PSYCHOMETRIKA—VOL. 56, NO. 4, 667–684
DECEMBER 1991

RELATIONSHIPS AMONG SEVERAL METHODS OF LINEARLY CONSTRAINED CORRESPONDENCE ANALYSIS

YOSHIO TAKANE

MCGILL UNIVERSITY

HARUO YANAI AND SHINICHI MAYEKAWA

THE NATIONAL CENTER FOR UNIVERSITY ENTRANCE EXAMINATION

This paper shows essential equivalences among several methods of linearly constrained correspondence analysis. They include Fisher's method of additive scoring, Hayashi's second type of quantification method, ter Braak's canonical correspondence analysis, Nishisato's ANOVA of categorical data, correspondence analysis of manipulated contingency tables, Böckenholt and Böckenholt's least squares canonical analysis with linear constraints, and van der Heijden and Meijerink's zero average restrictions. These methods fall into one of two classes of methods corresponding to two alternative ways of imposing linear constraints, the reparametrization method and the null space method. A connection between the two is established through Khatri's lemma.

Key words: canonical correlation analysis, generalized singular value decomposition (GSVD), the method of additive scoring, the second type of quantification method (Q2), canonical correspondence analysis (CCA), ANOVA of categorical data, canonical analysis with linear constraints (CALC), zero average restrictions, Khatri's lemma.

1. Introduction

Fisher's (1948) method of additive scoring has been discovered, and rediscovered in many different guises over the past forty years. Hayashi's (1950, 1952) second type of quantification method (Q2), Carroll's (1973) categorical conjoint measurement, special cases of ter Braak's (1986) canonical correspondence analysis (CCA), and Böckenholt and Böckenholt's (1990) canonical analysis with linear constraints (CALC) are but a few examples. For some of these methods, the relation to Fisher's original work is rather obvious. For others, particularly the ones more recently proposed, it is less obvious. This paper systematically investigates formal relationships among these methods. The methods to be discussed in this paper include, in addition to those cited above, Nishisato's (1971, 1980) ANOVA of categorical data, joint, conditional and marginal correspondence analysis (D'Ambra & Lauro, 1989; Israëls, 1987; Leclerc, 1975) and van der Heijden and Meijerink's (1989) zero average restrictions. While some of these methods may differ in their original formulations, intended data types, generalities, and so on, there is a common thread running through them. That is, they are all closely related to canonical discriminant analysis, or more generally to canonical correlation analysis.

These methods are roughly classified into two groups, corresponding to two alternative ways of restricting the parameter space. One is called the parameter reduction or

The work reported in this paper has been supported by grant A6394 from the Natural Sciences and Engineering Research Council of Canada to the first author. We wish to thank Carolyn Anderson, Ulf Böckenholt, Henk Kiers, Shizuhiko Nishisato, Jim Ramsay, Tadashi Shibayama, Cajo ter Braak, and Peter van der Heijden for their helpful comments on earlier drafts of this paper.

Requests for reprints should be sent to Yoshio Takane, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, Quebec H3A 1B1 CANADA.

0033-3123/91/1200-90223$00.75/0

reparametrization method, and specifies the space in which the original parameter vector should lie. That is, the original parameter vector, $\mathbf{u}$, is reparametrized as $\mathbf{u} = \mathbf{X}\mathbf{u}^*$, where $\mathbf{X}$ is a matrix of basis vectors, and $\mathbf{u}^*$ a reduced parameter vector. The other group, called the null space method, specifies the ortho-complement space of $\mathbf{X}$. Let $\mathbf{R}$ denote a matrix of basis vectors of this space. Then, $\mathbf{R}'\mathbf{u} = \mathbf{0}$. The space of all vectors, $\mathbf{u}$, such that $\mathbf{R}'\mathbf{u} = \mathbf{0}$, is called the kernel (null space) of $\mathbf{R}'$, and denoted by Ker ($\mathbf{R}'$), The space spanned by column vectors of $\mathbf{X}$ on the other hand, is denoted by Sp ($\mathbf{X}$). By appropriately choosing an $\mathbf{R}$ for a given $\mathbf{X}$, or vice versa, the two spaces, Sp ($\mathbf{X}$) and Ker ($\mathbf{R}'$), could be made identical, providing two equivalent ways of specifying linear constraints. The method of additive scoring, Q2, and canonical correspondence analysis (CCA), among others, belong to the first group of methods (the reparametrization method), while canonical analysis with linear constraints (CALC) and the zero average restrictions belong to the second group (the null space method). A specific relationship between the two approaches will be established through a simple lemma by Khatri (1966).

2. The Reparametrization Method

In this section, we discuss methods of linearly constrained correspondence analysis that use the reparametrization (parameter reduction) method for incorporating linear constraints.

2.1. *The Method of Additive Scoring*

Let $\mathbf{G}$ be an N by p superindicator matrix,

$$\mathbf{G} = [\mathbf{G}_1, \ldots, \mathbf{G}_{p^*}],$$

where N is the number of subjects (cases), each $\mathbf{G}_i$ is N by p_i (the number of observation categories), and $p = \Sigma_{i=1}^{p^*} p_i$. Suppose that nonnumerical (categorical) observations are made on the N subjects (cases), which are dummy-coded into an N by q indicator matrix, $\mathbf{H}$, where q is the number of observation categories. Fisher (1948, pp. 289-298) posed the problem of assigning scores to the q categories of observation which can be represented as an additive function of $\mathbf{G}_i'$s as much as possible. This amounts to finding a q-component vector, $\mathbf{v}$, that maximizes

$$\theta = \frac{\mathbf{v}'\mathbf{H}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-}\mathbf{G}'\mathbf{H}\mathbf{v}}{\mathbf{v}'\mathbf{H}'\mathbf{H}\mathbf{v}},$$

where $(\mathbf{G}'\mathbf{G})^{-}$ is a g-inverse of $\mathbf{G}'\mathbf{G}$. This quantity, θ, indicates the proportion of the total variation among the N subjects' assigned scores that can be accounted for by an additive function of $\mathbf{G}_i$'s. It turns out that the maximized value of θ is equal to the square of the largest (nontrivial) canonical correlation between $\mathbf{G}$ and $\mathbf{H}$. The problem thus reduces to one of canonical correlation analysis.

The canonical correlation analysis, in turn, is known to be equivalent to the generalized singular value decomposition (GSVD; e.g., Greenacre, 1984) of

$$\mathbf{M}_1 = (\mathbf{G}'\mathbf{G})^{+}\mathbf{G}'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}, \tag{1}$$

with column and row metrics, $\mathbf{G}'\mathbf{G}$ and $\mathbf{H}'\mathbf{H}$, respectively, where $(\mathbf{G}'\mathbf{G})^{+}$ is the Moore-Penrose inverse of $\mathbf{G}'\mathbf{G}$. The $\mathbf{G}'\mathbf{G}$ is called the column metric because it applies to columns of $\mathbf{M}_1$, and $\mathbf{L} = \mathbf{H}'\mathbf{H}$ the row metric because it applies to rows of $\mathbf{M}_1$. The

Moore-Penrose inverse of $\mathbf{G}'\mathbf{G}$ ensures uniqueness of $\mathbf{M}_1$ and uniqueness of the GSVD. The diagonal matrix, $\mathbf{H}'\mathbf{H}$, will be denoted as $\mathbf{L}$ hereafter.

Briefly, the GSVD of $\mathbf{M}_1$ with metrics $\mathbf{G}'\mathbf{G}$ and $\mathbf{L} = \mathbf{H}'\mathbf{H}$ is the decomposition of $\mathbf{M}_1$ into

$$\mathbf{M}_1 = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime}, \tag{2}$$

such that $\mathbf{U}^{*\prime}\mathbf{G}'\mathbf{G}\mathbf{U}^* = \mathbf{I}$, $\mathbf{V}^{*\prime}\mathbf{L}\mathbf{V}^* = \mathbf{I}$ and $\mathbf{D}^*$ is diagonal and positive definite. Takane and Shibayama (1991) describe an efficient way of obtaining the GSVD of a matrix of the above form. The largest singular value of $\mathbf{M}_1$ is unity, reflecting an overlap between Sp(G) and Sp(H). It represents a trivial solution, and corresponding singular vectors (both left and right), which are constant vectors, should be discarded from the solution. A way to eliminate the trivial solution from $\mathbf{M}_1$ will be discussed in section 3.1 (also, see Böckenholt and Böckenholt, 1990).

The method of additive scoring presented above is a somewhat generalized version of Fisher's original method, which was restricted to the unidimensional case. This generalized version is variously called Hayashi's (1950, 1952) second type of quantification method (Q2) in Japan, canonical analysis of categorical data (Johnson, 1950; Maxwell, 1961), categorical conjoint measurement (Carroll, 1973, pp. 339-348), and so on. The method can also be viewed as a special form of canonical discriminant analysis (Fisher, 1936), where not only the criterion but also the predictor variables are discrete.

When there is only a single predictor variable with p_1 observation categories ($p = p_1$, and $\mathbf{G} = \mathbf{G}_1$), the method reduces to ordinary correspondence analysis (OCA; Greenacre, 1984), which is also known as dual scaling (Nishisato, 1980), Hayashi's (1952) third type of quantification method (Q3), and so on. In this case, $\mathbf{G}'\mathbf{G} = \mathbf{K}$ is diagonal, and of full rank, so that (1) reduces to

$$\mathbf{M}_0 = \mathbf{K}^{-1}\mathbf{G}'\mathbf{H}\mathbf{L}^{-1}. \tag{3}$$

In OCA, the GSVD of $\mathbf{M}_0$ is obtained with metrics $\mathbf{K}$ and $\mathbf{L}$.

The above discussion suggests the method of additive scoring is an extension of OCA to more than one simple two-way contingency table. The next section shows it can also be viewed as a restricted form of OCA with linear constraints on the representation of row categories.

