Journal of the Korean Statistical Society (2006), **35**: *2, pp 143–156*

UNIFYING STATIONARY EQUATIONS FOR GENERALIZED CANONICAL CORRELATION ANALYSIS†

HYUNCHEOL KANG[1] AND KEEYOUNG KIM[2]

ABSTRACT

In the present paper, various solutions for generalized canonical correlation analysis (GCCA) are considered depending on the criteria and constraints. For the comparisons of some characteristics of the solutions, we provide with certain unifying stationary equations which might be also useful to obtain various generalized canonical correlation analysis solutions. In addition, we suggest an approach for the generalized canonical correlation analysis by exploiting the concept of maximum eccentricity originally designed to test the internal independence structure. The solutions, including new one, are compared through unifying stationary equations and by using some numerical illustrations. A type of iterative procedure for the GCCA solutions is suggested and some numerical examples are provided to illustrate several GCCA methods.

AMS 2000 subject classifications. Primary 62H20; Secondary 62G05.
Keywords. Unifying stationary equations, generalized canonical correlation, sets of variables, maximum eccentricity, internal independence.

1. INTRODUCTION

Generalized canonical correlation analysis (GCCA), which involves comparing $m(\geq 3)$ sets of variables after having removed linear dependencies within each of the sets, extends the canonical correlation analysis (CCA) of Hotelling (1936) to the case of more than two sets of variables. There have been many studies on how the two-set canonical solution can be generalized (Horst, 1961a, b, 1965;

Received November 2005; accepted March 2006.
†This work was supported by the Korea Research Foundation Grant (KRF-2004-003-C00041).
[1]Corresponding author. Department of Informational Statistics, Hoseo University, Asan 336-795, Korea (e-mail: hychkang@office.hoseo.ac.kr)
[2]Department of Statistics, Korea University, Seoul 136-701, Korea (e-mail: kykim@korea.ac.kr)

Carroll, 1968; Kettenring, 1971; Van de Geer, 1984; Ten Berge, 1988; Coppi and Bolasco, 1989; Gifi, 1990; Melzer *et al.*, 2001; Takane and Hwang, 2002) with some discussions on the criteria and constraints imposed.

Suppose we have m data matrices $\mathbf{X}_i$ $(i = 1, \ldots, m)$, each from a sample of size n on p_i variables. It is implicitly assumed that all the variables are standardized to have zero means and unit variances. The problem is to find linear composites $\mathbf{z}_i = \mathbf{X}_i \mathbf{a}_i$ $(i = 1, 2, \ldots, m)$ which optimize a certain function of the covariance matrix, $\mathbf{\Phi} = (\phi_{ij})$ $(i, j = 1, \ldots, m)$ of $\mathbf{z}_i$'s.

Criteria used to obtain the canonical loading vectors $\mathbf{a}_i$ $(i = 1, \ldots, m)$ are characterized by the associated object function defined in terms of $\mathbf{\Phi}$. Let $l_1 \geq l_2 \geq \cdots \geq l_m$ be the ordered eigenvalues of $\mathbf{\Phi}$. Kettenring (1971) considered the following five criteria for selecting $\mathbf{a}_i$'s: (i) SUMCOR [Maximize $\sum_{i,j}^m \phi_{ij}$]; (ii) MAXVAR [Maximize l_1]; (iii) MINVAR [Minimize l_m]; (iv) SSQCOR [Maximize $\sum_{i,j}^m \phi_{ij}^2$]; (v) GENVAR [Minimize $\det(\mathbf{\Phi})$]. In his discussion, factor-analytic models were used to motivate the criteria and comparisons among them were made. Some of the above criteria have also been suggested by Horst (1961a, b, 1965) and Steel (1951). Gifi (1990) studied other methods such as MINSUM [Minimize $\sum_{i,j}^m \phi_{ij}$] and PRINCALS [Maximize $\sum_{k=1}^q l_k$, $q < m$]. Also Gifi suggested OVERALS which is based on the MAXVAR criterion incorporating the nonlinear data transformation.

Among others, GENVAR of (v) above can be recognized to have a direct relation with the likelihood ratio statistic for testing the independence among $\mathbf{z}_i$'s. Under the union-intersection principle, Schuenemeyer and Bargmann (1978) developed a test statistic for the independence, which is the maximum attainable eccentricity of the correlation ellipsoid generated by the given data. Now we may add this union-intersection type of statistic applied to $\mathbf{z}_i$'s as a possible criterion for the development of GCCA solution.

(vi) MAXECC: Maximize the maximum eccentricity of $\mathbf{\Phi}$, $(l_1 - l_m)/(l_1 + l_m)$.

