AAECC 12, 391–419 (2001)

AAECC
Applicable Algebra in Engineering, Communication and Computing

Constrained Principal Component Analysis: A Comprehensive Theory

Yoshio Takane[1], **Michael A. Hunter**[2]

[1] Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montréal, Québec H3A 1B1 Canada (e-mail: takane@takane2.psych.mcgill.ca)
[2] University of Victoria, Department of Psychology, P.O. Box 3050 Victoria, British Columbia, V8W 3P5 (e-mail: mhunter@uvic.ca)

Received: June 23, 2000; revised version: July 9, 2001

Abstract. Constrained principal component analysis (CPCA) incorporates external information into principal component analysis (PCA) of a data matrix. CPCA first decomposes the data matrix according to the external information (external analysis), and then applies PCA to decomposed matrices (internal analysis). The external analysis amounts to projections of the data matrix onto the spaces spanned by matrices of external information, while the internal analysis involves the generalized singular value decomposition (GSVD). Since its original proposal, CPCA has evolved both conceptually and methodologically; it is now founded on firmer mathematical ground, allows a greater variety of decompositions, and includes a wider range of interesting special cases. In this paper we present a comprehensive theory and various extensions of CPCA, which were not fully envisioned in the original paper. The new developments we discuss include least squares (LS) estimation under possibly singular metric matrices, two useful theorems concerning GSVD, decompositions of data matrices into finer components, and fitting higher-order structures. We also discuss four special cases of CPCA; 1) CCA (canonical correspondence analysis) and CALC (canonical analysis with linear constraints), 2) GMANOVA (generalized MANOVA), 3) Lagrange's theorem, and 4) CANO (canonical correlation analysis) and related methods. We conclude with brief remarks on advantages and disadvantages of CPCA relative to other competitors.

The work reported in this paper has been supported by grant A6394 from the Natural Sciences and Engineering Research Council of Canada and by grant 410-89-1498 from the Social Sciences and Humanities Research Council of Canada to the first author.

Keywords: Projection, GSVD (generalized singular value decomposition), CCA, CALC, GMANOVA, Lagrange's theorem, CANO, CA (correspondence analysis).

1 Introduction

It is common practice in statistical data analysis to partition the total variability in a data set into systematic and error portions. Additionally, when the data are multivariate, dimension reduction becomes an important aspect of data analysis. Constrained principal component analysis (CPCA) combines these two aspects of data analysis into a unified procedure in which a given data matrix is first partitioned into systematic and error variation, and then each of these sources of variation is separately subjected to dimension reduction. By the latter we can extract the most important dimensions in the systematic variation as well as investigate the structure of the error variation, and display them graphically.

In short, CPCA incorporates external information into principal component analysis (PCA). The external information can be incorporated on both rows (e.g., subjects) and columns (e.g., variables) of a data matrix. CPCA first decomposes the data matrix according to the external information (external analysis), and then applies PCA to decomposed matrices (internal analysis). Technically, the former amounts to projections of the data matrix onto the spaces spanned by matrices of external information, and the latter involves the generalized singular value decomposition (GSVD). Since its original proposal (Takane and Shibayama, 1991), CPCA has evolved both conceptually and methodologically; it is now founded on firmer mathematical ground, allows a greater variety of decompositions, and includes a wider range of interesting special cases. In this paper we present a comprehensive theory and various extensions of CPCA, which were not fully envisioned in the original paper. The new developments we discuss include least squares (LS) estimation under non-negative definite (*nnd*) metric matrices which may be singular, two useful theorems concerning GSVD, decompositions of data matrices into finer components, and fitting higher-order structures.

The next section (Section 2) presents basic data requirements for CPCA. Section 3 lays down the theoretical ground work of CPCA, namely projections and GSVD. Section 4 describes two extensions of CPCA, decompositions of a data matrix into finer components and fitting of hierarchical structures. Section 5 discusses several interesting special cases, including 1) canonical correspondence analysis (CCA; ter Braak, 1986) and canonical analysis with linear constraints (CALC; Böckenholt and Böckenholt, 1990), 2) GMANOVA (Potthoff and Roy, 1964), 3) Lagrange's theorem on ranks of residual matrices and CPCA within the data spaces (Guttman, 1944), and 4) canonical correlation analysis (CANO) and related methods, such as CANOLC (CANO with linear constraints; Yanai and Takane, 1992) and CA (correspondence analysis; Greenacre, 1984; Nishisato, 1980). The paper concludes with a brief discussion on the relative merits and demerits of CPCA compared to other techniques (e.g., ACOVS; Jöreskog, 1970).

2 Data Requirements

PCA is often used for structural analysis of multivariate data. The data are, however, often accompanied by auxiliary information about rows and columns of a data matrix. CPCA incorporates such information in representing structures in the data. CPCA thus presupposes availability of meaningful auxiliary information. PCA usually obtains the best fixed-rank approximation to the data in the ordinary LS sense. CPCA, on the other hand, allows specifying metric matrices that modulate the effects of rows and columns of a data matrix. This in effect amounts to the weighted LS estimation. There are thus three important ingredients in CPCA; the main data, external information and metric matrices. In this section we discuss them in turn.

2.1 The Main Data

Let us denote an N by n data matrix by $\mathbf{Z}$. Rows of $\mathbf{Z}$ often represent subjects, while columns represent variables. The data in CPCA can, in principle, be any multivariate data. To avoid limiting applicability of CPCA, no distributional assumptions will be made. The data could be either numerical or categorical, assuming that the latter type of variables is coded into dummy variables. Mixing the two types of variables is also permissible. Two-way contingency tables, although somewhat unconventional as a type of multivariate data, form another important class of data covered by CPCA.

The data may be preprocessed or not preprocessed. Preprocessing here refers to such operations as centering, normalizing, both of them (standardizing), or any other prescribed data transformations. There is no cut-and-dry guideline for preprocessing. However, centering implies that we are not interested in mean tendencies. Normalization implies that we are not interested in differences in dispersion. Results of PCA and CPCA are typically affected by what preprocessing is applied, so the decision on the type of preprocessing must be made deliberately in the light of investigators' empirical interests.

When the data consist of both numerical and categorical variables, the problem of compatibility of scales across the two kinds of variables may arise. Although the variables are most often uniformly standardized in such cases, Kiers (1991) recommends orthonormalizing the dummy variables corresponding to each categorical variable after centering.

2.2 External Information

There are two kinds of matrices of external information, one on the row and the other on the column side of the data matrix. We denote the former by an N

by p matrix **G** and call it the row constraint matrix, and the latter by an n by q matrix **H** and call it the column constraint matrix. When there is no special row and/or column information to be incorporated, we may set $\mathbf{G} = \mathbf{I}_N$ and/or $\mathbf{H} = \mathbf{I}_n$.

When the rows of a data matrix represent subjects, we may use subjects' demographic information, such as IQ, age, level of education, etc, in **G**, and explore how they are related to the variables in the main data. If we set $\mathbf{G} = \mathbf{1}_N$ (N-component vector of ones), we see the mean tendency across the subjects. Alternatively, we may take a matrix of dummy variables indicating subjects' group membership, and analyze the differences among the groups. The groups may represent fixed classification variables such as gender, or manipulative variables such as treatment groups.

For **H**, we think of something similar to **G**, but for variables instead of subjects. When the variables represent stimuli, we may take a feature matrix or a matrix of descriptor variables of the stimuli as **H**. When the columns correspond to different within-subject experimental conditions, **H** could be a matrix of contrasts, or when the variables represent repeated observations, **H** could be a matrix of trend coefficients (coefficients of orthogonal polynomials). In one of the examples discussed in Takane and Shibayama (1991), the data were pair comparison preference judgments, and a design matrix for pair comparison was used for **H**.

Incorporating a specific **G** and **H** implies restricting the data analysis spaces to Sp(**G**) and Sp(**H**). This in turn implies specifying their null spaces. We may exploit this fact constructively, and analyze the portion of the main data that cannot be accounted for by certain variables. For example, if **G** contained subject's ages, then incorporating **G** into the analysis of **Z** and analyzing the null space would amount to analyzing that portion of **Z** that was independent of age. As another example, the columnwise centering of data discussed in the previous section is equivalent to eliminating the effect due to $\mathbf{G} = \mathbf{1}_N$, and analyzing the rest.

There are several potential advantages of incorporating external information (Takane et al., 1995). By incorporating external information, we may obtain more interpretable solutions, because what is analyzed is already structured by the external information. We may also obtain more stable solutions by reducing the number of parameters to be estimated. We may investigate the empirical validity of hypotheses incorporated as external constraints by comparing the goodness of fit of unconstrained and constrained solutions. We may predict missing values by way of external constraints which serve as predictor variables. In some cases we can eliminate incidental parameters (Parameters that increase in number as more observations are collected, are called incidental parameters.) by reparameterizing them as linear combinations of a small number of external constraints.

2.3 Metric Matrices

There are two kinds of metric matrices also, one on the row side, **K**, and the other on the column side, **L**. Metric matrices are assumed non-negative definite (*nnd*). Metric matrices are closely related to the criteria employed for fitting models to data. If coordinates that prescribe a data matrix are mutually orthogonal and have comparable scales, we may simply set $\mathbf{K} = \mathbf{I}$ and $\mathbf{L} = \mathbf{I}$, and use the simple unweighted LS criterion. However, when variables in a data matrix are measured on incomparable scales, such as height and weight, a special non-identity metric matrix is required, leading to a weighted LS criterion. It is common, when scales are incomparable, to transform the data to standard scores before analysis, but this is equivalent to using the inverse of the diagonal matrix of sample variances as **L**. A special metric is also necessary when rows of a data matrix are correlated. The rows of a data matrix can usually be assumed statistically independent (and hence uncorrelated) when they represent a random sample of subjects from a target population. They tend to be correlated, however, when they represent different time points in single-subject multivariate time series data. In such cases, a matrix of serial correlations has to be estimated, and its inverse be used as **K** (Escoufier, 1987). When differences in importance and/or in reliability among the rows are suspected, a special diagonal matrix is used for **K** that has the effect of differentially weighting rows of a data matrix. In correspondence analysis, rows and columns of a contingency table are scaled by the square root of row and column totals of the table. This, too, can be thought of as a special case of differential weighting reflecting differential reliability among the rows and columns.

When, on the other hand, columns of a data matrix are correlated, no special metric matrix is usually used, since PCA is applied to disentangle the correlational structure among the columns. However, when the columns of the residual matrix are correlated and/or have markedly different variances after a model is fitted to the data, the variance-covariance matrix among the residuals may be estimated, and its inverse be used as metric **L**. This has the effect of improving the quality (i.e., obtaining smaller expected mean square errors) of parameter estimates by orthonormalizing the residuals in evaluating the overall goodness of fit of the model to the data. Meredith and Millsap (1985) suggests to use reliability coefficients (e.g., test-retest reliability) or inverses of variances of anti-images (Guttman, 1953) as a non-identity **L**.

Although as typically used, PCA (and CPCA using identity metric matrices) are not scale invariant, Rao (1964, Section 9) has shown that specifying certain non-identity **L** matrices have the effect of attaining scale invariance. In maximum likelihood common factor analysis, scale invariance is achieved by scaling a covariance matrix (with communalities in the diagonal) by $\mathbf{D}^{-1}$, where $\mathbf{D}^2$ is the diagonal matrix of uniquenesses which are to be estimated simultaneously with other parameters of the model. This, however, is essentially the same as setting $\mathbf{L} = \mathbf{D}^{-1}$ in CPCA. CPCA, of course, assumes that $\mathbf{D}^2$

is known in advance, but a number of methods have been proposed to estimate $\mathbf{D}^2$ noniteratively (e.g., Ihara and Kano, 1986).

