Generalized Constrained Redundancy Analysis

Yoshio Takane, McGill University

Sunho Jung, McGill University

The work reported in this paper has been supported by research grants from the Natural Sciences and Engineering Research Council of Canada to the first author. Requests for reprints should be sent to Yoshio Takane, Department of Psychology, McGill University, 1205 Dr. Penfiled Avenue, Montréal, QC, H3A 1B1, Canada.

Mailing Address: Yoshio Takane

Department of Psychology, McGill Uinversity

1205 Dr. Penfield Ave., Montreal,

QC, H3A 1B1 CANADA

Email Address: takane@psych.mcgill.ca

Abstract

A method of generalized constrained redundancy analysis (GCRA) is proposed, which incorporates external information in redundancy analysis (RA). In this method both the criterion variables and the orthogonal projector defined by the predictor variables are first decomposed into several components according to the external information, and RA is applied to the decomposed matrices. By combining the terms in the two decompositions, a variety of existing and new methods of RA are realized including a variety of partial (non-partial, partial, semi-partial and bi-partial) and constrained (unconstrained, semi-constrained and bi-constrained) RA. An example is given to illustrate the method.

Keywords: Reduced rank regression, Row and column information matrices, Decompositions, Orthogonal projectors, Constrained RA, Partial RA, Permutation tests.

1 Introduction

There are countless situations in behavioral and social sciences in which a set of multivariate data are predicted from another set. In education, for example, the impact of various measures taken to improve the effectiveness of classroom teaching is evaluated on a number of performance measures on students. In ecology, frequencies of certain species living in soil are explained by a number of environmental conditions of the soil. In psychology, perceptual qualities of sounds (colors) evaluated on a number of subjective criteria are predicted from a number of physical attributes of the stimuli.

Redundancy analysis (RA: Van den Wollenberg, 1977) is designed for analyzing a directional relationship between two sets of multivariate data (Lambert, Wildt, and Durand, 1988), and as such offers an attractive technique for analyzing the data arising from the kinds of situations described above. RA extracts a series of components (called redundancy components) from predictor variables in such a way that they are mutually orthogonal and successively account for the maximum variability in criterion variables. It maximizes the proportion of the total sum of squares (SS) in the criterion variables that can be explained by each successive component. The first redundancy component explains a maximum proportion of the total SS in criterion variables. The second component is orthogonal to the first and accounts for the maximum proportion of unexplained SS after the first component is extracted, and so on. Because of its directionality, RA is often regarded as a more useful analytic tool than canonical correlation analysis which analyzes nondirectional relationships between two sets of variables (Lambert et al., 1988). The situation that calls for the former is much more prevalent than the latter in practical data analysis situations. Variants of RA are also known as reduced rank regression (Anderson, 1951) and principal component of instrumental variables (Rao, 1964).

Let X and Y be n by p and n by q matrices of predictor and criterion variables, respectively. (For simplicity, X is assumed to have full column rank throughout this paper.) The model for RA can be written as

$$Y = XB + E, \tag{1}$$

where the weight matrix B ($p \times q$) is assumed to satisfy the rank restiction, $\text{rank}(B) = r \leq \min(p, q)$, and E ($n \times q$) is a matrix of disturbance terms. RA finds the subspace of dimensionality r in $\text{Sp}(X)$ (where $\text{Sp}(X)$ indicates the space spanned by column vectors of X) which is most predictive of Y.

Let F and C denote p by r and q by r matrices, respectively, of full column rank such that $B = FC'$. Then the model for RA can be reparameterized as

$$Y = XFC' + E, \tag{2}$$

where F is called the matrix of weights (the weights applied to X to obtain redundancy components XF), and C is called the matrix of (cross) loadings. Computationally, RA amounts to obtaining the singular value decomposition (SVD) of $P_X Y$, where $P_X = X(X'X)^{-1}X'$ is the orthogonal projector onto $\text{Sp}(X)$. This solution is obtained by first splitting the least squares criterion into two additive terms,

$$\text{SS}(Y - XB) = \text{SS}(Y - X\hat{B}) + \text{SS}(X\hat{B} - XB), \tag{3}$$

where $\hat{B} = (X'X)^{-1}X'Y$ is the LS estimate of B without rank restriction. Since the first term on the right hand side of this criterion is unrelated to B, the whole criterion can be minimized by minimizing the second term, which can be achieved by the singular value decomposition of $X\hat{B} = P_X Y$. Let the SVD of $P_X Y$ be denoted by $P_X Y = U^* D^* V^{*\prime}$, and its reduced rank approximation by UDV'. Matrices U and V are obtained by retaining only those columns of U^* and V^* corresponding to the r largest singular values of $P_X Y$. Matrix D contains those r largest singular values as its diagonal elements, which are extracted from D^*. Then, XF, C and F are obtained by $n^{1/2}U$, $n^{-1/2}VD$, and $n^{1/2}(X'X)^{-1}X'U$, respectively. The matrix of correlations between predictor variables X and redundancy components XF, called the structure matrix (or the matrix of predictor loadings) is obtained by $n^{-1/2}X'U$. (We assume throughout this paper that both X and Y are standardized.)

The data matrices, X and Y, are often accompanied by auxiliary information. For example, subjects (or cases) representing rows of data matrices may have some demographic information such as age, gender, race, education level, and so on. Variables representing columns of data matrices may also have some specific structure or relationship among themselves; for example, the variables may be grouped into several distinct groups. In such situations, it may be desirable to incorporate the additional information in the analysis of relationships between the data sets. Including the external information may provide simpler interpretation of the analysis results. In this paper we propose a method of RA, called generalized constrained redundancy analysis (GCRA), that allows incorporation of external information in RA.