The rest of this paper is organized as follows. In Section 2, we suggest an approach for the GCCA by exploiting the concept of maximum eccentricity. In Section 3, we provide with certain unifying stationary equations which give GCCA solutions for various criteria mentioned in Section 1. In Section 4, a type of iterative procedure for the GCCA solutions is suggested and numerical examples are provided to illustrate several GCCA methods. The major findings through the stationary equations and numerical examples are summarized in Section 5.

2. MAXIMUM ECCENTRICITY SOLUTION

In order to achieve a valid optimization, the criterion is usually subjected to certain constraints. Let $\mathbf{R}_{ij}$ of size $p_i \times p_j$ denote the correlation matrix obtained from data matrices $(\mathbf{X}_i, \mathbf{X}_j)$ $(i, j = 1, 2, \ldots, m)$. A typical constraint is to take $\text{var}(\mathbf{z}_i) = 1$, that is

$$\mathbf{a}_i'\mathbf{R}_{ii}\mathbf{a}_i = 1, \quad i = 1, \ldots, m. \tag{2.1}$$

With this unit-variances constraint, $\mathbf{\Phi}$ can be interpreted as a correlation matrix among $\mathbf{z}_i$'s.

The derivation of the GCCA solution for MAXECC criterion under the unit-variances constraint reduces to finding the maximum of the maximum eccentricity $(l_1 - l_m)/(l_1 + l_m)$ subject to (2.1). Thus the object function $\text{g}(\cdot)$ to be optimized takes the form of

$$\text{g}(\mathbf{a}_1, \ldots, \mathbf{a}_m) = \frac{l_1 - l_m}{l_1 + l_m} - \sum_{i=1}^{m} \mu_i(\mathbf{a}_i'\mathbf{R}_{ii}\mathbf{a}_i - 1), \tag{2.2}$$

where μ_i's are the Lagrange multipliers. The partial derivative of $\text{g}(\cdot)$ with respect to $\mathbf{a}_i$ is given by

$$\frac{\partial \text{g}}{\partial \mathbf{a}_i} = \frac{2}{(l_1 + l_m)^2}\left(l_m \frac{\partial l_1}{\partial \mathbf{a}_i} - l_1 \frac{\partial l_m}{\partial \mathbf{a}_i}\right) - 2\mu_i \mathbf{R}_{ii}\mathbf{a}_i. \tag{2.3}$$

Since

$$\frac{\partial l_k}{\partial \mathbf{a}_i} = 2\sum_{j=1}^{m} e_{ik}e_{jk}\mathbf{R}_{ij}\mathbf{a}_j, \tag{2.4}$$

where e_{ik} is the i^{th} element of $\mathbf{e}_k$ which is the unit-normed eigenvector associated with the k^{th} eigenvalue l_k of $\mathbf{\Phi}$ (see Appendix), applying (2.4) to (2.3) and letting it equal to zero yield the MAXECC solutions for $\mathbf{a}_i$ which satisfies the following equations

$$\sum_{j=1}^{m} w_{ij}\mathbf{R}_{ij}\mathbf{a}_j = \mu_i^* \mathbf{R}_{ii}\mathbf{a}_i, \quad i = 1, \ldots, m, \tag{2.5}$$

where $w_{ij} = e_{i1}e_{j1}l_m - e_{im}e_{jm}l_1$ and $\mu_i^* = \mu_i(l_1 + l_m)^2/2$.

3. SOME COMPARISONS THROUGH UNIFYING STATIONARY EQUATIONS

In this section, we provide with certain unifying stationary equations which give GCCA solutions for various criteria mentioned in Section 1. These relationships can be used not only to develop the associated iterative routines but also to characterize the corresponding approaches. Solutions are derived under two different types of constraints.

3.1. Solutions with unit-variances constraint

Under the constraint (2.1), by differentiating the following object function $g(\cdot)$ with respect to $\mathbf{a}_i$,

$$g(\mathbf{a}_1, \ldots, \mathbf{a}_m) = f(\mathbf{a}_1, \ldots, \mathbf{a}_m) - \sum_{i=1}^{m} \mu_i(\mathbf{a}_i'\mathbf{R}_{ii}\mathbf{a}_i - 1), \tag{3.1}$$

where $f(\cdot)$ stands for a criterion to be considered, we have the following form of unifying stationary equations for each of the six different GCCA solutions,

$$\sum_{j=1}^{m} w_{ij}\mathbf{R}_{ij}\mathbf{a}_j = \mu_i^*\mathbf{R}_{ii}\mathbf{a}_i, \quad i = 1, \ldots, m, \tag{3.2}$$

where

$$w_{ij} = \begin{cases} \frac{1}{m}, & \text{for SUMCOR,} \\ e_{i1}e_{j1}, & \text{for MAXVAR,} \\ e_{im}e_{jm}, & \text{for MINVAR,} \\ \sum_{k=1}^{m} e_{ik}e_{jk}l_k, & \text{for SSQCOR,} \\ \sum_{k=1}^{m} \frac{e_{ik}e_{jk}}{l_k}, & \text{for GENVAR,} \\ e_{i1}e_{j1}l_m - e_{im}e_{jm}l_1, & \text{for MAXECC,} \end{cases} \tag{3.3}$$

respectively (see Appendix).