2.2 *Canonical Correspondence Analysis*

The method of additive scoring can be reformulated as special cases of canonical correspondence analysis (CCA; ter Braak, 1986). Let $\mathbf{F}$ be a data matrix that can take a variety of forms (ter Braak, 1988); for example, an indicator matrix, a superindicator matrix, a two-way contingency table, and so on. Let $\mathbf{X}$ be a matrix of predictor variables pertaining to the row structure of $\mathbf{F}$. It may include continuous as well as discrete variables. In CCA, representations of rows and columns of $\mathbf{F}$ are sought under the restriction that the row representation is a linear combination of $\mathbf{X}$. The method amounts to obtaining the GSVD of

$$\mathbf{M}_2 = (\mathbf{X}'\mathbf{K}\mathbf{X})^+\mathbf{X}'\mathbf{F}\mathbf{L}^{-1}, \tag{4}$$

with metrics $\mathbf{X}'\mathbf{K}\mathbf{X}$ and $\mathbf{L}$, where $\mathbf{K}$ and $\mathbf{L}$ are diagonal matrices of row totals and column totals of $\mathbf{F}$, respectively.

Two special cases of (4) are of interest to us, obtained by specializing $\mathbf{F}$ and $\mathbf{G}$. One is where $\mathbf{F}$ is a single indicator matrix. By setting $\mathbf{F} = \mathbf{H}$ and $\mathbf{X} = \mathbf{G}$, (4) reduces to (1). (That $\mathbf{K} = \mathbf{I}$ follows immediately.) The second case is in which $\mathbf{F}$ is a simple two-way contingency table, such that

$$\mathbf{F} = \mathbf{G}^{*\prime}\mathbf{H}, \tag{5}$$

where $\mathbf{G}^*$ is an indicator matrix obtained by interactive coding of p^* discrete variables, $\mathbf{G}_i$ (van der Burg, de Leeuw, & Verdegaal, 1988). The $\mathbf{G}^*$ is an N by $\tilde{p}$ matrix, where $\tilde{p} = \Pi_{i=1}^{p^*} p_i$. Let $\mathbf{X}$ be a $\tilde{p}$ by p matrix which turns $\mathbf{G}^*$ into $\mathbf{G}$. That is,

$$\mathbf{G} = \mathbf{G}^*\mathbf{X} \tag{6}$$

An example of $\mathbf{X}$ is given in Section 2.3. Since $\mathbf{G}'\mathbf{G} = \mathbf{X}'\mathbf{K}\mathbf{X}$, where $\mathbf{K} = \mathbf{G}^{*\prime}\mathbf{G}^*$, and $\mathbf{G}'\mathbf{H} = \mathbf{X}'\mathbf{G}^*\mathbf{H} = \mathbf{X}'\mathbf{F}$, (4) reduces to (1). Note that the first case is actually a special case of the second. The former is obtained from the latter by setting $\mathbf{G}^* = \mathbf{I}$.

When $\mathbf{X} = \mathbf{I}$ in (4), $\mathbf{M}_2$ reduces to $\mathbf{M}_0$, and consequently CCA reduces to OCA. Here, $\mathbf{X} = \mathbf{I}$ implies that rows of $\mathbf{F}$ are all regarded as distinct, and no special relationships are assumed among them.

For later reference, we note that the GSVD of $\mathbf{M}_2$ with metrics $\mathbf{X}'\mathbf{K}\mathbf{X}$ and $\mathbf{L}$ is simply related to the GSVD of

$$\mathbf{M}_2^* = \mathbf{X}\mathbf{M}_2 = \mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{X})^{-}\mathbf{X}'\mathbf{F}\mathbf{L}^{-1}, \tag{7}$$

with metrics $\mathbf{K}$ and $\mathbf{L}$. Note that in (7), the Moore-Penrose inverse of $\mathbf{X}'\mathbf{K}\mathbf{X}$ is replaced by a g-inverse, $(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}$. The $\mathbf{M}_2^*$ is unique no matter which g-inverse of $\mathbf{X}'\mathbf{K}\mathbf{X}$ is used, since $\mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{X})^{-}\mathbf{X}' = \mathbf{K}^{-1/2}\mathbf{K}^{1/2}\mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{X})^{-}\mathbf{X}'\mathbf{K}^{1/2}\mathbf{K}^{-1/2}$. Let the GSVD of $\mathbf{M}_2^*$ be denoted by $\mathbf{M}_2^* = \mathbf{U}\mathbf{D}\mathbf{V}'$. These $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{D}$ are related to $\mathbf{U}^*$, $\mathbf{V}^*$, and $\mathbf{D}^*$ in the GSVD of $\mathbf{M}_2$ (and of $\mathbf{M}_1$) with metrics $\mathbf{X}'\mathbf{K}\mathbf{X}$ and $\mathbf{L}$ by $\mathbf{U} = \mathbf{X}\mathbf{U}^*$ (or $\mathbf{U}^* = (\mathbf{X}'\mathbf{K}\mathbf{X})^{+}\mathbf{X}'\mathbf{K}\mathbf{U}$), $\mathbf{V} = \mathbf{V}^*$, and $\mathbf{D} = \mathbf{D}^*$ (Takane & Shibayama, 1991). The $\mathbf{M}_2^*$ in (7) rather than $\mathbf{M}_2$ in (4) will be directly related to $\mathbf{M}_3^*$ in section 3.1.

CCA was initially derived (ter Braak, 1986) as an approximation to the unfolding type of single-peaked response surface model called Gaussian ordination. Appendix A shows how CCA can be derived on a purely nonstochastic basis. Appendix A also describes an alternative generating mechanism for $\mathbf{F}$ due to Lebreton, Chessel, Prodon, and Yoccoz (1988).

2.3. *A Simple Illustrative Example*

An example will be given that illustrates the connection between the method of additive scoring and CCA. This example will also be used later to illustrate other methods of linearly constrained correspondence analysis (see section 2.5, and section 3.2).

Suppose there are two binary predictor variables, A and B. Suppose further that two observations each are sampled from each of the four cells obtained by factorial combinations of levels in variable A and those in variable B. The four cells are designated as a_1b_1, a_2b_1, a_1b_2, and a_2b_2, where a_j ($j = 1, 2$) indicates the j-th level of variable A, and b_k ($k = 1, 2$) the k-th level of variable B. Then, $\mathbf{G}^*$ may be

$$\mathbf{G}^* = \begin{array}{c} \begin{array}{cccc} a_1b_1 & a_2b_1 & a_1b_2 & a_2b_2 \end{array} \\ \left.\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}\right\} \text{8 observations.} \end{array}$$

If only the main effects of variables A and B are included in the predictor set, $\mathbf{X}$ may be

$$\mathbf{X} = \begin{array}{c} \begin{array}{cccc} a_1 & a_2 & b_1 & b_2 \end{array} \\ \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ a_1 b_1 \\ a_2 b_1 \\ a_1 b_2 \\ a_2 b_2 \end{array}. \tag{8}$$

Then,

$$\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2] = \mathbf{G}^*\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

Suppose that the observations are made in the form of classification into one of three criterion groups. Then, $\mathbf{H}$ may be

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

Then,

$$\mathbf{F} = \quad \mathbf{G}^{*\prime}\mathbf{H} = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

It is possible to redefine $\mathbf{X}$ using contrast vectors, so that it has a full column rank. For example,

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}. \tag{9}$$

This $\mathbf{X}$ spans the same column space as the one defined in (8), but has full column rank. Consequently, $\mathbf{X}'\mathbf{K}\mathbf{X}$ in (4) and (7) will be nonsingular, and the regular inverse can be calculated. Note that the first column of $\mathbf{X}$ in (9) pertains to the effect of row marginals, and the second and the third columns represent the contrast vectors for the overall main effects of variable A and variable B, respectively.

2.4 *ANOVA of Categorical Data*

Nishisato (1971, 1980, pp. 184-187) proposed ANOVA of multiple-choice categorical data, which can be put in a form similar to (4). A major difference is that in Nishisato's case, not only the predictor but also the criterion variables consist of more than a single categorical variable. This, however, should not cause any special problem. Multiple criterion variables can be treated in a manner similar to the multiple predictor variables. First, a single composite criterion variable, $\mathbf{H}^*$, is constructed by interactive coding of the criterion variables. This $\mathbf{H}^*$ is analogous to $\mathbf{G}^*$ introduced in section 2.2. Let

$$\mathbf{F} = \mathbf{G}^{*\prime}\mathbf{H}^*,$$

and let $\mathbf{Y}$ be such that

$$\mathbf{H} = \mathbf{H}^*\mathbf{Y}.$$

This $\mathbf{Y}$ is analogous to $\mathbf{X}$ for the predictor variables, and pertains to the column structure of $\mathbf{F}$ which is a simple two-way contingency table.

Canonical correlation analysis between $\mathbf{G}$ and $\mathbf{H}$ amounts to the GSVD of

$$(\mathbf{G}'\mathbf{G})^+\mathbf{G}'\mathbf{H}(\mathbf{H}'\mathbf{H})^+ = (\mathbf{X}'\mathbf{K}\mathbf{X})^+\mathbf{X}'\mathbf{F}\mathbf{Y}(\mathbf{Y}'\mathbf{L}\mathbf{Y})^+, \tag{10}$$

with column metric, $\mathbf{G}'\mathbf{G} = \mathbf{X}'\mathbf{K}\mathbf{X}$, and row metric, $\mathbf{H}'\mathbf{H} = \mathbf{Y}'\mathbf{L}\mathbf{Y}$, where $\mathbf{K} = \mathbf{G}^{*\prime}\mathbf{G}^*$ as before, and $\mathbf{L} = \mathbf{H}^{*\prime}\mathbf{H}^*$. Equation (10) is similar to (4), except that in (10), additional constraints, $\mathbf{Y}$, are imposed on columns of $\mathbf{F}$ as well as constraints, $\mathbf{X}$, on rows. This is a special case of the general model presented by Takane and Shibayama (1991), who extensively discuss principal component analysis of data matrices with linear constraints on both rows and columns (also, see Takane, 1990). This is also a special case of multiple-set homogeneity analysis of van der Burg et al. (1988), where the number of variable sets is restricted to two.

