Multivariate Behavioral Research, *37* (2), 163-195

Generalized Constrained Canonical Correlation Analysis

Yoshio Takane and Heungsun Hwang
McGill University

A method for generalized constrained canonical correlation analysis (GCCANO) is proposed that incorporates external information on both rows and columns of data matrices. In this method each set of variables is first decomposed into the sum of several submatrices according to the external information, and then canonical correlation analysis is applied to pairs of derived submatrices, one from each set, to explore linear relationships between them. Technically, the former amounts to projections of the data matrix onto the spaces spanned by matrices of external information, while the latter involves the generalized singular value decomposition of a matrix with certain metric matrices. GCCANO subsumes a number of existing methods as special cases. It generalizes various kinds of linearly constrained correspondence analysis as well as multivariate analysis of variance/canonical discriminant analysis. Permutation tests are applied to test the significance of canonical correlations obtained from GCCANO. Examples are given to illustrate the proposed method.

Introduction

Canonical correlation analysis (CANO) is used to explore linear relationships between two sets of multivariate data. Technically CANO amounts to extracting a series of linear combinations, called the *canonical variates*, from two sets of data, which are mutually orthogonal within each data set while are maximally correlated between the data sets.

Each data set is often accompanied by external information on its rows (corresponding to "cases" or "subjects") and columns (corresponding to "variables"). For example, subjects' demographic information (e.g., gender, age, level of education, etc.) may be available. Some relationships among variables may also be known in advance (e.g., no interaction between variables, equality of scaled variable scores, group membership of variables, etc.).

The work reported in this article was supported by Grant A6394 from the Natural Sciences and Engineering Research Council of Canada to the first author. We wish to thank Yun Joo Koh for kindly providing her Friendship data, Pierre Legendre for his helpful suggestions regarding permutation tests, and Jim Ramsay for his valuable comments on an earlier version of this article. A computer program for GCCANO written in MATLAB is available from the MBR data archive. Requests for reprints should be addressed to: Yoshio Takane, Department of Psychology, 1205 Dr. Penfield Avenue, Montréal, Québec H3A 1B1, Canada. Email: takane@takane2.psych.mcgill.ca.

When such information is incorporated into CANO, we may obtain simpler interpretations since the data to be analyzed are already structured by the external information. We may also look at the relationships between two sets of data from diverse perspectives, relating a variety of pairs of decomposed submatrices supplied by the external information. For more details about potential advantages of incorporating external information, see Takane, Kiers, and de Leeuw (1995).

In this article, we propose a method for generalized constrained canonical correlation analysis (GCCANO) that incorporates external information into CANO. The external information can be incorporated into both rows and columns of each data set. In GCCANO, each data set is first decomposed into several submatrices according to the external information, and then CANO is applied to pairs of the decomposed submatrices, one from each set, to explore linear relationships between them. Technically, the former amounts to projections of the data matrix onto the spaces spanned by matrices of external information, while the latter involves the generalized singular value decomposition (GSVD) of a matrix with certain metric matrices.

Takane and Shibayama (1991) proposed constrained principal component analysis (CPCA) that incorporated external information in dimension reduction within a single data set. They provided a comprehensive framework to incorporate external information in the form of linear constraints. We follow a similar approach for analyzing the relationship between two sets of data.

To illustrate further, suppose an investigator is interested in finding out the relationships between various kinds of food intake and the mortality rate by various kinds of cancer. She finds statistics from the UN reporting average intake of various food items by people living in various countries, and similar statistics on the mortality rates by various kinds of cancer. It is suspected that food variables are affected by the degree of economic development of the countries (i.e, how wealthy the countries are as measured by GDP per capita, average income, etc.) because what people can afford to eat depends on it, and similarly, the cancer mortality rates are affected by the overall health status of those countries as measured by the average life span, infant mortality rate, etcetera. In analyzing the relationships between food variables and the cancer mortality rates, we may wish to eliminate the effects of these variables from the corresponding data sets. This allows us to focus on more intrinsic aspects of the relationships between the food variables and the cancer mortality rates, not mediated by the extraneous variables. Furthermore, the food variables may be classified into several groups according to their nutritional profiles, and various kinds of cancer may also be grouped into several categories according to the closeness of their loci. We may incorporate this kind of prior information to simplify the analysis of the relationships between the two sets of variables. GCCANO is precisely designed to enable such an analysis by using the economic development and the health status variables as the row information, and the grouping variables for food and the cancer variables as the column information matrices.

This article is organized as follows. "The Method" section describes the proposed method in detail. It includes data requirements, data decompositions according to external information, and CANO of the decomposed submatrices of the data. Relations to other existing CANO methods are also discussed, including partial CANO (Rao, 1969), bipartial, and semi-partial (or part) CANO (Cohen, 1982; Timm & Carlson, 1976), and canonical correlation analysis with linear constraints (Yanai & Takane, 1992). The proposed method also generalizes various kinds of linearly constrained correspondence analysis (CA) and multivariate analysis of variance (MANOVA)/canonical discriminant analysis (CDA). The "Generalized Constrained Correspondence Analysis (GCCA) and Multivariate Analysis of Variance (MANOVA)/Canonical Discriminant Analysis (CDA)" section deals with relations to other CA methods, such as canonical correspondence analysis (ter Braak, 1986)/canonical analysis with linear constraints (Böckenholt & Böckenholt, 1990), partial correspondence analysis (Yanai, 1986), and MANOVA/CDA. The "Permutation Tests for Testing the Number of Significant Canonical Correlations" section presents permutation tests for testing the significance of canonical correlations obtained from GCCANO. The permutation tests in particular allow us to test the significance of individual canonical correlations. In the "Applications" section, two examples are given to illustrate the proposed method. The final section briefly summarizes the previous sections and discusses further prospects.

The Method

The Data

We denote an N-row by r-column data matrix by $\mathbf{X}$ and an N-row by c-column data matrix by $\mathbf{Y}$. Assume that there are two kinds of external information matrices for each data matrix, one on the row side and the other on the column side of the data matrix. We denote an $N \times a$ ($\leq N$) row information matrix on $\mathbf{X}$ by $\mathbf{G}_X$ and an $N \times b$ ($\leq N$) row information matrix on $\mathbf{Y}$ by $\mathbf{G}_Y$. We also denote an $r \times d$ ($\leq r$) column information matrix on

$\mathbf{X}$ by $\mathbf{H}_X$ and a $c \times f(\leq c)$ column information matrix on $\mathbf{Y}$ by $\mathbf{H}_Y$. As has been alluded to earlier, these external information matrices can take a variety of forms. For the row information matrix, we may, for example, use a matrix of subjects' attributes (age, gender, levels of education, interactions among these attributes, etc.), a matrix of dummy variables indicating subjects' group membership, or any other demographic information about the subjects. For the column information matrix, we may use an r- or c-component vector of ones, or any matrix capturing relationships among the columns of a data matrix (e.g., a design matrix for pair comparisons, an additivity constraint matrix, or an equality constraint matrix, a matrix of stimulus attributes, etc.). When no row and/or column information is available, we may simply set $\mathbf{G}_X$, $\mathbf{G}_Y$, $\mathbf{H}_X$, and $\mathbf{H}_Y$ equal to the identity matrices of appropriate size.

To illustrate further, consider the following example. This data set will be analyzed later ("The Friendship Data" section) to demonstrate the use of GCCANO. The data were originally collected by Koh, Mendelson, and Rhee (1998) (see Appendix C), who asked 420 South Korean college students (210 males and 210 females) positive and negative feelings toward their best same-sex friends. The degree of positive feelings for a friend was assessed by eight items based on a 6-point rating scale while that of negative feelings was described on eighteen items using a 9-point scale. We may use the eight items on positive feelings as one data set, $\mathbf{X}$, and the eighteen items on negative feelings as the other data set, $\mathbf{Y}$. Besides the items relevant to positive and negative feelings, there are also five demographic items available, so that we may use the demographic items for a single row information matrix, $\mathbf{G}$, common to both sets (that is, $\mathbf{G}_X = \mathbf{G}_Y = \mathbf{G}$). Matrix $\mathbf{G}$ includes items indicating the year in college (1 = freshman, 2 = sophomore, 3 = junior, 4 = senior), gender (1 = male, 2 = female), age, the number of failures to enter university (1 = no experience, 2 = once, 3 = twice, 4 = more than twice), and experience in military service (1 = not finished, 2 = not applicable, 3 = finished). Each demographic item is treated as a continuous variable with equally spaced levels. According to Koh et al. (1998), the eight items in $\mathbf{X}$ were used to assess two aspects of positive feelings toward a friend, satisfaction and affection. The eighteen items in $\mathbf{Y}$ were used to assess four aspects of negative feelings toward a friend, such as feelings of lack of closeness, feelings of conflict, feelings of worry, and feelings of incompetence. For simplification, we modify the original order of the eight items in $\mathbf{X}$, so that satisfaction was assessed by the first four items and affection by the following four items, and also modify the order of the eighteen items in $\mathbf{Y}$ in such a way that the first three items measured the degree of feelings of lack of closeness, the next eight items that of feelings of conflict, the next three items that of feelings of worry, and the last four items that of feelings of incompetence. We may then use two matrices of dummy variables, which indicate group membership of items, as the column information matrices, $\mathbf{H}_X$ and $\mathbf{H}_Y$. Matrices $\mathbf{H}_X$ and $\mathbf{H}_Y$ can be specifically defined as

$$\mathbf{H}_X = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}',$$

and

$$\mathbf{H}_Y = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}',$$

The first column of $\mathbf{H}_X$ indicates items assessing the degree of satisfaction toward a friend while the second column those measuring the degree of affection with a friend. The first column of $\mathbf{H}_Y$ indicates items on feelings of lack of closeness, the second column those on feelings of conflict, the third column those on feelings of worry, and the last column those on feelings of incompetence.

In this article we assume that $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{G}_X$ and $\mathbf{G}_Y$ are already columnwise-centered. In contrast, $\mathbf{H}_X$ and $\mathbf{H}_Y$ do not have to be centered, once $\mathbf{X}$ and $\mathbf{Y}$ are centered. The reason is as follows: In the case of continuous $\mathbf{X}$, $\mathbf{H}_X$ is usually not centered because row means of $\mathbf{X}$ are often interesting aspects of the data to be included in the analysis. When $\mathbf{X}$ is a matrix of dummy variables, the columnwise centering of $\mathbf{X}$ entails a rowwise centering of $\mathbf{X}$ in the metric of $\mathbf{X}'\mathbf{X}$. Let $\mathbf{X}$ be an N by r matrix of dummy variables (without missing data). Then, $\mathbf{X1}_r = \mathbf{1}_N$, where $\mathbf{1}_r$ and $\mathbf{1}_N$ are r-component and N-component vectors of ones. Consequently, $\mathbf{Q}_{1_N}\mathbf{X} = \mathbf{Q}_{1_N}\mathbf{X}\mathbf{Q}_{1_r/X'X} = \mathbf{X}\mathbf{Q}_{1_r/X'X}$, where $\mathbf{Q}_{1_N} = (\mathbf{I} - \mathbf{1}_N\mathbf{1}_N'/N)$, and $\mathbf{Q}_{1_r/X'X} = \left[\mathbf{I} - \mathbf{1}_r(\mathbf{1}_r'\mathbf{X}'\mathbf{X}\mathbf{1}_r)^{-1}\mathbf{1}_r'\mathbf{X}'\mathbf{X}\right]$. This implies that when the columnwise centered $\mathbf{X}$ is postmultiplied by $\mathbf{H}_X$, $\mathbf{H}_X$ is tacitly columnwise centered. This means that $\mathbf{H}_X$ does not have to be explicitly centered whether the data matrices are continuous or discrete. A similar argument holds for $\mathbf{H}_Y$.

