Higher-order solutions for generalized canonical correlation analysis

Hyuncheol Kang^{1,a}

^aDivision of Big Data and Management Engineering, Hoseo University, Korea

Abstract

Generalized canonical correlation analysis (GCCA) extends the canonical correlation analysis (CCA) to the case of more than two sets of variables and there have been many studies on how two-set canonical solutions can be generalized. In this paper, we derive certain stationary equations which can lead the higher-order solutions of several GCCA methods and suggest a type of iterative procedure to obtain the canonical coefficients. In addition, with some numerical examples we present the methods for graphical display, which are useful to interpret the GCCA results obtained.

Keywords: generalized canonical correlation analysis, higher-order solutions, canonical weights, goodness of approximation indices, canonical loadings, explained variance indices

1. Introduction

Generalized canonical correlation analysis (GCCA) involves comparing m sets of variables after removing linear dependencies within each set that can extend the canonical correlation analysis (CCA) of Hotelling (1936) to cases of more than two sets of variables. There have been many studies on how two-set canonical solutions can be generalized (Horst, 1961a, 1961b, 1965; Carroll, 1968; Kettenring, 1971; Van de Geer, 1984; Ten Berge, 1988; Coppi and Bolasco, 1989; Gifi, 1990; Park and Huh, 1996), with discussions on the criteria used and constraints imposed.

Suppose we have m data matrices $\mathbf{X}_i$ ($i = 1, \dots, m$), each from a sample of size n on p_i variables. It is assumed implicitly that all the variables are standardized to have zero means and unit variances. The problem is to find linear composites $\mathbf{z}_i = \mathbf{X}_i \mathbf{a}_i$ for each set in such a way that the matrix $\mathbf{Z} = (\mathbf{z}_1 \cdots \mathbf{z}_i \cdots \mathbf{z}_m) = (\mathbf{X}_1 \mathbf{a}_1 \cdots \mathbf{X}_i \mathbf{a}_i \cdots \mathbf{X}_m \mathbf{a}_m)$ optimizes a particular function of its covariance matrix, $\mathbf{\Phi} = \{\phi_{ij}\}$, $i, j = 1, \dots, m$.

Criteria, the functions to be optimized, for obtaining the canonical coefficient vectors $\mathbf{a}_i$ ($i = 1, \dots, m$) are characterized by the chosen function of $\mathbf{\Phi}$. Each object function corresponding to a criterion is 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 $\mathbf{Z}$: (i) SUMCOR (maximize $\sum_{i,j}^m \phi_{ij}$), (ii) MAXVAR (maximize l_1), (iii) SSQCOR (maximize $\sum_{i,j}^m \phi_{ij}^2$), (iv) MINVAR (minimize l_m), (v) GENVAR (minimize $\det(\mathbf{\Phi})$). Kang and Kim (2006) added a criterion for the GCCA solution: (vi) MAXECC (maximize $(l_1 - l_m)/(l_1 + l_m)$).

Some of the above have also been suggested by other studies (Horst, 1961a, 1961b, 1965; Steel, 1951) and Gifi (1990) mentioned other methods MINSUM (minimize the sum of the correlations)

¹ Division of Big Data and Management Engineering, Hoseo University, 20, Hoseo-ro 79beon-gil, Baebang-eup, Asan-si, Chungcheongnam-do 31499, Korea. E-mail: hychkang@hoseo.edu

and PRINCALS (maximize the sum of the q largest eigenvalues of Φ , or, equivalently, minimize the sum of the $m - q$ smallest ones). Gifi also suggested a nonlinear version of GCCA, called OVERALS based on the MAXVAR criterion that incorporates nonlinear transformations of the data.

The study of relations among the sets can be continued beyond the first order (stage) by considering higher order (stage) canonical variates $\mathbf{Z}_{(2)}, \mathbf{Z}_{(3)}, \dots$ to supplement the optimal $\mathbf{Z}_{(1)}$. The same criterion function is used at each order, but restrictions are added to assure that canonical variables for a particular order differ from the previous orders.

In this paper, we derive certain stationary equations that derive solutions for higher-order solutions of several GCCA methods and suggest a type of iterative procedure to obtain canonical coefficients. In addition, we present methods for graphical display with some numerical examples, which are useful to interpret GCCA results.

