An Improved Method for Generalized Constrained Canonical Correlation Analysis ⋆

Yoshio Takane *

McGill University, Department of Psychology, 1205 Dr. Penfield Avenue, Montreal, QC, H3A 1B1, Canada

Haruo Yanai

National Center for University Entrance Examinations, Research Division, 2-19-23 Komaba, Meguro-ku, Tokyo, 153-8501, Japan

Heungsun Hwang

HEC Montreal, Department of Marketing, 3000 Chemin de la Côte Ste Catherine, Montréal, Québec, H3T 2A7, Canada

Abstract

An improved method for generalized constrained canonical correlation analysis (GCCANO) is proposed. In the original GCCANO, data matrices were first decomposed into the sum of several matrices according to some external information on rows and columns of the data matrices. Decomposed matrices were then subjected to canonical correlation analysis (CANO). However, orthogonal decompositions of data matrices do not necessarily entail orthogonal decompositions of projectors defined by the data matrices. This latter property is crucial in additive partitionings of the total association between two sets of variables. Consequently, no additive partitionings of the total association was possible in the original GCCANO. In this paper two orthogonal decompositions of projectors were proposed that allow additive partitionings of the total association. Terms in the decompositions have straightforward interpretations. An improved method for GCCANO is developed based on the decompositions, while preserving the most important features of the original method. An example is given to illustrate the proposed method.

Key words: Row and column information matrices, Total association, Part associations, Constrained canonical correlation analysis (Constrained CANO), Partial CANO, The Bootstrap method, , Permutation tests

1 Introduction

Canonical correlation analysis (CANO) is often used to investigate the relationship between two sets of variables. Let X be an n (cases) by p_X (variables) matrix, and let Y be an n by p_Y matrix. Both matrices are assumed to be columnwise centered. CANO between them amounts to the generalized singular value decomposition (GSVD) of $(X'X)^{-}X'Y(Y'Y)^{-}$ with metric matrices, $X'X$ and $Y'Y$, where $(X'X)^{-}$ and $(Y'Y)^{-}$ are generalized inverses of $X'X$ and $Y'Y$. It can also be formulated as the singular value decomposition (SVD; GSVD with identity metrics) of $P_X P_Y$, where $P_X = X(X'X)^{-}X'$ and $P_Y = Y(Y'Y)^{-}Y'$ are orthogonal projectors onto the spaces spanned by column vectors of X and Y, respectively. The former obtains (canonical) weights to be applied to X and Y to derive canonical variates (scores), while the latter SVD obtains canonical variates (scores) directly. Let the GSVD of $(X'X)^{-}X'Y(Y'Y)^{-}$ with metrics $X'X$ and $Y'Y$ be denoted by $M^*D^*V^{*\prime}$, and let the SVD of $P_X P_Y$ be denoted by MDV'. Then, they are related by $U = XU^*$, $V = YV^*$, and $D = D^*$, or $U^* = (X'X)^{-}X'U$, $V^* = (Y'Y)^{-}Y'V$, and $D^* = D$ (e.g., Takane & Hunter, 2001). Throughout this paper we primarily use the latter formulation. Canonical variates are invariant over specific bases vectors used to characterize the relevant subspaces, while canonical weights are not. However, as shown above one can easily transform solutions in one form to the other. (See also Section 3.)

The total association between X and Y (the sum of squared canonical correlation coefficients) is given by $\mathrm{tr}(P_X P_Y)$, which is equal to $\mathrm{tr}(D^2) = \mathrm{tr}(D^{*2})$ (the sum of squared singular values of $P_X P_Y$, which in turn is equal to the sum of generalized singular values of $(X'X)^{-}X'Y(Y'Y)^{-}$ with metrics $X'X$ and $Y'Y$). This implies that CANO is a technique that decomposes the total association between two sets of variables into the sum of part associations (squared canonical correlations), each purported by a pair of canonical variates. Pairs of canonical variates are orthogonal across pairs, representing non-overlapping portions of the total association between the two data sets. The total association, as defined above, is symmetric (X and Y can be exchanged), and it is invariant over the linear transformations of X and Y of the form, $X^* = XA$ and $Y^* = YB$, where A and B are square nonsingular matrices of orders, p_X and p_Y, respectively. The total association is $s = \min(\mathrm{rank}(X), \mathrm{rank}(Y))$ times the average association called generalized coefficient of determination (G.C.D.) introduced by Yanai (1974).

⋆ The work reported in this paper has been supported by grant A6394 from the Natural Sciences and Engineering Research Council of Canada to the first author.

* Corresponding author (tel: 514-398-6125; fax: 514-398-4896).

Email addresses: takane@takane2.psych.mcgill.ca (Yoshio Takane), yanai@rd.dnc.ac.jp (Haruo Yanai), heungsun.hwang@hec.ca (Heungsun Hwang).

It seems customary in psychometrics to call symmetric measures of the relationship "association", and asymmetric ones "redundancy" (Cramer and Nicewander, 1979). Along the latter line, Stewart and Love (1968) introduced a redundancy index defined as $\mathrm{tr}(Y'P_XY)/\mathrm{tr}(Y'Y)$ or $\mathrm{tr}(X'P_YX)/\mathrm{tr}(X'X)$ depending on whether Y is the criterion variables (the former) or X is the criterion variables (the latter). The former reduces to $\mathrm{tr}(Y'P_XY)/p_Y$ (the average squared multiple correlation for predicting Y from X) if Y is standardized, and the latter to $\mathrm{tr}(X'P_YX)/p_X$ if X is standardized. The redundancy indices are asymmetric and not invariant over the linear transformations of the criterion variables. They indicate the average predictability of the criterion variables from the predictor set. Redundancy analysis (Van den Wollenberg, 1977) is designed to maximize the proportion of the total variance in the criterion set that can be explained by each successive redundancy component. See Lambert, Wildt, and Durand (1988) for an extensive discussion on relative merits and demerits of CANO and redundancy analysis for analyzing the relationship between two sets of variables.

The data matrices, X and Y, are often accompanied by auxiliary information. For example, subjects (or "cases") representing rows of the data matrices may have some demographic information. Variables representing columns of the data matrices may also have some specific structure or relationships among themselves. In such situations, it may be desirable to incorporate the additional information in the analysis of the relationship between the data sets. Additional structures supplied by the external information may provide simpler interpretations of the analysis results.

For illustration, suppose an investigator is interested in finding the relationship between intake of various kinds of food and susceptibility to various kinds of cancer. She finds relevant statistics reporting supplies of various food categories per capita in different countries of the world. She also finds information on the mortality rates by various kinds of cancer in these countries. She plans to apply CANO to investigate the relationship between the two sets of variables. However, she also suspects that part of the relationship between the two sets of variables is mediated by other variables such as the extent of economic development and the overall health status in these countries. After all, what people can afford to eat depends on how wealthy they are, and the chance of dying by cancer depends on how long people tend to live in these countries. In analyzing the relationship between the food variables and the cancer variables the investigator wishes to take into account the effects of these extraneous variables. This allows her to focus on more intrinsic aspects of the relationship between the two sets of variables. Furthermore, the food variables may be classified into several groups according to their nutritional profiles, and similarly various kinds of cancer may be grouped into several categories according to the proximity of their loci. This type of information may also be incorporated in the analysis of the relationship between the two sets of variables. (In

the example section, data arising from a similar situation will be analyzed by the proposed method.)

Generalized constrained CANO (GCCANO; Takane & Hwang, 2002) has been developed with this kind of situations in mind. (The word "generalized" in CANO is often used to refer to multiple-set CANO. However, in this paper we use the term to refer to a generalization of constrained CANO that can incorporate constraints on both rows and columns of a data matrix.) It allows CANO between two sets of variables incorporating external information on both the cases and the variables in the data sets. In GCCANO, data matrices are first decomposed into the sum of several matrices according to the external information on rows (corresponding to the cases) and columns (representing the variables) of the data matrices. Decomposed matrices are then subjected to CANO. In this way, we can look at the relationships between two data sets from diverse perspectives, relating a variety of pairs of matrices supplied by the external information.