To illustrate, suppose a personnel manager is interested in predicting work performance from aptitude test scores. There are several variables in the aptitude test (e.g., verbal, quantitative) and in the work performance (e.g., performance rating, compensation level, position in organization). It is speculated that both predictor variables (the aptitude test scores) and criterion variables (the work performance scores) are affected by extraneous variables (e.g., the former by gender and age, and the latter by previous job experience, education level, and organizational characteristics). Furthermore, two sets of variables could be classified into several distinct groups in terms of what they measure. By incorporating the extraneous variables into the analysis, GCRA can remove the effects that can be accounted for by the extraneous variables, allowing a better understanding of the relationship between the two sets of variables of interest. It can also impose the group structure among the variables to simplify the relationships among them.

Let G_X (n by a) and H_X (p by b) denote the row and column information matrices for predictor variables (X), and let G_Y (n by c) and H_Y (q by d) denote the same for criterion variables (Y). Takane and Shibayama (1991; also, see Takane and Hunter, 2001) proposed a comprehensive framework for various decompositions of the data matrix Y incorporating its row and column information matrices, G_Y and H_Y. Takane, Yanai, and Hwang (2006) also proposed a comprehensive framework for various decompositions of the orthogonal projector $P_{[X,G_X]}$ onto the range space of $[X, G_X]$, incorporating the column information matrix H_X on X. We then apply RA to any combinations of a term in the decompositions of Y and a term in the decompositions of $P_{[X,G_X]}$. (Technically, this amounts to setting any term in the decomposition of Y as the set of criterion variables, and any term in the decomposition of $P_{[X,G_X]}$ as the orthogonal projector onto the space of predictor variables in ordinary RA.) This allows us to predict some specific aspects of Y from some specific aspects of $[X, G_X]$. Consequently, more focused predictions are possible.

Our current approach is also comprehensive. By spelling out all the terms in the decompositions, we may position a particular analysis undertaken relative to other analyses that could have been undertaken. For example, we can immediately figure out which effects are left out of the analysis as well as the size of the effects that are left out, and may further investigate whether the left-out effects have any significant components, and so on.

2 Method

2.1 Decompositions of Data Matrix

Now let us describe how GCRA works. First, data matrix Y is decomposed into the sum of several matrices according to the external information on rows and columns. If both row and column information matrices on Y are available, the data matrix can be decomposed into the sum of four matrices (Takane and Shibayama, 1991),

$$
\begin{aligned}
Y &= (P_{G_Y} + Q_{G_Y})Y(P_{H_Y} + Q_{H_Y}) && (4) \\
&= P_{G_Y}YP_{H_Y} + P_{G_Y}YQ_{H_Y} + Q_{G_Y}YP_{H_Y} + Q_{G_Y}YQ_{H_Y}, && (5)
\end{aligned}
$$

where $P_{G_Y} = G_Y(G'_YG_Y)^-G'_Y$ (where $(G'_YG_Y)^-$ indicates a generalized inverse (g-inverse) of G'_YG_Y) represents the orthogonal projector onto $\mathrm{Sp}(G_Y)$, $Q_{G_Y} = I - P_{G_Y}$ represents its orthogonal complement, i.e., the orthogonal projector onto $\mathrm{Ker}(G'_Y)$ (where $\mathrm{Ker}(G'_Y)$ indicates the null space of G'_Y), and P_{H_Y} and Q_{H_Y} are similarly defined. (As is well known, these projectors are invariant over the choice of g-inverse $(G'_YG_Y)^-$.) The four terms in decomposition (5) are either column-wise or row-wise orthogonal. Two matrices, A and B, which are either column-wise or row-wise orthogonal are trace-orthogonal (i.e., $\mathrm{tr}(A'B) = \mathrm{tr}(BA') = 0$), so that the sum of squares on the left hand side (i.e., $\mathrm{SS}(Y) = \mathrm{tr}(Y'Y)$) is decomposed into the sum of sums of squares of the four terms on the right hand side. Note that each term on the right can be given a specific interpretation (Takane and Shibayama, 1991). The first term in this decomposition corresponds to the portion of Y that can be explained by both G_Y and H_Y. The second term represents the part of Y explained by G_Y, but not by H_Y, the third term by H_Y, but not by G_Y, and the last term by neither G_Y nor H_Y.

In general, we may also use terms in intermediary decompositions, and consider the following 9 terms as possible sets of criterion variables in RA.

$$
\begin{array}{ll}
[1] & Y \\
[2] & P_{G_Y}Y \\
[3] & Q_{G_Y}Y \\
[4] & YP_{H_Y} \\
[5] & YQ_{H_Y} \\
[6] & P_{G_Y}YP_{H_Y} \\
[7] & P_{G_Y}YQ_{H_Y} \\
[8] & Q_{G_Y}YP_{H_Y} \\
[9] & Q_{G_Y}YQ_{H_Y}
\end{array}
$$

Matrix [1] denotes the unconstrained data matrix. The next two matrices, represented by [2] and [3], are derived when only the row information matrix is incorporated. Matrix [2] indicates the portions of Y that can be explained by its row information matrix, while [3] indicates the residuals after eliminating the effects due to the row information matrix. Matrices [4] and [5], on the other hand, are derived when only the column information matrix is incorporated. The final four matrices ([6] through [9]) are the four terms in the full decomposition, as both row and column information matrices are incorporated. Later we denote the nine terms above by putting subscript Y on the term number (e.g., $[1]_Y$) to distinguish them from those in the decomposition of projector $P_{[X,G_X]}$.

