# A Method of Canonical Correlation Analysis for Ordinal Data

By

Hidekiyo ITAKURA* and Yoshikazu NISHIKAWA*

(Received August 14, 1981)

**Abstract**

A method of canonical correlation analysis is discussed, in which the order of magnitude is constrained among partial regression coefficients. Each variate is qualitative, having some categories which are on an ordinal scale. The problem is to solve a constrained nonlinear optimization problem, in which the objective function is the canonical correlation, and the constraints are linear ones imposed on the partial regression coefficients. Since the problem cannot be solved through a conventional eigenvalue problem, an iterative procedure is adopted with successively solving linear programs. This technique works effectively for some types of canonical correlation analysis.

## 1. Introduction

The techniques of linear canonical correlation analysis are very useful for multivariate analysis methods, but conventional techniques are sometimes unsuitable for analyzing certain types of statistical data. Among them is a type of data concerned with the qualitative attributes of ordered categories. It is sometimes found that results from a formal application of the conventional techniques to the data bewilder us in trying to interpret them. One of the bewildering points is that the order of values of partial regression coefficients given to the categories of each attribute seems unnatural, at least from the viewpoint of deriving a meaning of the canonical correlation under study.

This paper is concerned with a class of linear canonical correlation techniques. In order to avoid such unnaturalness, the order of magnitude is constrained among the values of the partial regression coefficients. Let the categories of an attribute be placed on an ordinal scale. Depending on the properties of the attributes and categories in the problem under study, although there may be a variety of order relations, one of the relatively typical types is considered in this paper.

In the next section, a conventional linear canonical correlation technique for quantitative variates is reviewed. Section 3, which is the principal part of this

* Department of Electrical Engineering.

paper, is concerned with a procedure of the canonical correlation analysis modified for ordered attributes. First, the standard quantification method for qualitative data is described, based on the canonical correlation. The significance of the presented modified technique is then explained. Finally, after an equational formulation for the technique is made, its iterative computational procedure is presented, along with the solving of successive linear programming problems. Section 4 illustrates an application of the method to real poll data associated with Kyoto City's citizens' evaluations regarding their satisfaction with public transportation facilities.

## 2. Conventional Canonical Correlation Analysis

This chapter reviews the technique of the canonical correlation analysis for quantitative variates. The technique is for analyzing the correlation between the statistical variates of one set and those of the other set. When we have the two sets of statistical variates, we may wish to study the interrelations between the sets. If the two sets are very large, we may wish to consider only a few functions of each set. Then we shall want to study those functions most highly correlated. The simplest form of the functions is linear in variates. A linear function implies that we find a new coordinate system in the space of each set of variates. Thus, we find the linear combinations of variates in each set that have a maximum correlation. The canonical correlation technique is one of the general multivariate techniques, and it includes various techniques for special cases: multiple regression analysis, discriminant function analysis, principal component analysis, and so on.

It is assumed that the $m$ and $n$ statistical variates are under study in the first and the second set, respectively, and that their values are observed from $N$ samples. We introduce the observation matrices

$$\begin{aligned} &X = \{x_{ij}\}, \quad Y = \{y_{ij'}\} \\ &(i = 1, 2, \cdots, N;\ j = 1, 2, \cdots, m;\ j' = 1, 2, \cdots, n) \end{aligned} \tag{1}$$

where $x_{ij}$ and $y_{ij'}$ are the observed values of variates $j$ and $j'$ from sample $i$, respectively. Since we are interested only in variances and covariances, we shall assume that

$$\sum_{i=1}^{N} x_{ij} = \sum_{i=1}^{N} y_{ij'} = 0 \tag{2}$$

for any $j$ and $j'$. The sample covariance matrix is then

$$S=\begin{bmatrix} S_{XX} & S_{XY} \\ S_{XY}^T & S_{YY} \end{bmatrix}=\frac{1}{N-1}\begin{bmatrix} X^TX & X^TY \\ Y^TX & Y^TY \end{bmatrix} \tag{3}$$

where the superscript $T$ denotes the transpose of a vector or a matrix.

Consider the arbitrary linear combinations

$$u=Xa, \quad v=Yb \tag{4}$$

in a vector-matrix form. The constant vectors $a$ and $b$ of $m$ and $n$ elements, respectively, are called the partial regression coefficients, and $u$ and $v$ are called the canonical variates. Unless otherwise mentioned, all vectors are in a column form throughout the paper. We need a $u$ and $v$ that have a maximum correlation. Since the correlation of a multiple of $u$ and a multiple of $v$ is the same as the correlation of $u$ and $v$, we therefore require $a$ and $b$ to be such that $u$ and $v$ have a unit variance:

$$1=\frac{1}{N-1}||u||^2=a^TS_{XX}a\,, \quad 1=\frac{1}{N-1}||v||^2=b^TS_{YY}b \tag{5}$$

where the symbol $||\cdot||$ denotes the Euclidean norm of a vector. We note that

$$\sum_{i=1}^{N}u_i=\sum_{i=1}^{N}v_i=0 \tag{6}$$

with the $i$th elements $u_i$ and $v_i$ of $u$ and $v$, respectively.

Then the correlation, called the canonical correlation, between $u$ and $v$ is

$$r=\frac{1}{N-1}u^Tv=a^TS_{XY}b \tag{7}$$

The problem is to find $a$ and $b$ to maximize $r$ subject to Eq. (5). It is well-known[1] to be reduced to the eigenvalue problem

$$(S_{XX}^{-1}S_{XY}S_{YY}^{-1}S_{XY}^T-\lambda I_m)a=0\,, \quad b=S_{YY}^{-1}S_{XY}^T a \tag{8}$$

under the condition of a positive definite $S$, where $I_k$ is the $k\times k$-identity matrix.