The stationary equations (3.2) imply that the GCCA methods differ only by applying different weights as shown in (3.3) to obtain the desired solutions. Therefore it could be one way to characterize each GCCA method in terms of w_{ij}. Clearly, MAXVAR and MINVAR emphasize the elements of eigenvectors associated with the largest and smallest eigenvalue of $\mathbf{\Phi}$, respectively. On the other hand, SSQCOR, GENVAR and MAXECC incorporate eigenvalues with eigenvector elements as the weights. In terms of the weight, SSQCOR emphasizes the largeness of larger eigenvalues, while GENVAR does to the smallness of smaller ones, which was also indicated by Kettenring (1971) and Gifi (1990). In the meantime, MAXECC combines the two extreme eigenvalues and the associated eigenvector elements simultaneously and thus introduces some intermediate possibility between MAXVAR and MINVAR. With this point, MAXECC should have similar propensity to PRINCALS of Gifi (1990), but the q in PRINCALS remains to be determined.

3.2. Solutions with constant-sum-variances constraint

The constant-sum-variances constraint is sometimes considered as an alternative to the unit-variances constraint. With this constraint the canonical variates are not restricted to have identical variances, thus imposing this constraint could result in different solutions even for the same criterion. The constant-sum-variances constraint is given by

$$\sum_{i=1}^{m} \mathbf{a}_i'\mathbf{R}_{ii}\mathbf{a}_i = \mathbf{a}'\mathbf{D}\mathbf{a} = m, \tag{3.4}$$

where $\mathbf{a}' = (\mathbf{a}_1' \ldots \mathbf{a}_m')$ and $\mathbf{D}$ is a block-diagonal matrix with $\mathbf{R}_{ii}$ as its i^{th} block.

There is a situation where the solutions with (3.4) could be very unfair (Van de Geer, 1984), since the solutions could be heavily dependent on the canonical variates which have relatively large or small variances. The constant-sum-variances constraint, however, would be sometimes preferred due to relatively easy computation. Moreover, the fact that the stationary equations with this constraint take explicit form for some criteria could be another merit to clarify the characteristics of the criteria.

Since the constant-sum-variances constraint needs only one Lagrange multiplier μ, the equations to solve for the solutions can be written as the following matrix form

$$\mathbf{R}\mathbf{D}_a\mathbf{W} = \mu\mathbf{D}\mathbf{D}_a, \tag{3.5}$$

where $\mathbf{R}$ denotes the correlation matrix obtained from matrix $\mathbf{X} = (\mathbf{X}_1 \ldots \mathbf{X}_m)$, $\mathbf{D}_a$ denotes a block diagonal matrix with $\mathbf{a}_i$ as its i^{th} block, and $\mathbf{W}$ is the matrix with w_{ij} of (3.3) as its $(i, j)^{th}$ element.

The equation (3.5) is related to multiple corresponding analysis (Lebart *et. al.*, 1984). It is worth noting that the weight matrix $\mathbf{W}$ can be written as $\mathbf{W} = \mathbf{v}\mathbf{v}'$ for some criteria. That is, $\mathbf{v}$ is $\mathbf{1}/\sqrt{m}$ for SUMCOR, $\mathbf{e}_1$ for MAXVAR, and $\mathbf{e}_m$ for MINVAR. Thus, for these three criteria, post-multiplying $\mathbf{v}$ to both sides of (3.5) and applying the specific value of $\mathbf{W}$ for each, we get

$$\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}\mathbf{D}^{1/2}\mathbf{D}_a\mathbf{v} = \mu\mathbf{D}^{1/2}\mathbf{D}_a\mathbf{v}. \tag{3.6}$$

Therefore, with the constant-sum-variances constraint, solutions for SUMCOR, MAXVAR and MINVAR can be obtained by performing a single eigen-analysis on $\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}$. Moreover, SUMCOR and MAXVAR solutions with (3.4) are equivalent where the solutions are proportional to the elements of the first eigenvector of $\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}$, and the same optimum value of Lagrange multiplier $\mu = \mathbf{1}'\mathbf{\Phi}\mathbf{1}/m = \mathbf{e}_1'\mathbf{\Phi}\mathbf{e}_1$.

