}_{21}\alpha\alpha'\mathbf{R}_{12})\\ &=p_2[1-R_2{}^{a}(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})],\end{aligned}\right\} \tag{25}$$

where RISK^{a} is the risk when $\mathbf{X}^{(2)}$ is regressed on any component $\mathbf{U}$ of $\mathbf{X}^{(1)}$. It shows that the expected loss, or risk, is equal to the product of variable-number p_2 in the set of $\mathbf{X}^{(2)}$ and the difference between unit and Van Den Wollenberg's redundancy. When α is chosen so as to maximize the redundancy $R_2^{a}(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$, the risk is minimized and the best representation of $\tilde{\mathbf{X}}^{(2)}$ furnished by $\mathbf{U}$ will be obtained. That is, $\mathrm{RISK}^{r}=p_2[1-R_2^{r}(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})]$.

The problem now becomes under what conditions or when the redundancy $R_2^{a}(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ will be maximized. It can be shown that when $\mathbf{U}=\alpha'\mathbf{X}^{(1)}$ and $\mathbf{V}=\beta'\mathbf{X}^{(2)}$ are canonical variables, the redundancy $R_2(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ will be maximized and the maximum is just $R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ which equals to ASMC, the average squared multiple correlations for each of the original $\mathbf{X}^{(2)}$ variables regressed on all the variables of the set $\mathbf{X}^{(1)}$ when $p_1\geq p_2=k$, [6]. As described in (15) – (18), when u_i and v_i are canonical factor pairs of $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$,

$$R_2^{c}(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}).$$

According to basic canonical equations (7), we have

$$\beta\Lambda^2=\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{R}_{11}{}^{-1}\mathbf{R}_{12}\beta.$$

Premultiplying both sides of above equation by $(\beta'\mathbf{R}_{22}\mathbf{R}_{22})$ and then taking the traces of them, it follows

$$\begin{aligned}&\mathrm{tr}[\beta'\mathbf{R}_{22}\mathbf{R}_{22})\ (\beta\Lambda^2)]\\&\quad=\mathrm{tr}[\beta'\mathbf{R}_{22}(\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12})\ \beta].\end{aligned}\tag{26}$$

Under the constraint $\beta'\mathbf{R}_{22}\beta=\mathbf{I}$ in (19), we have $\beta'\mathbf{R}_{22}=\beta^{-1}$ when $p_2=k$ (suppose β is nonsingular). So according to the cyclic permutation property of traces

$$\mathrm{tr}[(\beta'\mathbf{R}_{22}\mathbf{R}_{22}\beta)\ \Lambda^2]=\mathrm{tr}(\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}).\tag{27}$$

That is to say,

$$\begin{aligned}R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})&=[\mathrm{tr}((\beta'\mathbf{R}_{22}\mathbf{R}_{22}\beta)\ \Lambda^2)]/p_2\\&=[\mathrm{tr}(\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12})]/p_2=\mathrm{ASMC}.\end{aligned}\tag{28}$$

The diagonal elements of $\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}$ are squared multiple correlations between each variable of $\mathbf{X}^{(2)}$ and the vector $\mathbf{X}^{(1)}$. It is well known that the multiple correlation between $X_j^{(2)}$ and $\mathbf{X}^{(1)}$ is the maximum correlation between $X_j^{(2)}$ and any linear combination $u_i=\alpha'_i\mathbf{X}^{(1)}$, so we have [7]

$$\mathrm{ASMC}=\max(\sum_{i=1}^{p2}f_{x^{(2)}u_i}'f_{x^{(2)}u_i})/p_2\geq R_2^a(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}).\tag{29}$$

That is

$$\begin{aligned}R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})&=\max(\sum_{i=1}^{p2}f_{x^{(2)}u_i}'f_{x^{(2)}u_i})/p_2\geq R_2^r(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})\geq R_2^c(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})\\&=R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}).\end{aligned}\tag{30}$$