Observed data almost always contain a sizable amount of measurement errors. The effect of measurment errors in CANO has been investigated by Meredith (1964), who proposed a factor analysis like communality estimation procedure. If desirable, a similar approach can be taken in GCCANO as well. A simpler solution would be to apply PCA (principal component analysis) or FA (factor analysis) to screen out measurement errors from the data, which are then subjected to GCCANO (Gleason, 1973). Errors-in-variables regression (Gleser, 1981), in which the matrix of regression coefficients is either rank-reduced or confined to lie in a specific subspace, offers another interesting possibility. Other regularization techniques used in CANO (e.g., Ramsay & Silverman, 1997, chapter 12) may also be useful in this regard.

Decomposition of Data Matrices

When both row and column information matrices are available, the data matrices can be decomposed into four submatrices, following Takane and Shibayama (1991):

(1)
$$\begin{aligned}\mathbf{X} &= (\mathbf{P}_{G_X} + \mathbf{Q}_{G_X})\mathbf{X}(\mathbf{P}_{H_X} + \mathbf{Q}_{H_X}) \\ &= \mathbf{P}_{G_X}\mathbf{X}\mathbf{P}_{H_X} + \mathbf{P}_{G_X}\mathbf{X}\mathbf{Q}_{H_X} + \mathbf{Q}_{G_X}\mathbf{X}\mathbf{P}_{H_X} + \mathbf{Q}_{G_X}\mathbf{X}\mathbf{Q}_{H_X},\end{aligned}$$

and

(2)
$$\begin{aligned}\mathbf{Y} &= (\mathbf{P}_{G_Y} + \mathbf{Q}_{G_Y})\mathbf{Y}(\mathbf{P}_{H_Y} + \mathbf{Q}_{H_Y}) \\ &= \mathbf{P}_{G_Y}\mathbf{Y}\mathbf{P}_{H_Y} + \mathbf{P}_{G_Y}\mathbf{Y}\mathbf{Q}_{H_Y} + \mathbf{Q}_{G_Y}\mathbf{Y}\mathbf{P}_{H_Y} + \mathbf{Q}_{G_Y}\mathbf{Y}\mathbf{Q}_{H_Y},\end{aligned}$$

where

(3)
$$\mathbf{P}_{G_X} = \mathbf{G}_X(\mathbf{G}'_X\mathbf{G}_X)^{-}\mathbf{G}'_X,$$

(4)
$$\mathbf{Q}_{G_X} = \mathbf{I} - \mathbf{P}_{G_X},$$

(5)
$$\mathbf{P}_{H_X} = \mathbf{H}_X(\mathbf{H}'_X\mathbf{X}'\mathbf{X}\mathbf{H}_X)^{-}\mathbf{H}'_X\mathbf{X}'\mathbf{X},$$

and

(6)
$$\mathbf{Q}_{H_X} = \mathbf{I} - \mathbf{P}_{H_X}.$$

Symbol "⁻" in Equations 3 and 5 indicates a g-inverse. Note that $\mathbf{P}_{G_X} + \mathbf{Q}_{G_X} = \mathbf{I}$, $\mathbf{P}_{H_X} + \mathbf{Q}_{H_X} = \mathbf{I}$, $\mathbf{P}_{G_Y} + \mathbf{Q}_{G_Y} = \mathbf{I}$, and $\mathbf{P}_{H_Y} + \mathbf{Q}_{H_Y} = \mathbf{I}$ in Equation 1. Note also that $\mathbf{P}_{G_X}$ is the orthogonal projection operator (we call simply projector hereafter) onto Sp($\mathbf{G}_X$) (the space spanned by the column vectors of $\mathbf{G}_X$), and $\mathbf{Q}_{G_X}$ is its orthogonal complement [the orthogonal projector onto Ker($\mathbf{G}_X'$), where Ker($\mathbf{G}_X'$) denotes the null space of $\mathbf{G}_X'$], while $\mathbf{P}_{H_X}$ is the projector onto Sp($\mathbf{H}_X$) along Ker($\mathbf{H}_X'\mathbf{X}'\mathbf{X}$), and $\mathbf{Q}_{H_X}$ is the projector onto Ker($\mathbf{H}_X'\mathbf{X}'\mathbf{X}$) along Sp($\mathbf{H}_X$). Matrices $\mathbf{P}_{G_Y}$, $\mathbf{Q}_{G_Y}$, $\mathbf{P}_{H_Y}$, and $\mathbf{Q}_{H_Y}$ are analogously defined. Matrices $\mathbf{X}'\mathbf{X}$ in $\mathbf{P}_{H_X}$ and $\mathbf{Y}'\mathbf{Y}$ in $\mathbf{P}_{H_Y}$ are called metric matrices. (Projectors $\mathbf{P}_{H_X}$, $\mathbf{Q}_{H_X}$, $\mathbf{P}_{H_Y}$, and $\mathbf{Q}_{H_Y}$ might be denoted as $\mathbf{P}_{H_X/X'X}$, $\mathbf{Q}_{H_X/X'X}$, $\mathbf{P}_{H_Y/Y'Y}$, and $\mathbf{Q}_{H_Y/Y'Y}$, respectively, to explicitly indicate their dependence on metric matrices, $\mathbf{X}'\mathbf{X}$ and $\mathbf{Y}'\mathbf{Y}$. However, we suppress the metric matrices in these projectors to avoid notational clutter.) When they are not of full rank, the following rank conditions,

$$\text{rank}[(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\mathbf{H}_X] = \text{rank}\,(\mathbf{X}'\mathbf{X}\mathbf{H}_X), \tag{7}$$

$$\text{rank}[(\mathbf{Y}'\mathbf{Y})^{-}\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y] = \text{rank}\,(\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y) \tag{8}$$

have to be satisfied for $\mathbf{P}_{H_X}$, $\mathbf{Q}_{H_X}$, $\mathbf{P}_{H_Y}$, and $\mathbf{Q}_{H_Y}$ to be projectors without regard to the kind of g-inverses used (Yanai, 1990). However, it can be easily verified that these conditions are automatically satisfied in the present case. Equation 1 is obtained by setting the row information matrix in CPCA to $\mathbf{G}_X$ with metric $\mathbf{I}_N$, and the column information matrix in CPCA to $\mathbf{X}'\mathbf{X}\mathbf{H}_X$ with metric $(\mathbf{X}'\mathbf{X})^{-}$. Similarly, Equation 2 is obtained by setting the row information matrix to $\mathbf{G}_Y$ with metric $\mathbf{I}_N$, and the column information matrix to $\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y$ with metric $(\mathbf{Y}'\mathbf{Y})^{-}$. More detailed derivations of the decompositions are given in Appendix A.

The decompositions in Equations 1 and 2 are either columnwise or rowwise orthogonal in their respective metric matrices, and each term in Equations 1 and 2 can be given a specific interpretation (Takane & Shibayama, 1991). In Equation 1, the first term represents the portion of $\mathbf{X}$ that can be explained by both $\mathbf{G}_X$ and $\mathbf{H}_X$, the second term by $\mathbf{G}_X$, but not by $\mathbf{H}_X$, the third term by $\mathbf{H}_X$, but not by $\mathbf{G}_X$, and the final term by neither $\mathbf{G}_X$ nor $\mathbf{H}_X$. Similar interpretations can also be given to the four terms in Equation 2.

The four terms in Equations 1 and 2, however, can be derived only when both row and column information matrices are available. We may, more generally, consider the following submatrices of $\mathbf{X}$ and $\mathbf{Y}$:

$$
(9) \quad
\begin{array}{llll}
(\mathbf{X}_1) & \mathbf{X} & (\mathbf{Y}_1) & \mathbf{Y} \\
(\mathbf{X}_2) & \mathbf{P}_{G_X}\mathbf{X} & (\mathbf{Y}_2) & \mathbf{P}_{G_Y}\mathbf{Y} \\
(\mathbf{X}_3) & \mathbf{Q}_{G_X}\mathbf{X} & (\mathbf{Y}_3) & \mathbf{Q}_{G_Y}\mathbf{Y} \\
(\mathbf{X}_4) & \mathbf{X}\mathbf{P}_{H_X} & (\mathbf{Y}_4) & \mathbf{Y}\mathbf{P}_{H_Y} \\
(\mathbf{X}_5) & \mathbf{X}\mathbf{Q}_{H_X} & (\mathbf{Y}_5) & \mathbf{Y}\mathbf{Q}_{H_Y} \\
(\mathbf{X}_6) & \mathbf{P}_{G_X}\mathbf{X}\mathbf{P}_{H_X} & (\mathbf{Y}_6) & \mathbf{P}_{G_Y}\mathbf{Y}\mathbf{P}_{H_Y} \\
(\mathbf{X}_7) & \mathbf{P}_{G_X}\mathbf{X}\mathbf{Q}_{H_X} & (\mathbf{Y}_7) & \mathbf{P}_{G_Y}\mathbf{Y}\mathbf{Q}_{H_Y} \\
(\mathbf{X}_8) & \mathbf{Q}_{G_X}\mathbf{X}\mathbf{P}_{H_X} & (\mathbf{Y}_8) & \mathbf{Q}_{G_Y}\mathbf{Y}\mathbf{P}_{H_Y} \\
(\mathbf{X}_9) & \mathbf{Q}_{G_X}\mathbf{X}\mathbf{Q}_{H_X} & (\mathbf{Y}_9) & \mathbf{Q}_{G_Y}\mathbf{Y}\mathbf{Q}_{H_Y}
\end{array}
$$

Matrices $\mathbf{X}_1$ and $\mathbf{Y}_1$ simply indicate the unconstrained data matrices. The next two sets of submatrices, denoted by $\mathbf{X}_2$, $\mathbf{Y}_2$, $\mathbf{X}_3$ and $\mathbf{Y}_3$, can be derived when only the row information matrices are incorporated. Matrices $\mathbf{X}_2$ and $\mathbf{Y}_2$ indicate the portions of $\mathbf{X}$ and $\mathbf{Y}$ that can be explained by their respective row information matrices, while $\mathbf{X}_3$ and $\mathbf{Y}_3$ indicate the residuals after the effects due to the row information matrices are eliminated. Similarly, $\mathbf{X}_4$, $\mathbf{Y}_4$, $\mathbf{X}_5$ and $\mathbf{Y}_5$ are derived when only the column information matrices are incorporated. The last four submatrices ($\mathbf{X}_6$ and $\mathbf{Y}_6$ through $\mathbf{X}_9$ and $\mathbf{Y}_9$) correspond with the four terms in the full decompositions, Equations 1 and 2, when both row and column information matrices are incorporated.