Nishisato (1972) proposed the replacement of $\mathbf{Y}'\mathbf{L}\mathbf{Y}$ in (10) by diag $(\mathbf{H}'\mathbf{H})$ = diag $(\mathbf{Y}'\mathbf{L}\mathbf{Y})$ for computational convenience. This makes the method more like redundancy analysis (van den Wollenberg, 1977) than canonical correlation analysis. Ter Braak (1988) also proposed essentially the same approach.

Leclerc's (1975) "correspondence analysis of juxtaposed contingency tables" further replaces $\mathbf{X}'\mathbf{K}\mathbf{X}$ in (10) by diag $(\mathbf{G}'\mathbf{G})$ = diag $(\mathbf{X}'\mathbf{K}\mathbf{X})$. The same technique is called "composite" correspondence analysis by Israëls (1987).

2.5 *Joint, Conditional, and Marginal Correspondence Analysis*

Multi-way contingency tables may be subjected to OCA by rearranging them into two-way tables in various ways. Depending on how the rearrangement is done, a different type of correspondence analysis is possible.

For illustration, suppose there are only two predictor variables, A and B, with p_1 and p_2 categories, respectively, and there is only one criterion variable, C, with q categories. The p_1 by p_2 by q table may be rearranged into a p_1p_2 by q two-way contingency table by stacking p_2 slices of p_1 by q tables. This table is equivalent to $\mathbf{F}$ defined in (5), and is called a full or joint two-way table. This table may be subjected to OCA. OCA may also be applied to each of p_2 subtables of order p_1 by q separately (D'Ambra & Lauro, 1989; Israëls, 1987). These subtables are called conditional tables. (D'Ambra and Lauro call separate analyses of conditional tables partial correspondence analysis, but others—Israëls, 1987; Yanai, 1986, 1988—call a different technique by the same name. Also, see section 4.) The full joint table may be collapsed across p_1

categories of variable A to obtain the p_2 by q marginal table, or across p_2 categories of variable B to obtain the p_1 by q marginal table. OCA may also be applied to these marginal tables.

The OCA of the full joint table is equivalent to CCA of the same table with $\mathbf{X} = \mathbf{I}$. The basis vectors spanning the $p_1 p_2$-dimensional vector space of $\mathbf{I}$ may be changed without affecting the final solution. For example, assume, for simplicity, that $p_1 = p_2 = 2$. Instead of $\mathbf{X} = \mathbf{I}$, $\mathbf{X}$ may be taken to be

$$\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_{12}] = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}, \tag{11}$$

which spans the same space as $\mathbf{I}_4$. This $\mathbf{X}$ may, in a sense, be more informative, since each column of $\mathbf{X}$ has a specific meaning of its own. Equation (11) is similar to (9), except that $\mathbf{X}$ in (11) has an extra column, $\mathbf{x}_{12}$, pertaining to the interaction effect between variables A and B.

Separate analyses of conditional tables can also be expressed as a special form of CCA. Using the same 2 by 2 example as above, OCA of the 2 by q subtable for b_1 is equivalent to CCA of the 4 by q full table with

$$\mathbf{X}^{(b1)} = [\mathbf{x}_0^{(b1)}, \mathbf{x}_1^{(b1)}] = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \tag{12}$$

and the row metric, $\mathbf{L}_1$, which is a diagonal matrix of column totals of the subtable. The second column of $\mathbf{X}^{(b1)}$, that is, $\mathbf{x}_1^{(b1)}$, represents the simple main effect of variable A within the first level of variable B. Similarly, for b_2

$$\mathbf{X}^{(b2)} = [\mathbf{x}_0^{(b2)}, \mathbf{x}_1^{(b2)}] = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{13}$$

The $\mathbf{x}_1^{(b2)}$ represents the simple main effect of A within b_2. Note that $\mathbf{X}^{(b1)}$ and $\mathbf{X}^{(b2)}$ taken together span the same space as $\mathbf{X}$ defined in (11). Note also that $\mathbf{x}_1^{(b1)} = (1/2)(\mathbf{x}_1 + \mathbf{x}_{12})$ and $\mathbf{x}_1^{(b2)} = (1/2)(\mathbf{x}_1 - {}_{12})$, so that the simple main effects of A within B, $\mathbf{x}_1^{(b1)}$ and $\mathbf{x}_1^{(b2)}$ taken together, pertain to the overall main effect of A ($\mathbf{x}_1$) and the interaction between A and B ($\mathbf{x}_{12}$). Similarly, $\mathbf{x}_0^{(b1)} = (1/2)(\mathbf{x}_0 + \mathbf{x}_2)$ and $\mathbf{x}_0^{(b2)} = (1/2)(\mathbf{x}_0 - \mathbf{x}_2)$, so that $\mathbf{x}_0^{(b1)}$ and $\mathbf{x}_0^{(b2)}$, taken together, pertain to the row marginal effect ($\mathbf{x}_0$) and the overall main effect of B ($\mathbf{x}_2$).

OCA of the p_2 by q marginal table is equivalent to CCA of the full table with

$$\mathbf{X}^{(B)} = [\mathbf{x}_0, \mathbf{x}_2] = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix}, \tag{14}$$

where, as in (11), $\mathbf{x}_0$ represents the row marginal effect, and $\mathbf{x}_2$, the overall main effect of B. OCA of the p_1 by q marginal table can also be expressed analogously.

If the effect of $\mathbf{x}_0$ in (11) is eliminated, the remaining effects are A (the overall main effect of A), B (the overall main effect of B) and AB (the interaction between A and B).

(How particular effects can be eliminated will be discussed in section 3.1.) If the effects of $\mathbf{x}_0^{(b1)}$ and $\mathbf{x}_0^{(b2)}$ in (12) and (13) are eliminated, B as well as the row marginal effect will be eliminated. What remains are $A(b_1)$, the simple main effect of A within b_1, and $A(b_2)$, the simple main effect of A within b_2, which taken together, are equivalent to A plus AB. If the effect of $\mathbf{x}_0$ in (14) is eliminated, B is the only effect remaining. Thus, after eliminating the effects of the respective first columns in (11), (12), (13), and (14), what can be accounted for by (12), (13), and (14) add up to what can be accounted for by (11). It can also be readily verified that $\mathbf{x}_1^{(b1)}$, $\mathbf{x}_1^{(b2)}$, and $\mathbf{x}_2$ span the same space as do $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_{12}$.

The above discussion suggests an important property of correspondence analysis with linear constraints. As has been noted already, OCA presupposes Sp ($\mathbf{X}$) = Sp ($\mathbf{I}$). Specifying an $\mathbf{X}$ such that Sp ($\mathbf{X}$) $\subset$ Sp ($\mathbf{I}$) is equivalent to partialing out certain effects from $\mathbf{I}$. For example, using (14) for $\mathbf{X}$ implies including the overall main effect of B, which is equivalent to excluding the overall main effect of A and the interaction effect between A and B from Sp ($\mathbf{I}$). This observation will be made more rigorous in the next section.

3. The Null Space Method

In this section, we discuss methods that use the null space method for specifying linear constraints. We also discuss a specific relationship between these methods and those discussed in the previous sections.

3.1 Canonical Analysis with Linear Constraints

Böckenholt and Böckenholt (1990; also, see Cazes, Chessel, & Doledec, 1988) proposed least squares canonical analysis with linear constraints (CALC) on both rows and columns of contingency tables. The constraints are of the form, $\mathbf{R}'\mathbf{U} = \mathbf{0}$ and $\mathbf{C}'\mathbf{V} = \mathbf{0}$, where $\mathbf{R}$ and $\mathbf{C}$ are given matrices (called row and column constraint matrices), and $\mathbf{U}$ and $\mathbf{V}$ are matrices of row and column representations, respectively. The least squares CALC obtains the GSVD of

$$\mathbf{M}_{\ddagger} = \mathbf{K}^{-1}\mathbf{Q}_{R/K^{-1}}\mathbf{F}(\mathbf{Q}_{C/L^{-1}})'\mathbf{L}^{-1}, \tag{15}$$

with metrics $\mathbf{K}$ and $\mathbf{L}$, where

$$\mathbf{Q}_{R/K^{-1}} = \mathbf{I} - \mathbf{R}(\mathbf{R}'\mathbf{K}^{-1}\mathbf{R})^{-}\mathbf{R}'\mathbf{K}^{-1}, \tag{16}$$

and

$$\mathbf{Q}_{C/L^{-1}} = \mathbf{I} - \mathbf{C}(\mathbf{C}'\mathbf{L}^{-1}\mathbf{C})^{-}\mathbf{C}'\mathbf{L}^{-1}, \tag{17}$$

are projection operators onto Ker ($\mathbf{R}'\mathbf{K}^{-1}$) along Sp ($\mathbf{R}$) and onto Ker ($\mathbf{C}'\mathbf{L}^{-1}$) along Sp ($\mathbf{C}$), respectively. (Note that $\mathbf{Q}_{R/K^{-1}}$ is not the inverse of $\mathbf{Q}_{R/K}$, but the projection operator onto the ortho-complement space of $\mathbf{R}$ in metric $\mathbf{K}^{-1}$. The $\mathbf{Q}_{C/L^{-1}}$ is similar.) The method partials out the effects of $\mathbf{R}$ and $\mathbf{C}$ from their respective Sp ($\mathbf{I}$) before correspondence analysis is applied. The method may be considered a special case of canonical correlation analysis between $\mathbf{G}$ and $\mathbf{H}$, where the canonical weights, $\mathbf{U}$ and $\mathbf{V}$, are constrained by $\mathbf{R}'\mathbf{U} = \mathbf{0}$ and $\mathbf{C}'\mathbf{V} = \mathbf{0}$ (see Yanai & Takane, 1990).