Threrfore we have

$$\mathrm{ASMC}=R_2^r(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=R_2^c(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})\geq R_2^a(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}).$$

When $p_2\geq p_1=k$,

$$R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=\sum_{i=1}^{p1}\lambda_i^2 f_{x^{(2)}v_i}'f_{x^{(2)}v_i}/p_2,$$

$$R_2(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=\sum_{i=1}^{p1}f_{x^{(2)}u_i}'f_{x^{(2)}u_i}/p_2.$$

This time, $R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ is no longer equal to ASMC and we have

$$\mathrm{ASMC} \geq R_2^r(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}) \geq R_2^c(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}) = R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}).$$

So when $p_1 \geq p_2 = k$, the two definitions of redundancy are equivalent if $\mathbf{U}$, $\mathbf{V}$ in $R_2(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ are either redundancy variables or canonical variables, but when $p_2 \geq p_1 = k$, they are no longer the same. Similarly we have

$$\mathrm{ASMC} = R_2^r(\mathbf{X}^{(1)}/\mathbf{X}^{(2)}) = R_2^c(\mathbf{X}^{(1)}/\mathbf{X}^{(2)}) = R_1(\mathbf{X}^{(1)}/\mathbf{X}^{(2)}) \geq R_2^a(\mathbf{X}^{(1)}/\mathbf{X}^{(2)}), \qquad \text{if } p_2 \geq p_1 = k,$$

and

$$\mathrm{ASMC} \geq R_2^r(\mathbf{X}^{(1)}/\mathbf{X}^{(2)}) \geq R_2^c(\mathbf{X}^{(1)}/\mathbf{X}^{(2)}) = R_1(\mathbf{X}^{(1)}/\mathbf{X}^{(2)}), \qquad \text{if } p_1 \geq p_2 = k.$$

We have considered the expected loss of representing $\mathbf{X}^{(2)}$ by $\mathbf{X}^{(1)}$ via $\mathbf{U}$, which is $p_2[1-R_2(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})]$. Now let us take a look at the regression of $\mathbf{X}^{(1)}$ directiy on $\mathbf{X}^{(1)}$ without using $\mathbf{U} = \boldsymbol{\alpha}'\mathbf{X}^{(1)}$

$$\tilde{\mathbf{X}}^{(2)} = E(\mathbf{X}^{(2)}\mathbf{X}^{(1)\prime})\ [E(\mathbf{X}^{(1)}\mathbf{X}^{(1)\prime})]^{-1}\mathbf{X}^{(1)} = \mathbf{R}_{21}\mathbf{R}_{11}{}^{-1}\mathbf{X}^{(1)}, \tag{33}$$

$$\begin{aligned}
\mathrm{RISK}^d &= E[(\mathbf{X}^{(2)}-\tilde{\mathbf{X}}^{(2)})'(\mathbf{X}^{(2)}-\tilde{\mathbf{X}}^{(2)})] \\
&= \mathrm{tr}[E(\mathbf{X}^{(2)}-\tilde{\mathbf{X}}^{(2)})\ (\mathbf{X}^{(2)}-\tilde{\mathbf{X}}^{(2)})'] \\
&= \mathrm{tr}(\mathbf{R}_{22}) - \mathrm{tr}(\mathbf{R}_{21}\mathbf{R}_{11}{}^{-1}\mathbf{R}_{12}) \\
&= p_2(1-\mathrm{ASMC}), \qquad (34) \\
\text{or} \quad &= p_2(1-R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})), \quad \text{if } p_1 \geq p_2,
\end{aligned}$$

where RISK^d is the risk, when $\mathbf{X}^{(2)}$ is regressed directly on $\mathbf{X}^{(1)}$. Therefore the difference between the regression risks of $\mathbf{X}^{(2)}$ via $\mathbf{U}$ on $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ directly on $\mathbf{X}^{(1)}$ is

$$\begin{aligned}
(25)-(34) &= \mathrm{RISK}^a - \mathrm{RISK}^d \\
&= R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}) - R_2^a(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}) \geq 0, \quad \text{if } p_1 \geq p_2. \qquad (35)
\end{aligned}$$