Canonical Correlation Analysis of Pairs of the Decomposed Matrices

Once each data set is decomposed according to the external information, CANO can be applied to any pair of a submatrix from $\mathbf{X}$ and one from $\mathbf{Y}$, given in Equation 9, to explore relationships between them. It is well known that computationally CANO amounts to the generalized singular value decomposition (GSVD) of a matrix with certain metric matrices. The GSVD of matrix $\mathbf{Z}$ with metric matrices, $\mathbf{A}$ and $\mathbf{B}$ (temporarily assumed *pd*), is defined as

$$(10) \qquad \mathbf{Z} = \mathbf{UDV}',$$

where $\mathbf{U}'\mathbf{AU} = \mathbf{I}$ and $\mathbf{V}'\mathbf{BV} = \mathbf{I}$ and $\mathbf{D}$ is diagonal and *pd* (e.g., Greenacre, 1984; Takane & Shibayama, 1991). It is denoted as $\text{GSVD}(\mathbf{Z})_{A,B}$, and can be obtained as follows. Let $\mathbf{A} = \mathbf{R}_A\mathbf{R}_A'$ and $\mathbf{B} = \mathbf{R}_B\mathbf{R}_B'$ be arbitrary square root decompositions of **A** and **B**, respectively. Let the ordinary singular value decomposition (SVD) of $\mathbf{R}_A'\mathbf{Z}\mathbf{R}_B$ be denoted by

(11) $$\mathbf{R}_A'\mathbf{Z}\mathbf{R}_B = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime},$$

where $\mathbf{U}^{*\prime}\mathbf{U}^* = \mathbf{V}^{*\prime}\mathbf{V}^* = \mathbf{I}$, and $\mathbf{D}^*$ is diagonal and *pd*. Then, **U**, **V**, and **D** in Equation 10 are obtained by $\mathbf{U} = (\mathbf{R}_A')^{-1}\mathbf{U}^*$, $\mathbf{V} = (\mathbf{R}_B')^{-1}\mathbf{V}^*$, and $\mathbf{D} = \mathbf{D}^*$. GSVD of a matrix always exists (Takane & Shibayama, 1991). It is unique if **A** and **B** are nonsingular. If they are singular, we may use any *g*-inverses of **A** and **B** in the above formulae, but then the uniqueness of the solution is destroyed. If we still want to obtain a unique solution, we may use the Moore-Penrose inverses.

Let us briefly discuss some examples of applying CANO to pairs of the decomposed submatrices of **X** and **Y**, given in Equation 9. First, consider CANO($\mathbf{X}_1$, $\mathbf{Y}_1$), CANO of the pair of $\mathbf{X}_1$ and $\mathbf{Y}_1$. This is simply the ordinary CANO between **X** and **Y**, which amounts to obtaining

(12) $$\text{GSVD}[(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^-]_{X'X,Y'Y}.$$

Consider next CANO($\mathbf{X}_4$, $\mathbf{Y}_4$), which amounts to obtaining

(13) $$\text{GSVD}[(\mathbf{H}_X'\mathbf{X}'\mathbf{X}\mathbf{H}_X)^-\mathbf{H}_X'\mathbf{X}'\mathbf{Y}\mathbf{H}_Y(\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y)^-]_{H_X'X'XH_X,H_Y'Y'YH_Y},$$

or

(14) $$\text{GSVD}[\mathbf{H}_X(\mathbf{H}_X'\mathbf{X}'\mathbf{X}\mathbf{H}_X)^-\mathbf{H}_X'\mathbf{X}'\mathbf{Y}\mathbf{H}_Y(\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y)^-\mathbf{H}_Y']_{X'X,Y'Y}.$$

Let the GSVD in Equation 13 be denoted by $\mathbf{UVD}'$ and that in Equation 14 by $\hat{\mathbf{U}}\hat{\mathbf{D}}\hat{\mathbf{V}}'$. The two GSVD's are related by $\hat{\mathbf{U}} = (\mathbf{X}'\mathbf{X})^-(\mathbf{X}'\mathbf{X})\mathbf{H}_X\mathbf{U}$, $\hat{\mathbf{V}} = (\mathbf{Y}'\mathbf{Y})^-(\mathbf{Y}'\mathbf{Y})\mathbf{H}_Y\mathbf{V}$, and $\hat{\mathbf{D}} = \mathbf{D}$, or $\mathbf{U} = (\mathbf{H}_X'\mathbf{X}'\mathbf{X}\mathbf{H}_X)^-\mathbf{H}_X'\mathbf{X}'\mathbf{X}\hat{\mathbf{U}}$, $\mathbf{V} = (\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y)^-\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}\hat{\mathbf{V}}$, and $\mathbf{D} = \hat{\mathbf{D}}$. CANO($\mathbf{X}_4$, $\mathbf{Y}_4$) is a CANO of the portions of **X** and **Y** accounted for by $\mathbf{H}_X$ and $\mathbf{H}_Y$, respectively. This is equivalent to CANOLC proposed by Yanai & Takane (1992). A detailed derivation of how CANO($\mathbf{X}_4$, $\mathbf{Y}_4$) reduces to the GSVD given in Equation 13 is given in Appendix B.

We can also pair $\mathbf{X}_3$ and $\mathbf{Y}_3$ for CANO. These decomposed submatrices represent the residual portions of **X** and **Y** after partialing out the effects of the row information matrices. If there is a single row information matrix common to **X** and **Y**, that is, $\mathbf{G}_X = \mathbf{G}_Y$, CANO($\mathbf{X}_3$, $\mathbf{Y}_3$) amounts to partial CANO (Rao, 1969). On the other hand, if two distinct row information matrices are used for **X** and **Y**, that is, $\mathbf{G}_X \neq \mathbf{G}_Y$, CANO($\mathbf{X}_3$, $\mathbf{Y}_3$) amounts to bipartial CANO (Timm & Carlson, 1976).

We can also apply CANO to the pair of $\mathbf{X}_1$ and $\mathbf{Y}_3$ or of $\mathbf{X}_3$ and $\mathbf{Y}_1$. That is, one of the pair is the unconstrained data matrix, and the other is the residual after removing the effect of the row information matrix. CANO($\mathbf{X}_1$, $\mathbf{Y}_3$) or CANO($\mathbf{X}_3$, $\mathbf{Y}_1$) amounts to part CANO (Timm & Carlson, 1976) or semi-partial CANO (Cohen, 1982).

All the analyses mentioned above reduce to existing techniques. However, a combination such as $\mathbf{X}_8$ and $\mathbf{Y}_8$ is also possible, leading to a new technique which might be called (bi)partial CANOLC. In the example section we demonstrate the use of partial CANOLC. Let us reiterate that, in principle, any combination of $\mathbf{X}_1$ through $\mathbf{X}_9$ and $\mathbf{Y}_1$ through $\mathbf{Y}_9$ is possible, although some combinations such as $\mathbf{X}_2$ and $\mathbf{Y}_2$ with $\mathbf{G}_X = \mathbf{G}_Y$ may not be empirically meaningful.

Generalized Constrained Correspondence Analysis(GCCA) and Multivariate Analysis of Variance (MANOVA)/Canonical Discriminant Analysis (CDA)

When both **X** and **Y** consist of discrete data CANO reduces to correspondence analysis (CA) of contingency tables. GCCANO, in this case, generalizes a variety of linearly constrained correspondence analysis techniques. It also subsumes MANOVA/CDA and their constrained counterparts which are also special cases of CANO.

Generalized Constrained Correspondence Analysis (GCCA)

Let **X** and **Y** be *N* by *r* and *N* by *c* matrices of dummy variables. Let $\mathbf{G}_X$, $\mathbf{G}_Y$, $\mathbf{H}_X$, and $\mathbf{H}_Y$ denote the external information matrices as in "The Method" section. The data matrices can be decomposed according to the external information in a manner similar to Equation 3. As before, CANO can be applied to any pair of the components. All these analyses lead to special kinds of correspondence analysis.

Let $\mathbf{F} = \mathbf{X}'\mathbf{Y}$ denote a two-way contingency table with both row and column marginal effects removed. CANO($\mathbf{X}_1$, $\mathbf{Y}_1$) amounts to obtaining

$$\text{GSVD}[(\mathbf{X}'\mathbf{X})^{-}\mathbf{F}(\mathbf{Y}'\mathbf{Y})^{-}]_{X'X,Y'Y} \tag{15}$$

CANO($\mathbf{X}_1$, $\mathbf{Y}_1$) amounts to the unconstrained correspondence analysis of **F**. It is well known that in GCCA $\mathbf{X}'\mathbf{X}$ and $\mathbf{Y}'\mathbf{Y}$ can be replaced by $\mathbf{D}_R$ and $\mathbf{D}_C$, where $\mathbf{D}_R$ and $\mathbf{D}_C$ are diagonal matrices of row and column sums of the original contingency table (without the marginal effects removed).

CANO($\mathbf{X}_4$, $\mathbf{Y}_4$) amounts to obtaining

$$\text{(16)} \qquad \text{GSVD}[(\mathbf{H}_X'\mathbf{X}'\mathbf{X}\mathbf{H}_X)^{-}\mathbf{H}_X'\mathbf{F}\mathbf{H}_Y(\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y)^{-}]_{H_X'X'XH_X, H_Y'Y'YH_Y},$$

or

$$\text{(17)} \qquad \text{GSVD}[\mathbf{H}_X(\mathbf{H}_X'\mathbf{X}'\mathbf{X}\mathbf{H}_X)^{-}\mathbf{H}_X'\mathbf{F}\mathbf{H}_Y(\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y)^{-}\mathbf{H}_Y']_{X'X, Y'Y}.$$

It is a CANO of the portions of $\mathbf{X}$ and $\mathbf{Y}$ explained by $\mathbf{H}_X$ and $\mathbf{H}_Y$. It is analogous to canonical correspondence analysis (CCA) by ter Braak (1986) and canonical analysis with linear constraints (CALC) by Böckenholt and Böckenholt (1990).

CANO($\mathbf{X}_3$, $\mathbf{Y}_3$) amounts to

$$\text{(18)} \qquad \text{GSVD}[(\mathbf{X}'\mathbf{Q}_{G_X}\mathbf{X})^{-}\mathbf{X}'\mathbf{Q}_{G_X}\mathbf{Q}_{G_Y}\mathbf{Y}(\mathbf{Y}'\mathbf{Q}_{G_Y}\mathbf{Y})^{-}]_{X'Q_{G_X}X, Y'Q_{G_Y}Y}.$$

It is a CANO of the residuals obtained after the effects due to $\mathbf{G}_X$ and $\mathbf{G}_Y$ are removed from $\mathbf{X}$ and $\mathbf{Y}$. It may be called bipartial correspondence analysis, following the name, bipartial CANO. If $\mathbf{G}_X = \mathbf{G}_Y$, CANO($\mathbf{X}_3$, $\mathbf{Y}_3$) reduces to partial correspondence analysis by Yanai (1986). CANO($\mathbf{X}_1$, $\mathbf{Y}_3$) or CANO($\mathbf{X}_3$, $\mathbf{Y}_1$) can be called semi-partial or part correspondence analysis.

As in the case of continuous data, we can also consider many other combinations of $\mathbf{X}_1$ through $\mathbf{X}_9$ and $\mathbf{Y}_1$ through $\mathbf{Y}_9$, yielding a variety of new techniques for relating two sets of discrete data. In the example section, we demonstrate the use of CANO($\mathbf{X}_8$, $\mathbf{Y}_8$) with $\mathbf{G}_X = \mathbf{G}_Y = \mathbf{G}$, which might be called partial CCA. (ter Braak, 1986, uses the same terminology in a slightly different context, however.)

Multivariate Analysis of Variance (MANOVA)/Canonical Discriminant Analysis (CDA)

In MANOVA/CDA, one set of variables is discrete while the other set is continuous. Let $\mathbf{X}^*$ be an N by r matrix of raw dummy variables and let $\mathbf{Y}$ be an N by c matrix of (columnwise-centered) continuous multivariate observations. Assume that there is no additional information on both $\mathbf{X}^*$ and $\mathbf{Y}$ (i.e., $\mathbf{G}_X = \mathbf{G}_Y = \mathbf{I}$ and $\mathbf{H}_X = \mathbf{H}_Y = \mathbf{I}$). In this case, we columnwise-center only $\mathbf{Y}$. Technically, MANOVA/CDA of $\mathbf{X}^*$ and $\mathbf{Y}$ amounts to

$$\text{GSVD}[(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-}\mathbf{X}^{*\prime}\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-}]_{X^{*\prime}X^*, Y'Y}.$$

MANOVA/CDA is equivalent to CANO($\mathbf{X}^*$, $\mathbf{Y}$).