2.2 Decompositions of Orthogonal Projectors

Because orthogonal decompositions of a data matrix do not necessarily entail the corresponding orthogonal decomposition of projectors defined by the data matrix, we directly derive orthogonal decompositions of a projector defined by matrix $[X, G_X]$. Let X, G_X, and H_X be matrices of predictor variables, row information, and column information for X, respectively. We propose two decompositions of $P_{[X,G_X]}$ (Takane, Yanai, and Hwang, 2006). The first one is:

$$P_{[X,G_X]} = P_{G_X} + P_{Q_{G_X}X} \tag{6}$$

$$= P_{P_{G_X}X} + (P_{G_X} - P_{P_{G_X}X}) + P_{Q_{G_X}X} \tag{7}$$

$$= P_{P_{G_X}XH_X} + (P_{P_{G_X}X} - P_{P_{G_X}XH_X}) + (P_{G_X} - P_{P_{G_X}X}) + P_{Q_{G_X}XH_X} + (P_{Q_{G_X}X} - P_{Q_{G_X}XH_X}). \tag{8}$$

In (6), $P_{[X,G_X]}$ is split into P_{G_X} and $P_{Q_{G_X}X}$ by a well-known orthogonal decomposition formula,

$$P_{[X,G_X]} = P_{G_X} + P_{Q_{G_X}X}, \tag{9}$$

which derives from the fact that $\mathrm{Sp}([X, G_X]) = \mathrm{Sp}([Q_{G_X}X, G_X])$, and that G_X and $Q_{G_X}X$ are mutually orthogonal. To derive (7) from (6), P_{G_X} in (6) is further decomposed into $P_{P_{G_X}X}$ and $P_{G_X} - P_{P_{G_X}X}$ by another well-known orthogonal decomposition formula (Yanai and Takane, 1992),

$$P_A = P_{AX} + P_{A(A'A)^- X^*}, \tag{10}$$

where $A = G_X$ and X^* is such that $\mathrm{Ker}(X') = \mathrm{Sp}(X^*)$ and $X^* = A'W$ for some W (i.e., $\mathrm{Sp}(X^*) \subset \mathrm{Sp}(A')$). Note that $P_{G_X} - P_{P_{G_X}X}$ is also an orthogonal projector. This follows from the fact that $\mathrm{Sp}(G_X) \supset \mathrm{Sp}(P_{G_X}X)$. In fact, $P_{G_X} - P_{P_{G_X}X}$ can be re-expressed in the form of a single projector, P_{G_XW}, where W is such that $\mathrm{Sp}(W) = \mathrm{Ker}(X'G_X)$, using (10). (The derivation of (7) from (6) may look trivial because we are simply adding a quantity, which is then subtracted. However, this quantity is not arbitrary but is chosen in such a way that the first and the second terms in (7) are mutually orthogonal.) To derive (8) from (7), both $P_{P_{G_X}X}$ and $P_{Q_{G_X}X}$ are each split into two mutually orthogonal parts, $P_{P_{G_X}X} = P_{P_{G_X}XH_X} + (P_{P_{G_X}X} - P_{P_{G_X}XH_X})$ and $P_{Q_{G_X}X} = P_{Q_{G_X}XH_X} + (P_{Q_{G_X}X} - P_{Q_{G_X}XH_X})$, using (10). Again, the two matrices defined as the differences between two orthogonal projectors above are themselves orthogonal projectors. All five terms in (8) are mutually orthogonal.

The first term in decomposition (8) pertains to the space defined by the projection of $\mathrm{Sp}(XH_X)$ onto $\mathrm{Sp}(P_{G_X})$, the second term to the subspace of $\mathrm{Sp}(P_{G_X}X)$ orthogonal to the first term, and the third term to the subspace of $\mathrm{Sp}(G_X)$ orthogonal to $\mathrm{Sp}(X)$. The fourth term in the decomposition pertains to the space defined by the projection of $\mathrm{Sp}(XH_X)$ onto $\mathrm{Ker}(G'_X)$, and the last term to the subspace of $\mathrm{Sp}(Q_{G_X}X)$ orthogonal to the fourth. A geometric representation of the relationships among these subspaces has been given in Takane et al. (2006). The second term will be null if $\mathrm{rank}(P_{G_X}XH_X) = \mathrm{rank}(P_{G_X}X)$, the third term will be null if $\mathrm{rank}(X'G_X) = \mathrm{rank}(G_X)$ (or equivalently, $\mathrm{Ker}(X') \cap \mathrm{Sp}(G_X) = \{0\}$), and the fifth term will be null if $\mathrm{rank}(Q_{G_X}XH_X) = \mathrm{rank}(Q_{G_X}X)$.