3.3. Some additional comparisons of GCCA methods

Both the constant-sum-variances constraint and unit-variances constraint for MAXVAR and MINVAR criterion lead to the same results (Kettenring, 1971). For these two criteria, if $\mathbf{h}$, the first or the last eigenvectors of $\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}$, is partitioned into the successive $p_i \times 1$ subvectors, $\mathbf{h}_i$, then the canonical coefficient vectors $\mathbf{a}_i$ with any constraint are expressed as

$$\mathbf{a}_i = \pm \mathbf{R}_{ii}^{-1/2} \frac{\mathbf{h}_i}{||\mathbf{h}_i||}, \quad i = 1, \ldots, m. \tag{3.7}$$

It means that MAXVAR solution should be equivalent to the principal component loading of $\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}$, actually the correlation matrix for which dependencies within each of the sets are removed. Thus, a set of canonical variates for MAXVAR gives the best rank one (least squares) approximation of $\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}$ (Horst, 1961b). On the other hand, MINVAR tends to detect the gauge of some linear functional relation among the variables which are transformed by removing the within-set dependencies (Gifi, 1990).

Suppose λ_l and λ_s are the largest and the smallest eigenvalue of $\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}$ respectively. Then λ_l becomes the maximum among all possible largest eigenvalues and λ_s does the minimum among all possible smallest eigenvalues of $\mathbf{\Phi}$'s generated by any criterion, since λ_l and λ_s equal to the optimum value of MAXVAR and MINVAR respectively.

In summary, it is anticipated that there exists a tendency showing a certain ascending order such as MAXVAR (SUMCOR) - SSQCOR - MAXECC - GENVAR - MINVAR in terms of the degree of influence for which λ_s has on each criterion (MINVAR would be affected the most by λ_s). And for the case of λ_l, the order is reversed (MAXVAR would be affected the most by λ_l). Therefore this order could be thought of representing the affinity among the criteria. In practice, the results of all methods could be very similar when λ_l is relatively small but λ_s are far from zero. However the case that λ_s is close to zero gives somewhat different results. These points will be illustrated in Section 4 by examples. As shown previously, mathematical and computational convenience indicates that MAXVAR and MINVAR could be preferable, but the results of these two criteria sometimes take the most opposite positions to others. Since MAXECC, however, generalizes MAXVAR and MINVAR, it can be speculated that the results of MAXECC would take a midpoint among the methods.

4. NUMERICAL ILLUSTRATIONS

In this section, three numerical examples are provided to illustrate the six different GCCA methods. As mentioned in Section 3, the similarity or dissimilarity among the GCCA results relys heavily on the extreme eigenvalues, especially the smallest eigenvalue, of $\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}$. With this point in mind, three different types of correlation matrices (Table 4.1, 4.2, 4.3), for which λ_s is relatively large, moderately small, and extremely small against λ_l respectively, are constructed to see how the extreme eigenvalues of $\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}$ affect the final solutions. All examples have three sets of variables ($m = 3$).

The canonical weight vectors with the unit-variances constraint (2.1), $\mathbf{a}_i$, the correlation coefficients of canonical variates, ϕ_{ij}, and the two extreme elignvalues of $\mathbf{\Phi}$, l_1 and l_3, are presented in each table. As mentioned in Section 3.2, the solutions with constant-sum-variances constraint could be very unfair. Thus, for the simplicity of comparison, only the solutions with unit-variances constraint are displayed in each example.

The GCCA solutions with the unit-variances constraint is obtained by using the Gauss-Seidel type of iterative procedure. The following steps describe the routine.

Step 1. Obtain the initial values $\mathbf{a}_{i,0}$ $(i = 1, \ldots, m)$;

Step 2. At the t^{th} iteration $(t = 1, 2, \ldots)$, evaluate the stationary equations provided in (3.2) by exploiting $\mathbf{a}_{1,t}, \ldots, \mathbf{a}_{i-1,t}, \mathbf{a}_{i,t-1}, \ldots, \mathbf{a}_{m,t-1}$ and calculate the t^{th} updated values $\mathbf{a}_{i,t}$ in the due order of the subscript $i = 1, \ldots, m$;

Step 3. Repeat Step 2 until the convergence condition is satisfied.

4.1. Case 1: R *with relatively large* λ_s

This example (Table 4.1.A with $p_1 = 3$, $p_2 = 4$ and $p_3 = 2$) corresponds to the case where $\lambda_l = 1.996$ and $\lambda_s = 0.323$ for which λ_s is far from zero. Therefore most methods supposedly give very similar results overall for this data. MINVAR, however, looks somewhat different from others, resulting that ϕ_{12} is slightly large with relatively small both ϕ_{13} and ϕ_{23}.

Not many comparative works have been done on a large scale. Some of them are as follows; Horst (1961b) compared SUMCOR and MAXVAR using "Ability" data and Kettenring (1971) investigated SSQCOR, GENVAR and MINVAR on the same data. Gifi (1990) made some comparisons using "As years go by study" data, indicating that SUMCOR do not perform very well, MAXVAR and MINVAR are fair, and GENVAR and SSQCOR are possibly better, as compared with optimum values of criteria for these methods.