It is reasonable that the information in $\mathbf{X}^{(1)}$ about $\mathbf{X}^{(2)}$ is 'more' than that in the components $\mathbf{U} = \boldsymbol{\alpha}'\mathbf{X}^{(1)}$ about $\mathbf{X}^{(2)}$. When the components $\mathbf{U} = \boldsymbol{\alpha}'\mathbf{X}^{(1)}$ and $\mathbf{V} = \boldsymbol{\beta}'\mathbf{X}^{(2)}$ are canonical correlated, the information furnished by both cases becomes the same. This fact shows that the original Stewart and Love's redundancy index $R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ gives the extent of 'fit' from the set $\mathbf{X}^{(1)}$ to the set $\mathbf{X}^{(2)}$, which can be seen intuitively as the 'intersection' of the two sets. Similarly, by use of (22) to (25), we can also write

$$\mathrm{RISK}^c = p_2[1-R_2^c(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})] \quad \text{and} \quad \mathrm{RISK}^r = p_2[1-R_2^r(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})],$$

where RISK^c and RISK^r are risks, when $\mathbf{X}^{(2)}$ is regressed on redundancy variables of $\mathbf{X}^{(1)}$ and canonical variables of $\mathbf{X}^{(1)}$, respectively.

Now we sum up what we have shown foregoing in following theorem.

THEOREM 1. Let the linear combinations of $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ be

$$\mathbf{U} = \boldsymbol{\alpha}'\mathbf{X}^{(1)}, \ \mathbf{V} = \boldsymbol{\beta}'\mathbf{X}^{(2)}$$

with constraints as described in (2), (6), and (19). Let

$$R_2(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}) = \sum_{i=1}^{k} f_{x^{(2)}u_i}' f_{x^{(2)}u_i}/p_2, \quad k=\min(p_1, p_2),$$

where $f_{x^{(2)}u_i}$ is a column vector with correlations between variables $X_j^{(2)}$ in the set of $\mathbf{X}^{(2)}$ and u_i. Let the 'best' representations of $\mathbf{X}^{(2)}$ furnished by $\mathbf{X}^{(1)}$ and $\mathbf{U}$ be given by (33) and (22), respectively, and let RISK^d, RISK^a, RISK^c, RISK^r be expected losses or risks corresponding to regressions of $\mathbf{X}^{(2)}$ directly on $\mathbf{X}^{(1)}$, any component variables u_i, canonical variables u_i and redundancy variables u_i, respectively, and $R_2^a(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$, $R_2^c(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$, $R_2^r(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ have the similar meaning. Then the risks of above cases are as follows:

$$\mathrm{RISK}^d = p_2(1-\mathrm{ASMC}), \ \mathrm{RISK}^a = p_2[1-R_2^a(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})],$$

$$\mathrm{RISK}^c = p_2[1-R_2^c(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})] = p_2[1-R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}), \ \mathrm{RISK}^r = p_2[1-R_2^r(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})],$$

where $R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ is the Stewart and Love's redundancy given by (10), and the risks in various cases satisfy as follows,

$$\mathrm{RISK}^d = \mathrm{RISK}^r = \mathrm{RISK}^c \leq \mathrm{RISK}^a, \quad \text{if} \quad p_1 \geq p_2 = k,$$

$$\mathrm{RISK}^d \leq \mathrm{RISK}^r \leq \mathrm{RISK}^c, \quad \text{if} \quad p_2 \geq p_1 = k.$$

From theorem 1 and what is described above, we know that if $p_1 \geq p_2$, the variable number in regressing set $\mathbf{X}^{(1)}$ larger than that in regressed set $\mathbf{X}^{(2)}$, $R_2(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ is maximized both in canonical analysis and redundancy analysis, and the maximum is just $R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$, which equals to ASMC. Although the size-order of redundancies furnished by u_i in both analyses may be different, the total effects in representing $\mathbf{X}^{(2)}$ by regression are the same. So in some sense the two definitions of redundancy are equivalent.