The second decomposition of $P_{[X,G_X]}$ is stated as:

$$P_{[X,G_X]} = P_X + P_{Q_XG_X} \tag{11}$$

$$= P_{XH_X} + P_{XK_X} + P_{Q_XG_X} \tag{12}$$

$$\begin{aligned} &= P_{P_{XH_X}G_X} + (P_{XH_X} - P_{P_{XH_X}G_X}) + P_{P_{XK_X}G_X} \\ &\quad + (P_{XK_X} - P_{P_{XK_X}G_X}) + P_{Q_XG_X}, \end{aligned} \tag{13}$$

where K_X is such that $\mathrm{Ker}(K'_X) = \mathrm{Sp}(X'XH_X)$. In (11), $P_{[X,G_X]}$ is first split into P_X and $P_{Q_XG_X}$ by (9). In (12), P_X in (11) is further decomposed into the sum of P_{XH_X} and P_{XK_X} by (10). In (13), both P_{XH_X} and P_{XK_X} in (12) are each decomposed into two parts, $P_{XH_X} = P_{P_{XH_X}G_X} + (P_{XH_X} - P_{P_{XH_X}G_X})$ and $P_{XK_X} = P_{P_{XK_X}G_X} + (P_{XK_X} - P_{P_{XK_X}G_X})$ by (10). Again, the two matrices defined as the differences between two projectors are also projectors. All five terms in (13) are mutually orthogonal. All the terms in (13) except the last term are in $\mathrm{Sp}(X)$.

The first term in the decomposition (13) pertains to the space defined by the projection of $\mathrm{Sp}(G_X)$ onto $\mathrm{Sp}(XH_X)$, and the second term to the subspace of $\mathrm{Sp}(XH_X)$ orthogonal to the first term. The third term represents the space defined by the projection of $\mathrm{Sp}(G_X)$ onto $\mathrm{Sp}(XK_X)$, the fourth term to the subspace of $\mathrm{Sp}(XK_X)$ orthogonal to the third. The fifth term pertains to the space defined by the projection of $\mathrm{Sp}(G_X)$ onto $\mathrm{Ker}(X')$. This space is orthogonal to $\mathrm{Sp}(X)$. Again the geometric structure and relationships among these subspaces have been given in Takane et al. (2006). The second term of decomposition (13) will be null if $\mathrm{rank}(G'_X XH_X) = \mathrm{rank}(XH_X)$, and the fourth term will be null if $\mathrm{rank}(G'_X XK_X) = \mathrm{rank}(XK_X)$.

As before, we consider terms in intermediary decompositions as well. These amount to the following 17 terms:

[1] $P_{P_{G_X}XH_X}$
[2] $P_{P_{G_X}X} - P_{P_{G_X}XH_X}$
[3] $P_{Q_{G_X}XH_X}$
[4] $P_{Q_{G_X}X} - P_{Q_{G_X}XH_X}$
[5] $P_{G_X} - P_{P_{G_X}X}$
[6] $P_{P_{G_X}X}$
[7] P_{G_X}
[8] $P_{Q_{G_X}X}$
[9] $P_{[X,G_X]}$
[10] P_X
[11] P_{XH_X}
[12] P_{XK_X}
[13] $P_{P_{XH_X}G_X}$
[14] $P_{XH_X} - P_{P_{XH_X}G_X}$
[15] $P_{P_{XK_X}G_X}$
[16] $P_{XK_X} - P_{P_{XK_X}G_X}$
[17] $P_{Q_X G_X}$

Terms [1] through [5] correspond to those in (8), while terms [13] through [17] correspond to those in (13). Terms [6] through [12] are those in the intermediary decompositions. As has been noted previously, some of the terms in decompositions of $P_{[X,G_X]}$ may be null depending on the data sets. When this happens, some of the terms in intermediary decompositions will be identical to some of those in (8) and (13). Note also that we replace G_X and/or H_X with an identity matrix in the case of having no special row and/or column information matrices. Because terms [5] and [17] pertain to $\mathrm{Ker}(X')$, indicating the null space of X, they may not be of much empirical interest. The 17 terms above are more fully explained along with their geometric interpretations in Takane et al. (2006). Again, we later put subscript X on the term number above, as in $[1]_X$, to distinguish them from those in the decomposition of Y.

2.3 Redundancy Analysis with Pairs of the Decomposed Matrices

Once each data set is decomposed by the external information, we combine two terms, one each from decompositions of Y and $P_{[X,G_X]}$, and apply redundancy analysis to the pair to investigate the relationship between them. As a result, the total number of possible analyses amounts to 153 ($= 17 \times 9$). Note, however, that some combinations are more important than others, and some judicious choice of the analyses to be carried out is required.

In Table 1, we summarize some of the most important combinations. Ordinary RA corresponds with RA of the combination of $[10]_X$ and $[1]_Y$, where $[10]_X$ refers to term 10 in the decompositions of $P_{[X,G_X]}$ and $[1]_Y$ refers to term 1 in the decompositions of Y. The bi-partial bi-constraint RA is obtained by the combination of $[3]_X$ and $[8]_Y$. The bi-partial non-constraint RA could be realized by the combination of $[8]_X$ and $[3]_Y$, and so on. If G_X is equal to G_Y, both $[8]_X$, $[1]_Y$ and $[8]_X$, $[3]_Y$ represent the same portion of the relationship between two sets of variables. This is called partial (not bi-partial) RA. (Terminologies such as partial and bi-partial RA are analogous to those used in canonical correlation analysis by Timm and Carlson (1976).)

Insert Table 1 about here

After a constrained RA is conducted, it might be of interest to analyze the portion left out of the analysis. Suppose an RA with $[11]_X$ and $[1]_Y$ (P_{XH_X} and Y) is conducted. Then, it might be of interest to also conduct an RA of $[12]_X$ and $[1]_Y$ (P_{XK_X} and Y).