It is of interest to introduce external constraints into MANOVA/CDA. GCCANO provides a general framework for constrained and/or partial MANOVA/CDA including such familiar techniques as MANOCOVA (multivariate analysis of covariance), profile analysis (Morrison, 1990), GMANOVA (growth curve models), etcetera.

Permutation Tests for Testing the Number of Significant Canonical Correlations

It is important to be able to identify the number of significant canonical variates. We use permutation tests for testing the significance of canonical correlations obtained from GCCANO. The permutation tests have certain desirable properties for our purposes. First, they do not rely on any distributional assumptions on the data, so that they can be used even when the parametric (often the multivariate normality) assumptions fail. Secondly, they can be used for any kind of canonical correlation analysis realized by GCCANO, while parametric procedures are often designed only for a subset of analyses. The permutation tests can be applied for testing the independence model (i.e., independence between **X** and **Y**, which implies that all canonical correlations are zero) against the saturated model (which implies that not all canonical correlations are zero). However, models with a specific number of canonical variates can also be tested against the saturated model by eliminating the effects of previous canonical correlations from the data sets (see Legendre & Legendre, 1998; ter Braak, 1990).

Following Manly's (1997) procedure, the permutation method for testing the independence model against the saturated model can be stated as follows:

1. From the original data sets, **X** and **Y**, compute the observed value of Bartlett's (1938) statistic, given by

$$-\left[(N-1)-\frac{1}{2}(r+c+1)\right]\sum_{i=1}^{J}\log\left(1-\lambda_i^2\right),$$

where $J = min(r, c)$, and the λ_i are the sample canonical correlations in descending order. We denote the observed value by q_{obs}.

2. Construct a "permutation" sample only for one data set, say **Y**, by randomly permuting the cases (or randomly selecting one case at a time without replacement). We denote the sample by $\mathbf{Y}_{perm}$.

3. Apply GCCANO to **X** and $\mathbf{Y}_{perm}$, and calculate the Bartlett's (1938) statistic, denoted by q_{perm}.

4. Steps 2-3 are repeated B times (e.g., $B = 1{,}000$), yielding the null distribution of q_{perm} (the distribution of q_{perm} under the independence assumption).

5. Compute the so-called permutation achieved significance level (ASL_{perm}) defined to be the probability that q_{perm} is greater than or equal to q_{obs}, that is, $ASL_{perm} = \#\{q_{perm} \geq q_{obs}\}/B$.

If ASL_{perm} is less than.05, the null hypothesis of independence is rejected at a 5% level. To test the model with one canonical variate against the saturated model, we eliminate the effects of the largest canonical correlation from both **X** and **Y**. This can be done for **X** by $\mathbf{X} = \mathbf{XQ}_{w_1/X'X}$, where $\mathbf{w}_1$ is the vector of canonical weights used to derive the first canonical variate for **X**, and where $\mathbf{Q}_{w_1/X'X} = \mathbf{I} - \mathbf{w}_1(\mathbf{w}_1'\mathbf{X}'\mathbf{X}\mathbf{w}_1)^{-1}\mathbf{w}_1'\mathbf{X}'\mathbf{X}$. Essentially, the same procedure can be used for **Y** as well. The second largest canonical correlation then becomes the largest one, and the same permutational procedure can be used to test the null hypothesis that all the remaining canonical correlations are zero against the alternative hypothesis that it is false. The same strategy is applied until an unrejectable null hypoethsis is encountered, or the saturated model is found to be better than any other models. This procedure is essentially the same as that proposed by Legendre & Legendre (1998).

Although the above procedure employs Bartlett's (1938) statistic, other statistics such as Roy's max lambda ($N\lambda^2$) criterion can be used to construct similar permutation tests. We have tried this also. However, despite the known difference in power between the two tests under different conditions (Haberman, 1981), we have reached essentially the same conclusions in the examples to be presented in the next section. Consequently, we only report the results of Bartlett's tests.

Applications

We have written a MATLAB program implementing GCCANO. We applied GCCANO to actual data sets for empirical evaluation of the procedure. In this article, we present two examples of analysis. The first example deals with continuous data. The second example involves discrete data, which demonstrates that our method generalizes various kinds of linearly constrained correspondence analysis methods.

The Friendship Data

The first example pertains to Koh et al.'s (1998) Friendship data described in "The Data" section. The reader is encouraged to refer back to that section to remember the basic features of the data set as well as the kinds of external information available for $\mathbf{G} = \mathbf{G}_X = \mathbf{G}_Y$., $\mathbf{H}_X$, and $\mathbf{H}_Y$.

GCCANO was first applied with $\mathbf{G} = \mathbf{I}$ and $\mathbf{H}_X = \mathbf{H}_Y = \mathbf{I}$, that is, CANO($\mathbf{X}_1$, $\mathbf{Y}_1$), which is equivalent to the unconstrained CANO. The estimated squared canonical correlations and the corresponding empirical significance levels (ALS_{perm}) calculated by the permutation method with 1,000 permuted samples, are given in the first two columns of Table 1. It seems that the first three squared canonical correlations are significant at the 5% level. The empirical significance level of the smallest squared canonical correlation, λ_8^2, happens to be less than that of the second smallest one, λ_7^2. This may be due to the fact that they are very close to each other (i.e., $\lambda_7^2 = .023$ and $\lambda_8^2 = .021$), so that their order may have been reversed.

To interpret the three significant canonical variates, we looked at both pattern (the weights applied to the observed variables to derive the canonical variates) and structure (the correlations between the canonical variates and the observed variables). To increase their interpretability we also applied the normal varimax rotation method to both the pattern and the structure matrices (Cliff & Krus, 1976; a MATLAB code for the varimax rotation was kindly provided by Jim Ramsay).

Interpreting the pattern is like trying to characterize the canonical variates in terms of how they are constructed from the observed variables, while interpreting the structure is like explaining the nature of the canonical variates in terms of how they are related to the observed variables. The two modes of interpretation are often complementary, although it may also

Table 1
The Squared Canonical Correlations and the Corresponding Empirical Significance Levels in the Parenthesis Obtained from CANO($\mathbf{X}_1$, $\mathbf{Y}_1$), CANO($\mathbf{X}_4$, $\mathbf{Y}_4$), CANO($\mathbf{X}_3$, $\mathbf{Y}_3$), and CANO($\mathbf{X}_8$, $\mathbf{Y}_8$) of the Friendship Data

CANO($\mathbf{X}_1$, $\mathbf{Y}_1$)		CANO($\mathbf{X}_4$, $\mathbf{Y}_4$)		CANO($\mathbf{X}_3$, $\mathbf{Y}_3$)		CANO($\mathbf{X}_8$, $\mathbf{Y}_8$)	
.339	(.000)	.273	(.000)	.344	(.000)	.278	(.000)
.111	(.001)	.022	(.025)	.116	(.000)	.021	(.030)
.105	(.006)	0	0	.105	(.000)	0	0
.081	(.104)	0	0	.071	(.150)	0	0
.056	(.429)	0	0	.060	(.410)	0	0
.036	(.712)	0	0	.042	(.750)	0	0
.023	(.798)	0	0	.023	(.920)	0	0
.021	(.626)	0	0	.014	(.930)	0	0

happen that one is more interpretable than the other in some situations. It has turned out that the rotated structure matrix is the most interpretable in the present case. This matrix is presented in Table 2, along with the corresponding standard errors obtained by the bootstrap method (Efron, 1979; Efron & Tibshirani, 1998) with 100 bootstrap samples.

Table 2
Rotated Canonical Structure and the Corresponding Standard Errors in the Parenthesis Obtained from CANO($\mathbf{X}_1$, $\mathbf{Y}_1$) of the Friendship Data

Item	Dim = 1		Dim = 2		Dim = 3	
x_1 p_1	.637	(.110)	.379	(.142)	.476	(.127)
x_2 p_1	.804	(.174)	.333	(.243)	.115	(.215)
x_3 p_1	.783	(.151)	.311	(.208)	.279	(.205)
x_4 p_1	.451	(.179)	.754	(.233)	.381	(.221)
x_5 p_2	.630	(.139)	.246	(.187)	.597	(.155)
x_6 p_2	.187	(.215)	.468	(.258)	.711	(.280)
x_7 p_2	.356	(.197)	.740	(.272)	.269	(.250)
x_8 p_2	.326	(.211)	.619	(.262)	.487	(.229)
y_1 n_1	-.508	(.144)	-.310	(.204)	-.285	(.215)
y_2 n_1	-.570	(.127)	-.252	(.144)	-.401	(.159)
y_3 n_1	-.335	(.169)	-.436	(.202)	-.103	(.190)
y_4 n_2	-.244	(.168)	-.106	(.201)	-.544	(.217)
y_5 n_2	-.126	(.180)	.086	(.240)	-.440	(.260)
y_6 n_2	-.370	(.140)	-.116	(.195)	-.166	(.169)
y_7 n_2	-.698	(.187)	.065	(.254)	-.192	(.235)
y_8 n_2	-.144	(.135)	-.189	(.160)	-.406	(.153)
y_9 n_2	-.263	(.146)	.003	(.157)	-.026	(.180)
y_{10} n_2	-.189	(.172)	-.016	(.195)	-.504	(.215)
y_{11} n_2	-.170	(.154)	-.094	(.178)	-.237	(.185)
y_{12} n_3	-.004	(.160)	.097	(.192)	.203	(.175)
y_{13} n_3	-.206	(.176)	-.218	(.219)	.146	(.173)
y_{14} n_3	-.031	(.176)	.356	(.230)	.049	(.221)
y_{15} n_4	.274	(.167)	.010	(.205)	.109	(.218)
y_{16} n_4	.163	(.171)	.083	(.206)	.055	(.210)
y_{17} n_4	.153	(.169)	-.063	(.173)	-.246	(.244)
y_{18} n_4	.179	(.231)	-.376	(.265)	-.177	(.322)

In the table, among the eight items in **X** the four items on satisfaction are labeled as *p*1s, and the other four items on affection as *p*2s. Among the eighteen items in **Y**, the three items on feelings of lack of closeness are labeled as *n*1s, the eight items on feelings of conflict as *n*2s, the three items on feelings of worry as *n*3s, and the four items on feelings of incompetence as *n*4s. The rotated solution seems to indicate that the first canonical variate is closely related to satisfaction with a friend in **X**, and feelings of lack of closeness, and to a lesser extent, feelings of conflict in **Y**. The second and the third canonical variates are difficult to interpret, however, in terms of the a priori groupings of variables.

GCCANO was also applied with $\mathbf{G} = \mathbf{I}$, $\mathbf{H}_X \neq \mathbf{I}$, and $\mathbf{H}_Y \neq \mathbf{I}$ [CANO($\mathbf{X}_4$, $\mathbf{Y}_4$)]. The squared canonical correlations and their empirical significance levels estimated under CANO($\mathbf{X}_4$, $\mathbf{Y}_4$) are given in the third and the fourth columns of Table 1. It is found that all two nonzero squared canonical correlations are significant. The rotated canonical structure obtained from CANO($\mathbf{X}_4$, $\mathbf{Y}_4$) is reported in Table 3, along with the corresponding bootstrap standard errors. Those standard errors tend to be smaller than those found in the unconstrained CANO [CANO($\mathbf{X}_1$, $\mathbf{Y}_1$)] due to the constraints, $\mathbf{H}_X$ and $\mathbf{H}_Y$, incorporated.