But if $p_1 \leq p_2$, i.e. the variable number in regressing set $\mathbf{X}^{(1)}$ is smaller than that in regressed set $\mathbf{X}^{(2)}$, the maximum of $R_2(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ in redundancy analysis is no longer equal to $R_2^c(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ obtained in canonical analysis, which is still the same as $R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$. Of course in this case the total effect in representing $\mathbf{X}^{(2)}$ by regressing is different, and the two definitions are not equivalent. It is interested to see that the risk when regressed on redundancy variables is smaller than that when regressed on canonical variables. This shows that redundancy analysis not only gives the component variables furnishing the maximized redundancies in order, as in Dan Ven Wollenberg's article in 1977, but also provides better representation of $\mathbf{X}^{(2)}$ from component variables of $\mathbf{X}^{(1)}$ than canonical analysis does. Therefore the redundancy index proposed by Stewart and Love, the modified redundancy index by Van Den Wollenberg and his redundancy analysis, the canonacal/redundancy factoring analysis by Desarbo are all of importance both in theoretical and practical sense.

3. The relative efficiency of prediction

Now let us make a further consideration of the meanings of canonical correlation and redundancy in view of prediction. First let us consider two random variables u and v which satisfy the following conditions [8].

$$E(u)=E(v)=0,\quad V(u)=\sigma_u^2,\quad V(v)=\sigma_v^2,\quad (u, v)=\rho\,(\text{correlation}). \tag{36}$$

Now let us predict u from a linear predictor v as,

$$u \fallingdotseq bv \quad \text{or} \quad \tilde{u}=bv. \tag{37}$$

Then the variance of the error $u-\tilde{u}$ is given by

$$\left.\begin{aligned} E[(u-\tilde{u})^2]&=E[(u-bv)^2]=\sigma_u^2-2b\rho\sigma_u\sigma_v+b^2\sigma_v^2 \\ &=(1-\rho^2)\sigma_u^2+(b\sigma_v-\rho\sigma_u)^2. \end{aligned}\right\} \tag{38}$$

To make the best prediction in which the variance of the error is the minimum we must let the second term of the right side in (38) be zero. Then

$$b=\rho\sigma_u/\sigma_v. \tag{39}$$

Thereby predicting u from v by use of b, the variance of error of prediction is $(1-\rho^2)\sigma_u^2$, and the ratio of which to the variance of original u is

$$\frac{(1-\rho^2)\sigma_u^2}{\sigma_u^2}=1-\rho^2. \tag{40}$$

It can be seen that the nearer to 1 the squared correlation between u and v is, the better the effect of prediction of v on u will be. So we can define ρ^2 as the relative efficiency of prediction of v on u.

Now consider the efficiencies of prediction in the following four cases by the use of similar method given above. The notations represent the same meaning as that of the above.

1. The relative efficiency of prediction of **V** on **U**, or of **U** on **V**.
2. The relative efficiency of prediction of $\mathbf{X}^{(1)}$ on $\mathbf{X}^{(2)}$.
3. The relative efficiency of prediction of **U** on $\mathbf{X}^{(2)}$, where u_i is any component vari-

able of $\mathbf{X}^{(1)}$.

4. The relative efficiency of prediction of $\mathbf{U}$ on $\mathbf{X}^{(2)}$, where u_i and v_i are pairs of canonical variables between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$.

5. The relative efficiency of prediction of $\mathbf{U}$ on $\mathbf{X}^{(2)}$, where u_i and v_i are pairs of redundancy variables between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$.