Let G $(n \times q)$ and H $(p \times r)$ denote respectively the row and the column information matrices on X. Since similar decompositions are applied to both X and Y, only those for one, say X, need to be discussed in detail. Consequently, we do not distinguish G and H for X and those for Y until necessary (Section 3). Let $P_G = G(G'G)^{-}G'$ denote the orthogonal projector onto $\mathrm{Sp}(G)$ (the range space of G), and let $Q_G = I - P_G$ denote its orthogonal complement. Let $P_{H/X'X} = H(HX'XH)^{-}H'X'X$ denote the orthogonal projector onto $\mathrm{Sp}(H)$ in metric $X'X$, and let $Q_{H/X'X} = I - P_{H/X'X}$ be its orthogonal complement. The original GCCANO (Takane & Hwang, 2002) used the following decomposition of X (Takane & Shibayama, 1991):

$$\begin{aligned} X &= P_G X P_{H/X'X} + P_G X Q_{H/X'X} + Q_G X P_{H/X'X} \\ &\qquad + Q_G X Q_{H/X'X} \qquad (1) \\ &= P_G P_{XH} X + P_G Q_{XH} X + Q_G P_{XH} X + Q_G Q_{XH} X. \qquad (2) \end{aligned}$$

The two expressions of the above decomposition are term by term equal. The first term in the decomposition represents the portion of X that can be explained by both G and H, the second term to the portion that can be explained by G but not by H, the third term explained by H but not by G, and the last term by neither G nor H. An analogous decomposition was also applied to Y, and by combining the two decompositions, one for X and the other for Y, various kinds of CANO were devised, including constrained CANO (Yanai & Takane, 1992) and partial CANO (Timm & Carlson, 1976). Any term in the decomposition of X and that of Y can be paired, and CANO can be applied between them.

Orthogonal decompositions of data matrices, however, does not necessarily lead to the corresponding decomposition of projectors defined by the data

matrices. This is true even when the terms in the decompositions of the data matrices are columnwise orthogonal. This can be easily understood from the following observation: Let $A = B + C$ be an orthogonal decomposition of A (i.e., $B'C = 0$). In general, however, $P_A \neq P_B + P_C$, where P's are the orthogonal projectors defined by matrices A, B and C. This should not be confused with the situation in which $A = [B, C]$, and $B'C = 0$. In this case, $P_A = P_B + P_C$ indeed holds. (In fact, this latter relationship will be extensively used in this paper in the derivations of orthogonal decompositions of projectors. See two Lemmas below.) This means that the above decomposition of the data matrix does not entail the corresponding orthogonal decomposition of projectors, or that of the total association between two data sets. This in turn implies that in the original GCCANO we may inadvertently be analyzing the same portions of the total relationship over and over again. In this paper, we propose two orthogonal decompositions of projectors that allow additive partitionings of the total association. This guarantees that we only analyze non-overlapping portions of the relationship between two sets of variables (if we so wish), and when we intentionally analyze overlapping portions, we know exactly which portions are overlapping. Furthermore, terms in the proposed decompositions have simple straightforward interpretations. We develop an improved method for GCCANO based on the new decompositions and apply the method to an example data set.

As alluded to earlier, CANO decomposes the total association between two sets of variables into additive components. When external information is available, we may first decompose the total association according to the external information, and then apply CANO to each part. This notion of partitioning the total association is analogous to (and as important as) the partitionings of the sum of squares (SS) in ANOVA, in which the total variability in the data is decomposed into additive components that can be attributed to distinct sources. The proposed decompositions offer a comprehensive framework for the decompositions of the total association into non-overlapping portions, in which any kinds of linear CANO (both existing and those yet to be explored) can be placed in relation to other CANO's that might have been applied.

2 Two New Decompositions of Orthogonal Projectors

In this section, we derive two orthogonal decompositions of $P_{[X,G]}$, orthogonal projector defined by matrix $[X, G]$, obtained by juxtaposing X and G side by side. While both of these decompositions can be derived by combinations of some known decompositions (Lemmas 1 and 2 below), they possess a property particularly attractive in the context of canonical correlation analysis (CANO).

Throughout this paper, $\mathrm{Sp}(Z)$ indicates the range space of Z, and $\mathrm{Ker}(Z)$ indicates the null space of Z. As before, we use P_Z to indicate the orthogonal projector onto $\mathrm{Sp}(Z)$, and $Q_Z = I - P_Z$ to indicate its orthogonal complement, i.e., the orthogonal projector onto $\mathrm{Ker}(Z')$. The two projectors are mutually orthogonal in the identity metric (i.e., metric I). We use $P_{Z/M}$ and $Q_{Z/M}$ to indicate orthogonal projectors in metric M. Metric matrix M is assumed to be non-negative definite (*nnd*). However, for $P_{Z/M} = Z(Z'MZ)^{-}Z'M$ and $Q_{Z/M} = I - P_{Z/M}$ to be projectors (onto $\mathrm{Sp}(Z)$ along $\mathrm{Ker}(Z'M)$, and onto $\mathrm{Ker}(Z'M)$ along $\mathrm{Sp}(Z)$, respectively) for any choice of g-inverse, $(Z'MZ)^{-}$, the following rank condition, $\mathrm{rank}(MZ) = \mathrm{rank}(Z)$, must hold. This condition is automatically satisfied if M is positive definite. It may fail, however, if M is only positive semi-definite, in which case we may use a reflexive g-inverse for $(Z'MZ)^{-}$. This always ensures that $P_{Z/M}$ and $Q_{Z/M}$ are projectors (onto $\mathrm{Sp}(P_{Z/M})$ along $\mathrm{Ker}(P_{Z/M})$, and onto $\mathrm{Ker}(P_{Z/M})$ along $\mathrm{Sp}(P_{Z/M})$, respectively, because in this case $\mathrm{Sp}(P_{Z/M}) \subset \mathrm{Sp}(Z)$, and $\mathrm{Ker}(Z'M) \subset \mathrm{Ker}(P_{Z/M})$.)

The following decompositions are well known (e.g., Rao & Yanai, 1979) and are useful in deriving the proposed decompositions.

Lemma 1.

Let X and G be as introduced above. Then,

$$P_{[X,G]} = P_G + P_{Q_G X} \tag{3}$$
$$= P_X + P_{Q_X G}, \tag{4}$$

where $P_{Q_G X}$ and $P_{Q_X G}$ denote the orthogonal projectors defined by $Q_G X$ and $Q_X G$, respectively. The two terms in each of the above decompositions are mutually orthogonal.

Decomposition (3) splits $\mathrm{Sp}([X, G])$ into $\mathrm{Sp}(G)$ and $\mathrm{Sp}(Q_G X)$ (the subspace obtained by projecting $\mathrm{Sp}(X)$ onto $\mathrm{Sp}(Q_G) = \mathrm{Ker}(G')$). Likewise, (4) decomposes the same space into $\mathrm{Sp}(X)$ and $\mathrm{Sp}(Q_X G)$ (the subspace obtained by projecting $\mathrm{Sp}(G)$ onto $\mathrm{Sp}(Q_X) = \mathrm{Ker}(X')$). If X and G are orthogonal to begin with, we have a unique decomposition: $P_{[X,G]} = P_X + P_G$, that is, $Q_G X = X$ and $Q_X G = G$. However, X and G are usually not orthogonal. To "orthogonalize" them, either X is projected onto $\mathrm{Sp}(Q_G)$ or G is projected onto $\mathrm{Sp}(Q_X)$. We then have $G'Q_G X = 0$ and $X'Q_X G = 0$, and $\mathrm{Sp}([X, G]) = \mathrm{Sp}([Q_G X, G]) = \mathrm{Sp}([X, Q_X G])$. We use the simple decomposition formula given above for two orthogonal matrices to obtain $P_{[X,G]} = P_{[Q_G X,G]} = P_G + P_{Q_G X}$, and $P_{[X,G]} = P_{[X,Q_X G]} = P_X + P_{Q_X G}$.

Lemma 2.

Let X and H be as introduced earlier. Let K be a matrix such that $\mathrm{Sp}(K) =$

$\mathrm{Ker}(H'X'X)$. Then,

$$P_X = P_{XH} + P_{XK}. \tag{5}$$

The two terms on the right hand side of (5) are mutually orthogonal.

Proof. It is obvious that XK is columnwise orthogonal to XH. It remains to be seen that $\mathrm{rank}(XK) = \mathrm{rank}(X) - \mathrm{rank}(XH)$. From Corollary 6.2 (Eq. 3.14) of Marsaglia and Styan (1974), we have

$$\mathrm{rank}(XK) = \mathrm{rank}(X) - \dim(\mathrm{Sp}(X') \cap \mathrm{Ker}(K')). \tag{6}$$

We also see $\mathrm{Ker}(K') = \mathrm{Sp}(X'XH)$, so that the space in the argument of dim is equal to $\mathrm{Sp}(X'XH)$. We thus have $\dim(\mathrm{Sp}(X'XH)) = \dim(\mathrm{Sp}(XH)) = \mathrm{rank}(XH)$. One possible form of K is $K = Q_{X'XH} = I - X'XH(H'(X'X)^2 H)^- H'X'X$. QED.