2. Unifying stationary equations for GCCA

In order to construct a valid optimization problem, the criterion to be optimized is usually subjected to certain constraint. Let $\mathbf{R}$ denote the correlation matrix of all original variables, and $\mathbf{R}_{ij}$ of size $p_i \times p_j$ the correlation matrix between the i^{th} and the j^{th} set of variables, which is a submatrix of $\mathbf{R}$. A typical constraint (unit-variances constraint) is

$$\mathbf{a}_i' \mathbf{R}_{ii} \mathbf{a}_i = 1, \quad i = 1, \dots, m. \quad (2.1)$$

Under this constraint, Φ can be interpreted as a correlation matrix, which is advantage of considering the constraint.

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

$$g(\mathbf{a}_1, \dots, \mathbf{a}_m) = f(\mathbf{a}_1, \dots, \mathbf{a}_m) - \sum_{i=1}^m \mu_i (\mathbf{a}_i' \mathbf{R}_{ii} \mathbf{a}_i - 1), \quad (2.2)$$

where $f(\cdot)$ stands for a criterion to be considered, Kang and Kim (2006) derived the following form of unifying stationary equations for each of the six different GCCA solutions,

$$\sum_{j=1}^m w_{ij} \mathbf{R}_{ij} \mathbf{a}_j = \mu_i^* \mathbf{R}_{ii} \mathbf{a}_i, \quad i = 1, \dots, m, \quad (2.3)$$

where

$$w_{ij} = \begin{cases} \frac{1}{m}, & \text{for SUMCOR,} \\ e_{i1} e_{j1}, & \text{for MAXVAR,} \\ e_{im} e_{jm}, & \text{for MINVAR,} \\ \sum_{k=1}^m e_{ik} e_{jk} l_k, & \text{for SSQCOR,} \\ \sum_{k=1}^m e_{ik} e_{jk} / l_k, & \text{for GENVAR,} \\ e_{i1} e_{j1} l_m - e_{im} e_{jm} l_1, & \text{for MAXECC,} \end{cases} \quad (2.4)$$

respectively, and e_{ij} is the i^{th} elements of $\mathbf{e}_j$ which is the unit-normed eigenvector associated with the j^{th} eigenvalue l_j of Φ . The stationary equation (2.3) implies that GCCA methods differ only by

applying different weights as shown in (2.4) to obtain the desired solutions. Therefore it could be a way to characterize each GCCA method in terms of w_{ij} .

In general, imposing different constraints for canonical variates could mean deriving different solutions even for the same criterion. Another useful constraint (constant-sum-variances constraint), an alternative to the unit-variances constraint, is

$$\sum_{i=1}^m \mathbf{a}'_i \mathbf{R}_{ii} \mathbf{a}_i = \mathbf{a}' \mathbf{D} \mathbf{a} = m, \quad (2.5)$$

where $\mathbf{a}' = (\mathbf{a}'_1 \cdots \mathbf{a}'_m)$ 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 the constant-sum-variances constraint could be very unfair (Van de Geer, 1984). The canonical variates with this constraint have not identical variances in general and the criteria could be heavily dependent on canonical variates with relatively large or small variances. However, the constant-sum-variances constraint would sometimes be preferred due to relatively easy computations. Moreover, for some criteria, since the stationary equations with constant-sum-variances constraint have an explicit form (Kang and Kim, 2006), considering this constraint is helpful to clarify the characteristics of the criteria.

The constant-sum-variances constraint needs requires one Lagrange multiplier μ ; therefore, the equations to solve for the solutions can be written as the following matrix form of representation

$$\mathbf{R} \mathbf{D}_a \mathbf{W} = \mu \mathbf{D} \mathbf{D}_a, \quad (2.6)$$

where $\mathbf{D}_a$ denotes a block diagonal matrix with $\mathbf{a}_i$ as its i^{th} block and $\mathbf{W}$ is a matrix with w_{ij} of (2.4) as its $(i, j)^{th}$ element.

Further discussions for GCCA solutions through the stationary equations (2.3) and (2.6) are presented in Kang and Kim (2006) with some results of numerical illustrations.