The rotated structure matrix shows that the first canonical variate is more highly correlated with the four items related to the friendship satisfaction in **X**, and the items on feelings of lack of closeness and feelings of conflict in **Y**, while the second with the four items related to affection in **X**, and the items on feelings of worry in **Y**. It is noted that the relationships between the canonical variates and the items are much clearer in the constrained case than in the unconstrained case, leading to more confident interpretations.

We also applied GCCANO with $\mathbf{H}_X = \mathbf{H}_Y = \mathbf{I}$, and eliminating the effect of $\mathbf{G} \neq \mathbf{I}$ [CANO($\mathbf{X}_3$, $\mathbf{Y}_3$)]. This case is equivalent to partial CANO that obtains the set of canonical variates independent of the information in **G**. The squared canonical correlations and the corresponding empirical significance levels derived from partial CANO are presented in the fifth and the sixth columns of Table 1. The first three squared canonical correlations are found to be significant, as in CANO($\mathbf{X}_1$, $\mathbf{Y}_1$). The rotated three-dimensional canonical structure derived under CANO($\mathbf{X}_3$, $\mathbf{Y}_3$) is presented in Table 4. The canonical structure is similar to that obtained in CANO($\mathbf{X}_1$, $\mathbf{Y}_1$), though the corresponding canonical correlations are a bit smaller. This indicates that the effect of **G** on both **X** and **Y** is only minor. Thus, essentially the same interpretations of the canonical variates as in CANO($\mathbf{X}_1$, $\mathbf{Y}_1$) can be given.

Finally, GCCANO was applied with $\mathbf{H}_X \neq \mathbf{I}$, $\mathbf{H}_Y \neq \mathbf{I}$, and eliminating the effect of $\mathbf{G} \neq \mathbf{I}$ [CANO($\mathbf{X}_8$, $\mathbf{Y}_8$)]. This is called partial CANOLC. In this analysis, the relationship between the portions of **X** and **Y** explained by $\mathbf{H}_X$ and $\mathbf{H}_Y$, respectively, but not by **G**, is analyzed. The squared canonical correlations and their significance levels are presented in the last two columns of Table 1. It indicates that all two nonzero squared canonical

Table 3
Rotated Canonical Structure and the Corresponding Standard Errors in the Parenthesis Obtained from CANO($\mathbf{X}_4$, $\mathbf{Y}_4$) of the Friendship Data

Item		Dim = 1		Dim = 2	
x_1	p_1	.630	(.055)	.580	(.064)
x_2	p_1	.861	(.067)	.320	(.107)
x_3	p_1	.840	(.060)	.365	(.099)
x_4	p_1	.688	(.054)	.539	(.072)
x_5	p_2	.511	(.080)	.712	(.064)
x_6	p_2	.319	(.121)	.860	(.099)
x_7	p_2	.463	(.098)	.764	(.080)
x_8	p_2	.354	(.117)	.857	(.094)
y_1	n_1	-.757	(.087)	-.159	(.122)
y_2	n_1	-.770	(.074)	-.287	(.121)
y_3	n_1	-.683	(.112)	-.005	(.155)
y_4	n_2	-.515	(.117)	-.236	(.161)
y_5	n_2	-.292	(.151)	-.184	(.206)
y_6	n_2	-.394	(.125)	-.218	(.168)
y_7	n_2	-.486	(.115)	-.188	(.168)
y_8	n_2	-.426	(.138)	-.188	(.192)
y_9	n_2	-.353	(.134)	-.215	(.187)
y_{10}	n_2	-.394	(.137)	-.099	(.189)
y_{11}	n_2	-.314	(.162)	-.044	(.232)
y_{12}	n_3	-.222	(.183)	.656	(.221)
y_{13}	n_3	-.464	(.152)	-.411	(.204)
y_{14}	n_3	-.370	(.175)	.699	(.228)
y_{15}	n_4	.055	(.169)	.261	(.215)
y_{16}	n_4	.067	(.186)	.271	(.253)
y_{17}	n_4	-.079	(.194)	-.137	(.258)
y_{18}	n_4	-.220	(.177)	.103	(.235)

correlations are significant as in CANO($\mathbf{X}_4$, $\mathbf{Y}_4$). Table 5 presents the rotated canonical structure on the two dimensions. The pattern of correlations is very similar to that in CANO($\mathbf{X}_4$, $\mathbf{Y}_4$), again indicating that the

Table 4
Rotated Canonical Structure and the Corresponding Standard Errors in the Parenthesis Obtained from CANO($\mathbf{X}_3$, $\mathbf{Y}_3$) of the Friendship Data

Item		Dim = 1		Dim = 2		Dim = 3	
x_1	p_1	.601	(.129)	.395	(.157)	.490	(.114)
x_2	p_1	.807	(.203)	.327	(.273)	.137	(.197)
x_3	p_1	.823	(.163)	.299	(.194)	.278	(.173)
x_4	p_1	.470	(.174)	.743	(.231)	.382	(.172)
x_5	p_2	.595	(.139)	.270	(.185)	.605	(.156)
x_6	p_2	.184	(.245)	.468	(.264)	.730	(.254)
x_7	p_2	.358	(.211)	.738	(.267)	.280	(.212)
x_8	p_2	.310	(.214)	.661	(.282)	.482	(.210)
y_1	n_1	-.508	(.166)	-.274	(.185)	-.302	(.174)
y_2	n_1	-.549	(.137)	-.224	(.183)	-.445	(.251)
y_3	n_1	-.354	(.158)	-.441	(.192)	-.097	(.175)
y_4	n_2	-.253	(.181)	-.078	(.229)	-.574	(.201)
y_5	n_2	-.152	(.206)	.100	(.242)	-.464	(.230)
y_6	n_2	-.368	(.165)	-.080	(.202)	-.200	(.167)
y_7	n_2	-.714	(.206)	.086	(.229)	-.195	(.208)
y_8	n_2	-.155	(.142)	-.195	(.187)	-.407	(.161)
y_9	n_2	-.270	(.151)	.012	(.190)	-.039	(.162)
y_{10}	n_2	-.194	(.173)	-.044	(.202)	-.498	(.210)
y_{11}	n_2	-.167	(.163)	-.093	(.189)	-.255	(.184)
y_{12}	n_3	-.028	(.170)	.099	(.200)	.203	(.167)
y_{13}	n_3	-.214	(.158)	-.223	(.197)	.135	(.191)
y_{14}	n_3	-.041	(.203)	.394	(.241)	.029	(.196)
y_{15}	n_4	.299	(.191)	.084	(.234)	.078	(.213)
y_{16}	n_4	.126	(.182)	.114	(.221)	.057	(.204)
y_{17}	n_4	.132	(.186)	-.045	(.222)	-.242	(.218)
y_{18}	n_4	.165	(.212)	-.331	(.283)	-.191	(.263)

effect of **G** is quite minor, although the two canonical correlations are clearly smaller than those obtained in CANO($\mathbf{X}_4$, $\mathbf{Y}_4$). We can thus interpret the canonical variates in a manner similar to CANO($\mathbf{X}_4$, $\mathbf{Y}_4$).

Table 5

Rotated Canonical Structure and the Corresponding Standard Errors in the Parenthesis Obtained from CANO($\mathbf{X}_8$, $\mathbf{Y}_8$) of the Friendship Data

Item		Dim = 1		Dim = 2	
x_1	p_1	.634	(.054)	.582	(.058)
x_2	p_1	.871	(.083)	.318	(.118)
x_3	p_1	.850	(.075)	.363	(.110)
x_4	p_1	.693	(.055)	.539	(.067)
x_5	p_2	.511	(.090)	.716	(.080)
x_6	p_2	.315	(.145)	.867	(.119)
x_7	p_2	.462	(.110)	.769	(.099)
x_8	p_2	.350	(.137)	.864	(.134)
y_1	n_1	-.724	(.109)	-.197	(.145)
y_2	n_1	-.754	(.076)	-.311	(.120)
y_3	n_1	-.671	(.119)	-.028	(.174)
y_4	n_2	-.528	(.104)	-.237	(.158)
y_5	n_2	-.293	(.126)	-.198	(.194)
y_6	n_2	-.401	(.107)	-.225	(.168)
y_7	n_2	-.488	(.101)	-.200	(.163)
y_8	n_2	-.423	(.116)	-.206	(.178)
y_9	n_2	-.360	(.108)	-.223	(.174)
y_{10}	n_2	-.410	(.102)	-.099	(.172)
y_{11}	n_2	-.323	(.129)	-.053	(.212)
y_{12}	n_3	-.246	(.173)	.670	(.202)
y_{13}	n_3	-.481	(.142)	.418	(.200)
y_{14}	n_3	-.381	(.181)	.703	(.232)
y_{15}	n_4	.111	(.155)	.200	(.209)
y_{16}	n_4	.138	(.187)	.295	(.248)
y_{17}	n_4	-.012	(.192)	.061	(.252)
y_{18}	n_4	-.171	(.166)	.043	(.227)

The 1997 Canadian Election Data

The second example comes from a large-scale survey data set collected by the Institute for Social Research at York University to investigate political opinions of Canadians during the 1997 federal election campaign. Telephone interviews were given to randomly selected Canadian citizens of voting age, which began no later than four days after the election writs were issued and terminated at the last day of the campaign.

Three items were selected from the data set for our analysis. Item 1 asked the province where respondents live, item 2 asked what federal party respondents would vote for in the election, and item 3 asked the educational level of respondents. For simplicity, we removed rarely chosen categories from the items. Item 1 consisted of 10 Canadian provinces from East to West (1 = Newfoundland, 2 = Prince Edward Island, 3 = Nova Scotia, 4 = New Brunswick, 5 = Quebec, 6 = Ontario, 7 = Manitoba, 8 = Saskatchewan, 9 = Alberta, 10 = British Columbia). Item 2 comprised 5 federal parties (1 = Liberal, 2 = Progressive Conservative, 3 = New Democrats, 4 = Reform, and 5 = Bloc Quebecois). They may be ordered from politically right to left as follows: Reform (extreme-right), Progressive Conservative (right), Liberal (center), and New Democrats (central-left). However, it is difficult to classify Bloc Quebecois in terms of its political orientation since it is only organized to achieve sovereignty of Quebec from Canada. (Note that a large proportion of Quebec residents support independence of Quebec.) Item 3 included 13 different levels of education from no schooling to Ph.D. The sample size was 2121.

We used item 1 as $\mathbf{X}$ and item 2 as $\mathbf{Y}$ in order to explore relationships between respondents' provinces and their preference for federal parties. We used item 3 as $\mathbf{G}$ to investigate the effect of educational levels on the relationship between $\mathbf{X}$ and $\mathbf{Y}$.

In Canada, political orientations are often geographically linked. Residents of the Western provinces are known to be more inclined toward (extreme) right wing, while those of the East Coast provinces tend to lean toward liberals or political left. We thus presumed that the Western provinces might be grouped into a single category because they tend to share similar political attitudes. Likewise, the East Coast provinces might be combined into a single category. Then, we specified an equality constraint matrix for $\mathbf{X}$, denoted by $\mathbf{H}_X$, as follows

$$\mathbf{H}_X = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

The first column of $\mathbf{H}_X$ indicate the equality among four East Coast provinces (i.e., Newfoundland, Prince Edward Island, Nova Scotia, and New Brunswick) whereas the last column indicates the equality between two Western provinces (i.e., Alberta and British Columbia). Columns 2 to 5 indicate that the four remaining provinces (Quebec, Ontario, Manitoba, and Saskatchewan) are distinct. It was difficult to come up with any specific relationships among the parties due to their distinct political orientations. We therefore did not consider any constraint matrix for $\mathbf{Y}$, that is, $\mathbf{H}_Y = \mathbf{I}$.