Decomposition (5) splits $\mathrm{Sp}(X)$ into two orthogonal subspaces, $\mathrm{Sp}(XH)$ and $\mathrm{Sp}(XK)$. Lemma 2 generalizes Theorem 2.1 of Yanai and Takane (1992; see also Lemma 3(v) of Takane and Yanai (1999)). In these papers, XK is parameterized as $X^*\tilde{H}$, where $X^* = X(X'X)^-$ is a matrix of dual bases of X, and $\tilde{H} = X'XK$, and hence $\mathrm{Sp}(\tilde{H}) = \mathrm{Ker}(H')$. This parameterization was motivated by the following consideration: Let W_o represent the weight matrix applied to X. Matrix H imposes constraints on W_o by reparameterizing it by $W_o = HW_r$. Matrix $\tilde{H}$, on the other hand, specifies the same constraints in the form of $\tilde{H}'W_o = 0$ (Takane, Yanai, & Mayekawa, 1991).

We now present the first decomposition of $P_{[X,G]}$.

Theorem 1. Decomposition (A) of $P_{[X,G]}$:

Let X, G, and H be as introduced earlier. Further, let A, B, and W be such that

$$\mathrm{Sp}(A) = \mathrm{Ker}(H'X'P_GX), \tag{7}$$

$$\mathrm{Sp}(B) = \mathrm{Ker}(H'X'Q_GX), \tag{8}$$

and

$$\mathrm{Sp}(W) = \mathrm{Ker}(X'G). \tag{9}$$

Then, the following decomposition holds:

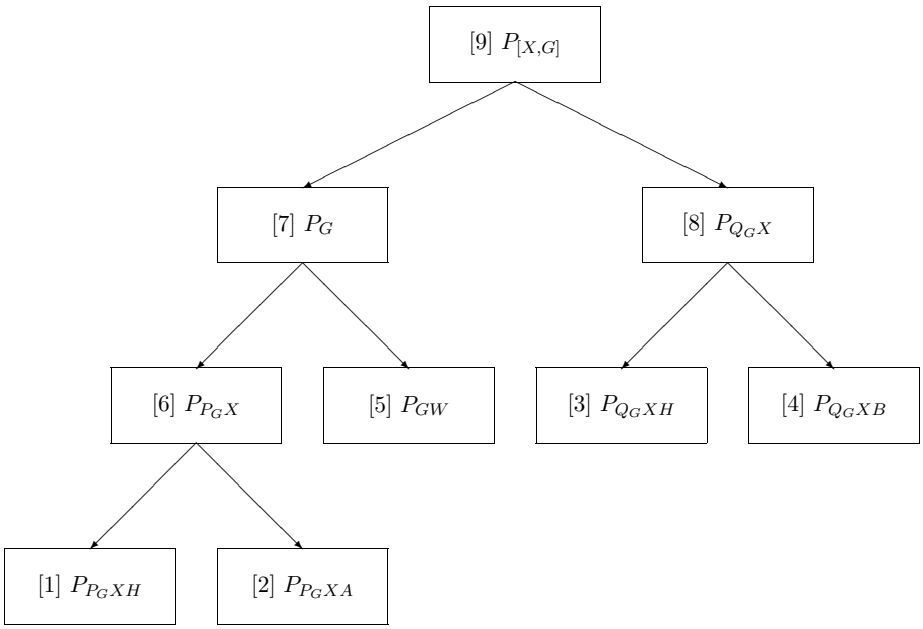

Fig. 1. The derivation of Decomposition (A). Numbers in square brackets indicate term numbers used in the improved method of GCCANO.

$$P_{[X,G]} = P_{P_GXH} + (P_{P_GX} - P_{P_GXH}) + P_{Q_GXH} + (P_{Q_GX} - P_{Q_GXH}) + (P_G - P_{P_GX}) \quad (10)$$

$$= P_{P_GXH} + P_{P_GXA} + P_{Q_GXH} + P_{Q_GXB} + P_{GW}. \quad (11)$$

The five terms in the above decomposition are mutually orthogonal, and the two expressions of the decomposition are term by term equal.

Proof. Orthogonalities among the five terms in the decomposition can be readily seen by tracing its derivation (see Figure 1). Projector $P_{[X,G]}$ is first split into P_G and P_{Q_GX} using (3). Then, P_G is split into P_{P_GX} and $P_G - P_{P_GX} = P_{GW}$ using (9) and Lemma 2. (Set $X = G$, $H = (G'G)^-G'X$, and $K = W$ in (5), and note that $\mathrm{Sp}(W) = \mathrm{Ker}(X'G(G'G)^-G'G) = \mathrm{Ker}(X'G)$.) Finally, P_{P_GX} is split into P_{P_GXH} and $P_{P_GX} - P_{P_GXH} = P_{P_GXA}$ using (7) and Lemma 2 (set $X = P_GX$, and $K = A$ in (5)), and P_{Q_GX} is split into P_{Q_GXH} and $P_{Q_GX} - P_{Q_GXH} = P_{Q_GXB}$ using (8) and Lemma 2 (set $X = Q_GX$, and $K = B$ in (5)). All of these decompositions are orthogonal, so that the resulting terms are all mutually orthogonal. The successive decompositions of projectors described above to derive Decomposition (A) is depicted in Figure 1. QED.

The first term in the above decomposition pertains to the space defined by the projection of $\mathrm{Sp}(XH)$ onto $\mathrm{Sp}(G)$ (the space in $\mathrm{Sp}(G)$ correlated with $\mathrm{Sp}(XH)$), and the second term to the portion of $\mathrm{Sp}(P_GX)$ (the space defined by the projection of $\mathrm{Sp}(X)$ onto $\mathrm{Sp}(G)$) orthogonal to the first term (i.e., the space in $\mathrm{Sp}(G)$ related to $\mathrm{Sp}(X)$ but orthogonal to $\mathrm{Sp}(XH)$). These

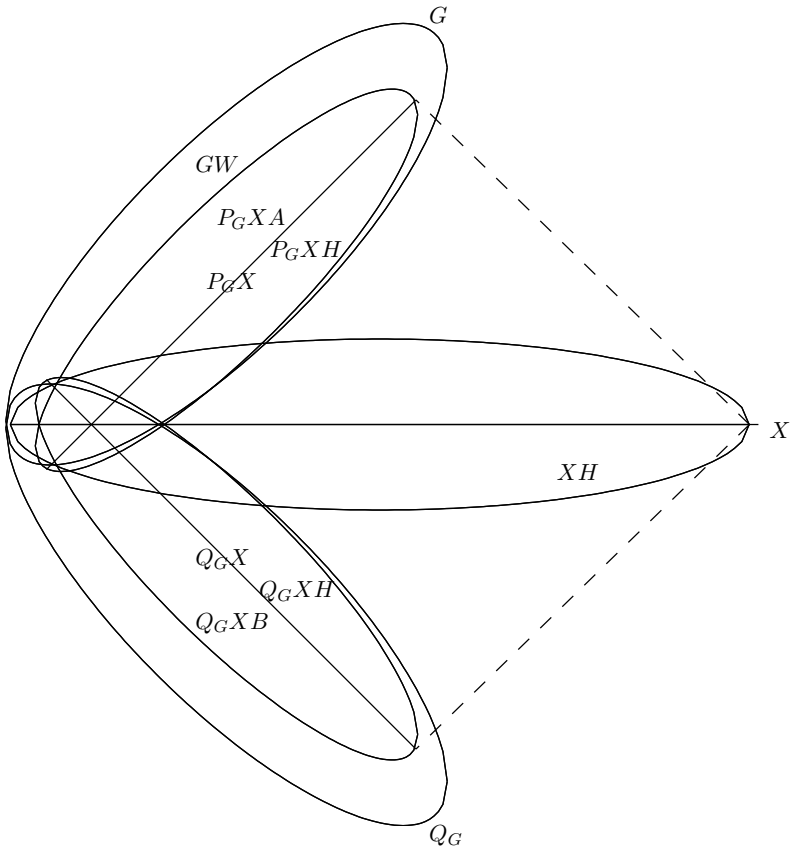

Fig. 2. The decomposition of $\mathrm{Sp}([X, G])$ corresponding to Decomposition (A) of $P_{[X,G]}$.