Each RA involves an SVD of the form SVD(P_AB), where P_A can be any one of the 17 terms in the decompositions of $P_{[X,G_X]}$ and B can be any one of the 9 terms in the decompositions of Y. The computation of the SVD of the above form can be efficiently carried out by the following procedure. Let $A = J_AR'_A$ denote the QR decomposition of A pertaining to Sp(A). Note that $J'_AJ_A = I$ and R'_A is upper triangular. Then, $P_A = J_AJ'_A$. We first calculate SVD(J'_AB). Let this SVD be denoted by $J'_AB = U^*D^*V^{*\prime}$. Then, SVD of $J_AJ'_AB$ is obtained by $J_AJ'_AB = UDV'$ where $U = J_AU^*$, $D = D^*$, and $V = V^*$. Note that matrix J'_AB is much smaller in size than $J_AJ'_AB$. The matrix of redundancy components is obtained by $n^{1/2}U$. The matrix of standardized redundancy weights, F, is obtained by $\text{diag}(R_AR'_A)^{1/2}(R_AR'_A)^-R_AU^*$. (To obtain a unique solution for F in the case of singularity of $R_AR'_A$, $(R_AR'_A)^-$ may be replaced by the Moore-Penrose inverse.) The standardized (cross) loading matrix, C, is obtained by $\text{diag}(B'B)^{-1/2}VD/n$. The structure matrix (the matrix of correlations between A and U) is obtained by $\text{diag}(R_AR'_A)^{-1/2}R_AU^*$.

Although GCRA does not make any distributional assumptions, the Bootstrap method (Efron and Tibshirani, 1998) can provide information about the reliability of parameter estimates. Permutation tests allow us to test the number of significant redundancy components (Takane and Hwang, 2002).

3 An Example of Application

Let us now apply the proposed method to an actual data set. The data were previously analyzed by generalized constrained canonical correlation analysis (Takane et al., 2006). However, as will be clear, RA seems more appropriate for this data set due to the obvious directionality in the relationship between the two sets of variables. More specifically, we would like to predict cancer mortality rates from the intake of various foods. The data on food intake were taken from the Food and Agriculture Organization's statistic archive (FAOSTAT). We have the average daily intake of the following foods per capita in 34 countries: (x_1) alcohol, (x_2) meat, (x_3) fish, (x_4) cereal, (x_5) vegetable, (x_6) milk products, and (x_7) the total calorie intake per day. For the criterion variables, the data on the cancer mortality rates were retrieved from the World Health Organization Statistical Information System (WHOSIS). There are four cancer variables indicating the locus of cancer: (y_1) esophagus, (y_2) stomach, (y_3) pancreas, and (y_4) liver. The effects of the predictor variables on the criterion variables are likely to be mediated by other variables, such as the degree of economic development and the overall health status in these countries. We consider eliminating these effects in predicting the cancer mortality rates from the food variables. This will more faithfully capture the true relationship between the cancer variables and the food variables. To do this, we take G_X to be GDP (gross domestic product) and G_Y to be DALE (disability adjusted life expectancy) and IM (infant mortality rate), and perform GCRA. The information on these auxiliary variables was retrieved from the United Nations' (UN) data archive.

The first RA is the ordinary RA which corresponds to the RA of the combination of $[10]_X$ and $[1]_Y$ (P_X and Y). Permutation tests showed that the largest (first) redundancy component was highly significant ($s_1^2 = 53.60$, $p < .0005$), while the second component was not ($s_2^2 = 12.72$, $p > .141$), where s_i indicates the i^{th} largest singular value. The s_i^2 indicates the sum of squares (SS) explained by component i, and is equal to n (the sample size) times the sum of squares (cross) loadings on the i^{th} component. The first component explains approximately 40% of the total SS of Y (which is $136 = 34 \times 4 = nq = \text{tr}(Y'Y)$, since Y is standardized). Still, it captures the majority (74.1%) of the total redundancy (72.33) between Y and $[X, G_X]$. (See Table 6 below.) Table 2 provides estimates of the weights, the structure vector, and (cross) loadings for the first redundancy component, as well as their standard error estimates obtained by the Bootstrap method. As indicated, the standard errors tend to be large since the sample size is very small. An asterisk indicates that the estimated coefficient is significant at the 5% level, while two asterisks indicate a significance level of 1%. The first redundancy component is significantly correlated with alcohol, meat, cereal, and the total calorie intake at the 1% level. We may call this component high-fat and high-cholesterol diet. This component is also significantly correlated with all cancer variables except y_2 (stomach cancer). The liver cancer has the strongest relationship with this component.

Insert Table 2 about here

The above analysis indicates which variables of X are positively or negatively correlated with the first redundancy component. The redundancy component is strongly positively correlated with alcohol, meat and the total calorie intake, moderately with milk products, and negatively correlated with the other food variables. The food variables could be grouped into two groups by the sign of their correlations with the first redundancy component. The two groups of variables are formally expressed by the following constraint matrix:

$$H_X' = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

We demonstrate that a use of the structural information about the predictor variables like this yields more reliable estimates of parameters. We incorporate the above H_X matrix in the following analysis. This is a semi-constrained (constraints only on X) RA, and represents RA of the combination of $[11]_X$ and $[1]_Y$ (P_{XH_X} and Y). Table 3 summarizes the results. Redundancy components ($s_1^2 = 45.07$, $s_2^2 = 1.16$) obtained in this analysis are uniformly smaller than those obtained in the first analysis due to the imposed constraint. Nevertheless, permutation tests indicate that the first redundancy component is still highly significant ($p < .0005$), while the second component remains non-significant ($p > .602$). Table 3 provides the estimated weights (here and hereafter all the weights presented are standardized weights analogous to the beta coefficients in multiple regression analysis), the structure vector, and cross loadings along with their standard errors estimated by the Bootstrap method. The standard errors tend to be smaller than those obtained from the first analysis, indicating that these coefficients are more reliably estimated. In fact, this analysis yielded more significant estimates of the coefficients at the 1% level than those obtained from the first analysis. The first redundancy component accounts for approximately 62% of the total redundancy. Although the proportion of the total redundancy explained is smaller than that in the first analysis, the first redundancy component is estimated much more reliably because of the external information that has been incorporated. This indicates the importance of incorporating proper constraints to get more reliable components, despite the fact that it may increase the bias slightly. Note that in calculating the structure vector (predictor loadings) in Table 3, correlations are taken between the redundancy component and the original predictor variables, X, rather than between the redundancy component and the constrained predictors, XH_X (or XP_{H_X}).