TABLE 4.1 *Case 1* ($\lambda_l = 1.996$, $\lambda_s = 0.323$)

A. Correlation matrix

Sets	I			II				III	
I	1	0.62	0.43	0.20	0.36	0.52	0.46	0.21	0.22
	0.62	1	0.62	0.13	0.37	0.58	0.46	0.40	0.19
	0.43	0.62	1	0.17	0.24	0.46	0.45	0.24	0.15
II	0.20	0.13	0.17	1	0.16	0.17	0.16	0.14	0.11
	0.36	0.37	0.24	0.16	1	0.48	0.36	0.16	0.03
	0.52	0.58	0.46	0.17	0.48	1	0.54	0.29	0.25
	0.46	0.46	0.45	0.16	0.36	0.54	1	0.28	0.15
III	0.21	0.40	0.24	0.14	0.16	0.29	0.28	1	0.08
	0.22	0.19	0.15	0.11	0.03	0.25	0.15	0.08	1

B. GCCA results

	SUMCOR	*MAXVAR*	*SSQCOR*	*MAXECC*	*GENVAR*	*MINVAR*
$\mathbf{a}_1$	0.313	0.331	0.355	0.433	0.404	0.446
	0.627	0.602	0.570	0.460	0.502	0.440
	0.211	0.223	0.238	0.287	0.269	0.297
$\mathbf{a}_2$	0.149	0.142	0.132	0.105	0.114	0.099
	0.017	0.031	0.050	0.110	0.088	0.120
	0.691	0.685	0.675	0.638	0.653	0.633
	0.381	0.383	0.385	0.397	0.392	0.397
$\mathbf{a}_3$	0.820	0.819	0.819	0.816	0.817	0.975
	0.510	0.511	0.513	0.516	0.516	−0.314
ϕ_{12}	0.666	0.668	0.671	0.677	0.676	0.677
ϕ_{13}	0.413	0.411	0.408	0.397	0.402	0.261
ϕ_{23}	0.397	0.396	0.395	0.389	0.391	0.247
l_1	1.996	1.996	1.995	1.989	1.992	1.832
l_3	0.334	0.331	0.328	0.323	0.324	0.323

In fact, all these results are easily expected since the two real data explored belong to the situation where *Case 1* takes. Specifically the smallest eigenvalue of $\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}$ turn out to be 0.235 for "Ability" data and 0.387 for "As years go by study" data. But, in Table 4.1.B, it is quite interesting to observe the fact that MAXECC gives basically proximate results as SUMCOR, MAXVAR, SSQCOR and GENVAR but it shows a tendency to lean slightly toward MINVAR.

4.2. Case 2: **R** *with moderately small* λ_s

For this example (Table 4.2.A with $p_1 = 3$, $p_2 = 3$, $p_3 = 3$, $\lambda_l = 2.459$ and $\lambda_s = 0.144$), the smallest eigenvalue of $\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}$ is moderately small. Thus one can expect that methods give quite different results.

TABLE 4.2 *Case 2* ($\lambda_l = 2.459$, $\lambda_s = 0.144$)

A. Correlation matrix

Sets	I			II			III		
	1	0.25	0.27	0.44	0.18	0.19	0.43	0.37	0.28
I	0.25	1	0.40	0.14	0.65	0.26	0.19	0.53	0.36
	0.27	0.40	1	0.18	0.41	0.61	0.23	0.47	0.61
	0.44	0.14	0.18	1	0.09	0.15	0.85	0.25	0.19
II	0.18	0.65	0.41	0.09	1	0.30	0.10	0.54	0.39
	0.19	0.26	0.61	0.15	0.30	1	0.18	0.44	0.50
	0.43	0.19	0.23	0.85	0.10	0.18	1	0.29	0.25
III	0.37	0.53	0.47	0.25	0.54	0.44	0.29	1	0.43
	0.28	0.36	0.61	0.19	0.39	0.50	0.25	0.43	1

B. GCCA results

	SUMCOR	*MAXVAR*	*SSQCOR*	*MAXECC*	*GENVAR*	*MINVAR*
	0.318	0.319	0.323	0.745	0.391	−0.026
$\mathbf{a}_1$	0.426	0.425	0.424	0.221	0.393	0.728
	0.590	0.589	0.588	0.341	0.565	0.464
	0.414	0.417	0.422	0.948	0.536	0.990
$\mathbf{a}_2$	0.558	0.557	0.554	0.111	0.494	−0.024
	0.497	0.497	0.495	0.143	0.460	0.067
	0.299	0.301	0.306	0.917	0.415	0.948
$\mathbf{a}_3$	0.557	0.556	0.555	0.153	0.515	0.105
	0.462	0.461	0.459	0.081	0.413	0.052
ϕ_{12}	0.712	0.712	0.711	0.503	0.691	0.188
ϕ_{13}	0.729	0.729	0.729	0.529	0.719	0.312
ϕ_{23}	0.748	0.748	0.749	0.851	0.771	0.846
l_1	2.459	2.459	2.459	2.270	2.454	1.975
l_3	0.250	0.249	0.248	0.148	0.226	0.144