We can see $\lambda=r^2$. Since we want a maximum $r$, we take the eigenvector $a$ and associated $b$ normalized by Eq. (5), corresponding to the maximum eigenvalue $\lambda$ of the matrix $S_{XY}^{-1}S_{XY}S_{YY}^{-1}S_{XY}^T$. Then $u$ and $v$ are the optimal linear combinations of $X$ and $Y$, respectively, in the sense of their having a maximum correlation.

We can consider the second linear combinations $u'$ and $v'$, using the second maximum $\lambda$. They are such that, of all linear combinations uncorrelated with $u$ and $v$, these have a maximum correlation. This procedure can be continued step by step for any positive $\lambda$.

## 3. Canonical Correlation Analysis for Ordinal Data

### 3.1 Quantification method based on canonical correlation

In the last chapter, the values of the variates have been assumed to be quantitative by a tacit understanding. The statistical data to be analyzed are sometimes qualitative, for example, data concerned with measurements of human's attitudes or opinions in a questionnaire polling, or measurements of some kinds of subjective evaluations. Sometimes generically called the quantification method, there are various special statistical methods[2-4] for analyzing such qualitative data. Notwithstanding, each of several methods can be formulated as a special case of the quantification method based on a canonical correlation.

Hereafter, let us use the term attribute instead of variate, as is usually done in the statistical method for qualitative data. It is assumed that, as in the last chapter, the $m$ and $n$ attributes in a type of qualitative measurement are under study in the two sets, and that each one of the attributes consists of $k_j$ and $l_{j'}$ categories ($k_j$, $l_{j'} \geqq 2$; $j=1, 2, \cdots, m$; $j'=1, 2, \cdots, n$), respectively. Each one of the $N$ samples is to respond to some of the categories in each attribute. This response for the first set is expressed, for convenience, by the dummy variable

$$\delta_i(j, k) = \begin{cases} 1, & \text{if sample } i \text{ responds to category } k \text{ of attribute } j \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

for $k=1, 2, \cdots, k_j$; $j=1, 2, \cdots, m$; and $i=1, 2, \cdots, N$. In the same way, $\varepsilon_i(j', l)$ for the second set is defined with $l=1, 2, \cdots, l_{j'}$; $j'=1, 2, \cdots, n$.

We introduce the $N\times K$- and $N\times L$-matrices

$$D = (D_1, D_2, \cdots, D_m)\ , \quad E = (E_1, E_2, \cdots, E_n) \tag{10}$$

where $D_j$ is the $N\times k_j$-matrix whose $(i, k)$ element is $\delta_i(j, k)-\bar{\delta}(j, k)$, and $E_{j'}$ is the $N\times l_{j'}$-matrix whose $(i, l)$ element is $\varepsilon_i(j', l)-\bar{\varepsilon}(j', l)$, and

$$\begin{aligned} \bar{\delta}(j, k) &= \frac{1}{N}\sum_{i=1}^{N}\delta_i(j, k)\ , \quad K = \sum_{j=1}^{m} k_j \\ \bar{\varepsilon}(j', l) &= \frac{1}{N}\sum_{i=1}^{N}\varepsilon_i(j', l)\ , \quad L = \sum_{j'=1}^{n} l_{j'} \end{aligned} \tag{11}$$

Consider the linear combinations

$$\xi = D\alpha\ , \quad \eta = E\beta \tag{12}$$

The constant vectors $\alpha$ and $\beta$ are of $K$ and $L$ unknown elements, respectively, written as

$$\begin{aligned} \alpha &= (\alpha_1^T, \alpha_2^T, \cdots, \alpha_m^T)^T, \quad \alpha_j = (\alpha_{j1}, \alpha_{j2}, \cdots, \alpha_{jk_j})^T \\ \beta &= (\beta_1^T, \beta_2^T, \cdots, \beta_n^T)^T, \quad \beta_{j'} = (\beta_{j'1}, \beta_{j'2}, \cdots, \beta_{j'l_{j'}})^T \end{aligned} \tag{13}$$

The scalar $\alpha_{jk}$ is called the category score for the category $k$ of attribute $j$ and, unlike the variables $\delta_i$ and $\varepsilon_i$, has a quantitative value; and the same is said for $\beta_{j'l}$. Each element of $\xi$ and $\eta$ is called the sample score. As is understood, the combination coefficient, $\alpha$ or $\beta$, is assigned, not to each attribute, but to each category of the attributes, differing from the case of $a$ and $b$. We require $\xi$ and $\eta$ that have a maximum correlation. As has been mentioned for $u$ and $v$ in the last chapter, it can be required that $\xi$ and $\eta$ have a unit variance:

$$1 = \frac{1}{N-1}||\xi||^2 = \alpha^T F \alpha, \quad 1 = \frac{1}{N-1}||\eta||^2 = \beta^T G \beta \tag{14}$$

where

$$F = \frac{1}{N-1} D^T D, \quad G = \frac{1}{N-1} E^T E \tag{15}$$

Note that, from the definitions of $D$ and $E$, $\xi$ and $\eta$ have a zero mean.