3. Higher-order solutions

In this section, we derive certain stationary equations which can lead the higher-order GCCA solutions.

3.1. Higher-order solutions and computing procedure

For higher-order GCCA solutions, we need to place further restrictions on the additional canonical variates in order to achieve mutual orthogonality and compute them successively.

Suppose $\mathbf{Z}_{(k)} = (\mathbf{z}_{1(k)} \cdots \mathbf{z}_{i(k)} \cdots \mathbf{z}_{m(k)}) = (\mathbf{X}_1 \mathbf{a}_{1(k)} \cdots \mathbf{X}_i \mathbf{a}_{i(k)} \cdots \mathbf{X}_m \mathbf{a}_{m(k)})$ denotes the set of the k^{th} order canonical variates for $k \leq q$ and $q = \min(p_1, p_2, \dots, p_m)$. Then the problem, under the unit-variance constraint, reduces to optimizing the following object function $g(\cdot)$:

$$g(\mathbf{a}_{(k)}) = f(\mathbf{a}_{(k)}) - \sum_{i=1}^m \mu_i (\mathbf{a}'_{i(k)} \mathbf{R}_{ii} \mathbf{a}_{i(k)} - 1) - 2 \sum_{i=1}^m \mathbf{a}'_{i(k)} \mathbf{C}_{i(k)} \boldsymbol{\gamma}_i, \quad (3.1)$$

where the μ_i and $\boldsymbol{\gamma}_i$ are the Lagrange multipliers, $f(\mathbf{a}_{(k)})$ is the function chosen according to the criterion, and $\mathbf{C}_{i(k)}$ is a matrix that satisfies the following form of restrictions for the orthogonality

$$\mathbf{a}'_{i(k)} \mathbf{C}_{i(k)} = \mathbf{0}, \quad i = 1, \dots, m. \quad (3.2)$$

In the context of m -set analysis, following two restrictions are usually used (Gifi, 1990). A reasonable restriction on $\mathbf{Z}_{(k)}$, the strong orthogonality restriction, could be

$$\text{corr}(\mathbf{z}_{i(k)}, \mathbf{z}_{i(l)}) = \mathbf{a}'_{i(k)} \mathbf{R}_{ii} \mathbf{a}_{i(l)} = 0, \quad l = 1, \dots, k-1; i = 1, \dots, m, \quad (3.3)$$

which means the canonical variates obtained at each successive order are required to be uncorrelated to each other within the same set. In this case, $\mathbf{C}_{i(k)} = \mathbf{R}_{ii}(\mathbf{a}_{i(1)} \cdots \mathbf{a}_{i(k-1)})$. Another restriction, the weak orthogonality restriction, is

$$\sum_{i=1}^m \text{corr}(\mathbf{z}_{i(k)}, \mathbf{z}_{i(l)}) = \mathbf{a}'_{(k)} \mathbf{D} \mathbf{a}_{(l)} = 0, \quad l = 1, \dots, k-1, \quad (3.4)$$

which is useful when joined with constant-sum-variances constraint. With this restriction, solutions of every order for MAXVAR and MINVAR can be obtained simultaneously by performing a single eigen-analysis on $\mathbf{D}^{-1/2} \mathbf{R} \mathbf{D}^{-1/2}$ which is a considerable advantage, both computationally and theoretically (Kang and Kim, 2006).

The two restrictions mentioned above yield the same canonical variates when $m = 2$ (Anderson, 1984; Kettenring, 1971).

We now derive stationary equations for solutions with any form of (3.2), under the unit-variance constraint. Taking the derivative of $g(\cdot)$ with respect to $\mathbf{a}_i$ in (3.1) and setting it to zero, applying appendix of Kang and Kim (2006) and after some mathematical computations, yield the following system of stationary equations

$$\sum_{j=1}^m w_{ij} \mathbf{E}_i \mathbf{R}_{ij} \mathbf{a}_{j(k)} = \mu_i \mathbf{R}_{ii} \mathbf{a}_{i(k)}, \quad i = 1, \dots, m, \quad (3.5)$$

where $\mathbf{E}_i = \mathbf{I} - \mathbf{C}_{i(k)}(\mathbf{C}'_{i(k)} \mathbf{R}_{ii}^{-1} \mathbf{C}_{i(k)})^{-1} \mathbf{C}'_{i(k)} \mathbf{R}_{ii}^{-1}$ and the weights w_{ij} are the same as those given in (2.4).