The two sets of data were decomposed according to their external information. CANO was then applied to a number of pairs of components from each data set. As in Example 1, we present four cases, CANO($\mathbf{X}_1$, $\mathbf{Y}_1$), CANO($\mathbf{X}_4$, $\mathbf{Y}_4$), CANO($\mathbf{X}_3$, $\mathbf{Y}_3$), and CANO($\mathbf{X}_8$, $\mathbf{Y}_8$), corresponding to the unconstrained CA, CCA, partial CA, and partial CCA, respectively. The squared canonical correlations estimated from each case are presented in Table 6, along with their significance levels obtained from the permutation method.

Table 6 shows that all four canonical correlations are significant in all four cases, despite the fact that the last two squared canonical correlations

Table 6

The Squared Canonical Correlations and the Corresponding Empirical Significance Levels in the Parenthesis Obtained from CANO($\mathbf{X}_1$, $\mathbf{Y}_1$), CANO($\mathbf{X}_4$, $\mathbf{Y}_4$), CANO($\mathbf{X}_3$, $\mathbf{Y}_3$), and CANO($\mathbf{X}_8$, $\mathbf{Y}_8$) of the 1997 Canadian Election Data

CANO($\mathbf{X}_1$, $\mathbf{Y}_1$)		CANO($\mathbf{X}_4$, $\mathbf{Y}_4$)		CANO($\mathbf{X}_3$, $\mathbf{Y}_3$)		CANO($\mathbf{X}_8$, $\mathbf{Y}_8$)	
.436	(.000)	.435	(.000)	.436	(.000)	.432	(.000)
.110	(.000)	.103	(.000)	.111	(.000)	.104	(.000)
.036	(.000)	.024	(.000)	.036	(.000)	.023	(.000)
.020	(.000)	.011	(.000)	.019	(.000)	.011	(.000)

look quite small. This is likely due to a large sample size. In fact, the first two squared canonical correlations explained about 90% or more of the sum of squared canonical correlations in all cases. This indicates that a two-dimensional solution captures the most of the important variations among item categories in all cases.

In CA, the structure of the associations between two sets of variables, which correspond with rows and columns of a contingency table, is conventionally examined with a graphical display of the rows and columns. The two-dimensional configuration for the estimated category points in CANO($\mathbf{X}_1$, $\mathbf{Y}_1$) is presented in Figure 1.

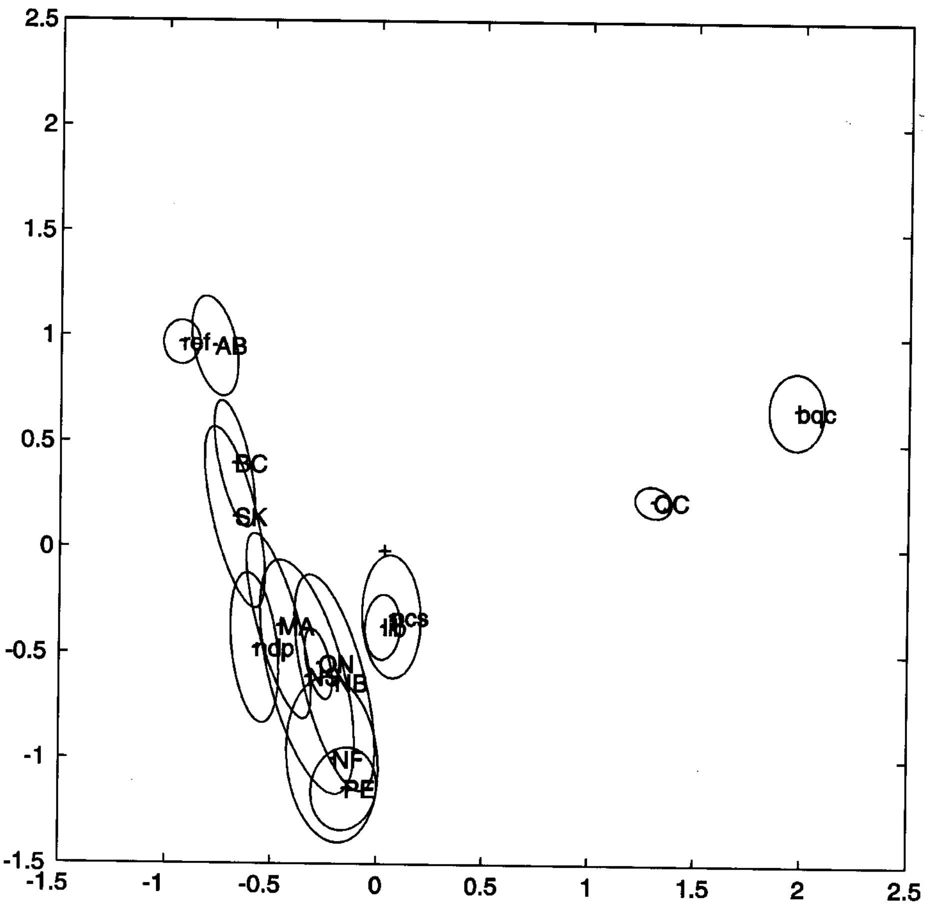

Figure 1
The Two-dimensional Category Point Configuration of the 1997 Canadian Election Data Derived from CANO($\mathbf{X}_1$, $\mathbf{Y}_1$), Along with 95% Confidence Regions

This case implies $\mathbf{G} = \mathbf{I}$, $\mathbf{H}_X = \mathbf{I}$, and $\mathbf{H}_Y = \mathbf{I}$. The estimated category points of $\mathbf{X}$ are labeled as NF, PE, NS, NB, QC, ON, MA, SK, AB, and BC, and those of $\mathbf{Y}$ as lib, pcs, ndp, ref, and bqc. The order of the labels is equivalent to that of the categories in items 1 and 2.

The upper right-hand portion of the configuration seems to indicate con-federalism, represented by QC and bqc. It may be safe to say that residents of Quebec are more likely to vote for Bloc Quebecois than other parties in the election. The upper left-hand portion seems to represent a politically federalistic and extreme-rightist position, characterized by ref. Residents of Alberta (AB) are most likely to vote for the party. Those of British Columbia (BC) are also more likely to support this party than other parties. We may say that Reform party is mainly preferred by the Western provinces. On the other hand, the lower left-hand portion seems to stand for a federalistic and more central or central-left position, featured by pcs, lib, and ndp. Ontario (ON) (that plays a crucial role in deciding a majority party due to its largest population in Canada) seems to support the three parties equally, while Manitoba (MA) seems to lean slightly more toward New Democrats party. Other East Coast provinces, such as Newfoundland (NF), Prince Edward Island (PE), Nova Scotia (NS), and New Brunswick (NB), also show fair support for the three parties.

Figure 2 displays the two-dimensional configuration of the estimated category points from CANO($\mathbf{X}_4$, $\mathbf{Y}_4$).

This case implies that $\mathbf{G} = \mathbf{I}$, $\mathbf{H}_X \neq \mathbf{I}$, and $\mathbf{H}_Y = \mathbf{I}$. In the figure, due to the imposition of the equality constraints on $\mathbf{X}$, such categories as NF, PE, NS, and NB were assigned the same value, and they were displayed under a single label, that is, EC (that stands for East Coast). Similarly, AB and BC were single-labeled as WE (that stands for West). Interpretations of the figure seem to be essentially the same as the unconstrained case. This indicates that the equality restrictions do not dramatically change the solution. This may help us simplify our interpretations, significantly reducing the number of parameters. In addition, 95% confidence regions (Ramsay, 1978) obtained by the bootstrap method (Efron, 1979) on average tended to be smaller than those from the unconstrained case. This provides additional evidence that our assumptions regarding the column structure of $\mathbf{X}$ seem to be reasonable.

GCCANO was also applied with $\mathbf{H}_X = \mathbf{I}$, $\mathbf{H}_Y = \mathbf{I}$, and $\mathbf{G} \neq \mathbf{I}$ [i.e., CANO($\mathbf{X}_3$, $\mathbf{Y}_3$)]. This case is equivalent to partial CA. Figure 3 presents the two-dimensional display of the category points estimated under partial CA.

It is found that Alberta and British Columbia show loyalty to Reform party, and Quebec to Bloc Quebecois, regardless of educational levels. We may say that educational backgrounds have little effects on party

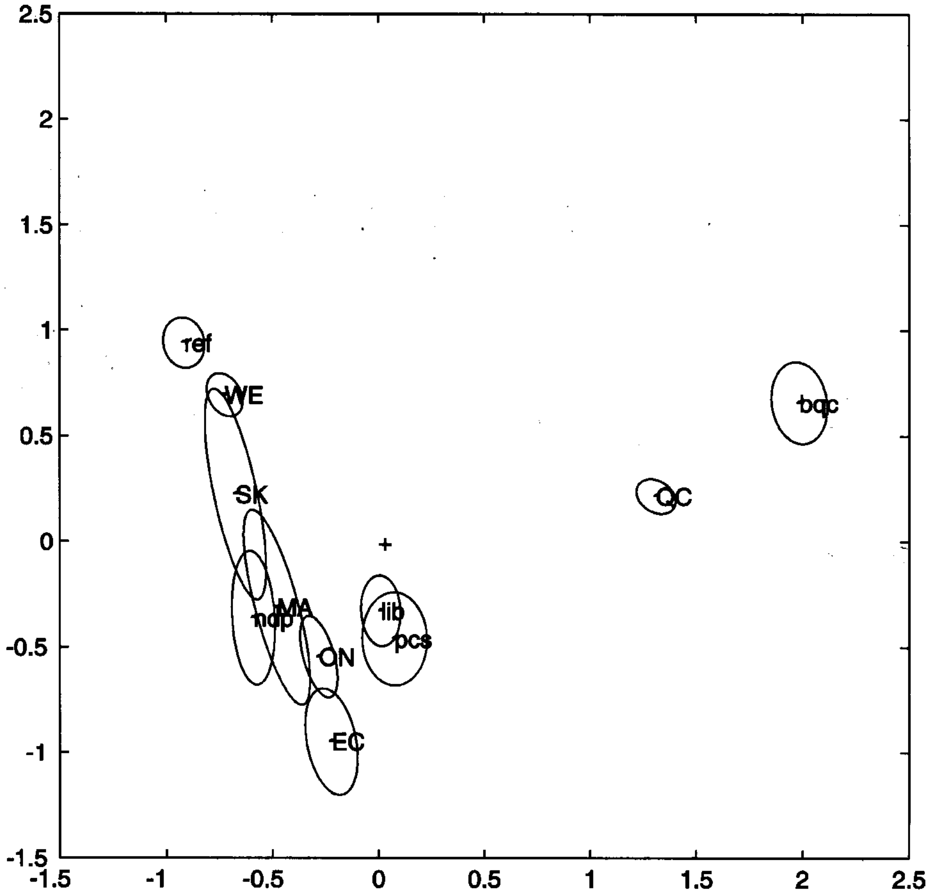

Figure 2
The Two-dimensional Category Point Configuration of the 1997 Canadian Election Data Derived from CANO($\mathbf{X}_4$, $\mathbf{Y}_4$), Along with 95% Confidence Regions

preferences of the residents of those provinces. East Coast provinces such as Nova Scotia and New Brunswick, on the other hand, turn out to be more inclined toward Liberal and Progressive Conservative parties, after eliminating the effect of levels of education. It implies that party preferences of the residents of some East Coast provinces are more closely related to their educational levels.