Then the canonical correlation between $\xi$ and $\eta$ is

$$\rho = \frac{1}{N-1} \xi^T \eta = \alpha^T H \beta \tag{16}$$

where

$$H = \frac{1}{N-1} D^T E \tag{17}$$

Thus, the problem is to find $\alpha$ and $\beta$ to maximize $\rho$ subject to Eq. (14). In the same way for Eq. (8), the problem is reduced to the eigenvalue problem

$$(F^{-1}HG^{-1}H^T - \lambda I_K)\alpha = 0, \quad \beta = G^{-1}H^T\alpha \tag{18}$$

Let $\alpha'$ be any eigenvector corresponding to the maximum eigenvalue of $F^{-1}HG^{-1}H^T$. Then optimal $\alpha = \alpha^*$ and $\beta = \beta^*$ satisfying Eq. (14) are uniquely given by

$$\alpha^* = \frac{\alpha'}{\sqrt{\alpha'^T F \alpha'}}, \quad \beta^* = \frac{\beta'}{\sqrt{\beta'^T G \beta'}} \tag{19}$$

with $\beta' = G^{-1}H^T\alpha'$. In addition, it is possible, as shown in the last chapter, to consider the second linear combinations and so on, corresponding to the second maximum eigenvalue and so on.

Equation (18) is meaningful, of course, under the condition that both $F$ and

$G$ are nonsingular. However, this is not the case when the categories in each of some attributes in the first and/or the second set are mutually exclusive and exhaustive. That is to say, it is not the case when each of the samples has to respond to one and only one of the categories in each of the attributes, or equivalently in equational forms

$$\sum_{k=1}^{k_j} \delta_i(j, k) = 1 \quad \text{and/or} \quad \sum_{l=1}^{l_{j'}} \varepsilon_i(j', l) = 1 \tag{20}$$

for some $j$'s and/or $j'$'s and for any $i$. This causes $D$ or $E$ to be degenerate, and then $F$ or $G$ becomes inevitably singular, because the column vectors in $D_j$ or $E_{j'}$ are linearly dependent on one another. This situation takes place, not only when the categories are designed to be mutually exclusive and exhaustive, but also when the resulting data of the responses from the samples happen to satisfy Eq. (20).

At any rate, let Eq. (20) be satisfied for attributes $j=j_1, j_2, \cdots, j_\mu$ and $j'=j'_1, j'_2, \cdots, j'_\nu$. Then, we put

$$\begin{aligned} \alpha_{j_p1} &= 0 \qquad (p = 1, 2, \cdots, \mu) \\ \beta_{j'_q1} &= 0 \qquad (q = 1, 2, \cdots, \nu) \end{aligned} \tag{21}$$

and remove the first column from $D_{j_p}$, denoting the new $N\times(k_{j_p}-1)$-matrix by $\tilde{D}_{j_p}$, for $p=1, 2, \cdots, \mu$. By replacing $D_{j_p}$ in Eq. (10) with $\tilde{D}_{j_p}$, we define anew the $N\times\tilde{K}$-reduced matrix $\tilde{D}$ with $\tilde{K}=K-\mu$, as a substitute for $D$. We can do the same for $E_{j'_q}$ and produce the associated $N\times\tilde{L}$-reduced matrix $\tilde{E}$ with $\tilde{L}=L-\nu$.

Thus, under the assumption that $\tilde{F}$ and $\tilde{G}$ defined by

$$\tilde{F} = \frac{1}{N-1}\tilde{D}^T\tilde{D}\,, \quad \tilde{G} = \frac{1}{N-1}\tilde{E}^T\tilde{E} \tag{22}$$

are nonsingular, the eigenvalue problem in question becomes

$$(\tilde{F}^{-1}\tilde{H}\tilde{G}^{-1}\tilde{H}^T - \lambda I_{\tilde{K}})\tilde{\alpha} = 0\,, \quad \tilde{\beta} = \tilde{G}^{-1}\tilde{H}^T\tilde{\alpha} \tag{23}$$

where

$$\tilde{H} = \frac{1}{N-1}\tilde{D}^T\tilde{E} \tag{24}$$

and $\tilde{\alpha}$ and $\tilde{\beta}$ are the $\tilde{K}$- and $\tilde{L}$-vectors such that the elements $\alpha_{j_p1}$ and $\beta_{j'_q1}$ are all removed from $\alpha$ and $\beta$, respectively.

Let $\tilde{\alpha}'$ be any eignevector corresponding to the maximum eigenvalue of $\tilde{F}^{-1}\tilde{H}\tilde{G}^{-1}\tilde{H}^T$ and $\tilde{\beta}'$ be $\tilde{G}^{-1}\tilde{H}^T\tilde{\alpha}'$. Define the $K$-vector $\alpha'$, by using $\tilde{\alpha}'$ and the first of Eq. (21), and similarly define the $L$-vector $\beta'$. Then, substituting $\alpha'$ and $\beta'$ into Eq. (19), we obtain the optimal $\alpha^*$ and $\beta^*$.

### 3.2 Attributes on an ordinal scale

The method described in the last section is originally intended to deal with the attributes on a nominal scale. The category scrores $\alpha_{jk}$ and $\beta_{j'l}$ are determined formally by maximizing the canonical correlation (16). Accordingly, there is no constraint on the values which the category scores may have. If, however, we can let our *a priori* knowledge about the features of designed attributes and categories reflect the resulting category scores, it may be somewhat sensible. A typical specific feature is an attribute with ordered categories, for example, "good or bad", "necessary or unnecessary", "satisfactory or unsatisfactory", "sufficient or insufficient", and so on. In fact, when we try to analyze real data, we often come across such a type of ordered categories. The original method, without any consideration of the special features, could be applied formally to such ordinal data. However, it is sometimes natural to presuppose some ordinal relations among the category scores $\alpha_{jk}$ for each of some $j$'s and/or $\beta_{j'l}$ for each of some $j'$'s.