3 Basic Theory

We present CPCA in its general form, with metric matrices other than identity matrices. The provision of metric matrices considerably widens the scope of CPCA. In particular, it makes correspondence analysis of various kinds (Greenacre, 1984; Nishisato, 1980; Takane et al., 1991) a special case of CPCA. As has been noted, a variety of metric matrices can be specified, and by judicious choices of metric matrices a number of interesting analyses become possible. It is also possible to allow metric matrices to adapt to the data iteratively, and construct a robust estimation procedure through iteratively reweighted LS.

3.1 External Analysis

Let $\mathbf{Z}$, $\mathbf{G}$ and $\mathbf{H}$ be the data matrix and matrices of external constraints, as defined earlier. We postulate the following model for $\mathbf{Z}$:

$$\mathbf{Z} = \mathbf{GMH}' + \mathbf{BH}' + \mathbf{GC} + \mathbf{E}, \tag{1}$$

where $\mathbf{M}$ (p by q), $\mathbf{B}$ (N by q), and $\mathbf{C}$ (p by n) are matrices of unknown parameters, and $\mathbf{E}$ (N by n) a matrix of residuals. The first term in model (1) pertains to what can be explained by both $\mathbf{G}$ and $\mathbf{H}$, the second term to what can be explained by $\mathbf{H}$ but not by $\mathbf{G}$, the third term to what can be explained by $\mathbf{G}$ but not by $\mathbf{H}$, and the last term to what can be explained by neither $\mathbf{G}$ nor $\mathbf{H}$. Although model (1) is the basic model, some of the terms in the model may be combined and/or omitted as interest dictates. Also, there may be only row constraints or column constraints, in which case some of the terms in the model will be null.

Let $\mathbf{K}$ (N by N) and $\mathbf{L}$ (n by n) be metric matrices. We assume that they are *nnd*, and that

$$\text{rank}(\mathbf{KG}) = \text{rank}(\mathbf{G}), \tag{2}$$

and

$$\text{rank}(\mathbf{LH}) = \text{rank}(\mathbf{H}). \tag{3}$$

These conditions are necessary for $\mathbf{P}_{G/K}$ and $\mathbf{P}_{H/L}$, to be defined below, to be projectors.

Model (1) is under-identified. To identify the model, it is convenient to impose the following orthogonality constraints:

$$\mathbf{G}'\mathbf{KB} = \mathbf{0}, \tag{4}$$

and

$$\mathbf{H'LC'} = \mathbf{0}. \tag{5}$$

Model parameters are estimated so as to minimize the sum of squares of the elements of **E** in the metrics of **K** and **L**, subject to the identification constraints, (4) and (5). That is, we obtain min $SS(\mathbf{E})_{K,L}$ with respect to **M**, **B**, and **C**, where

$$f \equiv SS(\mathbf{E})_{K,L} \equiv \mathrm{tr}(\mathbf{E'KEL}) = SS(\mathbf{R}'_K\mathbf{E}\mathbf{R}_L)_{I,I} \equiv SS(\mathbf{R}'_K\mathbf{E}\mathbf{R}_L). \tag{6}$$

Here, "≡" means "defined as", and $\mathbf{R}_K$ and $\mathbf{R}_L$ are square root factors of **K** and **L**, respectively, i.e., $\mathbf{K} = \mathbf{R}_K\mathbf{R}'_K$ and $\mathbf{L} = \mathbf{R}_L\mathbf{R}'_L$. This leads to the following LS estimates of **M**, **B**, **C**, and **E**: By differentiating f in (6) with respect to **M** and setting the result equal to zero, we obtain

$$-\frac{1}{2}\frac{\partial f}{\partial \mathbf{M}} = \mathbf{G'K}(\mathbf{Z} - \mathbf{G}\hat{\mathbf{M}}\mathbf{H}' - \hat{\mathbf{B}}\mathbf{H}' - \mathbf{G}\hat{\mathbf{C}})\mathbf{LH} \equiv \mathbf{0}. \tag{7}$$

This leads to, taking into account the orthogonality constraints, (4) and (5),

$$\hat{\mathbf{M}} = (\mathbf{G'KG})^{-}\mathbf{G'KZLH}(\mathbf{H'LH})^{-}, \tag{8}$$

where superscript "−" indicates a g-inverse of a matrix. This estimate of **M** is not unique, unless **G'KG** and **H'LH** are nonsingular. Similarly,

$$-\frac{1}{2}\frac{\partial f}{\partial \mathbf{B}} = \mathbf{K}(\mathbf{Z} - \mathbf{G}\hat{\mathbf{M}}\mathbf{H}' - \hat{\mathbf{B}}\mathbf{H}' - \mathbf{G}\hat{\mathbf{C}})\mathbf{LH} \equiv \mathbf{0}, \tag{9}$$

which leads to

$$\begin{aligned}\hat{\mathbf{B}} &= \mathbf{K}^{-}\mathbf{KZLH}(\mathbf{H'LH})^{-} - \mathbf{K}^{-}\mathbf{KG}\hat{\mathbf{M}}\\ &= \mathbf{K}^{-}\mathbf{KZLH}(\mathbf{H'LH})^{-} - \mathbf{K}^{-}\mathbf{KG}(\mathbf{G'KG})^{-}\mathbf{G'KZLH}(\mathbf{H'LH})^{-}\\ &= \mathbf{K}^{-}\mathbf{KQ}_{G/K}\mathbf{ZLH}(\mathbf{H'LH})^{-},\end{aligned} \tag{10}$$

where $\mathbf{Q}_{G/K} = \mathbf{I} - \mathbf{P}_{G/K}$ and $\mathbf{P}_{G/K} = \mathbf{G}(\mathbf{G'KG})^{-}\mathbf{G'K}$. This estimate of **B** is not unique, unless **K** and **H'LH** are nonsingular. Similarly,

$$\hat{\mathbf{C}} = (\mathbf{G'KG})^{-}\mathbf{G'KZQ}'_{H/L}\mathbf{LL}^{-}, \tag{11}$$

where $\mathbf{Q}_{H/L} = \mathbf{I} - \mathbf{P}_{H/L}$ and $\mathbf{P}_{H/L} = \mathbf{H}(\mathbf{H'LH})^{-}\mathbf{H'L}$. This estimate of **C** is likewise non-unique, unless **L** and **G'KG** are nonsingular. Finally, the estimate of **E** is obtained by

$$\hat{\mathbf{E}} = \mathbf{Z} - \mathbf{P}_{G/K}\mathbf{ZP}'_{H/L} - \mathbf{K}^{-}\mathbf{KQ}_{G/K}\mathbf{ZP}'_{H/L} - \mathbf{P}_{G/K}\mathbf{ZQ}'_{H/L}\mathbf{LL}^{-}. \tag{12}$$

This estimate of **E** is again not unique, unless **K** and **L** are nonsingular. Under (2) and (3), $\mathbf{P}_{G/K}$, $\mathbf{P}_{H/L}$, $\mathbf{Q}_{G/K}$, and $\mathbf{Q}_{H/L}$ are projectors such that $\mathbf{P}^2_{G/K} = \mathbf{P}_{G/K}$,

$\mathbf{Q}^2_{G/K} = \mathbf{Q}_{G/K}$, $\mathbf{P}_{G/K}\mathbf{Q}_{G/K} = \mathbf{Q}_{G/K}\mathbf{P}_{G/K} = \mathbf{0}$, $\mathbf{P}'_{G/K}\mathbf{K}\mathbf{P}_{G/K} = \mathbf{P}'_{G/K}\mathbf{K} = \mathbf{K}\mathbf{P}_{G/K}$, and $\mathbf{Q}'_{G/K}\mathbf{K}\mathbf{Q}_{G/K} = \mathbf{Q}'_{G/K}\mathbf{K} = \mathbf{K}\mathbf{Q}_{G/K}$. $\mathbf{P}_{G/K}$ is the projector onto Sp(G) along Ker(G′K). Note that $\mathbf{P}_{G/K}\mathbf{G} = \mathbf{G}$ and $\mathbf{G}'\mathbf{K}\mathbf{P}_{G/K} = \mathbf{G}'\mathbf{K}$. $\mathbf{Q}_{G/K}$ is the projector onto Ker(G′K) along Sp(G). That is, $\mathbf{G}'\mathbf{K}\mathbf{Q}_{G/K} = \mathbf{0}$ and $\mathbf{Q}_{G/K}\mathbf{G} = \mathbf{0}$. Similar properties hold for $\mathbf{P}_{H/L}$ and $\mathbf{Q}_{H/L}$. These projectors reduce to the usual I-orthogonal projectors when $\mathbf{K} = \mathbf{I}$ and $\mathbf{L} = \mathbf{I}$. Note also that $\tilde{\mathbf{Q}}_{G/K} \equiv \mathbf{K}^{-}\mathbf{K}\mathbf{Q}_{G/K}$ is also a projector, where $\mathbf{K}\mathbf{Q}_{G/K} = \mathbf{K}\tilde{\mathbf{Q}}_{G/K}$. A similar relation also holds for $\tilde{\mathbf{Q}}_{H/L} \equiv \mathbf{L}^{-}\mathbf{L}\mathbf{Q}_{H/L}$.

The effective numbers of parameters are pq in **M**, $(N-p)q$ in **B**, $p(n-q)$ in **C** and $(N-p)(n-q)$ in **E**, assuming that **Z**, **G**, and **H** all have full column ranks, and **K** and **L** are nonsingular. These numbers add up to Nn. The effective numbers of parameters in **B**, **C**, and **E** are less than the actual numbers of parameters in these matrices, because of the identification restrictions, (4) and (5).

Putting the LS estimates of **M**, **B**, **C**, and **E** given above in model (1) yields the following decomposition of the data matrix, **Z**:

$$\mathbf{Z} = \mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L} + \mathbf{K}^{-}\mathbf{K}\mathbf{Q}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L} + \mathbf{P}_{G/K}\mathbf{Z}\mathbf{Q}'_{H/L}\mathbf{L}\mathbf{L}^{-} + \left(\mathbf{Z} - \mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L} - \mathbf{K}^{-}\mathbf{K}\mathbf{Q}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L} - \mathbf{P}_{G/K}\mathbf{Z}\mathbf{Q}'_{H/L}\mathbf{L}\mathbf{L}^{-}\right). \quad (13)$$

This decomposition is not unique, unless **K** and **L** are nonsingular. To make it unique, we may use the Moore-Penrose inverses, $\mathbf{K}^{+}$ and $\mathbf{L}^{+}$, for $\mathbf{K}^{-}$ and $\mathbf{L}^{-}$. The four terms in (13) are mutually orthogonal in the metrics of **K** and **L**, so that

$$\mathrm{SS}(\mathbf{Z})_{K,L} = \mathrm{SS}(\mathbf{G}\hat{\mathbf{M}}\mathbf{H}')_{K,L} + \mathrm{SS}(\hat{\mathbf{B}}\mathbf{H}')_{K,L} + \mathrm{SS}(\mathbf{G}\hat{\mathbf{C}})_{K,L} + \mathrm{SS}(\hat{\mathbf{E}})_{K,L}. \quad (14)$$

That is, sum of squares of **Z** (in the metrics of **K** and **L**) is uniquely decomposed into the sum of sums of squares of the four terms in (13).