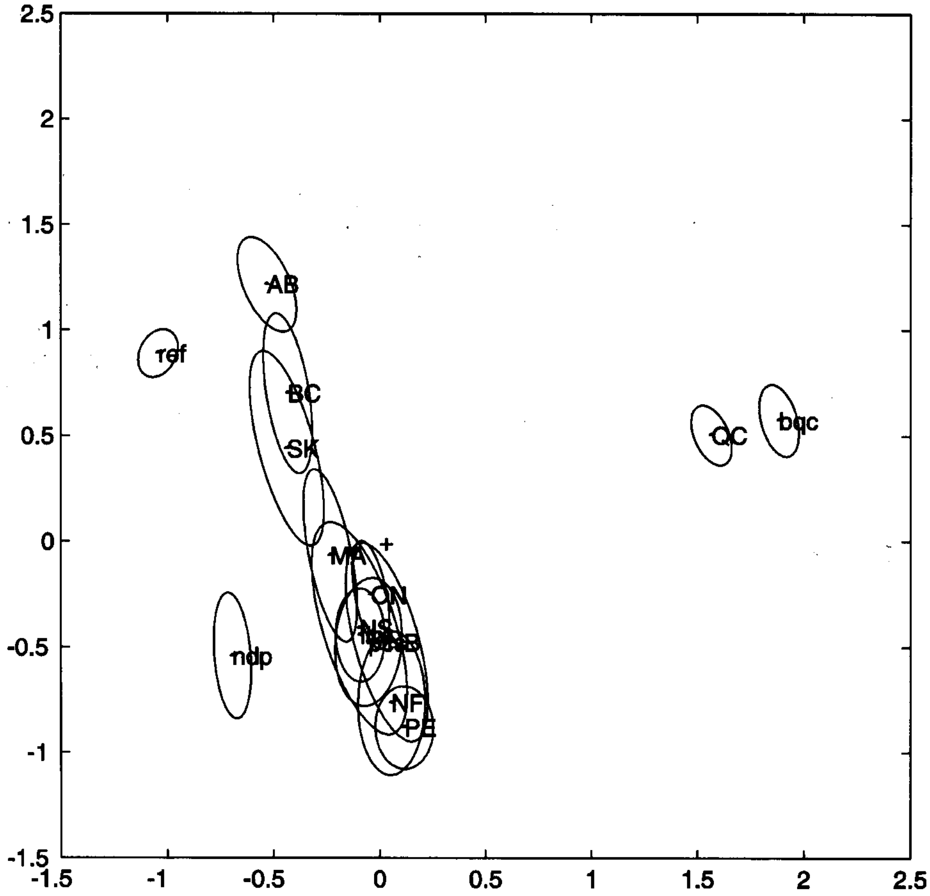

Figure 3
The Two-dimensional Category Point Configuration of the 1997 Canadian Election Data Derived from CANO($\mathbf{X}_3$, $\mathbf{Y}_3$), Along with 95% Confidence Regions

Finally, we applied GCCANO with $\mathbf{H}_X \neq \mathbf{I}$, $\mathbf{H}_Y = \mathbf{I}$ and $\mathbf{G} \neq \mathbf{I}$ [CANO($\mathbf{X}_8$, $\mathbf{Y}_8$)]. Figure 4 shows that the two-dimensional configuration of the estimated category points obtained from CANO($\mathbf{X}_8$, $\mathbf{Y}_8$).

In the display, the categories corresponding to AB and BC were assigned the same values due to $\mathbf{H}_X$, and were labeled in the same way as CANO($\mathbf{X}_4$, $\mathbf{Y}_4$). It was the case for the categories corresponding to NF, PE, NS, and NB. Due to the partialing out of the effect of $\mathbf{G}$, in addition, the single-labeled

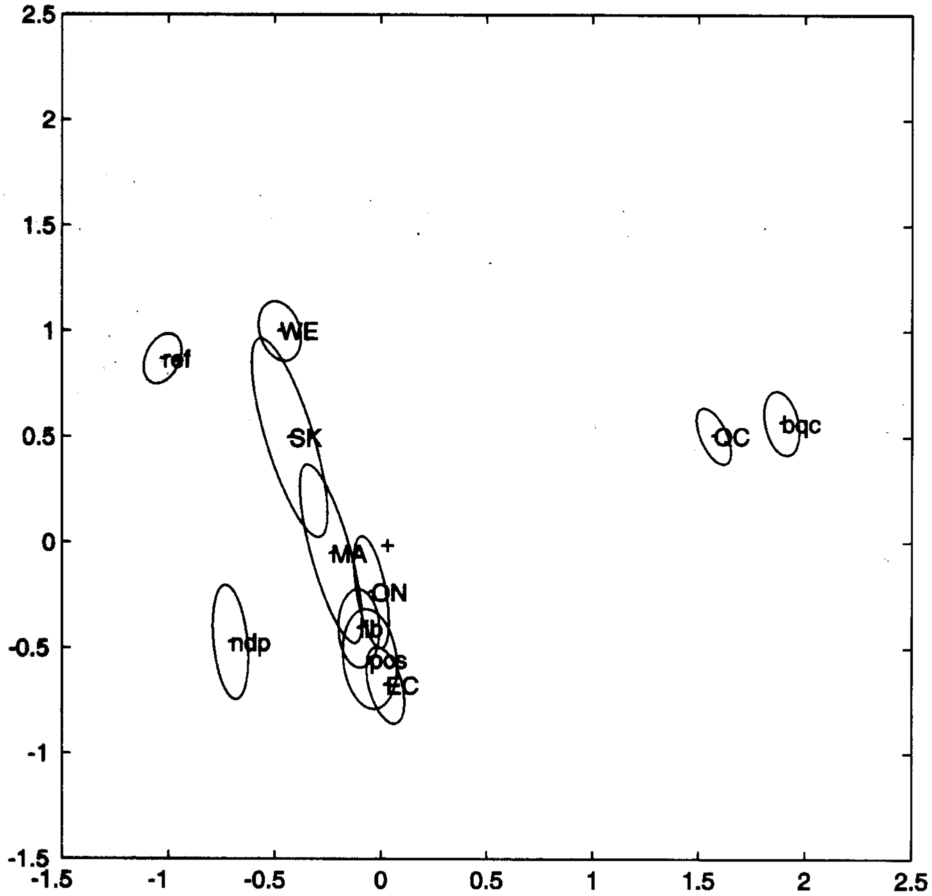

Figure 4
The Two-dimensional Category Point Configuration of the 1997 Canadian Election Data Derived from CANO($\mathbf{X}_8$, $\mathbf{Y}_8$), Along with 95% Confidence Regions

category point to represent the East Coast provinces are more closely located with Liberal and Progressive Conservative parties than New Democrats party. This is consistent with the results of partial CA.

Discussion

In this article, we proposed GCCANO, a method for CANO that could incorporate external information on both rows and columns of two data matrices. This method is quite general, and subsumes a broad range of existing methods. The usefulness of the method was demonstrated with two examples.

As in the usual CANO, the interpretation of the canonical variates obtained from GCCANO can be difficult, and a simple structure rotation of either the canonical pattern or the structure matrix (e.g., by varimax) may be useful to make it easier to interpret. We can also combine other techniques, such as canonical ridge regression (Vinod, 1976), complementary uses of PCA or factor analysis for CANO (e.g., Stevens, 1996), etcetera with GCCANO in order to obtain more interpretable and reliable solutions.

GCCANO is primarily a descriptive method to examine the association between two sets of variables. However, certain types of statistical inferences (e.g., hypothesis testing, assessment of stability, etc.) are possible with the use of the bootstrap method. For instance, interaction effects between variables, which are ignored in the usual CA, can be empirically tested under the additivity hypothesis, as illustrated in the second example of the "Applications" section. Permutation tests can be applied to test the significance of canonical correlations obtained from GCCANO, without any distributional assumptions on the data sets. It allows us to determine the number of significant canonical correlations. When certain distributional assumptions such as multivariate normality, independence of observations, etcetera, are met by the data, traditional parametric methods exist for such significance tests (e.g., Wilks' Λ based on the LR statistic, Roy's max λ, trace criterion, etc.). However, these parametric procedures are still limited, even if the distributional assumptions are satisfied, in that they can only be applied to a subset of analyses of GCCANO (e.g., Suzukawa, 1997).

In the present method, only one set of linear constraints is incorporated into each side (row and column) of a data matrix. However, it may be useful to impose different sets of constraints on different dimensions (DCDD). This type of constraints may, in some cases, yield more meaningful analyses since it often better captures the essence of original empirical hypotheses. Takane et al. (1995) presented the use of the DCDD type of constraints in principal component analysis, which may be extended to CANO. GCCANO follows a similar approach to CPCA in analyzing the relationship between

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

ter Braak, C. J. F. (1986). Canonical correspondence analysis: A new eigenvalue technique for multivariate direct gradient analysis. *Ecology, 67*, 1167-1179.

ter Braak, C. J. F. (1990). *Update notes: CANOCO version 3.10.* Wageningen, The Netherlands: Agricultural Mathematics Group.

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

Vinod, H. D. (1976). Canonical ridge and econometrics of joint production. *Journal of Econometrics, 4*, 147-166.

Yanai, H. (1986). Some generalizations of correspondence analysis in terms of projection operators. In E. Diday, Y. Escoufier, L. Lebart, J. P. Pagès, Y. Schektman, & R. Thomassone (Eds.), *Data analysis and informatics IV* (pp. 193-207). Amsterdam: North Holland.

Yanai, H. (1990). Some generalized forms of least squares g-inverse, minimum norm g-inverse and Moore-Penrose inverse matrix. *Computational Statistics and Data Analysis, 10*, 251-260.

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

Accepted March, 2001.

Appendix A

We derive decomposition Equation 1. Consider estimating **M**, **B**, **C** and **E** in the following model,

$$\text{(A1)} \qquad \mathbf{X} = \mathbf{G}_X\mathbf{M}\mathbf{H}_X'\mathbf{X}'\mathbf{X} + \mathbf{G}_X\mathbf{B} + \mathbf{C}\mathbf{H}_X'\mathbf{X}'\mathbf{X} + \mathbf{E}$$

by minimizing $f = \mathrm{tr}[\mathbf{E}'\mathbf{E}(\mathbf{X}'\mathbf{X})^-]$ under the identification restrictions

$$\text{(A2)} \qquad \mathbf{B}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}\mathbf{H}_X = 0$$

and

$$\text{(A3)} \qquad \mathbf{G}_X'\mathbf{C} = 0.$$

By differentiating f with respect to **M** and setting the result equal to zero, we obtain

$$-\frac{1}{2}\frac{\partial f}{\partial \mathbf{M}} = \mathbf{G}_X'\left(\mathbf{X} - \mathbf{G}_X\hat{\mathbf{M}}\mathbf{H}_X'\mathbf{X}'\mathbf{X} - \mathbf{G}_X\hat{\mathbf{B}} - \hat{\mathbf{C}}\mathbf{H}_X'\mathbf{X}'\mathbf{X}\right)(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}\mathbf{H}_X \equiv 0.$$

This leads to

(A4)
$$\hat{\mathbf{M}} = (\mathbf{G}'_X\mathbf{G}_X)^- \mathbf{G}'_X\mathbf{X}(\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{X}\mathbf{H}_X\left[\mathbf{H}'_X\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{X}\mathbf{H}_X\right]^- \\ = (\mathbf{G}'_X\mathbf{G}_X)^- \mathbf{G}'_X\mathbf{X}\mathbf{H}_X(\mathbf{H}'_X\mathbf{X}'\mathbf{X}\mathbf{H}_X)^-.$$