We shall examine an example. Let the respondents be asked about the "satisfaction" of a certain attribute, and let the attribute consist of five categories: "very satisfactory", "satisfactory", "neutral", "unsatisfactory", and "very unsatisfactory". By the original method, let the category scores for this attribute have been obtained as $\alpha_{j1}=-1$, $\alpha_{j2}=2$, $\alpha_{j3}=1$, $\alpha_{j4}=-2$, and $\alpha_{j5}=0$. From this result, we can conclude neither that, the more satisfactory the attribute, the larger the respondents' sample score nor that, the more satisfactory the attribute, the smaller the respondent's sample score. The utilization of *a priori* knowledge in such a case is to add to the original formulation the constraint that one of the following two conditions holds:

$$\begin{array}{l} \alpha_{j1} \leqq \alpha_{j2} \leqq \cdots \leqq \alpha_{j5} \\ \alpha_{j1} \geqq \alpha_{j2} \geqq \cdots \geqq \alpha_{j5} \end{array} \tag{25}$$

Such a constraint is called the order condition in this paper. Although the order condition may be a very strong one, it can be sometimes reasonable and it can be often imagined to be due to some type of transitivity in human judgment.

We shall consider the generalization of the above example. There could be various types of order relation among the category scores, depending on the properties of attributes and categories in the problem under study. Here, let all the attributes in the two sets be on an ordinal scale. Then we can set up the order condition as follows:

$$C = C_{\alpha 1} \cap C_{\alpha 2} \cap \cdots \cap C_{\alpha m} \cap C_{\beta 1} \cap \cdots \cap C_{\beta n} \tag{26}$$

where

$$C_{\alpha j} = \{\alpha_j\colon \alpha_{j1} \leqq \alpha_{j2} \leqq \cdots \leqq \alpha_{jk_j}\} \tag{27}$$

or

$$C_{\alpha j} = \{\alpha_j\colon \alpha_{j1} \geqq \alpha_{j2} \geqq \cdots \geqq \alpha_{jk_j}\} \tag{28}$$

and

$$C_{\beta j'} = \{\beta_{j'}\colon \beta_{j'1} \leqq \beta_{j'2} \leqq \cdots \leqq \beta_{j'l_{j'}}\} \tag{29}$$

or

$$C_{\beta j'} = \{\beta_{j'}\colon \beta_{j'1} \geqq \beta_{j'2} \geqq \cdots \geqq \beta_{j'l_{j'}}\} \tag{30}$$

Let the conditions (27) and (28) be said to be supplementary to each other; and let the same be said for $C_{\beta j'}$. Since either Eq. (27) or (28) for each $j$, and either Eq. (29) or (30) for each $j'$, can be chosen, $2^{m+n}$ different $C$'s can be considered. But the number of essentially different $C$'s are $2^{m+n-1}$, because simultaneously replacing each $C_{\alpha j}$ and $C_{\beta j'}$ with their corresponding supplementary conditions brings about no substantial difference.

Here, let us restrict the problem to one where all the attributes have a set of categories with the same features. For example, the respondents are asked about the "satisfaction" of each attribute. It is sufficient, in this case, to consider only the conditions (27) and (29) for all $j$'s and $j'$'s. Only these conditions, but not the conditions (28) and (30), will be dealt with in what follows.

Besides, we understand that the categories of an attribute on an ordinal scale are necessarily mutually exclusive and exhaustive. Therefore, let each attribute be mutually exclusive and exhaustive, that is, let Eq. (20) hold for any $i$, $j$, and $j'$. In its turn, we have $\mu = m$ and $\nu = n$ in Eqs. (21) through (24) in the following.

### 3.3 Solution procedure

In the canonical correlation analysis for ordinal data, the problem to be solved is formulated in short as the optimization problem: maximize the objective function

$$\Psi = \tilde{\alpha}^T \tilde{H} \tilde{\beta} \tag{31}$$

subject to the constraints

$$\tilde{\alpha}^T \tilde{F} \tilde{\alpha} = 1\,, \quad \tilde{\beta}^T \tilde{G} \tilde{\beta} = 1 \tag{32}$$

$$A\tilde{\alpha} \geqq 0\,, \quad B\tilde{\beta} \geqq 0 \tag{33}$$

with $\tilde{F}$, $\tilde{G}$, and $\tilde{H}$ defined by Eqs. (22) and (24), where

$$\begin{aligned}
&\tilde{\alpha} = (\tilde{\alpha}_1^T, \tilde{\alpha}_2^T, \cdots, \tilde{\alpha}_m^T)^T, \quad \tilde{\alpha}_j = (\alpha_{j2}, \alpha_{j3}, \cdots, \alpha_{jk_j})^T \\
&\tilde{\beta} = (\tilde{\beta}_1^T, \tilde{\beta}_2^T, \cdots, \tilde{\beta}_n^T)^T, \quad \tilde{\beta}_{j'} = (\beta_{j'2}, \beta_{j'3}, \cdots, \beta_{j'l_{j'}})^T \\
&A = \mathrm{diag}\,(P_{k_1}, P_{k_2}, \cdots, P_{k_m}) \\
&B = \mathrm{diag}\,(P_{l_1}, P_{l_2}, \cdots, P_{l_n}) \\
&P_q = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ -1 & 1 & \ddots & & & \vdots \\ 0 & -1 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 1 & 0 \\ 0 & \cdots & \cdots & 0 & -1 & 1 \end{pmatrix} : (q-1)\times(q-1)\text{-matrix}
\end{aligned} \tag{34}$$

The inequality (33) in vector-matrix forms implies element-wise the conditions (27) and (29) with Eq. (21). As mentioned in Section 3.2, the problem without these inequalities is reduced to the simple eigenvalue problem (23). However, with these inequalities, Eq. (23) is meaningless, and there is no systematic procedure to solve the problem.