The k^{th} order solutions for GCCA methods can be obtained using a Gauss-Seidel type of iterative procedure. The following steps describe the iterative routine.

Step 1. Obtain the initial values $\mathbf{a}_i^0$ ($i = 1, \dots, m$).

Step 2. At the t^{th} iteration ($t = 1, 2, \dots$), evaluate the stationary equations provided in (3.5) by exploiting $\mathbf{a}_1^t, \dots, \mathbf{a}_{i-1}^t, \mathbf{a}_i^{t-1}, \dots, \mathbf{a}_m^{t-1}$ and calculate the t^{th} updated values $\mathbf{a}_i^t$ in the due order of the subscript $i = 1, \dots, m$. In this step, $\mathbf{a}_i^t$ is obtained from a vector proportional to $\sum_{j=1}^m w_{ij} \mathbf{R}_{ii}^{-1} \mathbf{E}_i \mathbf{R}_{ij} \mathbf{a}_j$ as follows.

- (1) calculate Φ , the correlation matrix of canonical variates
- (2) all weights of w_{ij} ($j = 1, \dots, m$) are updated from Φ
- (3) $\mathbf{a}_i \leftarrow \sum_{j=1}^m w_{ij} \mathbf{E}_i \mathbf{R}_{ij} \mathbf{a}_j$
- (4) $\mathbf{a}_i \leftarrow \mathbf{a}_i / \sqrt{\mathbf{a}_i' \mathbf{a}_i}$
- (5) $\mathbf{a}_i \leftarrow \mathbf{R}_{ii}^{-1/2} \mathbf{a}_i$

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

Table 1: Correlation matrix

Sets of variables		I			II			III		
		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
I	x_1	1.00	0.25	0.27	0.44	0.18	0.19	0.43	0.37	0.28
	x_2	0.25	1.00	0.40	0.14	0.65	0.26	0.19	0.53	0.36
	x_3	0.27	0.40	1.00	0.18	0.41	0.61	0.23	0.47	0.61
II	x_4	0.44	0.14	0.18	1.00	0.09	0.15	0.85	0.25	0.19
	x_5	0.18	0.65	0.41	0.09	1.00	0.30	0.10	0.54	0.39
	x_6	0.19	0.26	0.61	0.15	0.30	1.00	0.18	0.44	0.50
III	x_7	0.43	0.19	0.23	0.85	0.10	0.18	1.00	0.29	0.25
	x_8	0.37	0.53	0.47	0.25	0.54	0.44	0.29	1.00	0.43
	x_9	0.28	0.36	0.61	0.19	0.39	0.50	0.25	0.43	1.00

Table 2: The first and the second-order solutions of GCCA

Sets	SUMCOR		MAXVAR		SSQCOR		MAXECC		GENVAR		MINVAR	
$\mathbf{a}_1$	0.318	1.002	0.319	1.001	0.323	0.999	0.745	0.731	0.391	0.967	-0.026	0.995
	0.426	-0.360	0.425	-0.368	0.424	-0.388	0.221	-0.638	0.393	-0.486	0.728	-0.416
	0.590	-0.376	0.589	-0.370	0.588	-0.355	0.341	-0.485	0.565	-0.328	0.464	0.131
$\mathbf{a}_2$	0.414	0.920	0.417	0.919	0.422	0.916	0.948	0.341	0.536	0.853	0.990	0.178
	0.558	-0.384	0.557	-0.385	0.554	-0.392	0.111	-0.696	0.494	-0.468	-0.024	-0.675
	0.497	-0.289	0.497	-0.290	0.495	-0.290	0.143	-0.526	0.460	-0.342	0.067	-0.576
$\mathbf{a}_3$	0.299	1.008	0.301	1.009	0.306	1.009	0.917	0.524	0.415	0.971	0.948	0.464
	0.557	-0.325	0.556	-0.343	0.555	-0.364	0.153	-0.714	0.515	-0.457	0.105	-0.693
	0.462	-0.395	0.461	-0.380	0.459	-0.362	0.081	-0.505	0.413	-0.375	0.052	-0.543
ϕ_{12}	0.712	0.415	0.712	0.416	0.711	0.417	0.503	0.555	0.691	0.433	0.188	0.005
ϕ_{13}	0.729	0.308	0.729	0.307	0.729	0.306	0.529	0.416	0.719	0.305	0.312	-0.082
ϕ_{23}	0.748	0.778	0.748	0.779	0.749	0.778	0.851	0.675	0.771	0.757	0.846	0.669
l_1	2.459	2.033	2.459	2.033	2.459	2.032	2.270	2.104	2.454	2.024	1.975	1.673
l_3	0.250	0.212	0.249	0.212	0.248	0.212	0.148	0.298	0.226	0.229	0.144	0.326