The processes are almost the same. So we will not discuss every case in detail. The corresponding results for the 5 cases are as following.

1) Let $\tilde{\mathbf{U}}=\mathbf{BV}$, then by the usual mean squared regression formula,

$$\mathbf{B}=E(\mathbf{UV}')\ [E(\mathbf{VV}')]^{-1}=\Lambda\mathbf{I}^{-1}=\Lambda, \tag{41}$$

$$\left.\begin{aligned} E[(\mathbf{U}-\tilde{\mathbf{U}})'(\mathbf{U}-\tilde{\mathbf{U}})]&=\mathrm{tr}[E(\mathbf{U}-\tilde{\mathbf{U}})\ (\mathbf{U}-\tilde{\mathbf{U}})']=\mathrm{tr}\mathbf{I}-\mathrm{tr}\Lambda^2 \\ &=k(1-\sum_{i=1}^{k}\lambda_i^2/k), \end{aligned}\right\} \tag{42}$$

$$\frac{E[(\mathbf{U}-\tilde{\mathbf{U}})'(\mathbf{U}-\tilde{\mathbf{U}})]}{E(\mathbf{U}'\mathbf{U})}=1-\sum_{i=1}^{k}\lambda_i^2/k, \tag{43}$$

$$\text{relative efficiency}=\sum_{i=1}^{k}\lambda_i^2/k. \tag{44}$$

2) According to (33) and (34), we have
relative efficiency$=$ASMC.
According to theorem 1, we have

3) relative efficiency$=R_2^a(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$,

4) relative efficiency$=R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=R_2^c(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$,

5) relative efficiency$=R_2^r(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$.

If we use r.e.$(\mathbf{X}/\mathbf{Y})$ to show the relative efficiency of prediction of $\mathbf{Y}$ on $\mathbf{X}$ and the same notations used forgoing, we have the following theorem.

THEOREM 2. The relative efficiencies of prediction of $\mathbf{V}$ on $\mathbf{U}$, $\mathbf{X}^{(1)}$ on $\mathbf{X}^{(2)}$ and $\mathbf{U}$ on $\mathbf{X}^{(2)}$ and so on are

$$\left.\begin{aligned} &\text{r.e.}(\mathbf{U}/\mathbf{V})=\text{r.e.}\ (\mathbf{V}/\mathbf{U})=\sum_{i=1}^{k}\lambda_i^2/k, \\ &\text{r.e.}(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=\text{ASMC}, \\ &\text{r.e.}(\mathbf{X}^{(2)}/\mathbf{U})=R_2^a(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}), \\ &\text{r.e.}(\mathbf{X}^{(2)}/\mathbf{U}^c)=R_2^c(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}), \\ &\text{r.e.}(\mathbf{X}^{(2)}/\mathbf{U}^r)=R_2^r(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}). \end{aligned}\right\} \tag{45}$$

From theorem 2 we can see more clearly that both canonical correlation and redundancy are measures which represent the relationships between two sets of $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$. While canonical correlations give more directly the links between the canonical factors, redundancy shows the relationships between original variables of $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ or between original variables $\mathbf{X}^{(2)}$ (or $\mathbf{X}^{(1)}$) and canonical factors $\mathbf{U}$ (or $\mathbf{V}$) or any other component factors, including redundancy factors. Another difference between them is that canonical correlation is a symmetric index, while redundancy is a nonsymmetric one. Therefore when we deal with two sets of variables, it would be better to consider both canonical correlations and redundancies instead of just the former one as usual the case.

4. Summary

We have the following points about two definitions of redundancy, canonical analysis and redundancy analysis.

1. Canonical correlations show the highest correlations between canonical factors of two sets, but they do not necessarily tell anything about the communality of the two sets of variables, because a squared canonical correlation represents the variance shared by linear composits of two sets of variables, and not the shared variance of the two sets. Therefore canonical correlations can not be interpreted as correlations between sets of variables.