subspaces are both in $\mathrm{Sp}(G)$. One possible form of A is: $A = Q_{X'P_GXH} = I - X'P_GXH(H'(X'P_GX)^2H)^-H'X'P_GX$. The third term in the decomposition pertains to the space defined by the projection of $\mathrm{Sp}(XH)$ onto $\mathrm{Sp}(Q_G) = \mathrm{Ker}(G')$ (the subspace of $\mathrm{Sp}(Q_G)$ correlated with $\mathrm{Sp}(XH)$), and the fourth term to the portion of $\mathrm{Sp}(Q_GX)$ (the space defined by the projection of $\mathrm{Sp}(X)$ onto $\mathrm{Ker}(G')$) orthogonal to the third (i.e., the subspace of $\mathrm{Sp}(Q_G)$ related to $\mathrm{Sp}(X)$ but orthogonal to $\mathrm{Sp}(XH)$). These two subspaces are both in $\mathrm{Sp}(Q_G)$. One form of B is: $B = Q_{X'Q_GXH} = I - X'Q_GXH(H'(X'Q_GX)^2H)^-H'XQ_GX$. The fifth term represents the space in $\mathrm{Sp}(G)$ orthogonal to $\mathrm{Sp}(X)$ (Rao, 1964, section 11). One form of W is: $W = Q_{G'X} = I - G'X(X'GG'X)^-X'G$. Note that the second term in the decomposition will be null if $\mathrm{rank}(P_GXH) = \mathrm{rank}(P_GX)$, the fourth term will be null if $\mathrm{rank}(Q_GXH) = \mathrm{rank}(Q_GX)$, and the fifth term will be null if $\mathrm{rank}(X'G) = \mathrm{rank}(G)$.

The relationships among the various subspaces described above are diagrammatically depicted in Figure 2. To avoid cluttering of symbols, we omitted the symbol Sp and the parentheses enclosing matrices. Thus, for example, X in the figure indicates $\mathrm{Sp}(X)$. Subspaces represented by distinct regions within ellipses (marked by lines or by inscribed ellipses) are assumed to be

orthogonal. Subspaces belonging to different ellipses are assumed to subtend the angles as shown in the figure (either orthogonal or nonorthogonal). The only exception to this rule is the relationship between $\mathrm{Sp}(GW)$ and $\mathrm{Sp}(X)$ which are orthogonal despite their look.

In Decomposition (A) of $P_{[X,G]}$, the subspaces generated by the first four terms in the decomposition are either in $\mathrm{Sp}(G)$ or in $\mathrm{Ker}(G') = \mathrm{Sp}(Q_G)$. Neither of them usually reside in $\mathrm{Sp}(X)$. The projections took X out of $\mathrm{Sp}(X)$. In the second decomposition we propose, the subspaces all remain in $\mathrm{Sp}(X)$ except the last term.

Theorem 2. Decomposition (B) of $P_{[X,G]}$:

Let X, G, and H be as defined in Theorem 1, and let K be as defined in Lemma 2. Further, let U and V be such that

$$\mathrm{Sp}(U) = \mathrm{Ker}(G'XH), \tag{12}$$

and

$$\mathrm{Sp}(V) = \mathrm{Ker}(G'XK). \tag{13}$$

Then, the following decomposition holds:

$$\begin{aligned} P_{[X,G]} &= P_{XH}P_{G/P_{XH}} + P_{XH}Q_{G/P_{XH}} + P_{XK}P_{G/P_{XK}} \\ &\qquad + P_{XK}Q_{G/P_{XK}} + P_{Q_XG} \qquad (14) \\ &= P_{P_{XH}G} + (P_{XH} - P_{P_{XH}G}) + P_{P_{XK}G} \\ &\qquad + (P_{XK} - P_{P_{XK}G}) + P_{Q_XG} \qquad (15) \\ &= P_{P_{XH}G} + P_{XHU} + P_{P_{XK}G} + P_{XKV} + P_{Q_XG}. \qquad (16) \end{aligned}$$

The five terms in the above decomposition are mutually orthogonal, and the three expressions of the decomposition are term by term equal.

Proof. Orthogonalities among the five terms in the above decomposition can be easily seen by tracing its derivation. (See Figure 3.) Projector $P_{[X,G]}$ is first split into P_X and P_{Q_XG} (the last term) by (4). Then, P_X is split into P_{XH} and P_{XK} by Lemma 2. Then, G is projected onto both $\mathrm{Sp}(XH)$ and $\mathrm{Sp}(XK)$. The former splits P_{XH} into $P_{XH}P_{G/P_{XH}} = P_{P_{XH}G}$ and $P_{XH}Q_{G/P_{XH}} = P_{XH} - P_{P_{XH}G} = P_{XHU}$, which are orthogonal by (12) and Lemma 2. (Set X, H, and K in (5) equal to XH, $(H'X'XH)^-H'X'G$, and U, respectively, and note that $\mathrm{Sp}(U) = \mathrm{Ker}(G'XH(H'X'XH)^-H'X'XH) = \mathrm{Ker}(G'XH)$.) The latter splits P_{XK} into $P_{XK}P_{G/P_{XK}} = P_{P_{XK}G}$ and $P_{XK}Q_{G/P_{XK}} = P_{XK} - P_{P_{XK}G} = P_{XKV}$, which can again be shown to be orthogonal by (13) and Lemma 2. (Set X, H, and K in (5) equal to XK, $(K'X'XK)^-K'X'G$, and V, respectively, and note

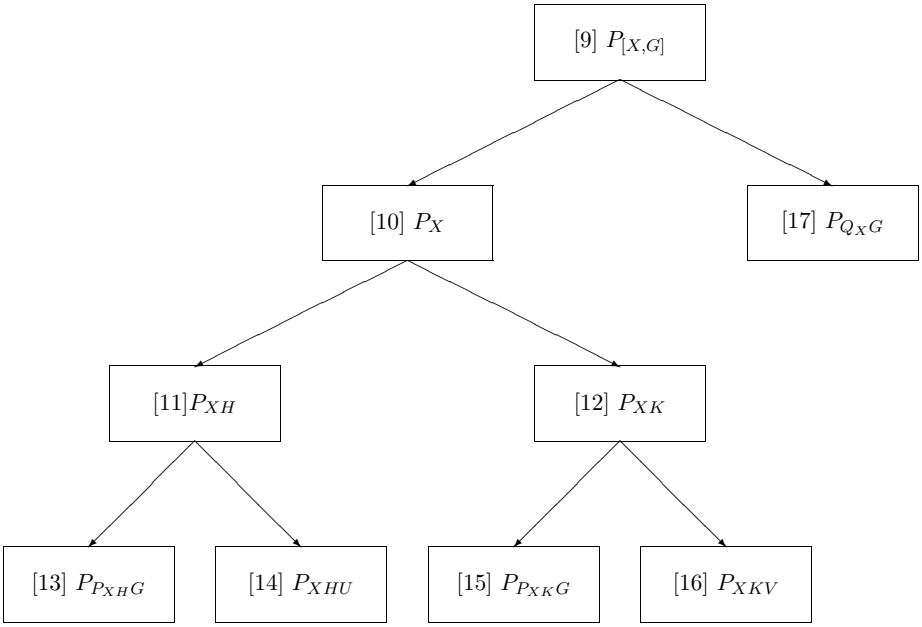

Fig. 3. The derivation of Decomposition (B). Numbers in square brackets indicate term numbers used in the improved method of GCCANO.

that $\mathrm{Sp}(V) = \mathrm{Ker}(G'XK(K'X'XK)^{-}K'X'XK) = \mathrm{Ker}(G'XK)$.) The successive decompositions of projectors described above leading to Decomposition (B) are depicted in Figure 3. QED.

The first term in the above decomposition pertains to the space defined by the projection of $\mathrm{Sp}(G)$ onto $\mathrm{Sp}(XH)$ (the space in $\mathrm{Sp}(XH)$ correlated with $\mathrm{Sp}(G)$), and the second term to the portion of $\mathrm{Sp}(XH)$ orthogonal to the first term (the space in $\mathrm{Sp}(XH)$ orthogonal to $\mathrm{Sp}(G)$). These spaces are both in $\mathrm{Sp}(XH)$. One form of U is: $U = Q_{H'X'G} = I - H'X'G(G'XHH'X'G)^{-}G'XH$. The third term pertains to the space defined by the projection of $\mathrm{Sp}(G)$ onto $\mathrm{Sp}(XK)$ (the space in $\mathrm{Sp}(XK)$ correlated with $\mathrm{Sp}(G)$), and the fourth term to the portions of $\mathrm{Sp}(XK)$ orthogonal to the third (the space in $\mathrm{Sp}(XK)$ orthogonal to $\mathrm{Sp}(G)$). Both of these spaces are in $\mathrm{Sp}(XK)$. One form of V is: $V = Q_{K'X'G} = I - K'X'G(G'XKK'X'G)^{-}G'XK$. (The second and the fourth terms are orthogonal to $\mathrm{Sp}(G)$ and are again special cases of the decomposition given by Rao (1964, section 11).) All these four subspaces are in $\mathrm{Sp}(X)$. The fifth term, on the other hand, pertains to the space defined by the projection of $\mathrm{Sp}(G)$ onto $\mathrm{Ker}(X')$ (the subspace in $\mathrm{Sp}(Q_X) = \mathrm{Ker}(X')$ related to $\mathrm{Sp}(G)$), and hence is orthogonal to $\mathrm{Sp}(X)$. The second term will be null if $\mathrm{rank}(G'XH) = \mathrm{rank}(XH)$, and the fourth term will be null if $\mathrm{rank}(G'XK) = \mathrm{rank}(XK)$.