GCCA = generalized canonical correlation analysis; SUMCOR = maximize $\sum_{i,j}^m \phi_{ij}$; MAXVAR = maximize l_1 ; SSQCOR = maximize $\sum_{i,j}^m \phi_{ij}^2$; MAXECC = maximize $(l_1 - l_m)/(l_1 + l_m)$; GENVAR = minimize $\det(\Phi)$; MINVAR = minimize l_m .

The weight vectors $\mathbf{a}_i$ from above procedure can be called standardized canonical weights (coefficients), because those are obtained based on $\mathbf{R}$, the correlation matrix of original variables. If data matrices $\mathbf{X}_i$ are not standardized, the raw canonical weights (coefficients) can be obtained by

$$\mathbf{b}_i = \mathbf{V}_i^{-\frac{1}{2}} \mathbf{a}_i, \quad i = 1, \dots, m, \quad (3.6)$$

where $\mathbf{V}_i$ is a diagonal matrix with the variances of original variables as its diagonal elements.

3.2. Numerical illustration

In this section, a numerical example is constructed to illustrate contrast among the six different GCCA methods. This example has three sets of variables ($m = 3$) and the correlation matrix with $p_1 = p_2 = p_3 = 3$ is presented in Table 1. For several GCCA methods, it can be expected that the similarities and dissimilarities of the results rely on the extreme eigenvalues, especially the smallest eigenvalue, of $\mathbf{D}^{-1/2} \mathbf{R} \mathbf{D}^{-1/2}$. For this example, the smallest eigenvalue of $\mathbf{D}^{-1/2} \mathbf{R} \mathbf{D}^{-1/2}$ is 0.144 which is moderately small. Thus one can expect that the methods give very different results.

Table 2 displays the first and the second-order solutions of six GCCA methods, which is obtained by above iterative procedure: the canonical weight vectors ($\mathbf{a}_i$), the correlation coefficients of canonical variates (ϕ_{ij}), and the two extreme eigenvalues of Φ (l_1 and l_3).

Table 3: Average correlations of canonical variates and GAI

k	SUMCOR		MAXECC		MINVAR	
	$\phi_{(k)}$	GAI	$\phi_{(k)}$	GAI	$\phi_{(k)}$	GAI
1	0.730	61.1%	0.628	50.1%	0.449	62.0%
2	0.500	89.8%	0.549	88.4%	0.197	74.0%
3	0.299	100%	0.302	100%	0.291	100%

GAI = goodness-of-approximation indices; SUMCOR = maximize $\sum_{i,j}^m \phi_{ij}$; MAXECC = maximize $(l_1 - l_m)/(l_1 + l_m)$; MINVAR = minimize l_m .

Table 2 indicates that MINVAR shows very different results of canonical weights and canonical correlation coefficients. The canonical weights of SUMCOR, MAXVAR, and SSQCOR are indistinguishable on the whole and provide almost the same level of canonical correlations; consequently, GENVAR appears similar to those of SUMCOR, MAXVAR, and SSQCOR. However, MAXECC gives intermediate results that seem noticeably different from SUMCOR, MAXVAR, and SSQCOR.

4. Graphical display of GCCA results

In this section, we present the methods for graphical display which are useful to interpret the GCCA results.