We note that the correlation coefficient between the first variable for the second (II) and the third (III) set in Table 4.2 looks very high (0.85). Firstly it is observed, in Table 4.2.B, MINVAR shows very much different results. In the last column of Table 4.2.B, the canonical weights corresponding to the first variable for the second (II) and the third (III) set are dominant, 0.990 and 0.948 respectively, resulting that the canonical correlation coefficients between the second and the third set, $\phi_{23} = 0.846$, is very large. While, the results of SUMCOR, MAXVAR and SSQCOR are indistinguishable on the whole and provide almost the same level of canonical correlations, and GENVAR appears similar to those of SUMCOR, MAXVAR and SSQCOR. Thus, for this case, SUMCOR, MAXVAR, SSQCOR and GENVAR would not provide sufficiently reasonable results to outline the structure of **R**.

On the other hand, MAXECC again gives intermediate results which seem to summarize properly the characteristics of **R** in the view of having dominantly large ϕ_{23}=0.851. Here ϕ_{12} and ϕ_{13} of MAXECC are of moderate size but noticeably smaller than those of SUMCOR, MAXVAR and SSQCOR. Thus, for this case, MAXECC could be a preferable choice for the good representation on the information contained in **R**.

4.3. Case 3: R *with near zero* λ_s

Table 4.3.A (with $p_1 = 4$, $p_2 = 2$, $p_3 = 3$, $\lambda_l = 2.160$ and $\lambda_s = 0.003$) is the case where the smallest eigenvalue of $\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}$ is extremely small.

Table 4.3.B shows that the six methods separated into two clearly different groups. The first three methods (SUMCOR, MAXVAR and SSQCOR) form one group, and the rest (MINVAR, MAXECC and GENVAR) of them another. Notable difference between those two groups lie in the composition of $\mathbf{a}_1$ and $\mathbf{a}_2$. This distinguishable patterns of the canonical weights may, in turn, cause large differences in the corresponding values of the ϕ_{ij}'s. For the second group of methods, all ϕ_{12}'s take negative, even they are relatively small while ϕ_{13}'s are a little inflated. In all, observing that the differences in the elements of $\mathbf{R}_{12}$ and $\mathbf{R}_{13}$ are negligible, the gap between ϕ_{12} and ϕ_{13} in the second group of methods seems to be unreasonable.

5. REMARKS

Considering the characteristics given by three examples above, the followings are summarized: (1) When λ_s is relatively large, no significant difference exists among methods. (2) Typically used methods such as SUMCOR, MAXVAR, SSQCOR and GENVAR do not always provide sufficiently reasonable results to outline the structure of **R**, especially for the case of moderately small λ_s. This situation can occur quite often in reality. (3) MINVAR generally goes to an extreme in all cases. (4) MAXECC tends to give the acceptable range of results unless λ_s is very close to zero. And thus MAXECC could be a preferable choice for the good representation on the information contained in **R**, especially when

TABLE 4.3 *Case 3* ($\lambda_l = 2.160$, $\lambda_s = 0.003$)

A. Correlation matrix

Sets	I				II		III		
I	1	0.36	0.18	0.02	0.02	0.05	0.40	0.63	0.34
	0.36	1	0.12	0.18	0.33	0.24	0.69	0.42	0.17
	0.18	0.12	1	0.36	0.38	0.63	0.10	0.15	0.14
	0.02	0.18	0.36	1	0.64	0.44	0.10	0.08	0.07
II	0.02	0.33	0.38	0.64	1	0.88	0.14	0.07	0.13
	0.05	0.24	0.63	0.44	0.88	1	0.17	0.14	0.20
III	0.40	0.69	0.10	0.10	0.14	0.17	1	0.85	0.30
	0.63	0.42	0.15	0.08	0.07	0.14	0.85	1	0.08
	0.34	0.17	0.14	0.07	0.13	0.20	0.30	0.08	1