The relationships among the various subspaces described above and associated with the terms in Decomposition (B) are illustrated in Figure 4. Again

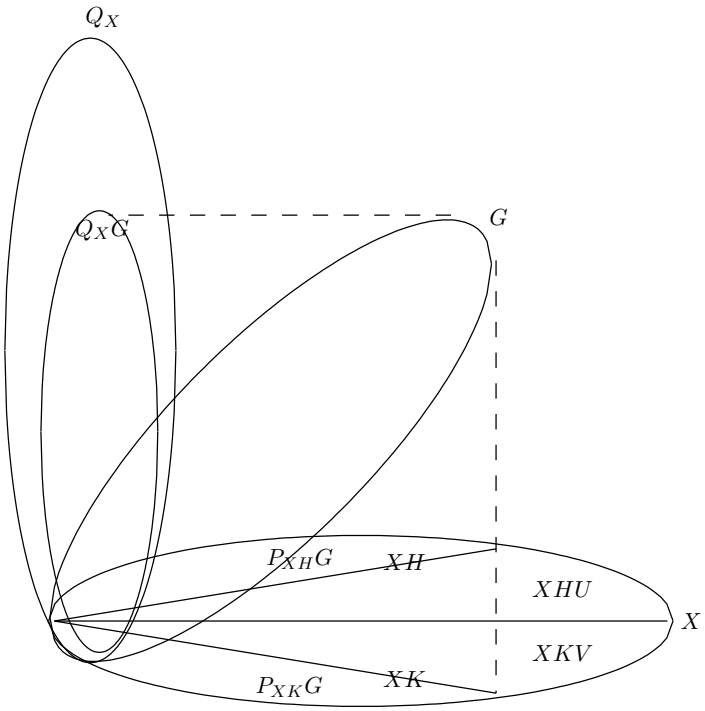

Fig. 4. The decomposition of $\mathrm{Sp}([X, G])$ corresponding to Decomposition (B) of $P_{[X,G]}$.

we omitted the symbol Sp. As in Figure 2, subspaces represented by distinct regions within bigger ellipses (marked by lines or by smaller ellipses) are assumed to be orthogonal. Subspaces belonging to different ellipses, on the other hand, are assumed to subtend the angles as shown in the figure. Exceptions are the relationship between $\mathrm{Sp}(G)$ and $\mathrm{Sp}(XHU)$, and that between $\mathrm{Sp}(G)$ and $\mathrm{Sp}(XKV)$, both of which are orthogonal, despite the fact that they do not look orthogonal in the figure.

3 An Improved Method for GCCANO

We are now in a position to apply the results presented in the previous section to develop an improved method for GCCANO. The idea is simple, now that all important mathematical results have been laid out. We apply the decompositions described in the previous section to both sets of variables, combine a term from each decomposition, and perform a CANO between them.

In what follows, we illustrate this process further. For this purpose, it is convenient to distinguish various quantities for X and those for Y explicitly. We do this by putting subscripts X and Y. Thus, G_X $(n \times q_X)$, H_X $(p_X \times r_X)$, K_X,

A_X, B_X, W_X, U_X and V_X indicate G, H, K, A, B, W, U, and V, respectively, for X, and G_Y $(n \times q_Y)$, H_Y $(p_Y \times r_Y)$, K_Y, A_Y, B_Y, W_Y, U_Y and V_Y denote the same for Y.

When all relevant information is available, both Decompositions (A) and (B) give five-term decompositions. In addition to these decompositions, we may more generally consider all intermediary decompositions obtained in the process of deriving these decompositions. This results in the following 17 terms to be considered for each of the two sets of variables:

[1] $P_{P_{G_X}XH_X}$	[1] $P_{P_{G_Y}YH_Y}$
[2] $P_{P_{G_X}XA_X}$	[2] $P_{P_{G_Y}YA_Y}$
[3] $P_{Q_{G_X}XH_X}$	[3] $P_{Q_{G_Y}YH_Y}$
[4] $P_{Q_{G_X}XB_X}$	[4] $P_{Q_{G_Y}YB_Y}$
[5] $P_{G_XW_X}$	[5] $P_{G_YW_Y}$
[6] $P_{P_{G_X}X}$	[6] $P_{P_{G_Y}Y}$
[7] P_{G_X}	[7] P_{G_Y}
[8] $P_{Q_{G_X}X}$	[8] $P_{Q_{G_Y}Y}$
[9] $P_{[X,G_X]}$	[9] $P_{[Y,G_Y]}$
[10] P_X	[10] P_Y
[11] P_{XH_X}	[11] P_{YH_Y}
[12] P_{XK_X}	[12] P_{YK_Y}
[13] $P_{P_{XH_X}G_X}$	[13] $P_{P_{YH_Y}G_Y}$
[14] $P_{XH_XU_X}$	[14] $P_{YH_YU_Y}$
[15] $P_{P_{XK_X}G_X}$	[15] $P_{P_{YK_Y}G_Y}$
[16] $P_{XK_XV_X}$	[16] $P_{YK_YV_Y}$
[17] $P_{Q_XG_X}$	[17] $P_{Q_YG_Y}$

Terms [1] through [5] correspond with those in Decomposition (A), while [13] through [17] correspond with those in Decomposition (B). Terms [6] through [12] represent those in the intermediary decompositions. Readers are referred to Figures 1 and 3 for the relationships among these terms.

Note that some of the terms listed above may be null, while others may be identical. As has been noted earlier, some of the terms in Decompositions (A) and (B) 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 Decompositions (A) and (B). In some cases, no special row and/or column information matrices are available. In such cases we set G and/or H equal to an identity matrix of appropriate order. This gives rise to more identical terms in the list. Note also that terms [5] and [17] may not be of direct interest for GCCANO, since they both pertain to $\mathrm{Ker}(X')$. They are included here for the sake of completeness.

Once the decompositions are made, CANO may be applied to any pair of terms in the decompositions to investigate the relationships between them. This allows a wide variety of canonical correlation analysis, including both existing and new ones. Some of the representative analyses that can be realized include: The bi-partial bi-constraint CANO obtained by the combination of [3]'s for both sets, the bi-partial (no-constraint) CANO realized by the combination of [8]'s for both sets, the (non-partial) bi-constraint CANO realized by the combination of (11)'s, the ordinary (non-partial, no-constraint) CANO obtained by the combination of [10]'s for both sets, etc. We may also "partial" and/or "constrain" only one set of variables to obtain various semi-partial and/or semi-constrained CANO, e.g., the combinations of [3] and [8], [3] and [10], [8] and [11], etc. In each case, we can assess what proportion of the total association is linked to the portion of the relationship between two sets of variables analyzed in the particular analysis.

Imposing constraints may also invoke our interest in analyzing "the other side of the same coin." As an example, suppose we have performed a semi-constrained CANO with some constraints on X, i.e., CANO between [11] for X and [10] for Y. Then, it may also be of interest to analyze the complementary part of [11], that is, [12] for X in combination with [10] for Y. This allows us to look at what is left out in the first analysis, i.e., what is left unaccounted for by the constraints on X. As another example, suppose we have performed a bi-partial CANO between X and Y eliminating the effect of G_X from X and the effect of G_Y from Y. What may seem interesting in this case is CANO's of the effects left out from the previous analysis, the combinations of [7] for both, [7] for X and [8] for Y, and vice versa. These combinations may be analyzed separately, or summed into a lumped effect, which is then subjected to CANO.

In the proposed framework, CANO amounts to obtaining the SVD of a product of two orthogonal projectors associated with the two sets of variables. This may pose some computational difficulty. The size of the matrix whose SVD is obtained, being equal to the number of cases n in the data set, could be quite large. Fortunately, the following procedure significantly cut down the size of the matrix whose SVD is to be computed (Takane & Hunter, 2002). Let T_1 and T_2 represent any orthonormal (i.e., $T_1'T_1 = I$ and $T_2'T_2 = I$) bases of the spaces spanned by the two projectors. Then, the product of the two projectors can be expressed as $T_1T_1'T_2T_2'$. Let the SVD of $T_1'T_2$ be denoted by $\tilde{M}D\tilde{N}'$.