4.1. Quantification plots with canonical weights

Let $\mathbf{A}_i$ denote the canonical weight matrix of the i^{th} set of variables under the unit-variance constraint,

$$\mathbf{A}_i = (\mathbf{a}_{i(1)}, \dots, \mathbf{a}_{i(k)}, \dots, \mathbf{a}_{i(q)}), \quad i = 1, \dots, m, \quad (4.1)$$

and $\mathbf{D}_{\bar{\phi}}$ a diagonal matrix with the average correlations of canonical variates as its k^{th} elements,

$$\mathbf{D}_{\bar{\phi}} = \text{diag}(\bar{\phi}_{(1)}, \dots, \bar{\phi}_{(k)}, \dots, \bar{\phi}_{(q)}), \quad (4.2)$$

where $\bar{\phi}_{(k)} = \sum \sum_{i \neq j} \phi_{ij(k)} / [m(m-1)]$, the average correlation coefficients of the k^{th} order canonical variates.

Park and Huh (1996) suggest an r -dimensional quantification plot obtained by plotting $\mathbf{A}_{i(r)} \mathbf{D}_{\bar{\phi}(r)}$, the first r columns of $\mathbf{A}_i \mathbf{D}_{\bar{\phi}}$, for $r \leq q$ and $q = \min(p_1, p_2, \dots, p_m)$. This plot is useful to investigate the structural relationships of variables. Now, as an extend of two sets of variables, we may define a ‘goodness-of-approximation indices (GAI)’ by

$$\text{GAI}_{(r)} = \frac{\text{tr}(\mathbf{D}_{\bar{\phi}(r)}^2)}{\text{tr}(\mathbf{D}_{\bar{\phi}}^2)} = \frac{\sum_{k=1}^r \bar{\phi}_{(k)}^2}{\sum_{k=1}^q \bar{\phi}_{(k)}^2}, \quad (4.3)$$

Table 3 presents the average correlation coefficients of the k^{th} order canonical variates ($\bar{\phi}_{(k)}$) and GAI for selected three GCCA methods. Figure 1 displays the 2-dimensional quantification plots with canonical weights.

4.2. Canonical loading plots

The canonical loadings, the correlations between the original variables and their canonical variables, can be obtained by

$$\mathbf{A}_i = \mathbf{R}_{il} \mathbf{A}_i, \quad i = 1, \dots, m \quad (4.4)$$

Figure 1: Quantification plots with canonical weights. SUMCOR = maximize $\sum_{i,j} \phi_{ij}$; MAXECC = maximize $(l_1 - l_m)/(l_1 + l_m)$; MINVAR = minimize l_m .

Table 4: The first and the second-order canonical loadings

	SUMCOR		MAXVAR		SSQCOR		MAXECC		GENVAR		MINVAR	
λ_1	0.584	0.811	0.585	0.810	0.587	0.806	0.892	0.441	0.642	0.756	0.281	0.926
	0.741	-0.260	0.741	-0.266	0.740	-0.280	0.543	-0.649	0.717	-0.376	0.907	-0.115
	0.846	-0.249	0.846	-0.247	0.845	-0.241	0.631	-0.543	0.827	-0.262	0.748	0.233
λ_2	0.539	0.842	0.541	0.840	0.546	0.837	0.979	0.199	0.649	0.760	0.998	0.031
	0.745	-0.388	0.743	-0.390	0.741	-0.396	0.239	-0.823	0.680	-0.494	0.085	-0.832
	0.727	-0.266	0.726	-0.268	0.725	-0.270	0.318	-0.684	0.688	-0.355	0.209	-0.752
λ_3	0.576	0.815	0.578	0.815	0.582	0.813	0.982	0.190	0.668	0.744	0.992	0.127
	0.842	-0.203	0.842	-0.213	0.841	-0.227	0.454	-0.779	0.813	-0.337	0.402	-0.791
	0.776	-0.283	0.776	-0.275	0.774	-0.267	0.376	-0.681	0.738	-0.329	0.334	-0.724

SUMCOR = maximize $\sum_{i,j} \phi_{ij}$; MAXVAR = maximize l_1 ; SSQCOR = maximize $\sum_{i,j} \phi_{ij}^2$; MAXECC = maximize $(l_1 - l_m)/(l_1 + l_m)$; GENVAR = minimize $\det(\Phi)$; MINVAR = minimize l_m .

and thus the k^{th} column of Λ_i is $\lambda_{i(k)} = \mathbf{R}_{ii} \mathbf{a}_{i(k)}$. The explained variance of the i^{th} set of variables by their k^{th} canonical variate is

$$EV_{i(k)} = \frac{1}{p_i} \lambda'_{i(k)} \lambda_{i(k)}, \quad i = 1, \dots, m; \quad k = 1, \dots, q, \quad (4.5)$$

which satisfies $\sum_{k=1}^q EV_{i(k)} = 1$.