Insert Table 3 about here

To make sure that the imposed constraint did not miss out any systematic portions of the total predictability, an RA of the complementary part was also conducted. This corresponds to the RA of $[12]_X$ and $[1]_Y$ (P_{XK_X} and Y). Permutation tests indicated that the first redundancy component is not significant ($s_1^2 = 14.12$, $p > .514$). This result confirms that no significant parts of the relationship between X and Y have been left out by imposing the constraint, H_X. As has been demonstrated, investigating complementary parts of a relationship between two sets of variables can be carried out with GCRA in a straightforward manner.

As previously mentioned, it is suspected that the relationship between X and Y is mediated by the degree of economic development and the overall health status in the countries. In the next two analyses, we eliminated the effect due to G_X (GDP) from X and the effect due to G_Y (DALE and IM) from Y. The first one of these two analyses corresponds to an RA of $[8]_X$ and $[3]_Y$ ($P_{Q_{G_X}}X$ and $Q_{G_Y}Y$). The results are summarized in Table 4. Since some portion of the redundancy between X and Y could be explained by G_X and G_Y, the redundancy components ($s_1^2 = 25.84$, $s_2^2 = 10.51$) are substantially smaller than those of the first analysis. However, permutation tests indicated that the first redundancy component was still highly significant ($p < .001$), while the second component was not ($p > .144$). This implies that a substantial amount of redundancy still remains even after the effects due to the extraneous variables were eliminated. By partialing out the portion of variability explained by external information, the standard errors were greater than those obtained from the ordinary redundancy analysis. The structure vector shows that x_1 (alcohol) and x_3 (fish) are correlated with the first redundancy component significantly (at the 5% level). For the cancer variables, y_1 (esophagus) and y_4 (liver) are the only variables significantly correlated with this component (again at the 5% level). The first component explains approximately 36% of the total redundancy.

Insert Table 4 about here

In order to obtain more stable parameter estimates, we incorporated constraints, this time on both X and Y. The following constraint matrix was constructed for Y:

$$H_Y' = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

The constraint matrix H_Y' was based on the sign of correlations between Y variables and the first component in the previous analyses. This analysis corresponds to an RA of $[3]_X$ and $[8]_Y$ ($P_{Q_{G_X}XH_X}$ and $Q_{G_Y}YP_{H_Y}$), bi-partial and bi-constraint RA. The size of redundancy components was moderately decreased ($s_1^2 = 20.25$, $s_2^2 = .60$) due to the imposed constraints. However, permutation tests indicate that the first redundancy component is still highly significant ($p < .0005$), while the second component is not ($p > .422$). The first component explains approximately 28% of the total redundancy. The standard errors of estimated coefficients were much smaller than those obtained from the previous analysis. The first component represents the high-fat, low-fiber diet unaffected by GDP. It correlated positively with x_1 (alcohol), x_2 (meat), x_6 (milk products) and x_7 (total calorie intake), and negatively with x_3 (fish), x_4 (cereal) and x_5 (vegetables). While this was also true in the previous analysis, the constraint on X bore this out more clearly. This component is also highly correlated with the aspects of y_1 (esophagus) and y_4 (liver) cancer unaffected by IM and DALE. Note that the structure vector and standardized cross loadings reported in Table 5 are correlations between redundancy component and the original predictor (X) and criterion variables (Y) rather than their constrained counterparts (XH_X and YH_Y).

Insert Table 5 about here

Table 6 provides part redundancies (Timm and Carlson, 1976) for all 153 combinations. The total redundancy is 72.33 for the combination of $[9]_X$ and $[1]_Y$. In the table those analyses whose results have been reported in the present paper are marked by asterisks. In this way we immediately know where those analyses are positioned in relation to other possible analyses, and the magnitude of the redundancies that were analyzed.

Insert Table 6 about here

Notice that in general we have the following additivity relations on the predictor side: $[6]_X = [1]_X + [2]_X$, $[8]_X = [3]_X + [4]_X$, $[7]_X = [5]_X + [6]_X$, $[9]_X = [7]_X + [8]_X = [10]_X + [17]_X$, $[10]_X = [11]_X + [12]_X$, $[11]_X = [13]_X + [14]_X$, and $[12]_X = [15]_X + [16]_X$, where the equality sign here means that the sum of the part redundancies on the right hand side is equal to the part redundancy on the left hand side. It is readily noticed that $[2]_X$ and $[5]_X$ are null in this particular instance because $\text{rank}(P_{G_X}XH_X) = \text{rank}(P_{G_X}X)$ and $\text{rank}(X'G_X) = \text{rank}(G_X)$, respectively, and consequently these terms are omitted from the table. This also implies that