For illustration, let us temporarily assume that only the row constraints exist. Then, (15) becomes

$$\mathbf{M}_{\ddagger} = \mathbf{K}^{-1}(\mathbf{I} - \mathbf{R}(\mathbf{R}'\mathbf{K}^{-1}\mathbf{R})^{-}\mathbf{R}'\mathbf{K}^{-1})\mathbf{F}\mathbf{L}^{-1}. \tag{15'}$$

The following lemma by Khatri (1966) establishes a direct relationship between $\mathbf{M}_3^*$ in (15′) and $\mathbf{M}_2^*$ in (7).

Lemma (Khatri, 1966; also, see Rao, 1973, p. 77 and Seber, 1984, p. 536). Let $\mathbf{A}$ ($p^* \times r_A$; $r_A \le p^*$) and $\mathbf{B}$ ($p^* \times r_B$; $r_B \le p^*$) be such that $\mathbf{A}'\mathbf{B} = \mathbf{0}$ and rank ($\mathbf{A}$) + rank ($\mathbf{B}$) = p^*. Let $\mathbf{S}$ be a p^* by p^* symmetric positive-definite (metric) matrix. Then,

$$\mathbf{S}^{-1}\mathbf{Q}_{A/S}^{-1} = \mathbf{S}^{-1}(\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{S}^{-1}),$$

$$= \mathbf{B}(\mathbf{B}'\mathbf{S}\mathbf{B})^{-}\mathbf{B}' = \mathbf{S}^{-1}\mathbf{P}_{SB/S^{-1}}. \tag{18}$$

The definition of $\mathbf{Q}_{A/S^{-1}}$ is analogous to $\mathbf{Q}_{R/K^{-1}}$ in (16), and $\mathbf{P}_{SB/S^{-1}}$ is the projection operator defined by column vectors of $\mathbf{SB}$ in the metric of $\mathbf{S}^{-1}$.

Note that the above lemma is stated in slightly more general terms than Khatri's original lemma in that neither $\mathbf{A}$ nor $\mathbf{B}$ is assumed to have full column ranks. This generalization, however, is rather trivial, since $\mathbf{A}$ and $\mathbf{B}$ can always be made to have full column ranks, in which case $r_A + r_B = p^*$. Proofs of the lemma in a slightly more restrictive form (where the g-inverses are replaced by the regular inverses) can be found in the articles cited above. It is fairly obvious that the g-inverses can be used where they are in (18), because of the invariance property of orthogonal projectors with respect to the type of inverse. In (18), however, $\mathbf{S}$ is assumed nonsingular. Equation (18) can be further generalized to allow a singular $\mathbf{S}$. See Appendix B.

Now showing the relationship between (15′) and (7) is straightforward. Let $\mathbf{A} = \mathbf{R}$, $\mathbf{S} = \mathbf{K}$ and $\mathbf{B} = \mathbf{X}$ and apply the lemma to (15′). Then,

$$\mathbf{M}_3^* = \mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{X})^{-}\mathbf{X}'\mathbf{F}\mathbf{L}^{-1} = \mathbf{M}_2^*,$$

where $\mathbf{X}$ is chosen such that $\mathbf{R}'\mathbf{X} = \mathbf{0}$ and p^* = rank ($\mathbf{R}$) + rank ($\mathbf{X}$). Given an $\mathbf{R}$, any $\mathbf{X}$ satisfying these conditions will suffice. An easy way to obtain such an $\mathbf{X}$ is by a square root decomposition of $\mathbf{Q}_R = \mathbf{I} - \mathbf{R}(\mathbf{R}'\mathbf{R})^{-}\mathbf{R}'$ into $\mathbf{XX}'$, such that $\mathbf{X}'\mathbf{X} = \mathbf{I}$. Alternatively, a square root decomposition of $\mathbf{Q}_{K^{-1/2}R}$ (the orthogonal projection operator onto Ker ($\mathbf{R}'\mathbf{K}^{-1/2}$)) into $\mathbf{X}^*\mathbf{X}^{*\prime}$ such that $\mathbf{X}^{*\prime}\mathbf{X}^* = \mathbf{I}$ may be obtained, and $\mathbf{X}$ is set to $\mathbf{X} = \mathbf{K}^{-1/2}\mathbf{X}^*$. This $\mathbf{X}$ has the property that $\mathbf{X}'\mathbf{K}\mathbf{X} = \mathbf{I}$. Similarly, for a given $\mathbf{X}$, an $\mathbf{R}$ satisfying these conditions can be obtained by a square root decomposition of $\mathbf{Q}_x = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ into $\mathbf{RR}'$, such that $\mathbf{R}'\mathbf{R} = \mathbf{I}$, or by $\mathbf{R} = \mathbf{K}^{1/2}\mathbf{R}^*$, where $\mathbf{R}^*$ is obtained by a square root decomposition of $\mathbf{Q}_{K^{1/2}X}$ (the orthogonal projection operator onto Ker ($\mathbf{X}'\mathbf{K}^{1/2}$)) into $\mathbf{R}^*\mathbf{R}^{*\prime}$ such that $\mathbf{R}^{*\prime}\mathbf{R}^* = \mathbf{I}$. The latter $\mathbf{R}$ has the property that $\mathbf{R}'\mathbf{K}^{-1}\mathbf{R} = \mathbf{I}$. A similar relationship also holds between $\mathbf{Y}$ and $\mathbf{C}$ for column restrictions. In general,

$$\mathbf{K}^{-1}\mathbf{Q}_{R/K^{-1}}\mathbf{F}(\mathbf{Q}_{C/L^{-1}})'\mathbf{L}^{-1} = \mathbf{K}^{-1}\mathbf{P}_{KX/K^{-1}}\mathbf{F}(\mathbf{P}_{LY/L^{-1}})'\mathbf{L}^{-1}. \tag{19}$$

Equation (19) shows the equivalence between CALC and Nishisato's ANOVA of categorical data discussed in section 2.4. It also suggests that whether $\mathbf{R}$ and $\mathbf{C}$ are specified in the form of $\mathbf{R}'\mathbf{U} = \mathbf{0}$ and $\mathbf{C}'\mathbf{V} = \mathbf{0}$ (the null space specification) or $\mathbf{X}$ and $\mathbf{Y}$ are specified in the form of $\mathbf{U} = \mathbf{X}\mathbf{U}^*$ and $\mathbf{V} = \mathbf{Y}\mathbf{V}^*$, identical results can be obtained by correspondence analysis of either the first term or the fourth term in the decomposition of a data matrix according to external information proposed by Takane and Shibayama (1991). Also, see Takane (1990).

If $\mathbf{X}$ and $\mathbf{R}$ are obtained by a square root decomposition of $\mathbf{Q}_R$ and $\mathbf{Q}_X$ or by an alternative method, they allow arbitrary (but nonsingular) linear transformations of the form $\mathbf{XW}$ and $\mathbf{RW}$, which do not change the column spaces of $\mathbf{X}$ and $\mathbf{R}$. Consequently, each column of $\mathbf{X}$ and $\mathbf{R}$ thus obtained is not likely to have any specific meaning. In certain special cases, $\mathbf{X}$ and $\mathbf{R}$ with more interpretable columns may be obtained. Suppose $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2]$ is known, such that $\mathbf{T}_1' \mathbf{T}_2 = \mathbf{0}$, $\mathbf{T}'\mathbf{1} = \mathbf{0}$ and Sp $(\mathbf{T})$ = Ker $(\mathbf{1}')$. Such a $\mathbf{T}$ may be obtained by a set of mutually orthogonal contrasts specified on row categories of $\mathbf{F}$. Suppose, as an example, that $\mathbf{R} = \mathbf{K1}$. Then, one choice of $\mathbf{X}$ is $\mathbf{X} = \mathbf{K}^{-1}\mathbf{T}$. It can also be $\mathbf{X} = (\mathbf{I} - \mathbf{1}(\mathbf{1}'\mathbf{K1})^{-1}\mathbf{1}'\mathbf{K})\mathbf{T}$. (This eliminates the unit singular value and constant singular vectors.) Suppose, as another example, that $\mathbf{R} = [\mathbf{K1}, \mathbf{T}_1]$. Then, one choice of $\mathbf{X}$ is $\mathbf{X} = (\mathbf{I} - \mathbf{1}(\mathbf{1}'\mathbf{K1})^{-1}\mathbf{1}'\mathbf{K})\mathbf{T}_2$. Suppose $\mathbf{R} = \mathbf{T}_1$, as a third example. Then, $\mathbf{X} = [\mathbf{1}, \mathbf{T}_2]$. Essentially the same holds between $\mathbf{Y}$ and $\mathbf{C}$. It is interesting to note that in CALC, it is not necessary to explicitly impose $\mathbf{C}'\mathbf{V} = \mathbf{1}'\mathbf{LV} = \mathbf{0}'$ to eliminate column marginal effects, so far as $\mathbf{R}$ contains Sp $(\mathbf{K1})$. The $\mathbf{1}'\mathbf{LV} = \mathbf{0}'$ and $\mathbf{M}_3^*\mathbf{L1} = \mathbf{0}$ still hold, since $\mathbf{F1} = \mathbf{K1}$.