Let

$$\mathbf{Z}^* = \mathbf{R}'_K\mathbf{Z}\mathbf{R}_L, \quad (15)$$

$$\mathbf{G}^* = \mathbf{R}'_K\mathbf{G}, \quad (16)$$

and

$$\mathbf{H}^* = \mathbf{R}'_L\mathbf{H}, \quad (17)$$

where $\mathbf{K} = \mathbf{R}_K\mathbf{R}'_K$, and $\mathbf{L} = \mathbf{R}_L\mathbf{R}'_L$ are, as before, square root decompositions of **K** and **L**. We then have, corresponding to decomposition (13),

$$\mathbf{Z}^* = \mathbf{P}_{G^*}\mathbf{Z}^*\mathbf{P}_{H^*} + \mathbf{Q}_{G^*}\mathbf{Z}^*\mathbf{P}_{H^*} + \mathbf{P}_{G^*}\mathbf{Z}^*\mathbf{Q}_{H^*} + \mathbf{Q}_{G^*}\mathbf{Z}^*\mathbf{Q}_{H^*}, \quad (18)$$

where $\mathbf{P}_{G^*} = \mathbf{G}^*(\mathbf{G}^{*\prime}\mathbf{G}^*)^-\mathbf{G}^{*\prime}$, $\mathbf{Q}_{G^*} = \mathbf{I} - \mathbf{P}_{G^*}$, $\mathbf{P}_{H^*} = \mathbf{H}^*(\mathbf{H}^{*\prime}\mathbf{H}^*)^-\mathbf{H}^{*\prime}$, and $\mathbf{Q}_{H^*} = \mathbf{I} - \mathbf{P}_{H^*}$ are orthogonal projectors. This decomposition is unique, while (13) is not. Note that $\mathbf{R}'_K\mathbf{K}^-\mathbf{K} = \mathbf{R}'_K$ and $\mathbf{R}'_L\mathbf{L}^-\mathbf{L} = \mathbf{R}'_L$. Again, four terms in (18) are mutually orthogonal, so that we obtain, corresponding to (14),

$$\begin{aligned}\mathrm{SS}(\mathbf{Z}^*)_{I,I} = \mathrm{SS}(\mathbf{Z}^*) &= \mathrm{SS}\left(\mathbf{P}_{G^*}\mathbf{Z}^*\mathbf{P}_{H^*}\right) + \mathrm{SS}\left(\mathbf{Q}_{G^*}\mathbf{Z}^*\mathbf{P}_{H^*}\right)\\ &+ \mathrm{SS}\left(\mathbf{P}_{G^*}\mathbf{Z}^*\mathbf{Q}_{H^*}\right) + \mathrm{SS}\left(\mathbf{Q}_{G^*}\mathbf{Z}^*\mathbf{Q}_{H^*}\right). \quad (19)\end{aligned}$$

Equations (18) and (19) indicate how we reduce the non-identity metrics, $\mathbf{K}$ and $\mathbf{L}$, to identity metrics in external analysis.

When $\mathbf{K}$ and $\mathbf{L}$ are both nonsingular (and consequently, *pd*), $\mathbf{K}^-\mathbf{K} = \mathbf{I}$ and $\mathbf{L}^-\mathbf{L} = \mathbf{I}$, so that decomposition (13) reduces to

$$\mathbf{Z} = \mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L} + \mathbf{Q}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L} + \mathbf{P}_{G/K}\mathbf{Z}\mathbf{Q}'_{H/L} + \mathbf{Q}_{G/K}\mathbf{Z}\mathbf{Q}'_{H/L}, \quad (20)$$

and (14) to

$$\begin{aligned}\mathrm{SS}(\mathbf{Z})_{K,L} &= \mathrm{SS}\left(\mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L}\right)_{K,L} + \mathrm{SS}\left(\mathbf{Q}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L}\right)_{K,L}\\ &+ \mathrm{SS}\left(\mathbf{P}_{G/K}\mathbf{Z}\mathbf{Q}'_{H/L}\right)_{K,L} + \mathrm{SS}\left(\mathbf{Q}_{G/K}\mathbf{Z}\mathbf{Q}'_{H/L}\right)_{K,L}. \quad (21)\end{aligned}$$

Decomposition (20) is unique.

3.2 Internal Analysis

In the internal analysis, the decomposed matrices in (13) or (20) are subjected to PCA either separately or some of the terms combined. Decisions as to which term or terms are subjected to PCA, and which terms are to be combined, are dictated by researchers' own empirical interests. For example, PCA of the first term in (13) reveals the most prevailing tendency in the data that can be explained by both $\mathbf{G}$ and $\mathbf{H}$, while that of the fourth term is meaningful as a residual analysis (Gabriel, 1978; Rao, 1980; Yanai, 1970).

PCA with non-identity metric matrices requires the generalized singular value decomposition (GSVD) with metrics $\mathbf{K}$ and $\mathbf{L}$, as defined below:

Definition (GSVD) *Let* $\mathbf{K}$ *and* $\mathbf{L}$ *be metric matrices. Let* $\mathbf{A}$ *be an* N *by* n *matrix of rank* r. *Then,*

$$\mathbf{R}'_K\mathbf{A}\mathbf{R}_L = \mathbf{R}'_K\mathbf{U}\mathbf{D}\mathbf{V}'\mathbf{R}_L \quad (22)$$

is called GSVD of $\mathbf{A}$ under metrics $\mathbf{K}$ and $\mathbf{L}$, and written as $\mathrm{GSVD}(\mathbf{A})_{K,L}$, where $\mathbf{R}_K$ and $\mathbf{R}_L$ are, as before, square root factors of $\mathbf{K}$ and $\mathbf{L}$, $\mathbf{U}$ (N by r)

is such that $\mathbf{U}'\mathbf{K}\mathbf{U} = \mathbf{I}$, $\mathbf{V}$ (n by r) is such that $\mathbf{V}'\mathbf{L}\mathbf{V} = \mathbf{I}$, and $\mathbf{D}$ (r by r) is diagonal and pd. When $\mathbf{K}$ and $\mathbf{L}$ are nonsingular, (22) reduces to

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}', \tag{23}$$

where $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{D}$ have the same properties as above. We write the usual SVD of $\mathbf{A}$ (i.e., GSVD$(\mathbf{A})_{I,I}$) simply as SVD($\mathbf{A}$).

GSVD$(\mathbf{A})_{K,L}$ can be obtained as follows. Let the usual SVD of $\mathbf{R}'_K\mathbf{A}\mathbf{R}_L$ be denoted as

$$\mathbf{R}'_K\mathbf{A}\mathbf{R}_L = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime}. \tag{24}$$

Then, $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{D}$ in GSVD$(\mathbf{A})_{K,L}$ are obtained by

$$\mathbf{U} = (\mathbf{R}'_K)^-\mathbf{U}^*, \tag{25}$$

$$\mathbf{V} = (\mathbf{R}'_L)^-\mathbf{V}^*, \tag{26}$$

and

$$\mathbf{D} = \mathbf{D}^*. \tag{27}$$

It can easily be verified that these $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{D}$ satisfy the required properties of GSVD. However, $\mathbf{U}$ or $\mathbf{V}$ given above is not unique, unless $\mathbf{K}$ and $\mathbf{L}$ are nonsingular. When $\mathbf{K}$ and $\mathbf{L}$ are singular, we may still obtain unique $\mathbf{U}$ and $\mathbf{V}$ by using the Moore-Penrose inverses of $\mathbf{R}'_K$ and $\mathbf{R}'_L$ in (25) and (26), respectively.

GSVD plays an important role in CPCA. The following two theorems are extremely useful in facilitating computations of SVD and GSVD in CPCA.

Theorem 1 *Let* $\mathbf{T}$ (N *by* t; $N \geq t$) *and* $\mathbf{W}$ (n *by* w; $n \geq w$) *be columnwise orthogonal matrices, i.e.,* $\mathbf{T}'\mathbf{T} = \mathbf{I}$ *and* $\mathbf{W}'\mathbf{W} = \mathbf{I}$. *Let the SVD of* $\mathbf{A}$ (t *by* w) *be denoted by* $\mathbf{A} = \mathbf{U}_A\mathbf{D}_A\mathbf{V}'_A$, *and that of* $\mathbf{T}\mathbf{A}\mathbf{W}'$ *by* $\mathbf{T}\mathbf{A}\mathbf{W}' = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime}$. *Then,* $\mathbf{U}^* = \mathbf{T}\mathbf{U}_A$ ($\mathbf{U}_A = \mathbf{T}'\mathbf{U}^*$), $\mathbf{V}^* = \mathbf{W}\mathbf{V}_A$ ($\mathbf{V}_A = \mathbf{W}'\mathbf{V}^*$), *and* $\mathbf{D}_A = \mathbf{D}^*$.

Proof of Theorem 1. Pre- and postmultiplying both sides of $\mathbf{A} = \mathbf{U}_A\mathbf{D}_A\mathbf{V}'_A$ by $\mathbf{T}$ and $\mathbf{W}'$, we obtain $\mathbf{T}\mathbf{A}\mathbf{W}' = \mathbf{T}\mathbf{U}_A\mathbf{D}_A\mathbf{V}'_A\mathbf{W}'$. By setting $\mathbf{U}^* = \mathbf{T}\mathbf{U}_A$, $\mathbf{V}^* = \mathbf{W}\mathbf{V}_A$ and $\mathbf{D}^* = \mathbf{D}_A$, we obtain $\mathbf{T}\mathbf{A}\mathbf{W}' = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime}$. It remains to be seen that the above $\mathbf{U}^*$, $\mathbf{V}^*$ and $\mathbf{D}^*$ satisfy the required properties of SVD (i.e., $\mathbf{U}^{*\prime}\mathbf{U} = \mathbf{I}$, $\mathbf{V}^{*\prime}\mathbf{V} = \mathbf{I}$, and $\mathbf{D}^*$ is diagonal and positive definite (pd)). Since $\mathbf{T}$ is columnwise orthogonal, and $\mathbf{U}_A$ is a matrix of left singular vectors, $\mathbf{U}^{*\prime}\mathbf{U}^* = \mathbf{U}'_A\mathbf{T}'\mathbf{T}\mathbf{U}_A = \mathbf{I}$. Similarly, $\mathbf{V}^{*\prime}\mathbf{V}^* = \mathbf{V}'_A\mathbf{W}'\mathbf{W}\mathbf{V}_A = \mathbf{I}$. Since $\mathbf{D}_A$ is diagonal and pd, so is $\mathbf{D}^*$.

Conversely, by pre- and postmultiplying both sides of $\mathbf{T}\mathbf{A}\mathbf{W}' = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime}$ by $\mathbf{T}'$ and $\mathbf{W}$, we obtain $\mathbf{T}'\mathbf{T}\mathbf{A}\mathbf{W}'\mathbf{W} = \mathbf{A} = \mathbf{T}'\mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime}\mathbf{W}'$. By setting $\mathbf{U}_A = \mathbf{T}'\mathbf{U}^*$, $\mathbf{V}_A = \mathbf{W}'\mathbf{V}^*$, and $\mathbf{D}_A = \mathbf{D}^*$, we obtain $\mathbf{A} = \mathbf{U}_A\mathbf{D}_A\mathbf{V}'_A$. It must be shown that $\mathbf{U}'_A\mathbf{U}_A = \mathbf{I}$, $\mathbf{V}'_A\mathbf{V}_A = \mathbf{I}$, and $\mathbf{D}_A$ is diagonal and pd. That $\mathbf{D}_A$ is diagonal and pd is trivial (note that $\mathbf{D}^*$ is pd). That $\mathbf{U}'_A\mathbf{U}_A = \mathbf{I}$, $\mathbf{V}'_A\mathbf{V}_A = \mathbf{I}$ can easily be shown by noting that $\mathbf{T}\mathbf{T}'\mathbf{U}^* = \mathbf{P}_T\mathbf{U}^* = \mathbf{U}^*$ and $\mathbf{W}\mathbf{W}'\mathbf{V}^* = \mathbf{P}_W\mathbf{V}^* = \mathbf{V}^*$, where

$\mathbf{P}_T$ and $\mathbf{P}_W$ are orthogonal projectors onto Sp(**T**) and Sp(**W**), respectively, and $Sp(\mathbf{U}^*) \subset Sp(\mathbf{T})$ and $Sp(\mathbf{V}^*) \subset Sp(\mathbf{W})$.