$[1]_X = [6]_X = [7]_X$. On the criterion side, $[1]_Y = [4]_Y + [5]_Y$, $[2]_Y = [6]_Y + [7]_Y$, and $[3]_Y = [8]_Y + [9]_Y$. However, $[1]_Y = [2]_Y + [3]_Y$, $[4]_Y = [6]_Y + [8]_Y$, or $[5]_Y = [7]_Y + [9]_Y$ does not hold. To see, for example, $[1]_Y = [2]_Y + [3]_Y$ does not hold for $[10]_X$, note that $P_{G_Y} Y$ and $Q_{G_Y} Y$ are no longer orthogonal when they are premultiplied by P_X. There is a way to make all these six equalities hold. Let P^* represent the projector (on the X side) associated with each row of the table. For example, $P^* = P_X$ for $[10]_X$. Then we replace P_{G_Y} by $P_{G_Y/P^*} = G_Y(G_Y' P^* G_Y)^- G_Y' P^*$. However, this implies that different decompositions of Y are involved for different rows of the table, and that the additivity relations on the X side mentioned above no longer hold, except for the first, fourth and fifth columns. Detailed properties of this new decomposition as well as the relationship between P_{G_Y} and P_{G_Y/P^*} are being investigated, but are beyond the scope of the present paper. However, there does not seem to be any decomposition in which the additivity holds in both ways (across rows and columns).

4 Conclusion

In the present article, we proposed GCRA where external information can be directly incorporated in both rows and columns of the two sets of variables in RA. In this method, $P_{[X,G_X]}$ was decomposed into several orthogonal components. The criterion variables were also orthogonally decomposed into several components according to the external information. RA was then applied to a number of combinations of the terms in the decompositions of $P_{[X,G_X]}$ and those in the decompositions of Y. This enabled us to predict specific aspects of Y from specific aspects of $[X, G_X]$. Taken together, we can conduct more comprehensive analyses of what can be predicted of one set of variables from another. Due to the orthogonalities of the decompositions, the total redundancy $\mathrm{tr}(P_{[X,G_X]}Y)$ can be partially decomposed into additive components, each pertaining to a specific portion of the total redundancy between $[X, G_X]$ and Y.

As a future extension we might consider incorporating regularization into GCRA in a manner similar to Takane and Hwang (2006a). Let A and B denote the generic predictor and criterion variables, respectively. Nonregularized RA obtains $\mathrm{SVD}(P_A B)$, where $P_A = A(A'A)^- A'$. In the regularized RA, we obtain SVD of $A(A'A+\lambda I)^{-1}A'B$, where λ represents a regularization parameter. A small positive value of λ is known to lead to estimates of model parameters associated with a smaller MSE, that is, on average closer to true parameter values (Takane and Hwang, 2006b). An optimal value of λ may be determined in such a way that the resultant estimates yield model predictions that cross validate best. Either the bootstrap or the K-fold cross validation method may be used for this purpose.

References

Anderson T.W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, **22**, 327-351.

Efron B. & Tibshirani R. J. (1998). An introduction to the Bootstrap. *CRC Press*, Boca Raton, Florida.

Hunter M. A. & Takane Y. (2002). Constrained principal component analysis: Various applications. *Journal of Educational and Behavioral Statistics*, **27**, 41-81.

Lambert Z. V., Wildt A. R., & Durand R. M. (1988). Redundancy analysis: An alternative to canonical correlation and multivariate multiple regression in exploring interset associations. *Psychological Bulletin*, **104**, 282-289.

Rao C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya A*, **26**, 329-358.

Takane Y. & Hunter M. A. (2001). Constrained principal component analysis: A comprehensive theory. *Applicable Algebra in Engineering, Communication and Computing*, **12**, 391-419.

Takane Y. & Hwang H. (2002). Generalized constrained canonical correlation analysis. *Multivariate Behavioral Research*, **37**, 163-195.

Takane Y. & Hwang H. (2006a). Regularized multiple correspondence analysis. In J. Blasius and M. J. Greenacre (Eds.), *Multiple correspondence analysis and related methods*, (pp. 259-279). London: Chapman and Hall.

Takane Y. & Hwang H. (2006b). Regularized linear and kernel redundancy analysis. A paper submitted for publication.

Takane Y. & Shibayama T. (1991). Principal component analysis with external information on both subjects and variables. *Psychometrika*, **56**, 97-120.

Takane Y., Yanai H., & Hwang H. (2006). An improved method for generalized constrained canonical correlation analysis. *Computational Statistics and Data Analysis*, **50**, 221-241.

Timm N. & Carlson J. (1976). Part and bipartial canonical correlation analysis, *Psychometrika*, **41**, 159-176.

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

Yanai H. & Takane Y. (1992). Canonical correlation analysis with linear constraints. *Linear Algebra and Its Applications*, **176**, 75-89.

Table 1: Representative combinations

	X			
Y	neither	partial	constrained	both
neither	$[10]_X$,$[1]_Y$	$[8]_X$,$[1]_Y$	$[11]_X$,$[1]_Y$	$[3]_X$,$[1]_Y$
partial	$[10]_X$,$[3]_Y$	$[8]_X$,$[3]_Y$	$[11]_X$,$[3]_Y$	$[3]_X$,$[3]_Y$
constrained	$[10]_X$,$[4]_Y$	$[8]_X$,$[4]_Y$	$[11]_X$,$[4]_Y$	$[3]_X$,$[4]_Y$
both	$[10]_X$,$[8]_Y$	$[8]_X$,$[8]_Y$	$[11]_X$,$[8]_Y$	$[3]_X$,$[8]_Y$

Table 2: RA of $[10]_X$,$[1]_Y$: Ordinary (non-partial, non-constrained) RA. (An asterisk indicates a significance at the 5% level, and two asterisks at the 1% level.)