Böckenholt and Böckenholt (1990) has the example of $\mathbf{R} = [\mathbf{K1}, \mathbf{T}_1]$, where

$$\mathbf{T}_1' = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 \\ 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

(Also, see Gilula & Haberman, 1988.) The first three columns of $\mathbf{T}_1$ stipulate that there are no quadratic trends over three levels of variable A at any of three levels of variable B. The fourth column, on the other hand, stipulates that $a_1 - a_2$ at b_1 is the same as $a_1 - a_2$ at b_2, implying no interaction between A and B for the first two levels of both variables. The $\mathbf{T}_2$ in this case might be

$$\mathbf{T}_2' = \begin{bmatrix} 1 & 0 & -1 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & -2 & -2 & -2 \end{bmatrix}.$$

The first column specifies the same linear trend over the three levels of variable A at both b_1 and b_2. The second column specifies a different linear trend over the levels of A at b_3. The last two columns pertain to the overall main effect of variable B. Note that $\mathbf{T}_1'\mathbf{T}_2 = \mathbf{0}$, $\mathbf{T}'\mathbf{1} = \mathbf{0}$ and Sp $(\mathbf{T})$ = Ker $(\mathbf{1}')$. An $\mathbf{X}$, therefore, could be $\mathbf{X} = (\mathbf{I} - \mathbf{1}(\mathbf{1}'\mathbf{K1})^{-1}\mathbf{1}'\mathbf{K})\mathbf{T}_2$, which is orthogonal to both $\mathbf{K1}$ and $\mathbf{T}_2$.

There are thus two alternative ways of specifying linear constraints, one via $\mathbf{X}$ and the other via $\mathbf{R}$ (Schmoyer, 1984). They closely parallel two ways of obtaining constrained least squares estimates in linear models, the projection method, and the Lagrangian multiplier method. This will be shown in Appendix C.

3.2 The Zero Average Restrictions

Van der Heijden and his collaborators (van der Heijden, de Falguerolles, & de Leeuw, 1989; van der Heijden & de Leeuw, 1985; van der Heijden & Meijerink, 1989) proposed correspondence analysis to analyze residuals from certain loglinear models for contingency tables. This procedure was initially motivated by the fact that OCA analyzes residuals from the loglinear independence model; that is, interactions between rows and columns after row and column marginal effects are eliminated. The loglinear model and the maximum likelihood estimation thereof are not always compatible with our least squares framework. However, certain loglinear models are equivalent to CCA. This is the case when the estimates of the fitted loglinear model can be obtained noniteratively, as in the cases of independence and conditional independence models.

Van der Heijden et al.'s analysis and CALC are completely equivalent in this case. When the fitted loglinear model requires an iterative solution, however, the two methods are similar, but not identical. However, even in this case, it is possible to develop the CALC version of "zero average" restrictions analogous to those derived from the loglinear model (van der Heijden & Meijerink, 1989).

Let us illustrate using the simple example considered in section 2.5, and the loglinear models fitted by van der Heijden and Meijerink (1989). Let

$$\mathbf{R}^* = [\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_{12}] = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}. \tag{20}$$

(This $\mathbf{R}^*$ is the same as $\mathbf{X}$ defined in (11), and the column vectors of $\mathbf{R}^*$ are mutually orthogonal.) Recall that $\mathbf{x}_0$ pertains to the marginal effect of rows, $\mathbf{x}_1$, the overall main effect of variable A, $\mathbf{x}_2$, that of variable B, and $\mathbf{x}_{12}$, the interaction between A and B. The criterion variable is denoted by C.

Fitting the (loglinear) independence model, often denoted as (AB) (C), is equivalent to CCA with $\mathbf{X} = \mathbf{1}$ and $\mathbf{Y} = \mathbf{1}$. Analyzing the residuals is equivalent to applying correspondence analysis subject to the restrictions that $\mathbf{x}_0'\mathbf{KU} = \mathbf{1}'\mathbf{KU} = \mathbf{0}$ and $\mathbf{1}'\mathbf{LV} = \mathbf{0}$, where $\mathbf{U}$ and $\mathbf{V}$ are row and column representations, respectively, and this is equivalent to CALC with $\mathbf{R} = \mathbf{K1}$ and $\mathbf{C} = \mathbf{L1}$. This, in turn, is equivalent to CCA with an $\mathbf{X}$, say, $\mathbf{X} = \mathbf{K}^{-1}[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_{12}]$ in (7). Note that $\mathbf{1}'\mathbf{KM}_3^* = \mathbf{R}'\mathbf{M}_3^* = \mathbf{0}$ and $\mathbf{M}_3^*\mathbf{L1} = \mathbf{M}_3^*\mathbf{C} = \mathbf{0}$, where $\mathbf{M}_3^*$ is defined in (15). These are called "zero average restrictions" by van der Heijden and Meijerink (1989).

Residuals from other loglinear models may be analyzed by correspondence analysis. For simplicity, only $\mathbf{R}$ is varied, while $\mathbf{C}$ is fixed at $\mathbf{C} = \mathbf{L1}$. Fitting the loglinear model, (AB) (BC), implies the overall main effect of B is fitted in addition to the row marginal effect (AB). This model is a conditional independence model (A and C independent, given B) for which a noniterative solution exists. This noniterative solution is obtained by taking appropriate conditional marginals of $\mathbf{F}$. In CCA, this amounts to choosing

$$\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

Note that the above $\mathbf{X}$ spans the same column space as $\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_2]$, and CCA with this latter $\mathbf{X}$ is known to be equivalent to OCA of the B by C marginal table, suggesting that this loglinear model corresponds with the saturated model for the B by C marginal table. The row representation, $\mathbf{U}$, obtained by OCA of residuals from this model would satisfy $[\mathbf{x}_0, \mathbf{x}_2]'\mathbf{KU} = \mathbf{0}$, and consequently, this analysis is equivalent to CALC with $\mathbf{R} = \mathbf{K}[\mathbf{x}_0, \mathbf{x}_2]$ and to CCA with an $\mathbf{X}$, say, $\mathbf{X} = \mathbf{K}^{-1}[\mathbf{x}_1, \mathbf{x}_{12}]$. The zero average restrictions will take the form of

$$[\mathbf{x}_0, \mathbf{x}_2]'\mathbf{KM}_3^* = \mathbf{R}'\mathbf{M}_3^* = \mathbf{0}.$$

Fitting the loglinear model, (AB) (AC) (BC), means the overall main effect of A is added to the previous model. This model cannot be fitted noniteratively, and the residual matrix from this model is not equal to $\mathbf{M}_3^*$. Still, the $\mathbf{U}$ obtained by OCA of the residual table satisfies $[\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2]'\mathbf{KU} = \mathbf{0}$, but this analysis is not equivalent (only analogous) to CALC with $\mathbf{R} = \mathbf{K}[\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2]$ or CCA with an $\mathbf{X} = \mathbf{K}^{-1}\mathbf{x}_{12}$. The CALC version of the zero average restrictions is given by

$$[\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2]'\mathbf{K}\mathbf{M}_3^* = \mathbf{R}'\mathbf{M}_3^* = \mathbf{0}.$$

Van der Heijden et al.'s approach can thus be reformulated in the least squares framework, which turns out to be a special case of CALC (Böckenholt & Böckenholt, 1990), and also of CCA (ter Braak, 1986).

Van der Heijden and Worsley (1988) not only analyze residuals but also attempt to model them. Their modelling approach to the residuals is analogous to the GSVD of

$$\mathbf{M}_4^* = \mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{X})^-\mathbf{X}'\mathbf{K}\mathbf{M}_3^*\mathbf{L}\mathbf{Y}(\mathbf{Y}'\mathbf{L}\mathbf{Y})^-\mathbf{Y}',$$

with metrics $\mathbf{K}$ and $\mathbf{L}$ and for some $\mathbf{X}$ and $\mathbf{Y}$. A similar analysis can also be done by ideal point discriminant analysis (Takane, 1987).

4. Concluding Remarks

This paper discussed two alternative ways of imposing linear constraints, the reparametrization method and the null space method, and demonstrated essential equivalences among the methods of linearly constrained correspondence analysis that fall into either one of these two classes of methods. Outside the linearly constrained correspondence analysis, the following methods, for example, use the reparametrization method: CANDELINC or canonical decomposition with linear constraints (Carroll, Pruzansky, & Kruskal, 1980), dual scaling with external criteria (Nishisato, 1980), multiattribute conjoint analysis (DeSarbo, Carroll, Lehmann, & O'Shaughnessy, 1982), GENFOLD2 or a restricted and unrestricted multidimensional unfolding procedure (DeSarbo & Rao, 1984; also, see Heiser, 1981, 1987), and ideal point discriminant analysis (Takane, 1987). On the other hand, the restricted eigenvalue problem (Rao, 1973, p.50), MULTISCALE, a maximum likelihood multidimensional scaling procedure (Ramsay, 1982), restricted maximum likelihood canonical analysis and association models (Gilula & Haberman, 1988), and so on, use the null space specification method. The results in this paper indicate there are alternative formulations to these methods, incorporating alternative ways of specifying linear constraints. Many other multivariate methods, such as multiple regression (e.g., Searle, 1971; Seber, 1977, 1984; also, see Appendix C), growth curve models (GMANOVA; e.g., Seber, 1984, pp. 474-492), multivariate linear hypotheses (e.g., Timm, 1975), and constrained principal component analysis (CPCA; Takane & Shibayama, 1991), use both the reparametrization and the null space specification methods.

It is difficult to make general remarks on the relative merits of the two approaches. Most often, natural forms of constraints follow from specific empirical questions posed by the investigator. Aside from empirical concerns, and in the context of linearly constrained correspondence analysis, however, the reparametrization method seems to have some computational advantage over the null space method. The former can solve for the GSVD of $\mathbf{M}_2$ (rather than $\mathbf{M}_2^*$), which is usually much smaller in size than $\mathbf{M}_3^*$ whose GSVD is to be obtained in the null space method.