Suppose we would like to obtain $\text{GSVD}(\mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L})_{K,L}$. This can be obtained from SVD of $\mathbf{R}'_K\mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L}\mathbf{R}_L = \mathbf{P}_{G^*}\mathbf{Z}^*\mathbf{P}_{H^*}$. Note that this is equal to the first term in decomposition (18). $\text{SVD}(\mathbf{P}_{G^*}\mathbf{Z}^*\mathbf{P}_{H^*})$, in turn, is obtained as follows: Let $\mathbf{G}^* = \mathbf{F}_{G^*}\mathbf{R}'_{G^*}$ and $\mathbf{H}^* = \mathbf{F}_{H^*}\mathbf{R}'_{H^*}$ be portions of the QR decompositions (e.g., Golub & Van Loan, 1989) of $\mathbf{G}^*$ and $\mathbf{H}^*$ pertaining to $Sp(\mathbf{G}^*)$ and $Sp(\mathbf{H}^*)$, respectively, where $\mathbf{G}^*$ and $\mathbf{H}^*$ are defined in (16) and (17). $\mathbf{F}_{G^*}$ and $\mathbf{F}_{H^*}$ are columnwise orthogonal, and $\mathbf{R}_{G^*}$ and $\mathbf{R}_{H^*}$ are upper trapezoidal. (When $\mathbf{G}^*$ and $\mathbf{H}^*$ have full column rank, $\mathbf{R}_{G^*}$ and $\mathbf{R}_{H^*}$ are upper triangular.) Then, $\mathbf{P}_{G^*} = \mathbf{F}_{G^*}\mathbf{F}'_{G^*}$ and $\mathbf{P}_{H^*} = \mathbf{F}_{H^*}\mathbf{F}'_{H^*}$. Define $\mathbf{J} \equiv \mathbf{F}'_{G^*}\mathbf{Z}\mathbf{F}_{H^*}$, and let $\mathbf{J} = \mathbf{U}_J\mathbf{D}_J\mathbf{V}'_J$ be SVD(**J**). Then, by Theorem 1, $\mathbf{U}^*$, $\mathbf{V}^*$, and $\mathbf{D}^*$ in the SVD of $\mathbf{P}_{G^*}\mathbf{Z}\mathbf{P}_{H^*}$ are obtained by $\mathbf{U}^* = \mathbf{F}_{G^*}\mathbf{U}_J$, $\mathbf{V}^* = \mathbf{F}_{H^*}\mathbf{V}_J$, and $\mathbf{D}^* = \mathbf{D}_J$. Once SVD of $\mathbf{P}_{G^*}\mathbf{Z}\mathbf{P}_{H^*}$ is obtained, **U**, **V** and **D** in $\text{GSVD}(\mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}'_{H/K})_{K,L}$ can be obtained by $\mathbf{U} = (\mathbf{R}'_K)^-\mathbf{U}^* = (\mathbf{R}'_K)^-\mathbf{F}_{G^*}\mathbf{U}_J$, $\mathbf{V} = (\mathbf{R}'_L)^-\mathbf{V}^* = (\mathbf{R}'_L)^-\mathbf{F}_{H^*}\mathbf{V}_J$, and $\mathbf{D} = \mathbf{D}^* = \mathbf{D}_J$. As before, the Moore-Penrose inverses may be used for $(\mathbf{R}'_K)^-$ and $(\mathbf{R}'_L)^-$ in these formulae to obtain unique **U** and **V**. Note that **J** is usually a much smaller matrix than either $\mathbf{P}_{G^*}\mathbf{Z}^*\mathbf{P}_{H^*}$ or $\mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L}$, and its SVD can be calculated much more quickly.

Theorem 2 *Let* **T** *and* **W** *be two matrices such that* $\mathbf{TAW}'$ *can be formed. Let* $GSVD(\mathbf{TAW}')_{K,L}$ *be denoted as* $\mathbf{UDV}'$ *and* $GSVD(\mathbf{A})_{T'KT,W'LW}$ *as* $\mathbf{U}_A\mathbf{D}_A\mathbf{V}'_A$. *Then,* $\mathbf{U} = \mathbf{K}^-\mathbf{KTU}_A$, $\mathbf{V} = \mathbf{L}^-\mathbf{LWV}_A$ *and* $\mathbf{D} = \mathbf{D}_A$, *and* $\mathbf{U}_A = (\mathbf{T}'\mathbf{KT})^-\mathbf{T}'\mathbf{KU}$, $\mathbf{V}_A = (\mathbf{W}'\mathbf{LW})^-\mathbf{W}'\mathbf{LV}$ *and* $\mathbf{D}_A = \mathbf{D}$.

Proof of Theorem 2. We have $\mathbf{R}'_K\mathbf{TAW}'\mathbf{R}_L = \mathbf{R}'_K\mathbf{UDV}'\mathbf{R}_L = \mathbf{R}'_K\mathbf{TU}_A\mathbf{D}_A\mathbf{V}'_A\mathbf{WR}_L$, so that $\mathbf{R}'_K\mathbf{U} = \mathbf{R}'_K\mathbf{TU}_A$, $\mathbf{R}'_L\mathbf{V} = \mathbf{R}'_L\mathbf{WV}_A$, and and $\mathbf{D} = \mathbf{D}_A$. Solving the first two equations for **U** and **V**, we obtain $\mathbf{U} = \mathbf{K}^-\mathbf{KTU}_A$, and $\mathbf{V} = \mathbf{L}^-\mathbf{LWV}_A$. Similarly, $\mathbf{U}_A = (\mathbf{T}'\mathbf{KT})^-\mathbf{T}'\mathbf{KU}$, and $\mathbf{V}_A = (\mathbf{W}'\mathbf{LW})^-\mathbf{W}'\mathbf{LV}$. It must be shown that $\mathbf{U}'\mathbf{KU} = \mathbf{I}$ and $\mathbf{V}'\mathbf{LV} = \mathbf{I}$, and that $\mathbf{U}'_A\mathbf{T}'\mathbf{KTU}_A = \mathbf{I}$ and $\mathbf{V}'_A\mathbf{W}'\mathbf{LWV}_A = \mathbf{I}$. $\mathbf{U}'\mathbf{KU} = \mathbf{U}'_A\mathbf{T}'\mathbf{KK}^-\mathbf{KK}^-\mathbf{KTU}_A = \mathbf{U}'_A\mathbf{T}'\mathbf{KTU}_A = \mathbf{I}$. $\mathbf{V}'\mathbf{LV} = \mathbf{I}$ can be similarly shown. Conversely, $\mathbf{U}'_A\mathbf{T}'\mathbf{KTU}_A = \mathbf{U}'\mathbf{P}'_{T/K}\mathbf{KP}_{T/K}\mathbf{U} = \mathbf{U}'\mathbf{P}'_{T/K}\mathbf{KK}^-\mathbf{KK}^-\mathbf{KP}_{T/K}\mathbf{U} = \mathbf{U}'\mathbf{KU} = \mathbf{I}$. (Note that $\mathbf{K}^-\mathbf{KP}_{T/K}\mathbf{U} = \mathbf{U}$.) $\mathbf{V}'_A\mathbf{W}'\mathbf{LWV}_A = \mathbf{I}$ can be similarly shown.

In some cases, $\text{GSVD}(\hat{\mathbf{M}})_{G^{*\prime}G^*,H^{*\prime}H^*}$, where $\hat{\mathbf{M}}$ is given in (5), and is part of the first term in decomposition (13), may be of direct interest. For example, Takane and Shibayama (1991) discussed vector preference models, in which $\mathbf{K} = \mathbf{I}$, $\mathbf{L} = \mathbf{I}$, $\mathbf{G} = \mathbf{I}$, and **H** is a design matrix for pair comparisons. In those models **M** contains scale values of stimuli, and consequently $\text{GSVD}(\hat{\mathbf{M}})_{I,H'H}$ is of direct interest, but not $\text{SVD}(\hat{\mathbf{M}}\mathbf{H}')$. $\text{GSVD}(\hat{\mathbf{M}})_{G^{*\prime}G^*,H^{*\prime}H^*}$ may be calculated directly, or from related SVD's or GSVD discussed above. In particular, if $\hat{\mathbf{M}} = \mathbf{U}_M\mathbf{D}_M\mathbf{V}'_M$ represents $\text{GSVD}(\hat{\mathbf{M}})_{G^{*\prime}G^*,H^{*\prime}H^*}$, then because of Theorem 2, $\mathbf{U}_M = (\mathbf{G}'\mathbf{KG})^-\mathbf{G}'\mathbf{KU}$, $\mathbf{V}_M = (\mathbf{H}'\mathbf{LH})^-\mathbf{H}'\mathbf{LV}$ and $\mathbf{D}_M = \mathbf{D}$, or $\mathbf{U} = \mathbf{K}^-\mathbf{KGU}_M$, $\mathbf{V} = \mathbf{L}^-\mathbf{LHV}_M$ and $\mathbf{D} = \mathbf{D}_M$. (Note that $\mathbf{U} = \mathbf{GU}_M$

and $\mathbf{V} = \mathbf{H}\mathbf{V}_M$, when $\mathbf{K}$ and $\mathbf{L}$ are nonsingular.) $\mathbf{U}_M$ and $\mathbf{V}_M$ are the regression weights applied to $\mathbf{G}$ and $\mathbf{H}$, respectively, to obtain $\mathbf{U}$ and $\mathbf{V}$, respectively. This is analogous to canonical correlation analysis between, say, $\mathbf{G}$ and $\mathbf{H}$, in which canonical weights are obtained by GSVD$((\mathbf{G}'\mathbf{G})^{-}\mathbf{G}'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-})_{G'G,H'H}$, whereas canonical variates are directly obtained by SVD$(\mathbf{P}_G\mathbf{P}_H)$.

The relationships among GSVD$(\mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}_{H/L})_{K,L}$, SVD$(\mathbf{P}_{G^*}\mathbf{Z}^*\mathbf{P}_{H^*})$, SVD $(\mathbf{J})$, and GSVD$(\hat{\mathbf{M}})_{G^{*\prime}G^*,H^{*\prime}H^*}$ are summarized in Table 1. In general, when we have a product of several matrices, say, $\mathbf{ABC}$, SVD$(\mathbf{ABC})$ can be related to a number of different GSVD's via Theorem 2: GSVD $(\mathbf{I})_{C'B'A'ABC,I}$, GSVD$(\mathbf{A})_{I,BCC'B'}$, GSVD$(\mathbf{AB})_{I,CC'}$, GSVD$(\mathbf{B})_{A'A,CC'}$, GSVD $(\mathbf{BC})_{A'A,I}$, and GSVD$(\mathbf{I})_{I,ABCC'B'A'}$. This extends to products of four or more matrices.