The SVD of $T_1T_1'T_2T_2'$ can then be obtained by $(T_1\tilde{M})D(\tilde{N}'T_2')$.

Let X^* and Y^* denote the matrices of specific bases vectors spanning $\mathrm{Sp}(T_1)$ and $\mathrm{Sp}(T_2)$, respectively. (These X^* and Y^* could be any one of the matrices that define the 17 projectors in the decompositions of $P_{[X,G_X]}$ and $P_{[Y,G_Y]}$.) Then, $X^* = T_1R_1'$ for some R_1, and $Y^* = T_2R_2'$ for some R_2. Matrices of canonical weights, M^* and N^*, can be obtained from $\tilde{M}$ and $\tilde{N}$ above by $M^* = (R_1R_1')^-R_1\tilde{M}$ and $N^* = (R_2R_2')^-R_2\tilde{N}$. Covariances (or correlations) between the bases vectors and the canonical variates are obtained by scaling $X^{*\prime}T_1\tilde{M} = R_1\tilde{M}$, and $Y^{*\prime}T_2\tilde{N}$ appropriately. These are called structure coefficients.

As has been alluded to above, we may sometimes wish to apply CANO to a sum of more than one product of projectors. This allows CANO of the joint effects of more than one source. In some cases, the sum of products of projectors can again be expressed as a single product of two projectors, in which case we may use the computational procedure just described. In other cases, however, the sum cannot be expressed in such a form. Suppose [1] and [3] are merged for both sets of variables, which are then subjected to CANO. This requires the SVD of the sum of $P_{P_{G_X}XH_X}P_{P_{G_Y}YH_Y}$ and $P_{Q_{G_X}XH_X}P_{Q_{G_Y}YH_Y}$. This can be efficiently calculated by the following procedure. Define

$$R = [P_{P_{G_X}XH_X}P_{G_Y}YH_Y, P_{Q_{G_X}XH_X}Q_{G_Y}YH_Y], \tag{17}$$

and

$$C = [P_{G_Y}YH_Y(H_Y'Y'P_{G_Y}YH_Y)^-, Q_{G_Y}YH_Y(H_Y'Y'Q_{G_Y}YH_Y)^-]. \tag{18}$$

Then, $RC' = P_{P_{G_X}XH_X}P_{P_{G_Y}YH_Y} + P_{Q_{G_X}XH_X}P_{Q_{G_Y}YH_Y}$. Let T_R and T_C be any orthonormal bases of $\mathrm{Sp}(R)$ and $\mathrm{Sp}(C)$, respectively. Then, $R = T_RR^*$ and $C = T_CC^*$ for some R^* and C^*, or $R^* = T_R'R$ and $C^* = T_C'C$. Let the SVD of $R^*C^{*\prime}$ be $\tilde{M}D\tilde{N}'$. Then, the SVD of RC' is obtained by $(T_R\tilde{M})D(\tilde{N}'T_C')$. This procedure can easily be extended to the sum of more than two terms. Note that other definitions of matrices R and C may be equally as good, since the primary purpose of defining them is to make the number of columns of these matrices as small as possible. In particular, $(H_Y'Y'P_{G_Y}YH_Y)^-$ and $(H_Y'Y'Q_{G_Y}YH_Y)^-$ could be either part of R or part of C. It may also be worthwhile using, whether SVD of products of projectors or that of their sums is to be computed, more elaborate techniques for SVD of a product of two or more matrices, such as the product SVD (e.g., Fernando & Hammarling, 1988; Bojanczyk, Ewerbring, Luk, & van Dooren, 1991; Zha, 1991), may be in order. The product SVD (PSVD) calculates the SVD of a product of two or more matrices without explicitly forming the product, thereby avoiding the accumulation of rounding errors.

As was the case with the original GCCANO, the proposed method for im-

proved GCCANO is primarily descriptive. No distributional assumptions were deliberately made to avoid limiting the applicability of the method. It should be emphasized, however, that information regarding the stability of the analysis results (e.g., standard errors of the estimates of canonical weights, etc.) can be readily obtained by adapting the Bootstrap method (Efron & Tibshirani, 1998) to canonical correlation analysis. The number of significant canonical correlations can also be tested using permutation tests. The construction of permutations tests for GCCANO has been described by Takane and Hwang (2002). See also Legendre and Legendre (1998) and ter Braak and Šmilaur (1991) for applications of permutation tests to similar situations. The use of these procedures will be demonstrated in the example section.

4 An Illustrative Example

In this section we present examples of analysis by the proposed method of GCCANO. The data to be analyzed are closely linked to the situation described in the introduction section. We are interested in finding the relationship between intake of various kinds of food and mortality rates by various kinds of cancer. We extracted the data on food supplies in 34 countries of the world from FAOSTAT (Food and Agriculture Organization's statistic archive), which were used as the X variables. Specific variables used were: (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 per day. All of the variables were on a per capita basis, and the data were mostly taken in 1994. We obtained the data on cancer mortality rates from WHOSIS (World Health Organization Statistical Information System). Cancer variables, which were used as the Y variables, consisted of the following four cancer sites: (y_1) esophagus, (y_2) stomach, (y_3) pancreas, and (y_4) liver. These data were also taken in 1994. In what follows, we present a series of analysis performed on this data set. A progression of ideas (about what the data could tell us) developed through these analyses is quite illuminating.

We first applied the ordinary CANO. (This is equivalent to the GCCANO of term [10] for both X and Y.) Permutation tests indicated that the largest canonical correlation was highly significant ($r_1^2 = .818$, $p < .000$), while the second one was not ($r_2^2 = .378$, $p > .290$). Table 1 provides estimates of canonical weights (weights applied to the observed data to obtain a canonical variate) as well as those of structure coefficients (correlations between a canonical variate and observed variables) corresponding to the first canonical variate, along with the standard errors of the estimates obtained by the Bootstrap method. The standard errors tend to be large because the sample size was rather small ($n = 34$). An asterisk indicates that the estimated coefficient is significant (i.e., significantly different from 0) at the 5% level, while two asterisks indicate a significance at the 1% level. This information was also obtained by the

Table 1
Weight and structure vectors from analysis [10] & [10]: Ordinary (non-partial, non-constraint) CANO. (An asterisk indicates a significance at the 5% level, and two asterisks at the 1% level.)

Variable	Weight	(SE)	Structure	(SE)
x1	-.076	(.287)	**.660	(.207)
x2	.198	(.229)	**.811	(.174)
x3	-.231	(.187)	-.115	(.262)
x4	**-.534	(.175)	**-.723	(.185)
x5	-.362	(.190)	-.169	(.202)
x6	-.079	(.136)	.176	(.245)
x7	*.675	(.220)	**.637	(.189)
y1	.267	(.172)	**.690	(.247)
y2	*-.402	(.214)	-.452	(.282)
y3	.254	(.254)	*.614	(.279)
y4	*.527	(.393)	**.908	(.320)

Bootstrap method. Comparing the weights and the structure coefficients, the latter seem to be more interpretable than the former. The canonical variate is highly positively correlated with alcohol, meat and high total calorie, and to a lesser degree with milk products, while it is negatively correlated with the other food variables (highly negatively correlated with cereal, and slightly negatively correlated with fish and vegetable). This profile is characteristic of the high-fat and high-cholesterol western European (and North American) style diet. The canonical variate is also highly correlated with three of the four cancer variables, esophagus, pancreas, and liver cancers. These are also the kinds of cancer prevalent in the western European countries. The first canonical variate thus suggests that the high-fat and high-cholesterol diet has some negative impacts on certain kinds of cancer in digestive organs. Stomach cancer is a bit peculiar among those cancers included in the analysis.

The above analysis has revealed which variables are positively or negatively correlated with the first canonical variate. We may utilize this information to obtain possibly more reliable estimates of parameters in CANO. Both food and cancer variables were classified into two groups according to the sign of their correlations with the canonical variate, and constraint matrices, H, were constructed accordingly. (Another idea was to classify the variables according to the absolute strength of their relationship with the canonical variate. That the idea implemented here makes more sense will become clearer later.) Specifically, the following constraint matrix was constructed for X:

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

and similarly, the following constraint matrix was constructed for Y:

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

This matrix puts esophagus, pancreas, and liver cancers in one group and stomach cancer in the other. This was again based on the sign of correlations between these variables and the first canonical variate. (Although in this particular example, H matrices are both indicator variables, more general forms of H matrices may be used. See Takane and Shibayama (1991), and Hunter and Takane (2002) for examples of more complex forms of H matrices in slightly different contexts.)