The canonical loading plot is obtained by plotting $\Lambda_{i(r)}$, the first r columns of Λ_i . This plot is useful to interpret the meaning of canonical variates. Now, as an extend of two sets of variables, we may define a 'explained-variance indices (EVI)' for i^{th} set of variables by

$$EVI_{i(r)} = \sum_{k=1}^r EV_{i(k)}, \quad i = 1, \dots, m. \quad (4.6)$$

Table 4 shows the first and the second-order canonical loadings of six GCCA methods; in addition, Table 5 presents the explained variances by the k^{th} order canonical variates and explained-variance indices for selected three GCCA methods. Figure 2 displays the 2-dimentional canonical loading plots. For example, from the results of SUMCOR, the first-order canonical variates of the first set can be interpreted as the overall mean of original variables, the second-order canonical variates of the first set can be interpreted as the contrast between x_1 and others.

Table 5: Explained-variance indices

Sets	k	SUMCOR		MAXECC		MINVAR	
		EV	EVI	EV	EVI	EV	EVI
I	1	0.535	53.5%	0.496	49.6%	0.487	48.7%
	2	0.262	79.8%	0.304	80.0%	0.308	79.6%
	3	0.202	100%	0.200	100%	0.204	100%
II	1	0.458	45.8%	0.372	37.2%	0.349	34.9%
	2	0.310	76.8%	0.395	76.7%	0.419	76.8%
	3	0.232	100%	0.233	100%	0.232	100%
III	1	0.548	54.8%	0.437	43.7%	0.419	41.9%
	2	0.262	81.0%	0.369	80.6%	0.389	80.8%
	3	0.190	100%	0.194	100%	0.192	100%

SUMCOR = maximize $\sum_{i,j}^m \phi_{ij}$; MAXECC = maximize $(l_1 - l_m)/(l_1 + l_m)$; MINVAR = minimize l_m ; EVI = explained-variance indices.

Figure 2: Canonical loading plots. SUMCOR = maximize $\sum_{i,j}^m \phi_{ij}$; MAXECC = maximize $(l_1 - l_m)/(l_1 + l_m)$; MINVAR = minimize l_m .

References

- Anderson TW (1984). *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York.
- Carroll JD (1968). Generalization of canonical correlation analysis to three or more sets of variables. In *Proceedings of American Psychology Association*, 227–228.
- Coppi R and Bolasco S (1989). *Multiway Data Analysis*, North-Holland, New York.
- Gifi A (1990). *Nonlinear Multivariate Analysis*, Wiley, New York.
- 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* (Monograph supplement), **14**, 331–347.
- Horst P (1965). *Factor Analysis of Data Matrices*, Holt, Rinehart and Winston, New York.
- Kang H and Kim K (2006). Unifying stationary equations for generalized canonical correlation analysis, *Journal of the Korean Statistical Society*, **35**, 143–156.
- Kettenring JR (1971). Canonical analysis of several sets of variables, *Biometrika*, **58**, 433–451.
- Park MR and Huh MH (1996). Quantification plots for several sets of variables, *Journal of the Korean Statistical Society*, **25**, 589–601.
- Steel GRD (1951). Minimum generalized variance for a set of linear functions, *Annals of Mathematical Statistics*, **22**, 456–460.

- Ten Berge JMF (1988). Generalized approaches to the MAXBET problem and the MAXDIFF problem, with applications to canonical correlations, *Psychometrika*, **53**, 487–494.
- Van de Geer JP (1984). Linear relations among k sets of variables, *Psychometrika*, **49**, 79–94.

Received February 27, 2019; Revised March 27, 2019; Accepted March 27, 2019