4 Some Extensions

Within the basic framework of CPCA, various extensions are possible. We discuss two major ones here; decompositions of a data matrix into finer components and incorporation of higher-order structures.

4.1 Decompositions into Finer Components

Decomposition (13) or (20) is a very basic one. When more than one set of external constraints are available on either side of a data matrix, it is possible to decompose the data matrix into finer components. This is akin to factorial ANOVA in which a data matrix may be decomposed into the main effect of Factor A, that of Factor B, the interaction effect between them, and the residual effect.

Table 1. Relationships among various SVD's and GSVD's

	(1) GSVD $(\mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L})_{K,L}$ $\mathbf{UDV}'$	(2) SVD$(\mathbf{P}_{G^*}\mathbf{Z}^*\mathbf{P}_{H^*})$ $\mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime}$	(3) SVD$(\mathbf{J})$ $\mathbf{U}_J\mathbf{D}_J\mathbf{V}'_J$	(4) GSVD$(\mathbf{M})_{G^{*\prime}G^*,H^{*\prime}H^*}$ $\mathbf{U}_M\mathbf{D}_M\mathbf{V}'_M$
(1)		$\mathbf{U}^* = \mathbf{R}'_K\mathbf{U}$ $\mathbf{V}^* = \mathbf{R}'_L\mathbf{V}$	$\mathbf{U}_J = \mathbf{F}'_{G^*}\mathbf{R}'_K\mathbf{U}$ $\mathbf{V}_J = \mathbf{F}'_{H^*}\mathbf{R}'_L\mathbf{V}$	$\mathbf{U}_M = (\mathbf{G}'\mathbf{G})^{-}\mathbf{G}'\mathbf{U}$ $\mathbf{V}_M = (\mathbf{H}'\mathbf{H})^{-}\mathbf{H}'\mathbf{V}$
(2)	$\mathbf{U} = \mathbf{R}_K'^{-}\mathbf{U}^*$ $\mathbf{V} = \mathbf{R}_L'^{-}\mathbf{V}^*$		$\mathbf{U}_J = \mathbf{F}'_{G^*}\mathbf{U}^*$ $\mathbf{V}_J = \mathbf{F}'_{H^*}\mathbf{V}^*$	$\mathbf{U}_M = (\mathbf{G}^{*\prime}\mathbf{G}^*)^{-}\mathbf{G}^{*\prime}\mathbf{U}^*$ $\mathbf{V}_M = (\mathbf{H}^{*\prime}\mathbf{H}^*)^{-}\mathbf{H}^{*\prime}\mathbf{V}^*$
(3)	$\mathbf{U} = \mathbf{R}_K'^{-}\mathbf{F}_{G^*}\mathbf{U}_J$ $\mathbf{V} = \mathbf{R}_L'^{-}\mathbf{F}_{H^*}\mathbf{V}_J$	$\mathbf{U}^* = \mathbf{F}_{G^*}\mathbf{U}_J$ $\mathbf{V}^* = \mathbf{F}_{H^*}\mathbf{V}_J$		$\mathbf{U}_M = \mathbf{R}_{G^*}'^{-}\mathbf{U}_J$ $\mathbf{V}_M = \mathbf{R}_{H^*}'^{-}\mathbf{V}_J$
(4)	$\mathbf{U} = \mathbf{G}\mathbf{U}_M$ $\mathbf{V} = \mathbf{H}\mathbf{V}_M$	$\mathbf{U}^* = \mathbf{G}^*\mathbf{U}_M$ $\mathbf{V}^* = \mathbf{H}^*\mathbf{V}_M$	$\mathbf{U}_J = \mathbf{R}'_{G^*}\mathbf{U}_M$ $\mathbf{V}_J = \mathbf{R}'_{H^*}\mathbf{V}_M$	

Notation: $\mathbf{K} = \mathbf{R}_K\mathbf{R}'_K$, $\mathbf{L} = \mathbf{R}_L\mathbf{R}'_L$, $\mathbf{G}^* = \mathbf{R}'_K\mathbf{G} = \mathbf{F}_{G^*}\mathbf{R}'_{G^*}$, $\mathbf{H}^* = \mathbf{R}'_L\mathbf{H} = \mathbf{F}_{H^*}\mathbf{R}'_{H^*}$, $\mathbf{Z}^* = \mathbf{R}'_K\mathbf{Z}\mathbf{R}_L$, $\mathbf{J} = \mathbf{F}'_{G^*}\mathbf{Z}^*\mathbf{F}_{H^*}$.
The "−" in the above table may be replaced by "+".

The problem of fitting multiple sets of constraints can be viewed as decompositions of a projector defined on the joint space of all constraints into the sum of projectors defined on subspaces corresponding to the different subsets of constraints. Suppose **G** consists of two constraint sets, **X** and **Y**; that is, $\mathbf{G} = [\mathbf{X}|\mathbf{Y}]$. Depending on the relationship between **X** and **Y** (Rao and Yanai, 1979), a variety of decompositions are possible.

When **X** and **Y** are mutually orthogonal (in the metric of **K**), we have

$$\mathbf{P}_{G/K} = \mathbf{P}_{X/K} + \mathbf{P}_{Y/K}. \tag{28}$$

This simply partitions the joint effect of **X** and **Y** into the sum of the separate effects of **X** and **Y**. Since **X** and **Y** are orthogonal, the decomposition is simple and unique. When **X** and **Y** are not completely orthogonal, but are orthogonal except in their intersection space, $\mathbf{P}_{X/K}$ and $\mathbf{P}_{Y/K}$ are still commutative (i.e., $\mathbf{P}_{X/K}\mathbf{P}_{Y/K} = \mathbf{P}_{Y/K}\mathbf{P}_{X/K}$), and

$$\mathbf{P}_{G/K} = \mathbf{P}_{X/K} + \mathbf{P}_{Y/K} - \mathbf{P}_{X/K}\mathbf{P}_{Y/K}. \tag{29}$$

This decomposition, when $\mathbf{K} = \mathbf{I}$, plays an important role in ANOVA for factorial designs. When **X** and **Y** are not mutually orthogonal in any sense, two decompositions are possible:

$$\begin{aligned}\mathbf{P}_{G/K} &= \mathbf{P}_{X/K} + \mathbf{P}_{Q_{Y/K}X/K} \\ &= \mathbf{P}_{Y/K} + \mathbf{P}_{Q_{X/K}Y/K},\end{aligned} \tag{30}$$

where $\mathbf{P}_{Q_{Y/K}X/K}$ and $\mathbf{P}_{Q_{X/K}Y/K}$ are projectors onto spaces of $\mathbf{Q}_{Y/K}\mathbf{X}$ (the portion of **X** that is unaccounted for by **Y**) and $\mathbf{Q}_{X/K}\mathbf{Y}$ (the portion of **Y** that is unaccounted for by **X**), respectively. The above decompositions are useful when one of **X** and **Y** is fitted first and the other is fitted to the residuals.

When Sp(**X**) and Sp(**Y**) are disjoint, but not orthogonal, we may use

$$\mathbf{P}_{G/K} = \mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{Q}_{Y/K}\mathbf{X})^{-}\mathbf{X}'\mathbf{K}\mathbf{Q}_{Y/K} + \mathbf{Y}(\mathbf{Y}'\mathbf{K}\mathbf{Q}_{X/K}\mathbf{Y})^{-}\mathbf{Y}'\mathbf{K}\mathbf{Q}_{X/K}. \tag{31}$$

Note that $\mathbf{K}\mathbf{Q}_{Y/K}$ and $\mathbf{K}\mathbf{Q}_{X/K}$ are both symmetric. This decomposition is useful when **X** and **Y** are fitted simultaneously. The first term on the right hand side of (31) is the projector onto Sp(**X**) along $\text{Sp}(\mathbf{Q}_{G/K}) \oplus \text{Sp}(\mathbf{Y})$ where $\oplus$ indicates the direct sum of two disjoint spaces, and the second term the projector onto Sp(**Y**) along $\text{Sp}(\mathbf{Q}_{G/K}) \oplus \text{Sp}(\mathbf{X})$. Note that unlike all the previous decompositions discussed in this section, the two terms in this decomposition are not mutually orthogonal. Takane and Yanai (1999), however, discuss a special metric $\mathbf{K}^*$ under which the two terms in (31) are mutually orthogonal, and are such that $\mathbf{P}_{G/K} = \mathbf{P}_{G/K^*}$, $\mathbf{P}_{X/KQ_{Y/K}} = \mathbf{P}_{X/K^*}$ and $\mathbf{P}_{Y/KQ_{X/K}} = \mathbf{P}_{Y/K^*}$. An example of such a metric is $\mathbf{K}^* = \mathbf{K}\mathbf{Q}_{X/K} + \mathbf{K}\mathbf{Q}_{Y/K}$.

When additional information is given as constraints on the weight matrix, $\mathbf{U}_G$, on $\mathbf{G}$, the following decomposition is useful. Suppose the constraints can be expressed as $\mathbf{U}_G = \mathbf{A}\mathbf{U}_A$ for a given matrix, $\mathbf{A}$. Then,

$$\mathbf{P}_{G/K} = \mathbf{P}_{GA/K} + \mathbf{P}_{G(G'KG)^{-}B/K}, \tag{32}$$

where $\mathbf{A}'\mathbf{B} = \mathbf{0}$, Sp($\mathbf{A}$) $\oplus$ Sp($\mathbf{B}$) = Sp($\mathbf{G}'$), and $\mathbf{B} = \mathbf{G}'\mathbf{K}\mathbf{W}$ for some $\mathbf{W}$ (Yanai and Takane, 1992). The first term in this decomposition is the projector onto Sp($\mathbf{GA}$), which is a subspace of Sp($\mathbf{G}$), and the second term onto the subspace of Sp($\mathbf{G}$) orthogonal to Sp($\mathbf{GA}$). Since $\mathbf{B}'(\mathbf{G}'\mathbf{K}\mathbf{G})^{-}\mathbf{G}'\mathbf{K}\mathbf{G}\mathbf{U}_A = \mathbf{0}$ for $\mathbf{B}$ such that $\mathbf{B} = \mathbf{G}'\mathbf{K}\mathbf{W}$, the constraint $\mathbf{U}_G = \mathbf{A}\mathbf{U}_A$ can also be expressed as $\mathbf{B}'\mathbf{U}_G = \mathbf{0}$. This decomposition is an example of higher-order structures to be discussed in the next section. It is often used when we have a specific hypothesis about $\mathbf{M}$ in model (1), for example, and we would like to obtain an estimate of $\mathbf{M}$ under the hypothesis. A detailed example of this will also be given in Section 5.2.

It is obvious that similar decompositions apply to $\mathbf{H}$ as well. It is also relatively straightforward to extend the decompositions to more than two sets of constraints on each side of a data matrix. The above decompositions can further be generalized to oblique projectors (Takane and Yanai, 1999) useful for the instrumental variable (IV) estimation often used in econometrics (e.g., Johnston, 1984).