	Variable	Weight	(SE)	Structure	(SE)
x_1	alcohol	.123	(.257)	**.740	(.106)
x_2	meat	.045	(.265)	**.773	(.094)
x_3	fish	-.154	(.175)	-.086	(.215)
x_4	cereal	**-.465	(.178)	**-.698	(.133)
x_5	vegetable	*-.358	(.188)	-.211	(.188)
x_6	milk	.016	(.173)	.228	(.236)
x_7	calorie	**.706	(.185)	**.648	(.123)

	Variable	Loading	(SE)
y_1	esophagus	**.661	(.155)
y_2	stomach	-.316	(.310)
y_3	pancreas	**.594	(.214)
y_4	liver	**.829	(.132)

Table 3: RA of $[11]_X$,$[1]_Y$: Non-partial semi-constrained RA with constraints only on X. (An asterisk indicates a significance at the 5% level, and two asterisks at the 1% level.)

	Variable	Weight	(SE)	Structure	(SE)
x_1	alcohol	**.213	(.023)	**.766	(.060)
x_2	meat	**.213	(.023)	**.807	(.061)
x_3	fish	**-.120	(.039)	-.316	(.176)
x_4	cereal	**-.120	(.039)	**-.548	(.109)
x_5	vegetable	**-.120	(.039)	-.223	(.224)
x_6	milk	**.213	(.023)	**.481	(.164)
x_7	calorie	**.213	(.023)	**.557	(.154)

	Variable	Loading	(SE)
y_1	esophagus	**.670	(.136)
y_2	stomach	-.262	(.257)
y_3	pancreas	**.534	(.226)
y_4	liver	**.724	(.139)

Table 4: RA of $[8]_X$,$[3]_Y$: Bi-partial no-constrained RA. (An asterisk indicates a significance at the 5% level, and two asterisks at the 1% level.)

	Variable	Weight	(SE)	Structure	(SE)
x_1	alcohol	-.069	(.393)	*.557	(.220)
x_2	meat	.007	(.429)	.530	(.258)
x_3	fish	*-.502	(.193)	*-.793	(.218)
x_4	cereal	-.219	(.246)	-.324	(.248)
x_5	vegetable	*-.682	(.288)	-.470	(.224)
x_6	milk	-.014	(.259)	.383	(.245)
x_7	calorie	*.738	(.329)	.345	(.231)

	Variable	Loading	(SE)
y_1	esophagus	*.584	(.188)
y_2	stomach	-.344	(.418)
y_3	pancreas	.474	(.319)
y_4	liver	*.624	(.200)

Table 5: RA of $[3]_X$,$[8]_Y$: Bi-partial bi-constrained RA. (An asterisk indicates a significance at the 5% level, and two asterisks at the 1% level.)

	Variable	Weight	(SE)	Structure	(SE)
x_1	alcohol	**.141	(.047)	**.686	(.115)
x_2	meat	**.141	(.047)	**.594	(.154)
x_3	fish	**-.235	(.056)	**-.732	(.132)
x_4	cereal	**-.235	(.056)	*-.576	(.164)
x_5	vegetable	**-.235	(.056)	*-.502	(.218)
x_6	milk	**.141	(.047)	**.504	(.133)
x_7	calorie	**.141	(.047)	.156	(.258)

	Variable	Loading	(SE)
y_1	esophagus	**.562	(.130)
y_2	stomach	-.317	(.268)
y_3	pancreas	*.370	(.186)
y_4	liver	**.555	(.128)

Table 6: Decompositions of the total redundancy

Term	Y	$[1]_Y$	$[2]_Y$	$[3]_Y$	$[4]_Y$	$[5]_Y$	$[6]_Y$	$[7]_Y$	$[8]_Y$	$[9]_Y$
X	rank	4	2	4	2	2	2	2	2	2
$[1,6,7]_X$	1	27.15	15.18	1.89	25.01	2.14	13.93	1.25	1.76	.14
$[3]_X$	2	28.48	1.85	22.52	27.10	1.37	1.76	.09	*20.84	1.68
$[4]_X$	5	16.70	4.45	17.07	13.98	2.73	4.29	.19	14.36	2.70
$[8]_X$	7	45.18	6.31	*39.59	41.08	4.10	6.03	.28	35.21	4.38
$[9]_X$	8	72.33	21.50	41.48	66.10	6.24	19.96	1.53	36.96	4.51
$[10]_X$	7	*69.66	19.87	40.38	64.86	4.79	18.60	1.26	36.90	3.48
$[11]_X$	2	*46.23	8.36	23.48	45.26	.97	7.83	.53	22.41	1.07
$[12]_X$	5	*23.43	11.51	16.89	19.61	3.82	10.78	.73	14.48	2.41
$[13]_X$	1	36.99	8.15	11.77	36.10	.89	7.66	.50	10.70	1.06
$[14]_X$	1	9.24	.21	11.72	9.20	.08	.17	.04	11.71	.01
$[15]_X$	1	7.55	8.33	.79	6.10	1.46	7.77	.55	.58	.22
$[16]_X$	4	15.87	3.18	16.10	13.51	2.36	3.00	.17	13.91	2.19
$[17]_X$	1	2.68	1.63	1.10	1.23	1.45	1.36	.27	.07	1.03

* indicates the analyses undertaken