A possibly easier way of obtaining the category scores satisfying the conditions (27) and (29) may be as follows: First, $\tilde{\alpha}$ and $\tilde{\beta}$ are obtained by solving the eigenvalue problem, discarding the conditions. Next, two neighboring categories whose scores do not satisfy the conditions are merged into one category. Then, the new problem with some merged categories is dealt with so as to obtain $\tilde{\alpha}$ and $\tilde{\beta}$. Again, the conditions are checked for the new $\tilde{\alpha}$ and $\tilde{\beta}$. The same process is repeated until all the conditions become satisfied. It can be readily seen that the process finishes after some finite iterations, but a set of the resulting category scores is not guaranteed to maximize $\Psi$ for the original problem.

The problem of Eqs. (31) through (33) is conceptually a type of constrained nonlinear optimization problem so that there is a variety of approaches to solving the problem numerically.[5] Nonetheless, an alternative procedure is adopted, taking into consideration the particularity of the problem.

Before that, in order to make the constraint (33) simpler, the transformation of variables can be introduced such that

$$\tilde{\alpha} = A^{-1}\phi, \quad \tilde{\beta} = B^{-1}\psi \tag{35}$$

because $P_q$ is nonsingular, as are $A$ and $B$. By these transformations, Eqs. (31) to (33) are rewritten as

$$\Psi = \phi^T H_0 \psi \tag{36}$$

$$\phi^T F_0 \phi = 1, \quad \psi^T G_0 \psi = 1 \tag{37}$$

$$\phi \geqq 0, \quad \psi \geqq 0 \tag{38}$$

where

$$F_0 = (A^{-1})^T \tilde{F} A^{-1}, \quad G_0 = (B^{-1})^T \tilde{G} B^{-1}, \quad H_0 = (A^{-1})^T \tilde{H} B^{-1} \tag{39}$$

and $A^{-1}$ and $B^{-1}$ can be written simply as

$$\begin{aligned} A^{-1} &= \mathrm{diag}(Q_{k_1}, Q_{k_2}, \cdots, Q_{k_m}) \\ B^{-1} &= \mathrm{diag}(Q_{l_1}, Q_{l_2}, \cdots, Q_{l_n}) \\ Q_p &= \begin{pmatrix} 1 & 0 \cdots\cdots & 0 \\ 1 & 1 \ddots & \vdots \\ \vdots & \ddots \ddots \ddots & \vdots \\ \vdots & \ddots \ddots & 0 \\ 1 & \cdots\cdots 1 & 1 \end{pmatrix} : (p-1)\times(p-1)\text{-matrix} \end{aligned} \tag{40}$$

Let $\phi=\phi^*$ and $\psi=\psi^*$ be the optimal solution of the problem (36) through (38). We consider any $\phi$ and $\psi$ which are near $\phi^*$ and $\psi^*$, respectively, and which satisfy Eqs. (37) and (38), namely

$$\phi = \phi^* + \Delta\phi, \quad \psi = \psi^* + \Delta\psi \tag{41}$$

with a sufficiently small $\Delta\phi$ and $\Delta\psi$. Substituting Eq. (41) into (36) to (38) and neglecting the higher order terms of $\Delta\phi$ and $\Delta\psi$ yields

$$\tilde{\Psi} = \phi^T H_0 \psi - \psi^T H_0^T \Delta\phi - \phi^T H_0 \Delta\psi \tag{42}$$

$$\phi^T F_0 \Delta\phi = 0, \quad \psi^T G_0 \Delta\psi = 0 \tag{43}$$

$$\Delta\phi \leqq \phi, \quad \Delta\psi \leqq \psi \tag{44}$$

First, when a set of $\phi=\phi^{(0)}$ and $\psi=\psi^{(0)}$ satisfying Eqs. (37) and (38) is given, we solve the problem of minimizing

$$z = \psi^T H_0^T \Delta\phi + \phi^T H_0 \Delta\psi \tag{45}$$

under the constraints (43) and (44), and obtain more approximate solutions than $\phi^{(0)}$ and $\psi^{(0)}$:

$$\phi' = \phi^{(0)} - \Delta\phi, \quad \psi' = \psi^{(0)} - \Delta\psi \tag{46}$$

The problem can be separated into the two independent problems for $\Delta\phi$ and $\Delta\psi$. Each of these problems is a conventional linear programming, so that it can be solved by the usual simplex method. By neglecting the higher order terms in Eq. (43), $\phi'$ and $\psi'$ will not always satisfy Eq. (37). Accordingly, we normalized the magnitude as

$$\phi^{(1)} = \frac{\phi'}{\sqrt{\phi'^T F_0 \phi'}}, \quad \psi^{(1)} = \frac{\psi'}{\sqrt{\psi'^T G_0 \psi'}} \tag{47}$$

Next, putting $\phi=\phi^{(1)}$ and $\psi=\psi^{(1)}$, we again solve the problem (45) with (43) and (44). Hereafter, we repeat in a similar way. When $|\Psi^{(k-1)}-\Psi^{(k)}|$ becomes sufficiently small where $\Psi^{(k)}$ is the value of $\Psi$ for $\phi=\phi^{(k)}$ and $\psi=\psi^{(k)}$, we regard them to be the most approximate solutions of $\phi^*$ and $\psi^*$.