Similarly,

$$-\frac{1}{2}\frac{\partial f}{\partial \mathbf{B}} = \mathbf{G}'_X\left(\mathbf{X} - \mathbf{G}_X\hat{\mathbf{M}}\mathbf{H}'_X\mathbf{X}'\mathbf{X} - \mathbf{G}_X\hat{\mathbf{B}} - \hat{\mathbf{C}}\mathbf{H}'_X\mathbf{X}'\mathbf{X}\right)(\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{X}\mathbf{H}_X \equiv 0.$$

which leads to

(A5)
$$\hat{\mathbf{B}} = (\mathbf{G}'_X\mathbf{G}_X)^- \mathbf{G}'_X\mathbf{X}\mathbf{Q}_{H_X},$$

where $\mathbf{Q}_{H_X} = \mathbf{I} - \mathbf{P}_{H_X}$ with $\mathbf{P}_{H_X} = \mathbf{H}_X(\mathbf{H}'_X\mathbf{X}'\mathbf{X}\mathbf{H}_X)^-\mathbf{H}'_X\mathbf{X}'\mathbf{X}$. (As in the main texts, the metric matrix $\mathbf{X}'\mathbf{X}$ was suppressed from $\mathbf{P}_{H_X/X'X}$ and $\mathbf{Q}_{H_X/X'X}$). Similarly,

(A6)
$$\hat{\mathbf{C}} = \mathbf{Q}_{G_X}\mathbf{X}\mathbf{H}_X(\mathbf{H}'_X\mathbf{X}'\mathbf{X}\mathbf{H}_X)^-.$$

and

(A7)
$$\hat{\mathbf{E}} = \mathbf{X} - \mathbf{G}_X\hat{\mathbf{M}}\mathbf{H}'_X\mathbf{X}'\mathbf{X} - \mathbf{G}_X\hat{\mathbf{B}} - \hat{\mathbf{C}}\mathbf{H}'_X\mathbf{X}'\mathbf{X} \\ = \mathbf{X} - \mathbf{P}_{G_X}\mathbf{X}\mathbf{P}_{H_X} - \mathbf{P}_{G_X}\mathbf{X}\mathbf{Q}_{H_X} - \mathbf{Q}_{G_X}\mathbf{X}\mathbf{P}_{H_X}.$$

If we put Equations A4 through A7 in Equation A1, we obtain Equation 1. Equation 2 can be derived similarly.

Appendix B

CANO($\mathbf{X}_4$, $\mathbf{Y}_4$) amounts to GSVD$[(\mathbf{X}'_4\mathbf{X}_4)^-\mathbf{X}'_4\mathbf{Y}_4(\mathbf{Y}'_4\mathbf{Y}_4)^-]_{X'_4X_4,Y'_4Y_4}$. We show that this GSVD reduces to the GSVD given in Equation 13. The following results are useful in the sequel.

Let $\mathbf{P}_{A/K} = \mathbf{A}(\mathbf{A}'\mathbf{K}\mathbf{A})^-\mathbf{A}'\mathbf{K}$ be the projector onto Sp(**A**) along Ker(**A**′**K**), where **A** is a matrix of predictor variables, and **K** is an *nnd* (non-negative definite) metric matrix satisfying the condition analogous to Equations 7 and 8. Then, (A8) $$\mathbf{P}'_{A/K}\mathbf{K}\mathbf{P}_{A/K} = \mathbf{P}'_{A/K}\mathbf{K} = \mathbf{K}\mathbf{P}_{A/K}$$

(A9) $$\mathbf{P}_{A/K} \in \{\mathbf{P}^-_{A/K}\}, \text{ and } \mathbf{P}'_{A/K} \in \{(\mathbf{P}'_{A/K})^-\}$$

(A10) $$(\mathbf{P}'_{A/K}\mathbf{K})^- = \mathbf{K}^-\mathbf{P}'_{A/K}, \text{ and } (\mathbf{K}\mathbf{P}_{A/K})^- = \mathbf{P}_{A/K}\mathbf{K}^-,$$

where $^-$ denotes a generalized inverse.

Equations A8 and A9 are trivial. Equation A10 may require a proof. A necessary and sufficient condition for $(\mathbf{BC})^- = \mathbf{C}^-\mathbf{B}^-$ is that $\mathbf{B}^-\mathbf{BCC}^-$ is idempotent (Searle, 1971, p.28). We check to see if $\mathbf{B} = \mathbf{P}'_{A/K}$ and $\mathbf{C} = \mathbf{K}$ satisfies this condition:

(A11) $$\begin{aligned}(\mathbf{P}'_{A/K}\mathbf{K}\mathbf{K}^-)^2 &= \mathbf{KA}(\mathbf{A}'\mathbf{KA})^-\mathbf{A}'\mathbf{KK}^-\mathbf{KA}(\mathbf{A}'\mathbf{KA})^-\mathbf{A}'\mathbf{KK}^- \\ &= \mathbf{KA}(\mathbf{A}'\mathbf{KA})^-\mathbf{A}'\mathbf{KK}^- \\ &= \mathbf{P}'_{A/K}\mathbf{KK}^-.\end{aligned}$$

It can be easily shown that

(A12) $$(\mathbf{X}_4'\mathbf{X}_4)^-\mathbf{X}_4'\mathbf{Y}_4(\mathbf{Y}_4'\mathbf{Y}_4)^- = (\mathbf{X}'\mathbf{X})^-\mathbf{P}'_{H_X/X'X}\mathbf{X}'\mathbf{Y}\mathbf{P}_{H_Y/Y'Y}(\mathbf{Y}'\mathbf{Y})^-,$$

(A13) $$\mathbf{X}_4'\mathbf{X}_4 = \mathbf{X}'\mathbf{X}\mathbf{H}_X(\mathbf{H}_X'\mathbf{X}'\mathbf{X}\mathbf{H}_X)^-\mathbf{H}_X'\mathbf{X}'\mathbf{X}\mathbf{H}_X(\mathbf{H}_X'\mathbf{X}'\mathbf{X}\mathbf{H}_X)^-\mathbf{H}_X'\mathbf{X}'\mathbf{X},$$

and

(A14) $$\mathbf{Y}_4'\mathbf{Y}_4 = \mathbf{Y}'\mathbf{Y}\mathbf{H}_Y(\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y)^-\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y(\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y)^-\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y},$$

using the above results. The GSVD$[(\mathbf{X}_4'\mathbf{X}_4)^-\mathbf{X}_4'\mathbf{Y}_4(\mathbf{Y}_4'\mathbf{Y}_4)^-]_{X_4'X_4, Y_4'Y_4}$ reduces to that of

(A15) $$\begin{aligned}&(\mathbf{H}_X'\mathbf{X}'\mathbf{X}\mathbf{H}_X)^-\mathbf{H}_X'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}\mathbf{H}_X(\mathbf{H}_X'\mathbf{X}'\mathbf{X}\mathbf{H}_X)^-\mathbf{H}_X'\mathbf{X}'\mathbf{Y}\mathbf{H}_Y \\ &\times(\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y)^-\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^-\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y(\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y)^- \\ &= (\mathbf{H}_X'\mathbf{X}'\mathbf{X}\mathbf{H}_X)^-\mathbf{H}_X'\mathbf{X}'\mathbf{Y}\mathbf{H}_Y(\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y)^-,\end{aligned}$$

with metric matrices $\mathbf{H}_X'\mathbf{X}'\mathbf{X}\mathbf{H}_X$ and $\mathbf{H}_Y'\mathbf{Y}'\mathbf{Y}\mathbf{H}_Y$. This is obtained by including parts of the metric matrices, $\mathbf{X}_4'\mathbf{X}_4$ and $\mathbf{Y}_4'\mathbf{Y}_4$, in the matrix whose GSVD is to be obtained. Similarly, when both $\mathbf{X}$ and $\mathbf{Y}$ consist of discrete data, GSVD$[(\mathbf{X}_4'\mathbf{X}_4)^-\mathbf{X}_4'\mathbf{Y}_4(\mathbf{Y}_4'\mathbf{Y}_4)^-]_{X_4'X_4, Y_4'Y_4}$ reduces to the GSVD in Equation 16 (and Equation 17).

Appendix C
The Items Used in the Friendship Data (Koh, Mendelson, & Rhee, 1998)

The following items concern your CLOSEST SAME-SEX FRIEND. To begin with, please select your closest same-sex friend and write the friend's number from the list here _____. Throughout the form, read "_____" as the name of your closest friend.

This part concerns your POSITIVE FEELINGS for your CLOSEST SAME-SEX FRIEND. On the scale to the right of each item circle the number that indicates how much you experience the feelings described in the item. The feelings for friends differ from person to person, so just honestly describe your feelings.

			not really	a little	quite a bit	a lot	very much	as much as possible
x_1	p_1	I am happy with my friendship with _____.	0	1	2	3	4	5
x_2	p_1	I feel my friendship with _____ is a great one.	0	1	2	3	4	5
x_3	p_1	I am satisfied with my friendship with _____.	0	1	2	3	4	5
x_4	p_1	I think my friendship with _____ is strong.	0	1	2	3	4	5
x_5	p_2	I like _____ a lot.	0	1	2	3	4	5
x_6	p_2	I want to stay friends with _____ for a long time.	0	1	2	3	4	5
x_7	p_2	I feel close to _____.	0	1	2	3	4	5
x_8	p_2	I hope _____ and I will stay friends.	0	1	2	3	4	5

Individuals may also experience some NEGATIVE FEELING for a friend. On the scale to the right of each item circle the number that indicates how often you have that NEGATIVE FEELING for your CLOSEST SAME-SEX FRIEND. Remember, adult's feelings for friends differ from person to person and from time to time, so just honestly describe your feelings.

			never		rarely		once in a while		fairly often		often
y_1	n_1	I feel aloof from _____.	0	1	2	3	4	5	6	7	8
y_2	n_1	I have ambivalent feelings about _____.	0	1	2	3	4	5	6	7	8
y_3	n_1	I am uncertain about _____.	0	1	2	3	4	5	6	7	8
y_4	n_2	I feel bothered by _____.	0	1	2	3	4	5	6	7	8
y_5	n_2	I feel controlled by _____.	0	1	2	3	4	5	6	7	8
y_6	n_2	I am disagreeable with _____.	0	1	2	3	4	5	6	7	8
y_7	n_2	I am dissatisfied with _____.	0	1	2	3	4	5	6	7	8
y_8	n_2	I feel inhibited by _____.	0	1	2	3	4	5	6	7	8
y_9	n_2	I feel insulted by _____.	0	1	2	3	4	5	6	7	8
y_{10}	n_2	I feel quarrelsome with _____.	0	1	2	3	4	5	6	7	8
y_{11}	n_2	I feel restricted by _____.	0	1	2	3	4	5	6	7	8
y_{12}	n_3	I feel responsible for _____.	0	1	2	3	4	5	6	7	8
y_{13}	n_3	I feel sorry for _____.	0	1	2	3	4	5	6	7	8
y_{14}	n_3	I am worried about _____.	0	1	2	3	4	5	6	7	8
y_{15}	n_4	I feel dependent on _____.	0	1	2	3	4	5	6	7	8
y_{16}	n_4	I am envious of _____.	0	1	2	3	4	5	6	7	8
y_{17}	n_4	I feel inferior to _____.	0	1	2	3	4	5	6	7	8
y_{18}	n_4	I am jealous of _____.	0	1	2	3	4	5	6	7	8