Ter Braak (1988) proposed partial canonical correspondence analysis (PCCA). This method eliminates the effect of extraneous variables $\mathbf{Z}$ from the predictor variables $\mathbf{X}$. Let $\mathbf{Q}_{Z/K} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{K}\mathbf{Z})^-\mathbf{Z}'\mathbf{K}$, and define $\mathbf{X}^* = \mathbf{Q}_{Z/K}\mathbf{X}$. This $\mathbf{X}^*$ is used in CCA to obtain PCCA. PCCA is interesting in two respects from our viewpoint. First, the original PCCA is based on the reparametrization method for incorporating the linear constraints. It obtains the GSVD of $\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{K}\mathbf{X}^*)^-\mathbf{X}^{*\prime}\mathbf{F}\mathbf{L}^{-1}$ with metrics $\mathbf{K}$ and $\mathbf{L}$. An equivalent formulation is possible based on the null space method, which obtains the GSVD of $\mathbf{K}^{-1}(\mathbf{I} - \mathbf{R}(\mathbf{R}'\mathbf{K}^{-1}\mathbf{R})^{-}\mathbf{R}'\mathbf{K}^{-1})\mathbf{F}\mathbf{L}^{-1}$ under the same metrics, but with an $\mathbf{R}$ such that $\mathbf{I} - \mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-}\mathbf{X}^{*\prime} = \mathbf{R}\mathbf{R}'$ and $\mathbf{R}'\mathbf{R} = \mathbf{I}$, or an $\mathbf{R}$ such that $\mathbf{R} = \mathbf{K}^{1/2}\mathbf{R}^*$, where $\mathbf{R}^*$ is such that $\mathbf{Q}_{K^{1/2}X^*} = \mathbf{R}^*\mathbf{R}^{*\prime}$ and $\mathbf{R}^{*\prime}\mathbf{R}^* = \mathbf{I}$. (The $\mathbf{Q}_{K^{1/2}X^*}$ is the orthogonal projection operator onto Ker $(\mathbf{X}^{*\prime}\mathbf{K}^{1/2})$). Secondly, PCCA has to reduce to CALC of Böckenholt and Böckenholt, when $\mathbf{X} = \mathbf{I}$ and $\mathbf{Z} = \mathbf{K}^{-1}\mathbf{R}$. This is indeed the case, since $\mathbf{X}^* = (\mathbf{Q}_{R/K^{-1}})'$, and $\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{K}\mathbf{X}^*)^{-}\mathbf{X}^{*\prime} = (\mathbf{Q}_{R/K^{-1}})'\,(\mathbf{Q}_{R/K^{-1}}\mathbf{K}(\mathbf{Q}_{R/K^{-1}})')^{-}\mathbf{Q}_{R/K^{-1}} = \mathbf{K}^{-1}\,\mathbf{Q}_{R/K^{-1}}$. Note that $\mathbf{K}(\mathbf{Q}_{R/K^{-1}})' = \mathbf{Q}_{R/K^{-1}}\mathbf{K}$, $(\mathbf{Q}_{R/K^{-1}})^2 = \mathbf{Q}_{R/K^{-1}}$, $\mathbf{Q}_{R/K^{-1}}$ is a g-inverse of itself, $(\mathbf{Q}_{R/K^{-1}}\mathbf{K})^{-} = \mathbf{K}^{-1}\mathbf{Q}_{R/K^{-1}}$ and $(\mathbf{Q}_{R/K^{-1}})'\mathbf{K}^{-1} = \mathbf{K}^{-1}\mathbf{Q}_{R/K^{-1}}$ (The last equation requires $(\mathbf{AB})^{-} = \mathbf{B}^{-}\mathbf{A}^{-}$ if $(\mathbf{A}^{-}\mathbf{A}\mathbf{B}\mathbf{B}^{-})^2 = \mathbf{A}\mathbf{A}\mathbf{B}\mathbf{B}^{-}$; see Yanai & Takeuchi, 1983, p. 76).

Partial correspondence analysis (Yanai, 1986, 1988; also, see Israëls, 1987) is distinct from any of the methods discussed in this paper. It obtains the GSVD of $(\mathbf{G}'\mathbf{Q}_J\mathbf{G})^{-1}\mathbf{G}'\mathbf{Q}_J\mathbf{H}(\mathbf{H}'\mathbf{Q}_J\mathbf{H})^{-1}$ with metrics $\mathbf{G}'\mathbf{Q}_J\mathbf{G}$ and $\mathbf{H}'\mathbf{Q}_J\mathbf{H}$ where $\mathbf{Q}_J = \mathbf{I} - \mathbf{J}(\mathbf{J}'\mathbf{J})^{-}\mathbf{J}'$ with $\mathbf{J}$ being the subject design matrix, whose effect is to be eliminated. It would be interesting to combine Yanai's approach with Nishisato's described in section 2.4, or Böckenholt and Böckenholt's approach described in section 3.1.

Appendix

In this appendix we present: (A) an alternative derivation of canonical correspondence analysis (CCA), (B) a generalization of Khatri's lemma, and (C) constrained least squares solutions in linear models.

A. *Alternative Derivation of Canonical Correspondence Analysis*

CCA was originally derived (ter Braak, 1986) as an approximation to the unfolding type of single-peaked response surface model called Gaussian ordination. In view of its relationship to OCA, CCA can be derived from a somewhat different perspective, but on the same unfolding rationale (Heiser, 1981; Takane, 1980).

In this appendix, CCA is first derived in its general form and then specialized. Let $\mathbf{F}$ be a given data matrix whose entries represent some sort of similarities between rows and columns. The similarities are assumed to be all nonnegative. In the unfolding model, the rows and the columns of $\mathbf{F}$ are represented as points in a multidimensional Euclidean space. Let $\mathbf{U}$ and $\mathbf{V}^*$ be row and column representations of $\mathbf{F}$, respectively. It is assumed that

$$\mathbf{U} = \mathbf{X}\mathbf{U}^*, \tag{21}$$

where $\mathbf{X}$ is the matrix of predictor variables, and $\mathbf{U}^*$ a matrix of weights. The predictor variables can include both continuous and discrete variables.

Consider finding $\mathbf{U}^*$ and $\mathbf{V}^*$ such that

$$\phi = \mathrm{tr}(\mathbf{F}'\mathbf{D}^{(2)}), \tag{22}$$

is minimized, subject to

$$\mathbf{U}'\mathbf{K}\mathbf{U} = \mathbf{U}^{*\prime}\mathbf{X}'\mathbf{K}\mathbf{X}\mathbf{U}^* = \mathbf{I}, \tag{23}$$

where $\mathbf{D}^{(2)}$ is the matrix of squared Euclidean distances between the row and column points, and $\mathbf{K}$ the diagonal matrix of row totals of $\mathbf{F}$. (Matrix $\mathbf{K}$ may be singular, provided that rank $(\mathbf{KX})$ = rank $(\mathbf{X})$. Nonsingularity of $\mathbf{K}$ is ensured, however, if no rows of $\mathbf{F}$ are entirely zeroes.) The above criterion requires that the squared Euclidean distances be inversely related to the corresponding entries of $\mathbf{F}$ as much as possible. Matrix $\mathbf{D}^{(2)}$ can be more explicitly written as

$$\mathbf{D}^{(2)} = \text{diag}\ (\mathbf{XU}^*\mathbf{U}^{*\prime}\mathbf{X}')\mathbf{1}_m\mathbf{1}_n' - 2\mathbf{XU}^*\mathbf{V}^{*\prime} + \mathbf{1}_m\mathbf{1}_n'\ \text{diag}\ (\mathbf{V}^*\mathbf{V}^{*\prime}), \tag{24}$$

where $\mathbf{1}_m$ and $\mathbf{1}_n$ are m- and n-component vectors of ones, respectively. (The m and n are the numbers of rows and columns of $\mathbf{F}$, respectively.) Equation (23) is a normalization restriction. While the specific form of the normalization restriction is somewhat arbitrary, it is convenient to use (23).

Using (24), (22) can be rewritten as

$$\phi = \text{tr}\ (\mathbf{U}^{*\prime}\mathbf{X}'\mathbf{KXU}^*) - 2\ \text{tr}\ (\mathbf{U}^{*\prime}\mathbf{X}'\mathbf{FV}^*) + \text{tr}\ (\mathbf{V}^{*\prime}\mathbf{LV}^*), \tag{22'}$$

where $\mathbf{L}$ is the diagonal matrix of column totals of $\mathbf{F}$. Since

$$\min_{u^*, v^*} \phi = \min_{v^*} \phi$$

where

$$\phi^* = \min_{v^*|u^*} \phi$$

(i.e., the minimum of ϕ with respect to $\mathbf{V}^*$ for $\mathbf{U}^*$ fixed), we first obtain ϕ^*. Differentiating ϕ in (22') with respect to $\mathbf{V}^*$ and setting the result equal to zero gives

$$\frac{1}{2}\frac{\partial \phi}{\partial \mathbf{V}^*} = \mathbf{LV}^* - \mathbf{F}'\mathbf{XU}^* = \mathbf{0}.$$

Hence,

$$\mathbf{V}^* = \mathbf{L}^{-1}\mathbf{F}'\mathbf{XU}^*, \tag{25}$$

where $\mathbf{L}^{-1}$ may be replaced by the Moore-Penrose inverse of $\mathbf{L}$, if $\mathbf{L}$ is not of full rank. Nonsingularity of $\mathbf{L}$ is ensured, if no columns of $\mathbf{F}$ are entirely zeroes. A general treatment of singular metric matrices has been given by de Leeuw (1984) (also, see Takane & Shibayama, 1991).

Using $\mathbf{V}^*$ in (25), ϕ^* can be expressed as

$$\phi^* = \text{tr}\ (\mathbf{U}^{*\prime}\mathbf{X}'\mathbf{KXU}^*) - \text{tr}\ (\mathbf{U}^{*\prime}\mathbf{X}'\mathbf{FL}^{-1}\mathbf{F}'\mathbf{XU}^*). \tag{26}$$

Since under (23), the first term in ϕ^* is a constant, minimizing ϕ^* with respect to $\mathbf{U}^*$ is equivalent to maximizing the second term of (26) under the same normalization restriction. The problem is solved by the generalized eigenvalue problem of

$$\mathbf{X}'\mathbf{FL}^{-1}\mathbf{F}'\mathbf{XU}^* = \mathbf{X}'\mathbf{KXU}^*\mathbf{\Delta}, \tag{27}$$

where $\mathbf{\Delta}$ is a diagonal matrix of eigenvalues. Once $\mathbf{U}^*$ is obtained by solving (27), $\mathbf{U}$ is obtained by (21), and $\mathbf{V}^*$, in turn, by (25). This is essentially equivalent to the GSVD of $(\mathbf{X}'\mathbf{KX})^+\mathbf{X}'\mathbf{FL}^{-1}$ with metrics $\mathbf{X}'\mathbf{KX}$ and $\mathbf{L}$ except that the normalization restriction on $\mathbf{V}^*$, that is, $\mathbf{V}^{*\prime}\mathbf{LV}^* = \mathbf{I}$ is not explicitly imposed in the above derivation.