At every repetition, $\phi^{(k)}$ and $\psi^{(k)}$ will be expected to get near $\phi^*$ and $\psi^*$ gradually, but Eqs. (42) to (44) themselves are meaningless for the large $||\Delta\phi||$ or $||\Delta\psi||$, owing to the neglect of the higher orders in these equations. Therefore, instead of Eqs. (44) and (46), we consider

$$-\bar{\phi}\leqq\Delta\phi\leqq\phi\,,\quad -\bar{\psi}\leqq\Delta\psi\leqq\psi \tag{48}$$

$$\phi' = \phi^{(k)}-\gamma\Delta\phi\,,\quad \psi' = \psi^{(k)}-\gamma\Delta\psi \tag{49}$$

where $\bar{\phi}$ has the same elements which are the mean values of the elements of $\phi^{(k)}$, $\bar{\psi}$ is similarly defined, and $\gamma$ ($\leqq 1$) is a scalar parameter. In each repetition, $\gamma$ is chosen to maximize $\Psi^{(k+1)}$ resulting from $\phi^{(k+1)}$ and $\psi^{(k+1)}$ with Eq. (49). This is a kind of the so-called linear search technique[5] in nonlinear optimization.

Here, let us mention a set of initial solutions $\phi^{(0)}$ and $\psi^{(0)}$ from which the iterative procedure starts. The eigenvalue problem equivalent to Eqs. (36) and (37) without (38) is

$$(F_0^{-1}H_0G_0^{-1}H_0^T-\lambda I_{\bar{K}})\phi = 0\,,\quad \psi = G_0^{-1}H_0^T\phi \tag{50}$$

First, let $\phi'$ be any eigenvector corresponding to the maximum $\lambda$ and $\psi'$ be by the second equation of (50). We replace the negative elements of $\phi'$ and $\psi'$ with zero, denote the modified $\phi'$ and $\psi'$ as $\phi''$ and $\psi''$, respectively, and have

$$\phi^{(0)} = \frac{\phi''}{\sqrt{\phi''^T F_0\phi''}}\,,\quad \psi^{(0)} = \frac{\psi''}{\sqrt{\psi''^T G_0\psi''}} \tag{51}$$

Next, we do the same for $-\phi'$ and $-\psi'$ and denote the modified $-\phi'$ and $-\psi'$ as $\phi''$ and $\psi''$, respectively. We have another $\phi^{(0)}$ and $\psi^{(0)}$ from Eq. (51). Finally, we adopt one of these two sets of $\phi^{(0)}$ and $\psi^{(0)}$ giving a larger $\Psi^{(0)}$ as an initial solution. This gives a good initial solution, when either $\phi'$ and $\psi'$ or $-\phi'$ and $-\psi'$ have a small number of negative elements.

## 4. Application to Real Data

The data treated here are those resulting from the summary of a poll, which was conducted by the Urban Planning Bureau of the Kyoto City Government,[6] with regard to citizens' evaluations of their everyday living environments. The purpose of the poll was to investigate how the citizens evaluate their physical and human environments, and their hopes for Kyoto City in the future. The poll was conducted throughout the whole city. 2157 persons, aged twenty years or more were chosen randomly, among which 1739 persons responded, meaning a high 80 % response.

Several inquiries were made in the questionnaire, concerning the above aims. The principal part of the questionnaire consisted of eighteen inquiries, concerning the interest in and the hope for municipal administration and the evaluations of living environments. One of those inquiries was about the public transportation facilities in the city. For thirteen items, the people were asked to indicate the degree of their feelings of satisfaction with railroads, buses, and taxies, in one of five specified categories: category 1 means "very satisfactory", 2 means "satisfactory", 3 "indifferent", 4 "unsatisfactory", and 5 "very unsatisfactory". The items were the following:

1. The distance from your home to a station or a stop.
2. Waiting time at a station or a stop.
3. Comfortableness of a ride.
4. Manners and services of the crews.
5. Understandableness of stations or stops and routes.
6. Fares.
7. The time when the first and the last cars start.
8. Crowdedness in a car.
9. Easiness of using facilities when you are in a hurry.
10. Convenience of changing cars.
11. As a vehicle suitable to Kyoto.
12. As an alternative to a private car.
13. Overall satisfaction.

For taxies, only items 3, 4, 6, 9, 11, 12, and 13 were evaluated.

The first twelve items were used as the attributes for the first set in the canonical correlation analysis; and the last item was for the second set, so that the set had

Table 1. Kendall's rank correlation coefficients between item 13 and the other twelve items (concerning railroads).

| Item $j$ | Rank correlation | Item $j$ | Rank correlation |
|---|---|---|---|
| 1 | 0.32 | 7 | 0.36 |
| 2 | 0.45 | 8 | 0.29 |
| 3 | 0.45 | 9 | 0.51 |
| 4 | 0.39 | 10 | 0.45 |
| 5 | 0.41 | 11 | 0.45 |
| 6 | 0.32 | 12 | 0.56 |

only one attribute. This analysis of the canonical correlation aimed at observing what the citizens conceive unconsciously would be factors of satisfaction alout public transportation facilities. In this analysis, it was assumed that, because every one of the first twelve or six items could be one of the factors constituting an overall satisfaction, the more satisfactory each of these items the more satisfactory the last item should be. A justification of this assumption is revealed by Table 1. The table shows the value of the so-called Kendall's rank correlation coefficients* between item 13 and each of items 1 to 12. All the coefficients were significant with the level of significance 0.1 %. From the table, it is seen that they are all positive. The table is concerned with railroads. The features were almost the same for buses and taxies. Therefore, on the average, the present assumption could be true.