Decompositions into finer components may generally be written as (Nishisato and Lawrence, 1989):

$$\mathbf{Z} = \Big(\sum_i \mathbf{P}_{G(i)/\tilde{K}}\Big)\mathbf{Z}\Big(\sum_j \mathbf{P}'_{H(j)/\tilde{L}}\Big), \tag{33}$$

where $\sum_i \mathbf{P}_{G_i/\tilde{K}} = \mathbf{I}$ and $\sum_j \mathbf{P}_{H_j/\tilde{L}} = \mathbf{I}$, and where $\mathbf{P}_{G_i/\tilde{K}}$ and $\mathbf{P}_{H_j/\tilde{L}}$ are projectors onto Sp($\mathbf{G}_i$) and Sp($\mathbf{H}_j$), respectively, in the metrics of $\tilde{\mathbf{K}}$ and $\tilde{\mathbf{L}}$. The $\tilde{\mathbf{K}}$ and $\tilde{\mathbf{L}}$ are orthogonalizing metrics, which are simply $\mathbf{K}$ and $\mathbf{L}$, except in (31) where $\tilde{\mathbf{K}} = \mathbf{K}^*$ and $\tilde{\mathbf{L}} = \mathbf{L}^*$. Because of the orthogonality of the terms in decomposition (33), the sum of squares (SS) in $\mathbf{Z}$ is uniquely partitioned into the sum of part SS's, each pertaining to each term in (33). The partitioning of SS in this manner is similar to the partitioning of deviance in maximum likelihood estimation.

4.2 Higher-Order Structures

External information other than $\mathbf{G}$ or $\mathbf{H}$ can also be incorporated into the model. This information often takes the form of a hypothesis about the parameters in the model, in which case we may be interested in obtaining an estimate of the parameters under that hypothesis. For example, a model similar to (1) may be assumed for $\mathbf{M}$ as well. Suppose $\mathbf{A}$(=$\mathbf{H}$) is a design matrix for pair comparisons, and suppose stimuli in the pair comparisons are constructed by systematically manipulating some basic factors. Let **S** denote the design matrix for the stimuli. It may be assumed that $\mathbf{M} = \mathbf{WS}' + \mathbf{E}^*$, where **W** is a matrix of weights applied to $\mathbf{S}'$. The entire model may then be written as

$$\begin{aligned}\mathbf{Z} &= \mathbf{G}(\mathbf{WS}' + \mathbf{E}^*)\mathbf{A}' + \mathbf{E} \\ &= \mathbf{GWS}'\mathbf{A}' + \mathbf{GE}^*\mathbf{A}' + \mathbf{E}. \end{aligned} \tag{34}$$

This model partitions **Z** into three parts: what can be explained by **G** and **AS**, what can be explained by **G** and **A** but not by **AS**, and the residuals. In Takane and Shibayama (1991), this model was treated as a special provision in CPCA. This, however, is an instance of partition (32).

Alternatively, **M** may be subjected to PCA first, and then some hypothesized structure may be imposed on its row representation, $\mathbf{U}_M$, or on $\mathbf{U} \equiv \mathbf{GU}_M$. In the former case, the model could be:

$$\begin{aligned}\mathbf{Z} &= \mathbf{G}(\mathbf{U}_M^*\mathbf{D}_M^*\mathbf{V}_M^{*\prime} + \mathbf{E}^*)\mathbf{H}' + \mathbf{E} \\ &= \mathbf{G}((\mathbf{TW} + \tilde{\mathbf{E}})\mathbf{D}_M^*\mathbf{V}_M^{*\prime} + \mathbf{E}^*)\mathbf{H}' + \mathbf{E}, \end{aligned} \tag{35}$$

where $\mathbf{U}_M^*\mathbf{D}_M^*\mathbf{V}_M^{*\prime}$ is the best fixed-rank approximation of **M** obtained by its PCA, $\mathbf{E}^*$ is its residuals, and **T** the design matrix for $\mathbf{U}_M^*$. In this model, $\mathbf{U}_M^*$ is modeled by $\mathbf{U}_M^* = \mathbf{TW} + \tilde{\mathbf{E}}$, but $\mathbf{D}_M^*$ and $\mathbf{V}_M^*$ are left unmodeled.

If, on the other hand, a model is assumed on **U**, the entire model might be:

$$\begin{aligned}\mathbf{Z} &= \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime} + \mathbf{GE}^*\mathbf{H}' + \mathbf{E} \\ &= (\mathbf{TW} + \tilde{\mathbf{E}})\mathbf{D}^*\mathbf{V}^{*\prime} + \mathbf{GE}^*\mathbf{H}' + \mathbf{E}, \end{aligned} \tag{36}$$

where **T** is an additional row information matrix. An LS estimate of **W** in this model, given the estimate of $\mathbf{U}^*$, is obtained by

$$\hat{\mathbf{W}} = (\mathbf{T}'\mathbf{KT})^-\mathbf{T}'\mathbf{KU}^*. \tag{37}$$

Rows of $\hat{\mathbf{W}}$ are linear combinations of rows of $\mathbf{U}^*$, and thus can be represented as vectors in the same space as row vectors of $\mathbf{U}^*$. The above $\hat{\mathbf{W}}$ can also be obtained directly by $\text{GSVD}(\mathbf{P}_{GT/K}\mathbf{ZP}'_{H/L})_{K,L}$. Note, however, that in general $\text{SVD}(\mathbf{PZ}) \neq \mathbf{P} \cdot \text{SVD}(\mathbf{Z})$, where **P** is any projector. That is, the order in which projection and SVD are performed is important. The LS estimate of **W** given above is thus contingent on the fact that SVD is applied to $\mathbf{GMH}'$ first.

Model (1) as well as its extensions discussed in this section can generally be expressed as

$$\mathbf{Z} = \Big(\prod_i \mathbf{G}_{(i)}\Big)\mathbf{R}\Big(\prod_j \mathbf{H}_{(j)}\Big), \tag{38}$$

where $\mathbf{R}$, $\mathbf{G}_{(i)}$ and $\mathbf{H}_{(j)}$ are specially defined matrices (see below for an example). This expression is similar to that of COSAN for structural equation models (McDonald, 1978; see also Faddeev and Feddeeva, 1963). The major difference between COSAN and (38) is that in the former, $\mathbf{Z}$ is a variance-covariance matrix, which is bound to be symmetric, so is $\mathbf{R}$, and $\mathbf{G}_{(i)} = \mathbf{H}_{(i)}$, whereas in (38) no such restrictions apply. In CPCA, $\mathbf{Z}$ is usually rectangular.

We show, as an example, how model (34) can be expressed in the above form. We define

$$\mathbf{R} = \begin{bmatrix} \mathbf{W} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{E} \end{bmatrix},$$

$$\mathbf{G}_{(1)} = [\mathbf{G} \quad \mathbf{I}],$$

$$\mathbf{G}_{(2)} = \begin{bmatrix} \mathbf{I} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix},$$

$$\mathbf{H}_{(1)} = [\mathbf{H} \quad \mathbf{I}],$$

and

$$\mathbf{H}_2 = \begin{bmatrix} \mathbf{S} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

It can easily be verified that these matrices yield model (34). Models (1), (35) and (36) can also be expressed in similar ways by defining $\mathbf{R}$, $\mathbf{G}_{(i)}$ and $\mathbf{H}_{(j)}$ appropriately.

5 Special Cases

CPCA subsumes a number of interesting special cases. Those already discussed by Takane and Shibayama (1991) are vector preference models (Bechtel et al., 1971; Takane, 1980; Heiser and de Leeuw, 1981; De Soete and Carroll, 1983), two-way CANDELINC (Carroll et al. 1980), dual scaling of categorical data (Nishisato, 1980), canonical correlation analysis (CANO), and redundancy analysis (van den Wollenberg, 1977), also known as PCA of instrumental variables (Rao, 1964) and reduced-rank regression (Anderson, 1951). In this paper we focus on other special cases. Specifically, we discuss four groups of methods; canonical correspondence analysis (CCA; ter Braak, 1986) and canonical analysis with linear constraints (CALC; Böckenholt and Böckenholt, 1990), which are both constrained versions of correspondence analysis (CA; Greenacre, 1984), which in turn is a special case of CANO; GMANOVA (Potthoff and Roy, 1964) and its extensions (Khatri, 1966; Rao, 1965; 1985); CPCA with components within row and column spaces of data matrices (Guttman, 1944; Rao, 1964); and relationships among CPCA, CANO and related methods. We close this section with some historical remarks on the development of CPCA.

5.1 CCA and CALC

We show that both CCA and CALC are special cases of CPCA. For illustration, we discuss the case in which there are only row constraints, $\mathbf{G}$, although CALC was originally proposed to accommodate both row and column constraints, and CCA, though not presented as such, can readily be extended to accommodate both.

Let $\mathbf{F}$ denote a two-way contingency table. CA of $\mathbf{F}$ obtains "optimal" row and column representations of $\mathbf{F}$. Technically, it amounts to obtaining $\text{GSVD}(\mathbf{D}_R^-\mathbf{F}\mathbf{D}_C^-)_{D_R,D_C}$, where $\mathbf{D}_R$ and $\mathbf{D}_C$ are diagonal matrices of row and column totals of $\mathbf{F}$, respectively. (All the g-inverses in this section may be replaced by the Moore-Penrose inverses.) Let $\mathbf{UDV}'$ denote the GSVD. The row and column representations of $\mathbf{F}$ are obtained by simple rescaling of $\mathbf{U}$ and $\mathbf{VD}$. In CA, a component corresponding to the largest singular value is eliminated as being trivial. This component can *a priori* be eliminated from the solution by replacing $\mathbf{F}$ by $\mathbf{Q}'_{1_R/D_R}\mathbf{F}\mathbf{Q}_{1_C/D_C} = \mathbf{Q}'_{1_R/D_R}\mathbf{F} = \mathbf{F}\mathbf{Q}_{1_C/D_C}$, where

$$\mathbf{Q}_{1_R/D_R} = \mathbf{I}_R - \mathbf{1}_R\mathbf{1}'_R\mathbf{D}_R/N, \tag{39}$$

and

$$\mathbf{Q}_{1_C/D_C} = \mathbf{I}_C - \mathbf{1}_C\mathbf{1}'_C\mathbf{D}_C/N. \tag{40}$$

Here, $N = \mathbf{1}'_R\mathbf{D}_R\mathbf{1}_R = \mathbf{1}'_C\mathbf{D}_C\mathbf{1}_C = \mathbf{1}'_R\mathbf{F}\mathbf{1}_C$ is the total number of observations, $\mathbf{I}_R$ and $\mathbf{I}_C$ are identity matrices of orders R and C, respectively, and $\mathbf{1}_R$ and $\mathbf{1}_C$ are R-element and C-element vectors of ones, respectively.

Suppose some external information is available on rows of $\mathbf{F}$. Let $\mathbf{X}$ denote the row constraint matrix. CCA by ter Braak (1986) obtains $\mathbf{U}$ under the restriction that $\mathbf{U} = \mathbf{XU}^*$, where $\mathbf{U}^*$ is a matrix of weights. This amounts to $\text{GSVD}((\mathbf{X}'\mathbf{D}_R\mathbf{X})^-\mathbf{X}'\mathbf{F}\mathbf{D}_C^-)_{X'D_RX,D_C}$ from which $\mathbf{U}^*$ is obtained (and then, $\mathbf{U}$ is derived by $\mathbf{U} = \mathbf{XU}^*$), or to $\text{GSVD}(\mathbf{X}(\mathbf{X}'\mathbf{D}_R\mathbf{X})^-\mathbf{X}'\mathbf{F}\mathbf{D}_C^-)_{D_R,D_C}$ from which $\mathbf{U}$ is directly obtained (Takane, Yanai, and Mayekawa, 1991). When $\mathbf{X}'\mathbf{D}_R\mathbf{X}$ is singular, $\mathbf{U}^*$ is not unique, but $\mathbf{U}$ is. CCA of $\mathbf{F}$ with row constraint matrix $\mathbf{X}$ will be denoted as CCA($\mathbf{F}$, $\mathbf{X}$), or simply CCA($\mathbf{X}$). Thus, CCA($\mathbf{F}$, $\mathbf{X}$) = $\text{GSVD}(\mathbf{X}(\mathbf{X}'\mathbf{D}_R\mathbf{X})^-\mathbf{X}'\mathbf{F}\mathbf{D}_C^-)_{D_R,D_C}$.