B. GCCA results

	SUMCOR	*MAXVAR*	*SSQCOR*	*MAXECC*	*GENVAR*	*MINVAR*
$\mathbf{a}_1$	0.639	0.699	0.741	0.944	0.944	0.944
	−0.686	−0.750	−0.816	−0.842	−0.842	−0.842
	0.546	0.456	0.353	−0.101	−0.101	−0.102
	−0.417	−0.340	−0.263	0.105	0.105	0.106
$\mathbf{a}_2$	−2.093	−2.102	−2.102	−0.631	−0.631	−0.633
	1.951	1.905	1.795	1.509	1.510	1.511
$\mathbf{a}_3$	−1.977	−2.012	−2.052	−1.987	−1.987	−1.987
	1.898	1.896	1.881	1.912	1.912	1.912
	0.692	0.668	0.636	0.669	0.669	0.669
ϕ_{12}	0.626	0.570	0.513	−0.114	−0.115	−0.115
ϕ_{13}	0.817	0.871	0.913	0.966	0.966	0.966
ϕ_{23}	0.246	0.239	0.228	0.132	0.132	0.133
l_1	2.156	2.160	2.155	1.966	1.966	1.966
l_3	0.080	0.057	0.037	0.003	0.003	0.003

there is no definite ground available telling us how to choose a criterion. (5) When λ_s is sufficiently small, any method of SUMCOR, MAXVAR or SSQCOR is recommended.

APPENDIX : PROOF OF THE EQUATIONS (2.4) AND (3.3)

Proof of the Equation (2.4)

Since l_k is a function of ϕ_{uv}'s ($u \geq v$), we have, by using the chain rule,

$$\frac{\partial l_k}{\partial a_{si}} = \sum_{u=1}^{m} \sum_{v<u}^{m} \frac{\partial l_k}{\partial \phi_{uv}} \frac{\partial \phi_{uv}}{\partial a_{si}}, \tag{A.1}$$

where a_{si} is the s^{th} element of $\mathbf{a}_i$. The first differential term in the right-hand side of (A.1) becomes (Magnus and Neudecker, 1988),

$$\begin{aligned}\frac{\partial l_k}{\partial \phi_{uv}} &= \mathbf{e}_k' \frac{\partial \mathbf{\Phi}}{\partial \phi_{uv}} \mathbf{e}_k \\ &= \begin{cases} \mathbf{e}_k' \mathbf{J}_{uu} \mathbf{e}_k = e_{uk}^2, & u = v, \\ \mathbf{e}_k'(\mathbf{J}_{uv} + \mathbf{J}_{vu})\mathbf{e}_k = 2e_{uk}e_{vk}, & u \neq v, \end{cases}\end{aligned} \tag{A.2}$$

where $\mathbf{J}_{uv}$ stands for a matrix with (A.1) in the $(u,v)^{th}$ place and zeroes elsewhere. By denoting ${}_{ht}r_{uv}$ the (h,t) element of $\mathbf{R}_{uv}$, we have

$$\phi_{uv} = \mathbf{a}_u' \mathbf{R}_{uv} \mathbf{a}_v = \sum_{h=1}^{p_u} \sum_{t=1}^{p_v} a_{hu}\ {}_{ht}r_{uv}\ a_{tv}. \tag{A.3}$$

Therefore the differential $\partial \phi_{uv} / \partial a_{si}$ takes the following form;

$$\frac{\partial \phi_{uv}}{\partial a_{si}} = \begin{cases} \sum_{t=1}^{p_i} {}_{st}r_{ii}\ a_{ti} + \sum_{h=1}^{p_i} a_{hi}\ {}_{hs}r_{ii}, & i = u = v, \\ \sum_{t=1}^{p_v} {}_{st}r_{iv}\ a_{tv}, & i = u \text{ and } i \neq v, \\ \sum_{h=1}^{p_u} a_{hu}\ {}_{hs}r_{ui}, & i \neq u \text{ and } i = v, \\ 0, & i \neq u \text{ and } i \neq v. \end{cases} \tag{A.4}$$

Substituting (A.2) and (A.4) into (A.1), we have

$$\begin{aligned}\frac{\partial l_k}{\partial a_{si}} &= \sum_{t=1}^{p_i} e_{ik}^2\ {}_{st}r_{ii}\ a_{ti} + \sum_{h=1}^{p_i} e_{ik}^2\ {}_{hs}r_{ii}\ a_{hi} \\ &\quad + 2 \sum_{v=1}^{i-1} \sum_{t=1}^{p_v} e_{ik}e_{vk}\ {}_{st}r_{iv}\ a_{tv} \\ &\quad + 2 \sum_{u=i+1}^{m} \sum_{h=1}^{p_u} e_{ik}e_{uk}\ {}_{hs}r_{ui}\ a_{hu}.\end{aligned} \tag{A.5}$$

Since ${}_{ts}r_{ji} = {}_{st}r_{ij}$, we finally obtain

$$\frac{\partial l_k}{\partial a_{si}} = 2 \sum_{j=1}^{m} \sum_{t=1}^{p_j} e_{ik}e_{jk}\ {}_{st}r_{ij}\ a_{tj}. \tag{A.6}$$

The proof is completed when we generalize the result (A.6).