We have conducted a number of possible semi- and bi-constraint CANO's using the above constraint matrices. We report only one of them here, that is, the results from semi-constraint CANO with constraints imposed only on X. Other results were similar. (Results involving the H_Y matrix will be given later.) This analysis corresponds with GCCANO between term [11] for X and [10] for Y. Canonical correlations ($r_1^2 = .677$, and $r_2^2 = .114$) were somewhat smaller than those obtained in the previous analysis due to the constraints imposed. However, permutation tests indicated that the first canonical correlation was still highly significant ($p < .000$), while the second one was not ($p > .295$). Table 2 provides estimates of canonical weights and structure coefficients for the first canonical variate along with their standard errors obtained by the Bootstrap method. As before, asterisks indicate the level of significance. The estimated coefficients were similar to those obtained in the previous analysis. Note, however, that the standard errors were consistently smaller than those given in Table 1, indicating that the coefficients were much more reliably estimated due to the constraints imposed. Biases (differences between the original estimates and the Bootstrap means of the estimated coefficients) were also found to be small. This suggests a potential utility of incorporating constraints; we may obtain more reliable estimates of parameters without introducing much biases. (However, we should also note the post-hoc nature of the constraints in the present analysis.)

To verify that the imposed constraints did not leave out any important aspects of the total relationship between the two sets of variables, CANO of the complementary part of [11] and [10], that of [12] and [10], was also conducted. The strength of the relationship between [12] and [10] (the part-association of .812) was found to be larger than that of .791 between [11] and [10] that was analyzed above. (See Table 5.) Permutation tests, however, indicated

Table 2
Weight and structure vectors from analysis [11] & [10]: Non-partial semi-constraint CANO 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)
x1	**.824	(.112)	**.765	(.116)
x2	**.824	(.112)	**.792	(.122)
x3	**-.402	(.135)	-.336	(.205)
x4	**-.402	(.135)	**-.561	(.128)
x5	**-.402	(.135)	-.257	(.218)
x6	**.824	(.112)	**.484	(.182)
x7	**.824	(.112)	**.526	(.199)
y1	**.499	(.173)	**.814	(.162)
y2	-.295	(.167)	-.326	(.294)
y3	.358	(.241)	**.650	(.250)
y4	.303	(.239)	**.874	(.152)

that none of the canonical correlations between [12] and [10] were significant ($r_1^2 = .123, p > .104$, and $r_2^2 = .269, p > .406$), confirming that no significant parts of the relationship between X and Y have been left out by imposing the constraints. Analysis of complementary parts of a relationship between two sets of variables is very straightforward with GCCANO, which can be viewed as an important advantage.

How much of the relationship between the food and cancer variables we have found is real, and how much is spurious? We suspected that at least part of the relationship between the two sets of variables was mediated by the degree of economic development and the overall health status in these countries. So, in addition to the variables mentioned above, we obtained data on GDP per capita, disability adjusted life expectancy (DALE), and infant mortality (IM) rate from the United Nation's data archive. We then partialled out the effect of GDP from the food variables, and the effect of DALE and IM from the cancer variables, and correlated the residuals. This is called bi-partial CANO (Timm & Carlson, 1976) and corresponds with GCCANO of [8] for both X and Y. (Bi-partial) canonical correlations ($r_1^2 = .670$, and $r_2^2 = .388$) were discernibly smaller than the (non-partial) canonical correlations obtained in the first analysis, indicating that considerable portions of the original relationship between the food and cancer variables could be explained by the economic and health status variables. Permutation tests, however, indicated that the first (bi-partial) canonical correlation was still highly significant ($p < .004$),

Table 3
Weight and structure vectors from analysis [8] & [8]: Bi-partial no-constraint CANO. (An asterisk indicates a significance at the 5% level, and two asterisks at the 1% level.)

Variable	Weight	(SE)	Structure	(SE)
x1	-.188	(.446)	.373	(.283)
x2	-.104	(.437)	.317	(.331)
x3	-.568	(.276)	-.709	(.279)
x4	-.166	(.281)	-.200	(.220)
x5	*-.735	(.317)	-.553	(.291)
x6	.092	(.299)	.493	(.294)
x7	*.718	(.338)	.193	(.258)
y1	.531	(.270)	.619	(.293)
y2	-.465	(.418)	-.366	(.457)
y3	.766	(.361)	.535	(.277)
y4	-.064	(.412)	.519	(.295)

while the second ($p > .150$) and subsequent ones were not. This means that a substantial amount of the relationship still remained even after the effects of the economic and the health status variables were eliminated. Table 3 shows estimates of weights and structure coefficients obtained in the present analysis. GDP tended to be more highly correlated with the first group of food variables (alcohol, meat, and total calorie) than the second (fish, cereal, and vegetables), and it seems that eliminating its effect from the food variables resulted in shifting the relative importance of the first group of food variables to the second in defining the canonical variate.

No structural coefficients were significant in the above analysis. This was partly because the partialling introduced additional instability in the estimates of parameters. To compensate for this effect, we combined the bi-partial CANO with constrained CANO which was found useful to obtain more reliable estimates of weights and structure coefficients in the second analysis. In the present analysis, we used the grouping information on the cancer variables (H_Y) as well as on the food variables. This lead to CANO of the combination of [3] for both X and Y, which may be called bi-partial bi-constraint CANO. Although the size of canonical correlations decreased ($r_1^2 = .505$, and $r_2^2 = .021$), permutation tests indicated that the first canonical correlation was still highly significant ($p < .001$), while the second one was not ($p > .427$). The empirical significance level for the largest canonical correlation improved somewhat from the previous analysis. Table 4 provides estimates of weights and structure co-

Table 4
Weight and structure vectors from analysis [3] & [3]: Bi-partial bi-constraint CANO. (An asterisk indicates a significance at the 5% level, and two asterisks at the 1% level.)

Variable	Weight	(SE)	Structure	(SE)
x1	*.555	(.228)	**.648	(.168)
x2	*.555	(.228)	*.523	(.205)
x3	**-.711	(.197)	**-.642	(.172)
x4	**-.711	(.197)	*-.467	(.131)
x5	**-.711	(.197)	*-.507	(.248)
x6	*.555	(.228)	*.502	(.174)
x7	*.555	(.228)	.130	(.262)
y1	**.897	(.153)	**.666	(.174)
y2	-.373	(.287)	-.419	(.334)
y3	**.897	(.153)	*.435	(.185)
y4	**.897	(.153)	**.636	(.133)

efficients obtained in this analysis. The overall pattern of structure coefficients remained essentially the same as before. The first canonical variate was positively correlated with alcohol, meat and milk products, and negatively with fish, cereal and vegetable on the food variables, and positively with esophagus, pancreas and liver cancers among cancer variables. Their standard errors were much smaller than those obtained in the previous analysis, leading to more significant estimates of parameters.

Table 5 gives a complete breakdown of the total association between $P_{[X,G_X]}$ and $P_{[Y,G_Y]}$ for the current data set. Rows of the table represent terms in the decompositions of $P_{[X,G_X]}$, while columns represent those in the decompositions of $P_{[Y,G_Y]}$. Both rows and columns of the table are grouped into three blocks labelled (A), (I), and (B). Block (A) represents terms ([1] through [5]) in Decomposition (A), Block (B) represents those ([13] though [17]) in Decomposition (B), and (I) represents terms ([6] through [12]) in intermediary decompositions. (Column blocks may not be clear in the table. Block (A) consists of three columns labelled [1, 6, 7], [3], and [4], Block (I) the next three columns labelled [8], [9], and [10], and Block (B) the last three columns labelled [11, 19], [12, 13], and [17].) Note that no null terms are included in the table. On the X side, [2] and [5] are null (because the number of variables in G_X is less than the number of variables in H_X, i.e., $q_X < r_X$). For the same reason, we also have [1] = [6] = [7]. Similarly, on the Y side, [2] and [5] are null (because $q_Y < r_Y$), and we again have [1] = [6] = [7]. Also, [14] and [16]

Table 5
Additive partitionings of the total association, tr ([9],[9]) = 2.402.