CALC by Böckenholt and Böckenholt (1990) is similar to CCA, but instead of restricting $\mathbf{U}$ by $\mathbf{U} = \mathbf{XU}^*$, it restricts $\mathbf{U}$ by $\mathbf{R}'\mathbf{U} = \mathbf{0}$, where $\mathbf{R}$ is a constraint matrix. That is, CALC specifies the null space of $\mathbf{U}$. CALC obtains $\text{GSVD}(\mathbf{D}_R^-(\mathbf{I} - \mathbf{R}(\mathbf{R}'\mathbf{D}_R^-\mathbf{R})^-\mathbf{R}'\mathbf{D}_R^-)\mathbf{F}\mathbf{D}_C^-)_{D_R,D_C}$, which will be denoted as CALC($\mathbf{F}$, $\mathbf{R}$) or simply CALC($\mathbf{R}$).

To eliminate the trivial solution in CCA we replace $\mathbf{X}$ by $\mathbf{Q}_{1_R/D_R}\mathbf{X}$. In CALC we simply include $\mathbf{D}_R\mathbf{1}_R$ in $\mathbf{R}$. Once $\mathbf{X}$ or $\mathbf{R}$ is adjusted this way, there is no longer any adjustment needed on $\mathbf{F}$.

Takane et al. (1991) have shown that CCA and CALC can be made equivalent by appropriately choosing an $\mathbf{R}$ for a given $\mathbf{X}$ or vice versa. More specifically, CCA(**X**) = CALC(**R**) if **X** and **R** are mutually orthogonal, and together they span the entire column space of **F**. That is, Sp(**X**) = Ker(**R**′) (or equivalently Sp(**R**) = Ker(**X**′)). For a given **R**, such an **X** can be obtained by a square root decomposition of $\mathbf{I} - \mathbf{R}(\mathbf{R}'\mathbf{R})^{-}\mathbf{R}'$ (i.e., **X** such that $\mathbf{I} - \mathbf{R}(\mathbf{R}'\mathbf{R})^{-}\mathbf{R}' = \mathbf{X}\mathbf{X}'$). Similarly, an **R** can be obtained from a given **X** by $\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' = \mathbf{R}\mathbf{R}'$. Neither **X** nor **R** are uniquely determined given the other. Only Sp(**X**) or Sp(**R**) can be uniquely determined from the other.

It can easily be shown that CCA and CALC are both special cases of CPCA. When $\mathbf{H} = \mathbf{I}$, decomposition (13) reduces to

$$\mathbf{Z} = \mathbf{P}_{G/K}\mathbf{Z} + \mathbf{Q}_{G/K}\mathbf{Z}, \tag{41}$$

where, as before, $\mathbf{P}_{G/K} = \mathbf{G}(\mathbf{G}'\mathbf{K}\mathbf{G})^{-}\mathbf{G}'\mathbf{K}$ and $\mathbf{Q}_{G/K} = \mathbf{I} - \mathbf{P}_{G/K}$. Note that the first term in (41) can be rewritten as

$$\mathbf{P}_{G/K}\mathbf{Z} = \mathbf{G}(\mathbf{G}'\mathbf{K}\mathbf{G})^{-}\mathbf{G}'(\mathbf{K}\mathbf{Z}\mathbf{L})\mathbf{L}^{-}, \tag{42}$$

which is equal to $\mathbf{X}(\mathbf{X}'\mathbf{D}_R\mathbf{X})^{-}\mathbf{X}'\mathbf{F}\mathbf{D}_C^{-}$, if $\mathbf{G} = \mathbf{X}$, $\mathbf{K} = \mathbf{D}_R$, $\mathbf{L} = \mathbf{D}_C$, and $\mathbf{Z} = \mathbf{D}_R^{-}\mathbf{F}\mathbf{D}_C^{-}$. This means that under these conditions, $\mathrm{GSVD}(\mathbf{P}_{G/K}\mathbf{Z})_{K,L} =$ CCA(**F**, **X**).

The residual matrix, $\mathbf{Q}_{G/K}\mathbf{Z}$, can be rewritten as

$$\begin{aligned}\mathbf{Q}_{G/K}\mathbf{Z} &= (\mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{K}\mathbf{G})^{-}\mathbf{G}'\mathbf{K})\mathbf{Z} \\ &= \mathbf{K}^{-}(\mathbf{I} - \mathbf{K}\mathbf{G}(\mathbf{G}'\mathbf{K}\mathbf{K}^{-}\mathbf{K}\mathbf{G})^{-}\mathbf{G}'\mathbf{K}\mathbf{K}^{-})(\mathbf{K}\mathbf{Z}\mathbf{L})\mathbf{L}^{-},\end{aligned} \tag{43}$$

which is equal to $\mathbf{D}_R^{-}(\mathbf{I} - \mathbf{R}(\mathbf{R}'\mathbf{D}_R^{-}\mathbf{R})^{-}\mathbf{R}'\mathbf{D}_R^{-})\mathbf{F}\mathbf{D}_C^{-}$, if $\mathbf{R} = \mathbf{K}\mathbf{G}$, $\mathbf{K} = \mathbf{D}_R$, $\mathbf{L} = \mathbf{D}_C$, and $\mathbf{Z} = \mathbf{D}_R^{-}\mathbf{F}\mathbf{D}_C^{-}$. Thus, $\mathrm{GSVD}(\mathbf{Q}_{G/K}\mathbf{Z})_{K,L}$ = CALC(**F**, **R**) under these conditions.

The above discussion shows that both CCA and CALC are special cases of CPCA, and that CCA(**X**) and CALC($\mathbf{D}_R\mathbf{X}$) analyze complementary parts of data matrix **Z**. CALC($\mathbf{D}_R\mathbf{X}$), in turn, is equivalent to CCA($\mathbf{X}^*$), where $\mathbf{X}^*$ is such that $\mathrm{Sp}(\mathbf{X}^*) = \mathrm{Ker}(\mathbf{X}'\mathbf{D}_R)$. The analysis of residuals from CCA($\mathbf{X}^*$) is equivalent to CALC($\mathbf{D}_R\mathbf{X}^*$), which in turn is equivalent to CCA(**X**), where **X** is such that $\mathrm{Sp}(\mathbf{X}) = \mathrm{Ker}(\mathbf{X}^{*\prime}\mathbf{D}_R)$. Such an **X** can be the **X** in the original CCA. This circular relationship is illustrated in Fig. 1.

5.2 GMANOVA

GMANOVA (growth curve models; Potthoff and Roy, 1964) postulates

$$\mathbf{Z} = \mathbf{G}\mathbf{M}\mathbf{H}' + \mathbf{E}. \tag{44}$$

This is a special case of model (1) in which only the first term is isolated from the rest. Under the assumption that rows of **E** are *iid* multivariate normal, a maximum likelihood estimate of **M** is obtained by

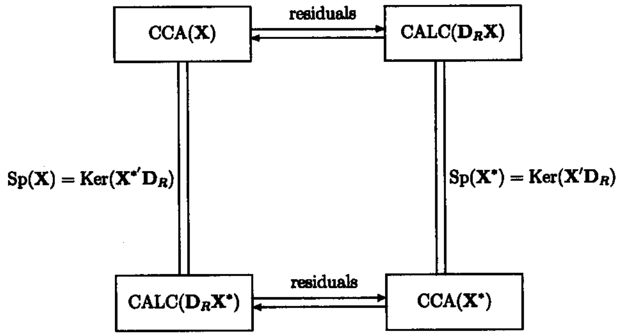

Fig. 1. Complementality and equivalence of CCA and CALC

$$\hat{\mathbf{M}} = (\mathbf{G}'\mathbf{G})^{-}\mathbf{G}'\mathbf{Z}\mathbf{S}^{-1}\mathbf{H}(\mathbf{H}'\mathbf{S}^{-1}\mathbf{H})^{-} \tag{45}$$

(Khatri, 1966; Rao, 1965), where $\mathbf{S} = \mathbf{Z}'(\mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-}\mathbf{G}')\mathbf{Z}$ which is assumed nonsingular. This estimate of $\mathbf{M}$ is equivalent to an LS estimate of $\mathbf{M}$ in (5) with $\mathbf{K} = \mathbf{I}$ and $\mathbf{L} = \mathbf{S}^{-1}$.

In GMANOVA, tests of hypotheses about $\mathbf{M}$ of the following form are typically of interest, rather than PCA of the structural part of model (44):

$$\mathbf{R}'\mathbf{M}\mathbf{C} = \mathbf{0}, \tag{46}$$

where $\mathbf{R}$ and $\mathbf{C}$ are given constraint matrices. We assume that $\mathbf{R} = \mathbf{G}'\mathbf{K}\mathbf{W}_R$ for some $\mathbf{W}_R$, and similarly $\mathbf{C} = \mathbf{H}'\mathbf{L}\mathbf{W}_C$ for some $\mathbf{W}_C$. These conditions are automatically satisfied if $\mathbf{G}$ and $\mathbf{H}$ have full column ranks. An LS estimate of $\mathbf{M}$ under the above hypothesis can be obtained as follows: Let $\mathbf{X}$ and $\mathbf{Y}$ be such that $\mathbf{R}'\mathbf{X} = \mathbf{0}$ and $\mathrm{Sp}[\mathbf{R}|\mathbf{X}] = \mathrm{Sp}(\mathbf{G}')$, and $\mathbf{C}'\mathbf{Y} = \mathbf{0}$ and $\mathrm{Sp}[\mathbf{C}|\mathbf{Y}] = \mathrm{Sp}(\mathbf{H}')$. (These conditions reduce to $\mathrm{Sp}(\mathbf{X}) = \mathrm{Ker}(\mathbf{R}')$ and $\mathrm{Sp}(\mathbf{Y}) = \mathrm{Ker}(\mathbf{C}')$, respectively, when $\mathbf{G}$ and $\mathbf{H}$ have full column ranks.) Then, $\mathbf{M}$ in (46) can be reparameterized as

$$\mathbf{M} = \mathbf{X}\mathbf{M}_{XY}\mathbf{Y}' + \mathbf{M}_Y\mathbf{Y}' + \mathbf{X}\mathbf{M}_X, \tag{47}$$

where $\mathbf{M}_{XY}$, $\mathbf{M}_Y$ and $\mathbf{M}_X$ are matrices of unknown parameters. This representation is not unique. For identification, we assume

$$\mathbf{X}'\mathbf{G}'\mathbf{K}\mathbf{G}\mathbf{M}_Y = \mathbf{0}, \tag{48}$$

(where $\mathbf{K} = \mathbf{I}$ in GMANOVA), and

$$\mathbf{Y}'\mathbf{H}'\mathbf{L}\mathbf{H}\mathbf{M}_X' = \mathbf{0}, \tag{49}$$

(where $\mathbf{L} = \mathbf{S}^{-1}$ in GMANOVA). These constraints are similar to (2) and (3). Putting (47) in model (44), we obtain

$$\mathbf{Z} = \mathbf{G}\mathbf{X}\mathbf{M}_{XY}\mathbf{Y}'\mathbf{H}' + \mathbf{G}\mathbf{M}_Y\mathbf{Y}'\mathbf{H}' + \mathbf{G}\mathbf{X}\mathbf{M}_X\mathbf{H}' + \mathbf{E}. \tag{50}$$

Note that this is an instance of higher-order structures discussed in Section 4.2.