The original method in Section 3.1 and the proposed modified method in Section 3.3 were applied to the data of 1414, 1483, and 1421 valid responses for railroads, buses, taxies, respectively. (The remaining 325, 256, and 318 responses for each of the transportation facilities gave no answer to at least one of the items

Table 2. Category scores computed by the two methods (concerning railroads).

| Item $j$ | Original method | | | | Modified method | | | |
|---|---|---|---|---|---|---|---|---|
| | $\alpha_{j2}$ | $\alpha_{j3}$ | $\alpha_{j4}$ | $\alpha_{j5}$ | $\alpha_{j2}$ | $\alpha_{j3}$ | $\alpha_{j4}$ | $\alpha_{j5}$ |
| 1 | 0.133 | 0.239 | 0.318 | $>$ 0.260 | 0.131 | 0.237 | 0.305 | $=$ 0.305 |
| 2 | 0.205 | 0.263 | 0.466 | 0.600 | 0.182 | 0.240 | 0.441 | 0.556 |
| 3 | 0.021 | 0.129 | 0.437 | $>$ 0.433 | 0.019 | 0.120 | 0.430 | $=$ 0.430 |
| 4 | 0.124 | 0.258 | 0.424 | $>$ 0.398 | 0.117 | 0.257 | 0.420 | $=$ 0.420 |
| 5 | 0.248 | 0.345 | 0.519 | 0.803 | 0.212 | 0.302 | 0.483 | 0.726 |
| 6 | 0.153 | 0.233 | 0.471 | $>$ 0.171 | 0.150 | 0.229 | 0.433 | $=$ 0.433 |
| 7 | 0.091 | 0.109 | 0.313 | $>$ 0.249 | 0.078 | 0.096 | 0.296 | $=$ 0.296 |
| 8 | $>-0.007$ | 0.054 | 0.149 | $>$ 0.106 | $=$ 0.0 | 0.057 | 0.146 | $=$ 0.146 |
| 9 | $>-0.037$ | 0.140 | 0.480 | 0.481 | $=$ 0.0 | 0.175 | 0.503 | 0.504 |
| 10 | 0.238 | $>$ 0.214 | 0.446 | $>$ 0.328 | 0.203 | $=$ 0.203 | 0.423 | $=$ 0.423 |
| 11 | $>-0.268$ | $-0.090$ | 0.089 | 0.287 | $=$ 0.0 | 0.152 | 0.328 | 0.497 |
| 12 | 0.211 | 0.492 | 1.049 | 1.216 | 0.180 | 0.459 | 1.031 | 1.174 |
| | $\beta_{12}$ | $\beta_{13}$ | $\beta_{14}$ | $\beta_{15}$ | $\beta_{12}$ | $\beta_{13}$ | $\beta_{14}$ | $\beta_{15}$ |
| 13 | 0.608 | 1.523 | 3.331 | 3.969 | 0.651 | 1.556 | 3.363 | 4.088 |

* Kendall's rank correlation[7] $\tau$ between attributes $j$ and $j'$ in the first and the second sets, respectively, is defined as follows. Let $k_1$ and $k_2$ be $k$ such that $\delta_i(j, k)=1$ and $\delta_{i'}(j, k)=1$ for two samples $i$ and $i'$, respectively, and let $l_1$ and $l_2$ be defined similarly for $\varepsilon_i(j', l)$ and $\varepsilon_{i'}(j', l)$. Thus, $\tau=(d-\bar{d})/\sqrt{[N(N-1)/2-t][N(N-1)/2-t']}$, where $d$ is the number of pairs $(i, i')$ with $k_1<k_2$ and $l_1<l_2$ or with $k_1>k_2$ and $l_1>l_2$, and $\bar{d}$ is the number of such pairs with $k_1<k_2$ and $l_1>l_2$ or with $k_1>k_2$ and $l_1<l_2$; $t$ and $t'$ are the numbers of such pairs with $k_1=k_2$ and $l_1=l_2$, respectively.

Table 3. Category scores computed by the two methods (concerning taxies).

| Item $j$ | Original method | | | | Modified method | | | |
|---|---|---|---|---|---|---|---|---|
| | $\alpha_{j2}$ | $\alpha_{j3}$ | $\alpha_{j4}$ | $\alpha_{j5}$ | $\alpha_{j2}$ | $\alpha_{j3}$ | $\alpha_{j4}$ | $\alpha_{j5}$ |
| 3 | 0.267 | 0.506 | 0.925 > | 0.791 | 0.228 | 0.457 | 0.871 = | 0.871 |
| 4 | >−0.210 | 0.097 | 0.675 | 0.928 | = 0.0 | 0.289 | 0.855 | 1.071 |
| 6 | 0.338 | 0.583 | 0.825 > | 0.715 | 0.332 | 0.581 | 0.787 = | 0.787 |
| 9 | >−0.198 | 0.043 | 0.242 | 0.671 | = 0.0 | 0.205 | 0.407 | 0.772 |
| 11 | 0.470 | 0.546 | 1.167 > | 1.100 | 0.409 | 0.478 | 1.094 = | 1.094 |
| 12 | 0.105 | 0.478 | 1.194 > | 0.787 | 0.089 | 0.459 | 1.165 = | 1.165 |
| | $\beta_{12}$ | $\beta_{13}$ | $\beta_{14}$ | $\beta_{15}$ | $\beta_{12}$ | $\beta_{13}$ | $\beta_{14}$ | $\beta_{15}$ |
| 13 | 0.584 | 1.635 | 3.597 | 5.353 | 0.665 | 1.724 | 3.671 | 5.668 |

under study.) The computed values of the category scores are shown in Tables 2 and 3 concerning railroads and taxies, respectively. It had been expected that $0 \leqq \alpha_{j2} \leqq \alpha_{j3} \leqq \alpha_{j4} \leqq \alpha_{j5}$ and $0 \leqq \beta_{j2} \leqq \cdots \leqq \beta_{j5}$. In both tables, the results by the orginal method do not necessarily meet this expectation. That is, there are pairs of two neighboring category scores where the order of values is reversed. (These cases are indicated by the symbol $>$ in the tables.) Besides, the results by the modified method are reasonable. Also, it is seen that, in the case of the original method, every pair of two neighboring category scores having the same value (indicated by the symbol $=$ in the tables) is in accord with the pair of the scores between which the symbol $>$ appears.