2. Stewart and Love's redundancy index $R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$ shows the proportion of variance of $\mathbf{X}^{(2)}$ predictable from $\mathbf{X}^{(1)}$, or the redundancy in $\mathbf{X}^{(2)}$ given $\mathbf{X}^{(1)}$, or the proportion of the set $\mathbf{X}^{(2)}$, which is in the intersection of sets $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$.

3. In canonical analysis, a relatively strong canonical correlation may be obtained for a pair of canonical variables, but redundancy given by it may be very low. Van Den Wollenberg developed a method called redundancy analysis in which the redundancies for the redundancy variables are maximized in size-order. That is the first pair of redundancy variables corresponds to the largest redundancy and the second pair the second largest, and so forth.

4. In order to perform redundancy analysis, Van Den Wollenberg rewrote the original definition into $R_2(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$, which is equal to $R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})$, when the component variables are canonical correlated.

5. The comparision between two definitions of redundancy are

$$\mathrm{ASMC}=R_2^r(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=R_2^c(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})\geq R_2^a(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}),\quad \text{if } p_1\geq p_2,$$

$$\mathrm{ASMC}\geq R_2^r(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})\geq R_2^c(\mathbf{X}^{(2)}/\mathbf{X}^{(1)})=R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}),\qquad \text{if } p_1\leq p_2.$$

6. In view of regression, we have following risks

$$\mathrm{RISK}^d=\mathrm{RISK}^r=\mathrm{RISK}^c,\qquad \text{if } p_1\geq p_2,$$

$$\mathrm{RISK}^d\leq\mathrm{RISK}^r\leq\mathrm{RISK}^c,\qquad \text{if } p_1\leq p_2.$$

That means when $p_1\leq p_2$, the total effect of regression of $\mathbf{X}^{(2)}$ on redundancy variables seems better than that on canonical variables.

7. In point of prediction, we have following relative efficiency

$$\text{r.e.}(\mathbf{U}/\mathbf{V}) = \text{r.e.}\ (\mathbf{V}/\mathbf{U}) = \sum_{i=1}^{k} \lambda_i^2/k, \qquad \text{r.e.}(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}) = \text{ASMC},$$

$$\text{r.e.}(\mathbf{X}^{(2)}/\mathbf{U}) = R_2^a(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}), \qquad \text{r.e.}(\mathbf{X}^{(2)}/\mathbf{U}^c) = R_2^c(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}) = R_1(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}),$$

$$\text{r.e.}(\mathbf{X}^{(2)}/\mathbf{U}^r) = R_2^r(\mathbf{X}^{(2)}/\mathbf{X}^{(1)}).$$

That means canonical analysis gives the relationship between component variables whereas redundancy analysis shows more the relationship between component variables and the original variables.

References

1) Stewart, D. K. and Love, W. A., A General Canonical Correlation Index, Psy. Bul., 1968, 70.
2) Van Den Wollenberg, A. L., Redundancy Analysis: an Alternative for Canonical Correlation Analysis, Psy., 1977, 42.
3) Desarbo, W. S., Canonical/Redundancy Factoring Analysis, Psy., 1981, 46.
4) Nicewander, W. A. and Wood, D. A., Comments on 'A General Canonical Correlation Index', Psy. Bul., 1974, 81.
5) Nicewander, W. A. and Wood, D. A., On the Mathematical Bases of the General Canonical Correlation Index: Rejionder to Miller, Psy. Bul., 1975, 82.
6) Miller, J. K., In Defense of the General Canonical Correlation Index: Reply to Nicewander and Wood, Psy. Bul., 1975, 82.
7) Anderson, T. W., Introduction to Multivariate Statistical Analysis, John Wiley, New York, 1984.
8) Asano, Ch., Theory and Applications of Factor Analysis (In Japanese), Kyoritsu Publisher, Tokyo, 1971.