The data matrix, $\mathbf{F}$, is now specialized. Let $\mathbf{F}$ be a two-way contingency table, such that $\mathbf{F} = \mathbf{G}^{*\prime}\mathbf{H}$, where $\mathbf{G}^*$ (N by m) and $\mathbf{H}$ (N by n) are both single indicator matrices. Such an $\mathbf{F}$ may arise, for example, as follows (Lebreton, et al., 1990). Suppose a sample of N children are drawn from m schools. Which schools children belong to are indicated by $\mathbf{G}^*$. The children are also classified into n groups according to some criterion (e.g., father's occupation). This is indicated by $\mathbf{H}$. Suppose that predictor variables, $\mathbf{X}$, are provided for schools rather than for children. For relating these predictor variables to father's occupation, we may count the numbers of children attending specific schools and having fathers of specific occupational categories, which yield **F**, and perform CCA of **F** with **X** as the predictor variables. Alternatively, we may define $\mathbf{G} = \mathbf{G^*X}$. Then, (22′) can be rewritten as

$$\phi = \mathrm{tr}\,(\mathbf{GU^*} - \mathbf{HV^*})'(\mathbf{GU^*} - \mathbf{HV^*}),$$

since $\mathbf{H1}_n = \mathbf{1}_m$, $\mathbf{L} = \mathbf{H'H}$ (diagonal) and $\mathbf{K} = \mathbf{G^{*\prime}G^*}$. Thus, the same problem can also be solved by canonical correlation analysis between **G** and **H**. CCA at the school level is thus equivalent to canonical correlation analysis at the children's level (Chessel, Lebreton, & Yoccoz, 1987; Lebreton, et al., 1988).

If further $\mathbf{G^*} = \mathbf{I}$ the above special case reduces to the one in which $\mathbf{F} = \mathbf{H}$ (a single indicator matrix), $\mathbf{X} = \mathbf{G}$ and $\mathbf{K} = \mathbf{I}$. The two special cases of CCA which are equivalent to the method of additive scoring follow by specializing **X** in the way described in Section 2.2.

B. A Generalization of Khatri's Lemma

The following theorem generalizes Khatri's lemma.

Theorem. Let **A** and **B** be as stated in Khatri's lemma. Further, let **M** and **N** be nonnegative definite matrices of order p^* satisfying the following conditions: (a) rank (**A**) = rank (**A′M**), (b) rank (**B**) = rank (**B′N**), and (c) $\mathbf{A'MNB} = \mathbf{0}$. Then,

$$\mathbf{I}_{p^*} = \mathbf{A(A'MA)^-A'M} + \mathbf{NB(B'NB)^-B'}, \tag{28}$$

where $\mathbf{I}_{p^*}$ is an identity matrix of order p^*.

The following lemma is useful to prove the theorem.

Lemma (Yanai, 1990). Let **A** and **M** be as defined in the above theorem. Then, the following three statements are equivalent: (1) rank (**A**) = rank (**A′M**), (2) $\mathbf{A(A'MA)^-A'MA} = \mathbf{A}$, and (3) Sp (**A**) + *Ker*(**A′M**) = Sp ($\mathbf{I}_{p^*}$), and Sp (**A**) ∩ Ker (**A′M**) is null.

Similar relations also hold for **B** and **N**.

Proof of the Theorem. Let $\mathbf{J} = \mathbf{A(A'MA)^-A'M} + \mathbf{NB(B'NB)^-B'} - \mathbf{I}_{p^*}$, and prove $\mathbf{J} = \mathbf{0}$. Observe that the condition rank (**A**) = rank (**A′M**) ensures the decomposition of Sp ($\mathbf{I}_{p^*}$) as the direct sum of Sp (**A**) and Ker (**A′M**). Then, for any $\mathbf{x} \in$ Sp (**A**), $\mathbf{Jx} = \mathbf{0}$ (i.e., $\mathbf{JA} = \mathbf{0}$) because of (2) in the above lemma. Furthermore, $\mathbf{A'MNB} = \mathbf{0}$ and rank (**NB**) = rank (**B**) imply Ker (**A′M**) = Sp (**NB**). Thus, for any $\mathbf{y} \in$ Ker (**A′M**), $\mathbf{y} = \mathbf{NBz}$ for some **z**, and $\mathbf{NB(B'NB)^-B'y} = \mathbf{NB(B'NB)^-B'NBz} = \mathbf{NBz} = \mathbf{y}$, so that $\mathbf{Jy} = \mathbf{0}$ (i.e., $\mathbf{JNB} = \mathbf{0}$). This implies $\mathbf{J} = \mathbf{0}$ and (28). □

Corollary (Khatri, 1988, Theorem 1).

(1) If Sp (**N**) ⊃ Sp (**A**), and choose $\mathbf{M} = \mathbf{N}^-$, then

$$\mathbf{N} = \mathbf{A(A'N^-A)^-A'} + \mathbf{NB(B'NB)^-B'N}. \tag{29}$$

(2) If Sp (**M**) ⊃ Sp (**B**), and choose $\mathbf{N} = \mathbf{M}^-$, then

$$\mathbf{M} = \mathbf{MA(A'MA)^-A'M} + \mathbf{B(B'M^-B)^-B'}. \tag{29'}$$

Proof. (1). Postmultiply (28) by **N**. Sp (**N**) ⊃ Sp (**A**) ensures $\mathbf{A'N^-N} = \mathbf{A'}$. Equation (2) can be proved similarly.

$\mathbf{M} = \mathbf{S}^{-1}$ in (29′), where **S** is positive definite, leads to Khatri's lemma as stated in (18). □

C. Constrained Least Squares Problem in Linear Models

Consider the following regression model

$$\mathbf{y} = \mathbf{Wm} + \mathbf{e} \tag{30}$$

where $E(\mathbf{e}) = \mathbf{0}$, and $V(\mathbf{e}) = \sigma^2\mathbf{I}$. Linear constraints on parameter vector $\mathbf{m}$ may be incorporated by

$$\mathbf{m} = \mathbf{Xm}^*, \tag{31}$$

where $\mathbf{X}$ is a known matrix, and $\mathbf{m}^*$ a reduced parameter vector. The ordinary least squares estimate of $\mathbf{m}^*$ is obtained by

$$\hat{\mathbf{m}}^* = (\mathbf{X}'\mathbf{W}'\mathbf{WX})^{-}\mathbf{X}'\mathbf{W}'\mathbf{y},$$

so that

$$\hat{\mathbf{m}} = \mathbf{X}(\mathbf{X}'\mathbf{W}'\mathbf{WX})^{-}\mathbf{X}'\mathbf{W}'\mathbf{y}, \tag{32}$$

and

$$\mathbf{W}\hat{\mathbf{m}} = \mathbf{P}_{X^*}\mathbf{y},$$

with $\mathbf{X}^* = \mathbf{WX}$ and $\mathbf{P}_{X^*} = \mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-}\mathbf{X}^{*\prime}$.

Linear constraints may also be specified by

$$\mathbf{R}'\mathbf{m} = \mathbf{0}. \tag{33}$$

There are two conventional ways of solving a linear least squares problem subject to the constraint of the above form (e.g., Seber, 1984, pp. 403-405). One is the projection method, and the other the Lagrangian multiplier method. In the former we define $\mathbf{Q} = \mathbf{I} - \mathbf{R}(\mathbf{R}'\mathbf{R})^{-}\mathbf{R}'$. Then $\mathbf{m} = \mathbf{Qa}$ for some $\mathbf{a}$. A least squares estimate of $\mathbf{a}$ is obtained by

$$\hat{\mathbf{a}} = (\mathbf{QW}'\mathbf{WQ})^{-}\mathbf{Q}'\mathbf{W}'\mathbf{y},$$

which leads to

$$\hat{\mathbf{m}} = \mathbf{Q}(\mathbf{QW}'\mathbf{WQ})^{-}\mathbf{QW}'\mathbf{y} = \mathbf{T}(\mathbf{T}'\mathbf{W}'\mathbf{WT})^{-}\mathbf{T}'\mathbf{W}'\mathbf{y}, \tag{34}$$

and

$$\mathbf{W}\hat{\mathbf{m}} = \mathbf{P}_{WT}\mathbf{y},$$

where $\mathbf{T}$ is such that $\mathbf{Q} = \mathbf{TT}'$ and $\mathbf{T}'\mathbf{T} = \mathbf{I}$. The second equation in (34) follows from the fact that $\mathbf{T}(\mathbf{T}'\mathbf{W}'\mathbf{WT})^{-}\mathbf{T}'$ is a g-inverse of $\mathbf{QW}'\mathbf{WQ}$. A proof of the latter is straightforward by showing $\mathbf{QW}'\mathbf{WQT}(\mathbf{T}'\mathbf{W}'\mathbf{WT})^{-}\mathbf{T}'\mathbf{QW}'\mathbf{WQ} = \mathbf{QW}'\mathbf{WQ}$. Equation (34) is identical to (32) by setting $\mathbf{X} = \mathbf{T}$. The projection method thus corresponds to the reparametrization method.