					(A)			(I)			(B)	
	Term	Y	[1,6,7]	[3]	[4]	[8]	[9]	[10]	[11,13]	[12,15]	[17]
	X	Rk	2	2	2	4	6	4	2	2	2
	[1,6,7]	1	.653	.045	.015	.060	.712	.457	.382	.075	.255
(A)	[3]	2	.082	.525	.101	.626	.708	.560	.422	.138	.148
	[4]	5	.225	.443	.314	.757	.982	.689	.365	.324	.293
	[8]	7	.306	.968	.415	1.383	1.690	1.249	.788	.462	.440
	[9]	8	.959	1.013	.430	1.443	2.402	1.707	1.170	.537	.695
(I)	[10]	7	.771	1.010	.365	1.375	2.146	1.603	1.149	.455	.543
	[11]	2	.312	.564	.072	.635	.948	.791	.693	.098	.156
	[12]	5	.459	.447	.293	.740	1.198	.812	.456	.356	.386
	[13]	1	.285	.265	.024	.289	.574	.554	.532	.022	.020
	[14]	1	.027	.299	.048	.346	.373	.237	.162	.076	.136
(B)	[15]	1	.304	.019	.018	.037	.342	.166	.097	.069	.175
	[16]	4	.154	.428	.275	.703	.857	.646	.359	.287	.211
	[17]	1	.188	.003	.065	.068	.256	.103	.021	.082	.153

are null (because $r_Y = q_Y$ and $p_Y - r_Y = q_Y$, respectively), so that [11] = [13], and [12] = [15]. Ranks of the projection matrices associated with the terms in various decompositions are given in the third row and the third column of the table.

The intersection of Block (A) for both rows and columns (which we call Block (A, A)) represents combinations of terms in Decomposition (A) for both $P_{[X,G_X]}$ and $P_{[Y,G_Y]}$. One can easily verify that the nine numbers (part associations) in Block (A, A) add up to the total association of 2.402 in row [9] and column [9]. Similar observations can also be made in Blocks (A, B), (B, A), and (B, B). As has been claimed, the decompositions proposed in this paper indeed provide part associations that add up to the total association. The ranks of projection matrices are also additive; in each block the row ranks add up to $p_X + q_X = 7+1 = 8$ and the column ranks add up to $p_Y + q_Y = 4+2 = 6$.

As noted earlier, Block (I) refers to intermediary decompositions. Terms in the intermediary decompositions are sums of certain subsets of terms in Decomposition (A) or (B). To recapitulate, [8] = [3] + [4], [9] = [7] + [8] = [10] + [17] and [10] = [11] + [12] for both X and Y, and [11] = [13] + [14] and [12]

= [15] + [16] for X ([11] = [13] and [12] = [15] for Y). Again, one can easily verify that these relationships hold among the relevant part associations.

In this section, we have seen a series of interesting analyses by GCCANO. They are, however, just a few examples of analyses that can be readily carried out by the proposed method. A lot of other possibilities exist that may be explored in the future.

5 Concluding Remarks

In this paper, we proposed an improved method of GCCANO, in which projectors associated with two sets of data were first orthogonally decomposed, and then CANO was applied to pairs of decomposed projectors. Because of the orthogonalities of terms in the decompositions, the total association was partitioned into the sum of part associations defined by pairs of decomposed projectors. This means that we can always (if we so wish) analyze non-overlapping portions of the total relationship between two sets of variables separately, which taken together represent the entire relationship between them. Partialling out certain effects and constraining variables in CANO often prompt our interest in analyzing complementary parts of the relationship between two sets of variables, the portions of the relationship that are partialled out and/or left out by the constraints. GCCANO makes this type of analysis extremely easy and straightforward. The part association analyzed in each analysis may have a simpler interpretation because it represents a specific aspect of the total relationship for which additional structural information is provided by the external information. The proposed decompositions of projectors provide a comprehensive framework for additive partitionings of the total association in CANO by external information, which nicely complements the notion of SS decompositions in ANOVA and CPCA (Constrained Principal Component Analysis; Takane & Hunter, 2001). A MATLAB program has been written that implements the proposed method. An example was given that demonstrates the use of the proposed method.

There are a variety of ways in which the proposed method may be further improved. We discuss only one of them here. We may incorporate a regularization procedure to improve the quality of parameter estimates in GCCANO. This may be implemented as follows: For any standardized data matrix A, an extended "projector" may be defined as

$$P_A(\lambda) = A(A'A + \lambda I)^{-1}A', \tag{19}$$

where λ represents a regularization parameter. We may use $P_A(\lambda)$ instead of the usual projector, $P_A(0) = A(A'A)^{-}A'$, in GCCANO. A small positive value

of λ is known to provide better predictions (associated with smaller prediction errors) than $\lambda = 0$ (e.g., Hoerl and Kennard, 1970), particularly when the sample size is small and variables within sets are highly correlated. An optimal value of λ may be determined in such a way as to maximize the association between two sets of variables in validation samples. A leave-one-out method (e.g., Takane & Hwang, 2003; Takane & Yanai, 2003) can easily be adapted for this purpose. The effectiveness of the regularization procedure has been demonstrated in the contexts of CANO by Vinod (1976), and Ramsay and Silverman (1997) among others. GCCANO is already quite flexible, making a variety of CANO's possible within a unified framework. With the kind of regularization described above, it will become an even more versatile technique as a practical data analysis tool.

References

[1] Bojanczyk, A.W., Ewerbring, L.M., Luk, F.T. and van Dooren, P., 1991. An accurate product SVD algorithm. In: R.J. Vaccaro (Ed.), SVD and Signal Processing II, Elsevier, Amsterdam, 113-131.

[2] Cramer, E.M. and Nicewander, W.A., 1979. Some symmetric, invariant measures of multivariate association. Psychometrika, 44, 43-54.

[3] Efron, B. and Tibshirani, R.J., 1998. An Introduction to the Bootstrap. CRC Press, Boca Raton, Florida.

[4] Fernando, K.V. and Hammarling, S., 1988. A product induced singular value decomposition (PSVD) for two matrices and balanced realization. In: Proc. Conf. Linear Algebra in Signals, Systems, and Control, SIAM, Philadelphia, 128-140.

[5] Hoerl, K.E. and Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12, 55-67.

[6] Hunter, M.A. and Takane Y., 2002. Constrained principal component analysis: Various applications. Journal of Educational and Behavioral Statistics, 27, 41-81.

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

[8] Legendre, P. and Legendre, L., 1998. Numerical ecology. The Second English Edition, Elsevier, Oxford.

[9] Ramsay, J.O. and Silverman, B.W., 1997. Functional Data Analysis. Springer, New York.

[10] Rao, C.R., 1964. The use and interpretation of principal component analysis in applied research. Sankhya A, 26, 329-358.

[11] Rao, C.R. and Yanai, H., 1979. General definition and decomposition of projectors and some applications to statistical problems. Journal of Statistical Planning and Inference, 3, 1-17.

[12] Takane, Y. and Hunter, M.A., 2001. Constrained principal component analysis: A comprehensive theory. Applicable Algebra in Engineering, Communication and Computing, 12, 391-419.
[13] Takane, Y. and Hwang, H., 2002. Generalized constrained canonical correlation analysis. Multivariate Behavioral Research, 37, 163-195.
[14] Takane, Y. and Hwang, H., 2004. Regularization in multiple-set canonical correlation analysis. In: J. Blasius and M.J. Greenacre (Eds.), Multiple correspondence analysis and related methods, Academic Press, New York, in press.
[18] Takane, Y. and Shibayama, T., 1991. Principal component analysis with external information on both subjects and variables. Psychometrika, 56, 97-120.
[16] Takane, Y. and Yanai H., 1999. On oblique projectors. Linear Algebra and its Applications, 289, 207-310.
[17] Takane, Y. and Yanai, H., 2003. A simple technique for regularization in redundancy analysis and kernel redundancy analysis. Paper presented at the IMPS-2003 Meeting, Sardinia, Italy.
[18] ter Braak, C.J.F. and Šmilauer, P.,1998. CANOCO reference manual and user's guide to Canoco for windows. Microcomputer Power, Ithaka, N.Y.
[19] Timm, N. and Carlson, J., 1976. Part and bipartial canonical correlation analysis. Psychometrika, 41, 159-176.
[20] Van den Wollenberg, A.L., 1977. Redundancy analysis: an alternative for canonical analysis. Psychometrika, 42, 207-219.
[21] Vinod, H.D., 1976. Canonical ridge and econometrics of joint production. Journal of Econometrics, 4, 147-166.
[22] Yanai, H., 1974. Unification of various techniques of multivariate analysis by means of generalized coefficient of determination (G.C.D.). Japanese Journal of Behaviormetrics, 1, 45-54.
[23] Yanai, H. and Takane, Y., 1992. Canonical correlation analysis with linear constraints. Linear Algebra and Its Applications, 176, 75-89.
[24] Zha, H., 1991. The product singular value decomposition of matrix triplets. BIT, 31, 711-726.