Proof of the Equation (3.3)

The criterion for each of six GCCA solutions can be written as

$$f(\cdot) = \begin{cases} \sum_{i,j}^{m} \phi_{ij} = \sum_{i,j}^{m} \mathbf{a}_i'\mathbf{R}_{ij}\mathbf{a}_j, & \text{for SUMCOR,} \\ l_1, & \text{for MAXVAR,} \\ l_m, & \text{for MINVAR,} \\ \sum_{i,j}^{m} \phi_{ij}^2 = \sum_{k=1}^{m} l_k^2, & \text{for SSQCOR,} \\ \det(\mathbf{\Phi}) = \prod_{k=1}^{m} l_k, & \text{for GENVAR,} \\ \frac{l_1 - l_m}{l_1 + l_m}, & \text{for MAXECC.} \end{cases} \tag{A.7}$$

Using the equation (2.4), for each of six GCCA solutions, the partial derivative of f(·) with respect to $\mathbf{a}_i$ is given by

$$\frac{\partial f}{\partial \mathbf{a}_i} = \begin{cases} 2\sum_{j=1}^{m} \mathbf{R}_{ij}\mathbf{a}_j, \\ 2\sum_{j=1}^{m} e_{i1}e_{j1}\mathbf{R}_{ij}\mathbf{a}_j, \\ 2\sum_{j=1}^{m} e_{im}e_{jm}\mathbf{R}_{ij}\mathbf{a}_j, \\ 4\sum_{j=1}^{m}\sum_{k=1}^{m} e_{ik}e_{jk}l_k\mathbf{R}_{ij}\mathbf{a}_j, \\ 2\sum_{j=1}^{m}\sum_{k=1}^{m} e_{ik}e_{jk}\frac{l_k}{\det(\mathbf{\Phi})}\mathbf{R}_{ij}\mathbf{a}_j, \\ 2\sum_{j=1}^{m} \frac{e_{i1}e_{j1}l_m - e_{im}e_{jm}l_1}{(l_1 + l_m)^2}\mathbf{R}_{ij}\mathbf{a}_j. \end{cases} \tag{A.8}$$

Thus letting the partial derivative of g(·) with respect to $\mathbf{a}_i$ equal to zero, after some calculation by applying the equation (A.8), yields the equation (3.3).

REFERENCES

CARROLL, J. D. (1968). "Generalization of canonical correlation analysis to three or more sets of variables", *Proceedings of the American Psychological Association*, 3 227–228.

COPPI, R. AND BOLASCO, S. (1989). *Multiway Data Analysis*, North-Holland, Amsterdam.

GIFI, A. (1990). *Nonlinear Multivariate Analysis*, John Wiley & Sons, Chichester.

HORST, P. (1961a). "Relations among m sets of measures", *Psychometrika* **26**, 129–149.

HORST, P. (1961b). "Generalized canonical correlations and their applications to experimental data", *Journal of Clinical Psychology*, **17**, 331–347.

HORST, P. (1965). *Factor Analysis of Data Matrices*, Holt, Rinehart and Winston, New York.

HOTELLING, H. (1936). "Relations between two sets of variates", *Biometrika*, **28**, 321–377.

KETTENRING, J. R. (1971). "Canonical analysis of several sets of variables", *Biometrika*, **58**, 433–451.

LEBART, L., MORINEAU, A. AND WARWICK, K. (1984). *Multivariate Descriptive Statistical Analysis. Correspondence Analysis and Related Techniques for Large Matrices*, John Wiley & Sons, New York.

MAGNUS, J. R. AND NEUDECKER, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons, Chichester.

MELZER, T., REITER, M. AND BISCHOF, H. (2001). "Nonlinear feature extraction using generalized canonical correlation analysis", *Proceedings of the International Conference on Artificial Neural Networks*, 353–360.

SCHUENEMEYER, J. H. AND BARGMANN R. E. (1978). "Maximum eccentricity as a union-intersection test statistic in multivariate analysis", *Journal of Multivariate Analysis*, **8**, 268–273.

STEEL, R. G. D. (1951). "Minimum generalized variance for a set of linear functions", *The Annals of Mathematical Statistics*, **22**, 456–460.

TAKANE, Y. AND HWANG, H. (2002). "Generalized constrained canonical correlation analysis", *Multivariate Behavioral Research*, **37**, 163–195.

TEN BERGE, J. M. F. (1988). "Generalized approaches to the Maxbet problem and the Maxdiff problem, with applications to canonical correlations", *Psychometrika*, **53**, 487–494.

VAN DE GEER, J. P. (1984). "Linear relations among k sets of variables", *Psychometrika*, **49**, 79–94.