In the Lagrangian multiplier method, we obtain

$$\hat{\mathbf{m}} = [(\mathbf{W}'\mathbf{W})^{-1} - (\mathbf{W}'\mathbf{W})^{-1}\mathbf{R}[\mathbf{R}'(\mathbf{W}'\mathbf{W})^{-1}\mathbf{R}]^{-}\mathbf{R}'(\mathbf{W}'\mathbf{W})^{-1}]\mathbf{W}'\mathbf{y}, \tag{35}$$

assuming $\mathbf{W}'\mathbf{W}$ is nonsingular, and

$$\mathbf{W}\hat{\mathbf{m}} = (\mathbf{P}_W - \mathbf{P}_{W^*})\mathbf{y},$$

where $\mathbf{P}_W = \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$ and $\mathbf{P}_{W^*}$ is defined similarly with $\mathbf{W}^* = \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{R}$. For a detailed derivation, see Searle (1971, pp. 110-123) or Seber (1977, pp. 84–87). Khatri's lemma can be used to establish the equivalence between (32) and (35) with $\mathbf{A} = \mathbf{R}$, $\mathbf{B} = \mathbf{X}$, and $\mathbf{S} = \mathbf{W}'\mathbf{W}$. The Lagrangian multiplier method thus corresponds to the null space specification method. The equivalence between (32) and (35) suggests a general decomposition of $\mathbf{P}_W$. That is,

$$\mathbf{P}_W = \mathbf{P}_{WT} + \mathbf{P}_{W^*}, \tag{36}$$

where $\mathbf{P}_{WT}\mathbf{P}_{W^*} = \mathbf{0}$, since $\mathbf{T}'\mathbf{W}'\mathbf{W}^* = \mathbf{T}'\mathbf{W}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{R} = \mathbf{0}$.

When $\mathbf{W}'\mathbf{W}$ is singular, but rank $(\mathbf{W})$ + rank $(\mathbf{R}')$ = rank $(\mathbf{W})$, where $\mathbf{W}' = [\mathbf{W}', \mathbf{R}]$, then $\mathbf{P}_{W^*} = \mathbf{0}$, or $\mathbf{P}_W = \mathbf{P}_{WT}$. This follows from $\mathbf{W}^* = \mathbf{W}(\mathbf{W}'\mathbf{W} + \mathbf{R}\mathbf{R}')^{-1}\mathbf{R} = \mathbf{0}$ (Rao, 1973, p. 34; Seber, 1977, p. 79). In this case, $\mathbf{R}$ is called an identification restriction (Scheffé, 1959, p. 17; Seber, 1977, p. 74).

References

Böckenholt, U., & Böckenholt, I. (1990). Canonical analysis of contingency tables with linear constraints. *Psychometrika*, *55*, 633–639.

Carroll, J. D. (1973). Models and algorithms for multidimensional scaling, conjoint measurement and related techniques. In P. E. Green & Y. Wind (Eds.), *Multiattribute decisions in marketing: A measurement approach* (pp. 299–387). New York: Dryden Press.

Carroll, J. D., Pruzansky, S., & Kruskal, J. B. (1980). CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, *45*, 3–24.

Cazes, P., Chessel, D., & Doledec, S. (1988). L'analyse des correspondances internes d'un tableau partitionné: Son usage en hydrobiologie [Internal correspondence analysis of a partitioned table: Its use in hydrobiology]. *Revue des Statistique Appliquée*, *36*, 39–54.

Chessel, D., Lebreton, J. D., & Yoccoz, N. (1987). Propriétés de l'analyse canonique des correspondances: Une illustration en hydrobiologie [Properties of canonical correspondence analysis: An illustration in hydrobiology.] *Revue des Statistique Appliquée*, *35*, 55–72.

D'Ambra, L., & Lauro, N. (1989). Nonsymmetrical analysis of three-way contingency tables. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 301–315). Amsterdam: Elsevier.

de Leeuw, J. (1984). Fixed rank matrix approximation with singular weights matrices. *Computational Statistics Quarterly*, *1*, 3–12.

DeSarbo, W. S., Carroll, J. D., Lehmann, D. R., & O'Shaughnessy, J. (1982). Three-way multivariate conjoint analysis. *Marketing Science*, *1*, 323–350.

DeSarbo, W. S., & Rao, V. R. (1984). GENFOLD2: A set of models and algorithms for the GENeral UnFOLDing analysis of preference/dominance data. *Journal of Classification*, *1*, 147–186.

Fisher, R. A. (1936). The use of multiple measurement in taxonomic problems. *Annals of Eugenics*, *7*, 179–188.

Fisher, R. A. (1948). *Statistical methods for research workers* (10th ed., Enlarged edition). London: Oliver and Boyd.

Gilula, Z., & Haberman, S. J. (1988). The analysis of contingency tables by restricted canonical and restricted association models. *Journal of the American Statistical Association*, *83*, 760–771.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.

Hayashi, C. (1950). On the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics*, *2*, 35–47.

Hayashi, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics*, *3*, 69–98.

Heiser, W. J. (1981). *Unfolding analysis of proximity data*. Leiden, The Netherlands: DSWO Press.

Heiser, W. J. (1987). Joint ordination of species and sites: The unfolding technique. In P. Legendre & L. Legendre (Eds.), *Developments in numerical ecology* (pp. 189–221). Berlin: Springer.

Israëls, A. (1987). *Eigenvalue techniques for qualitative data*. Leiden, The Netherlands: DSWO Press.

Johnson, P. O. (1950). The quantification of qualitative data in discriminant analysis. *Journal of the American Statistical Association*, *45*, 65–76.

Khatri, C. G. (1966). A note on a MANOVA model applied to problems in growth curves. *Annals of the Institute of Statistical Mathematics*, *18*, 75–86.

Khatri, C. G. (1988). *Some properties of BLUE in a linear model and canonical correlations associated with linear transformations* (Technical Report 88-20) Pittsburgh: University of Pittsburgh, Center for Multivariate Analysis.

Lebreton, J. D., Chessel, D., Prodon, R., & Yoccoz, N. (1988). L'analyse des relations espèces-milieu par l'analyse canonique des correspondences, I. Variables de milieu quantitatives. [Analysis of species-environment relations by canonical correspondence analysis, I. Quantitative environment variables.] *Acta Oecologica*, *9*, 53–67.

Leclerc, A. (1975). L'analyse des correspondence sur juxtaposition de tableaux de contingence [Correspondence analysis of juxtaposed contingency tables]. *Revue de Statistique Appliquée, 23*, 5–16.

Maxwell, A. E. (1961). Canonical variate analysis when the variables are dichotomous. *Educational and Psychological Measurement, 21*, 259–271.

Nishisato, S. (1971). Analysis of variance through optimal scaling. In C. S. Carter, D. W. Dwivedi, I. P. Felligi, D. A. S. Fraser, J. R. McGregor, & D. A. Sprott (Eds.), *Statistics '71 Canada, Proceedings of the first Canadian conference in applied statistics* (pp. 306–316). Montreal: Sir George Williams University Press.

Nishisato, S. (1972). *Analysis of variance of categorical data through selective scaling*. Abstract Guide, 20th International Congress of Psychology, Tokyo, p. 279.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.

Ramsay, J. O. (1982). Some statistical approaches to multidimensional scaling. *Journal of the Royal Statistical Society, Series A, 145*, 285–312.

Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: Wiley.

Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.

Schmoyer, R. L. (1984). Everyday application of the cell means model. *The American Statistician, 38*, 49–52.

Searle, S. R. (1971). *Linear models*. New York: Wiley.

Seber, G. A. F. (1977). *Linear regression analysis*. New York: Wiley.

Seber, G. A. F. (1984). *Multivariate observations*. New York: Wiley.

Takane, Y. (1980). Analysis of categorizing behavior by a quantification method. *Behaviormetrika, 8*, 75–86.

Takane, Y. (1987). Analysis of contingency tables by ideal point discriminant analysis. *Psychometrika, 52*, 493–513.

Takane, Y. (1990). *Constrained principal component analysis and its applications*. Paper submitted for publication.

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 eigenvector technique for multivariate direct gradient analysis. *Ecology, 67*, 1167–1179.

ter Braak, C. J. F. (1988). Partial canonical correspondence analysis. In H. H. Bock (Ed.), *Classification and related methods of data analysis* (pp. 551–558). Amsterdam: North-Holland.

Timm, N. H. (1975). *Multivariate analysis with applications in education and psychology*. Belmont, CA: Wadsworth.

van den Wollenberg, A. L. (1977). Redundancy analysis: An alternative for canonical correlation analysis. *Psychometrika, 42*, 207–219.

van der Burg, E., de Leeuw, J. & Verdegaal, R. (1988). Homogeneity analysis with k sets of variables: An alternating least squares method with optimal scaling features. *Psychometrika, 53*, 177–197.

van der Heijden, P. G. M., de Falguerolles, A., & de Leeuw, J. (1989). A combined approach to contingency table analysis using correspondence analysis and log-linear analysis. *Applied Statistics, 38*, 249–292.

van der Heijden, P. G. M., & de Leeuw, J. (1985). Correspondence analysis used complementary to loglinear analysis. *Psychometrika, 50*, 429–447.

van der Heijden, P. G. M., & Meijerink, F. (1989). Generalized correspondence analysis of multi-way contingency tables and multi-way (super-) indicator matrices. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 185–202). Amsterdam: Elsevier.

van der Heijden, P. G. M., & Worsley, K. J. (1988). Comment on "Correspondence analysis used complementary to loglinear analysis". *Psychometrika, 53*, 287–291.

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

Yanai, H. (1988). Partial correspondence analysis and its properties. In C. Hayashi, M. Jambu, E. Diday, & N. Ohsumi (Eds.), *Recent developments in clustering and data analysis* (pp. 259–266). Boston: Academic Press.

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

Yanai, H., & Takane, Y. (1990). Canonical correlation analysis with linear constraints. Paper submitted for publication.

Yanai, H., & Takeuchi, K. (1983). *Projection matrices, generalized inverse and singular value decomposition*. Tokyo: University of Tokyo Press. (in Japanese)

Manuscript received 3/12/90
Final version received 12/6/90