Finally, let us examine the partial correlation coefficient between the attributes $j$ and $j'$ in the first and the second sets, respectively. Define the $N$-vector $\zeta_p$, the sample score for each attribute, as

$$\begin{aligned} \zeta_j &= D_j \alpha_j \qquad (j = 1, 2, \cdots, m) \\ \zeta_{m+j'} &= E_{j'} \beta_{j'} \qquad (j' = 1, 2, \cdots, n) \end{aligned} \tag{52}$$

It is readily seen that the average of $\zeta_p$ with respect to all the samples is zero for each $p = 1, 2, \cdots, m+n$. We have the correlation between $\zeta_p$ and $\zeta_q$:

$$r_{pq} = \frac{\zeta_p^T \zeta_q}{||\zeta_p|| \, ||\zeta_q||} \tag{53}$$

The partial correlation coefficient is defined by

$$r_{jj'}^* = -\frac{r^{jj'}}{\sqrt{r^{jj} r^{j'j'}}} \tag{54}$$

$$(j = 1, 2, \cdots, m; \; j' = 1, 2, \cdots, n)$$

where $r^{jj'}$ is the $(j, j')$ element of the inverse of $(m+n)\times(m+n)$-matrix $R=\{r_{pq}\}$. The value of $r^*_{jj'}$ gives the degree of the direct relation of the attribute $j$ to $j'$, and the relation involves no indirect relations through the other attributes.

Table 4 shows the values of $r^*_{jj'}$, for our present application. The larger the value, the stronger is the degree of the influence of item $j$ on the overall satisfaction. It can be seen that, for all the public transportation facilities under study, the item "As an alternative to a private car" is the largest factor of the overall satisfaction. Successively, the items "Easiness of using facilities when you are in a hurry" and "Comfortableness of a ride" are larger factors for railroads, and "Waiting time at a stop" is larger for buses. On the other hand, item 8 for railroads, 6 and 10 for buses, and 9 for taxies are small factors.

Table 4. Partial correlation coefficients between item 13 and the other twelve items.

| Item $j$ | Railroads | Buses | Taxies | Item $j$ | Railroads | Buses | Taxies |
|---|---|---|---|---|---|---|---|
| 1 | 0.11 | 0.11 | | 7 | 0.11 | 0.12 | |
| 2 | 0.13 | 0.19 | | 8 | 0.05 | 0.10 | |
| 3 | 0.17 | 0.14 | 0.21 | 9 | 0.19 | 0.15 | 0.18 |
| 4 | 0.11 | 0.13 | 0.25 | 10 | 0.12 | 0.08 | |
| 5 | 0.14 | 0.16 | | 11 | 0.11 | 0.22 | 0.25 |
| 6 | 0.14 | 0.08 | 0.22 | 12 | 0.33 | 0.29 | 0.36 |

## 5. Concluding Remarks

A modified technique of linear canonical correlation analysis has been discussed, in which some linear constraints are imposed on the magnitude of category scores. While the conventional analysis with no constraint is reduced to a simple eigenvalue problem, the analysis proposed in this paper involves a kind of constrained nonlinear optimization problem. Although there is a variety of approaches to solving the problem numerically, an alternative procedure has been adopted, taking the particularity of the problem into consideration. The procedure consists of iteratively solving successive linear programming problems.

The analysis with the numerical procedure here presented has been applied to real data. The problem has twelve and one attributes of the first and the second sets, respectively, in the canonical correlation. The data had features suitable for the method. In fact, the rank correlation coefficients between the attributes in the two sets were all positive. As a consequence of the numerical computation, the expected results have been obtained: Pairs of two category scores, obtained by the original method, violating the presumed constraints, are turned into the same values by the modified method.

The idea of the modified method is fairly simple and, if the method is used carefully, it will be useful for a type of canonical correlation analysis. Although the procedure for the modified analysis has been formulated for the same type of constraints for all the attributes, the formulation is also possible for some other types of constraints: for example, a case where the category scores are constrained for some attributes but not for others.

## Acknowledgments

The authors would like to sincerely acknowledge the kind consideration of the members of the Urban Planning Bureau of the Kyoto City Government. Mr. Hideo Yoshida of the bureau has made available the polling data to the authors and has given his helpful advice to them. They would also like to thank Mr. Hisashi Takamatsu, presently with the Hitachi, Ltd., for his cooperation regarding the development of an idea while a graduate student at Kyoto University; and Mr. Hisashi Yasuda for his cooperation regarding computer programming codings and numerical computations.

### References

1) T.W. Anderson: Introduction to multivariate analysis, John Wiley, New York (1958).
2) C. Hayashi: On the quantification of qualitative data from the mathematico-statistical point of view, Annals of Institute of Statistical Mathematics, **2**, 35–47 (1950).
3) C. Hayashi: On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view, Annals of Institute of Statistical Mathematics, **3**, 69–98 (1952).
4) C. Hayashi: Multidimensional quantification—with the applications to analysis of social phenomena—, Annals of Institute of Statistical Mathematics, **5**, 121-143 (1954).
5) S.L.S. Jacoby, J.S. Kowalik, and J.T. Pizzo: Iterative methods for nonlinear optimization problems, Prentice-Hall (1972).
6) Urban Planning Bureau, Kyoto City Government: Report of investigation on citizen opinion poll for a new plan of city construction (1977).
7) H. M. Blalock: Social statistics, McGraw-Hill, New York (1960).