LS estimates of $\mathbf{M}_{XY}$, $\mathbf{M}_Y$ and $\mathbf{M}_X$ subject to (48) and (49) are obtained by

$$\hat{\mathbf{M}}_{XY} = (\mathbf{X}'\mathbf{G}'\mathbf{K}\mathbf{G}\mathbf{X})^{-}\mathbf{X}'\mathbf{G}'\mathbf{K}\mathbf{Z}\mathbf{L}\mathbf{H}\mathbf{Y}(\mathbf{Y}'\mathbf{H}'\mathbf{L}\mathbf{H}\mathbf{Y})^{-}, \quad (51)$$

$$\hat{\mathbf{M}}_Y = \mathbf{P}_{G(G'KG)^{-}R/K}\mathbf{Z}\mathbf{H}\mathbf{L}\mathbf{Y}(\mathbf{Y}'\mathbf{H}'\mathbf{L}\mathbf{H}\mathbf{Y})^{-}, \quad (52)$$

and

$$\hat{\mathbf{M}}_X = (\mathbf{X}'\mathbf{G}'\mathbf{K}\mathbf{G}\mathbf{X})^{-}\mathbf{X}'\mathbf{G}\mathbf{K}\mathbf{Z}\mathbf{P}'_{H(H'LH)^{-}C/L}, \quad (53)$$

where because of (32), $\mathbf{P}_{G(G'KG)^{-}R/K} = \mathbf{P}_{G/K} - \mathbf{P}_{GX/K}$ and $\mathbf{P}_{H(H'LH)^{-}C/L} = \mathbf{P}_{H/L} - \mathbf{P}_{HY/L}$. These are analogous to (5), (10) and (11). Putting (51) through (53) into (50) leads to

$$\mathbf{Z} = \mathbf{P}_{GX/K}\mathbf{Z}\mathbf{P}'_{HY/L} + \mathbf{P}_{G(G'KG)^{-}R/K}\mathbf{Z}\mathbf{P}'_{HY/L} + \mathbf{P}_{GX/K}\mathbf{Z}\mathbf{P}'_{H(H'LH)^{-}C/L} + \hat{\mathbf{E}}, \quad (54)$$

where $\hat{\mathbf{E}}$ is defined as $\mathbf{Z}$ minus the sum of the first three terms in (54).

The above partition suggests that Sp($\mathbf{Z}$) is split into three mutually orthogonal subspaces (in metric $\mathbf{K}$) with associated projectors, $\mathbf{P}_{GX/K}$, $\mathbf{P}_{G(G'KG)^{-}R}$ and $\mathbf{Q}_{G/K}$. The Sp($\mathbf{Z}'$) can be similarly partitioned. By combining the two partitionings we obtain the nine-term partition listed in Table 2. The first three terms in (54) correspond with (a), (b) and (d) in the table. The fourth term in (54), $\hat{\mathbf{E}}$, represents the sum of all the remaining terms ((c), (e), (f), (g), (h) & (i)) in Table 2. It will be interesting to obtain fixed-rank approximations (Internal Analysis) of not only the last term in (54), as was done by Rao (1985; to be described shortly), but also the first three terms in (54). Rao (1985) considered a slightly generalized version of the hypothesis (47), namely

$$\tilde{\mathbf{M}} = \mathbf{X}\mathbf{M}_{XY}\mathbf{Y}' + \mathbf{M}_Y\mathbf{Y}' + \mathbf{X}\mathbf{M}_X + \mathbf{E}^*, \quad (55)$$

where $\mathbf{E}^*$ is assumed to have a prescribed rank, and is such that

$$\mathbf{X}'\mathbf{G}'\mathbf{K}\mathbf{G}\mathbf{E}^* = \mathbf{0}, \quad (56)$$

Table 2. Decomposition in GMANOVA

Decomposition of Sp(Z)	Decomposition of Sp($\mathbf{Z}'$)		
	$\mathbf{P}_{HY/L}$	$\mathbf{P}_{H(H'LH)^{-}C/L}$	$\mathbf{Q}_{H/L}$
$\mathbf{P}_{GX/K}$	(a)	(b)	(c)
$\mathbf{P}_{G(G'KG)^{-}R/K}$	(d)	(e)	(f)
$\mathbf{Q}_{G/K}$	(g)	(h)	(i)

and

$$\mathbf{E}^*\mathbf{H}'\mathbf{L}\mathbf{H}\mathbf{Y} = \mathbf{0}. \tag{57}$$

Under (55), LS estimates of $\mathbf{M}_{XY}$, $\mathbf{M}_Y$, and $\mathbf{M}_X$ given in (51), (52) and (53) are still valid. The estimate of $\mathbf{E}^*$, on the other hand, can be obtained as follows: Let $\tilde{\mathbf{E}}^*$ be such that

$$\mathbf{G}\tilde{\mathbf{E}}^*\mathbf{H}' = \mathbf{P}_{G(G'KG)^-R/K}\mathbf{Z}\mathbf{P}'_{H(H'LH)^-C/L}, \tag{58}$$

which is the LS estimate of $\mathbf{G}\mathbf{E}^*\mathbf{H}'$ under no rank restriction on $\mathbf{E}^*$. This corresponds with term (e) in Table 2. The fixed-rank approximation of $\mathbf{G}\tilde{\mathbf{E}}^*\mathbf{H}'$ is obtained by the $\text{GSVD}(\mathbf{G}\tilde{\mathbf{E}}^*\mathbf{H}')_{K,L}$. Let $\hat{\mathbf{W}}$ represent the fixed-rank approximation of $\mathbf{G}\hat{\mathbf{E}}^*\mathbf{H}'$. Then, a fixed-rank approximation, $\hat{\mathbf{E}}^*$, of $\mathbf{E}^*$ is obtained by

$$\hat{\mathbf{E}}^* = (\mathbf{G}'\mathbf{K}\mathbf{G})^-\mathbf{G}'\mathbf{K}\hat{\mathbf{W}}\mathbf{L}\mathbf{H}(\mathbf{H}'\mathbf{L}\mathbf{H})^-, \tag{59}$$

or directly by the GSVD of

$$\tilde{\mathbf{E}}^* = (\mathbf{G}'\mathbf{K}\mathbf{G})^-\mathbf{R}(\mathbf{R}'(\mathbf{G}'\mathbf{K}\mathbf{G})^-\mathbf{R})^-\mathbf{R}'(\mathbf{G}'\mathbf{K}\mathbf{G})^-\mathbf{G}'\mathbf{K}\mathbf{Z}$$
$$\times \mathbf{L}\mathbf{H}(\mathbf{H}'\mathbf{L}\mathbf{H})^-\mathbf{C}(\mathbf{C}'(\mathbf{H}'\mathbf{L}\mathbf{H})^-\mathbf{C})^-\mathbf{C}'(\mathbf{H}'\mathbf{L}\mathbf{H})^- \tag{60}$$

with metrics $\mathbf{G}'\mathbf{K}\mathbf{G}$ and $\mathbf{H}'\mathbf{L}\mathbf{H}$. Rao's hypothesis, (55), can be expressed in the form of a conventional GMANOVA hypothesis (like (47)) as

$$\mathbf{R}'(\tilde{\mathbf{M}} - \mathbf{E}^*)\mathbf{C} = \mathbf{0}, \tag{61}$$

where $\mathbf{E}^*$ is, as before, assumed to have a prescribed rank.

5.3 Lagrange's Theorem

It is well known (e.g., Yanai, 1990) that

$$(\mathbf{A}')^-_{ZB} = \mathbf{Z}\mathbf{B}(\mathbf{A}'\mathbf{Z}\mathbf{B})^- \tag{62}$$

and

$$\mathbf{B}^-_{A'Z} = (\mathbf{A}'\mathbf{Z}\mathbf{B})^-\mathbf{A}'\mathbf{Z} \tag{63}$$

are reflexive g-inverses of $\mathbf{A}'$ and $\mathbf{B}$, respectively, under

$$\text{rank}(\mathbf{A}'\mathbf{Z}\mathbf{B}) = \text{rank}(\mathbf{A}') \tag{64}$$

and

$$\text{rank}(\mathbf{A}'\mathbf{Z}\mathbf{B}) = \text{rank}(\mathbf{B}), \tag{65}$$

respectively. A reflexive g-inverse $\mathbf{X}^-$ of $\mathbf{X}$ satisfies $\mathbf{XX}^-\mathbf{X} = \mathbf{X}$ and $\mathbf{X}^-\mathbf{XX}^- = \mathbf{X}^-$. Define

$$\mathbf{Q}_{ZB,A} = \mathbf{I} - (\mathbf{A}')^-_{ZB}\mathbf{A}', \quad (66)$$

and

$$\mathbf{Q}_{Z'A,B} = \mathbf{I} - \mathbf{BB}^-_{A'Z}. \quad (67)$$

Then, $\mathbf{Q}_{ZB,A}$ is the projector onto Ker($\mathbf{A}'$) along Sp($\mathbf{ZB}$), and $\mathbf{Q}_{Z'A,B}$ onto Ker($\mathbf{A}'\mathbf{Z}$) along Sp($\mathbf{B}$). Define

$$\mathbf{Z}_1 = \mathbf{Q}_{ZB,A}\mathbf{Z} = \mathbf{ZQ}_{Z'A,B}. \quad (68)$$

Then, under both (64) and (65),

$$\text{rank}(\mathbf{Z}_1) = \text{rank}(\mathbf{Z}) - \text{rank}(\mathbf{A}'\mathbf{ZB}). \quad (69)$$

This is called Lagrange's theorem (Rao, 1973, p. 69). Note that (64) and (65) are sufficient, but not necessary, conditions for (69).

Rao (1964, Section 11) considered extracting components within Sp($\mathbf{Z}$) but orthogonal to a given $\mathbf{G}$. This amounts to SVD of

$$\begin{aligned}\mathbf{ZQ}_{Z'G} &= \mathbf{Z}(\mathbf{I} - \mathbf{Z}'\mathbf{G}(\mathbf{G}'\mathbf{ZZ}'\mathbf{G})^-\mathbf{G}'\mathbf{Z}) \\ &= (\mathbf{I} - \mathbf{ZZ}'\mathbf{G}(\mathbf{G}'\mathbf{ZZ}'\mathbf{G})^-\mathbf{G}')\mathbf{Z} = \mathbf{Q}_{G/ZZ'}\mathbf{Z}. \end{aligned} \quad (70)$$

This reduces to $\mathbf{Z}_1$ in (68) by setting $\mathbf{A} = \mathbf{G}$ and $\mathbf{B} = \mathbf{Z}'\mathbf{G}$. It is obvious that this is also a special case of $\mathbf{ZQ}_H$ with $\mathbf{H} = \mathbf{Z}'\mathbf{G}$, and of $\mathbf{Q}_{G/K}\mathbf{Z}$ with $\mathbf{K} = \mathbf{ZZ}'$. Rao's method is thus a special case of CPCA in two distinct ways. (It can easily be verified that $\mathbf{ZQ}_{Z'G}$ and $\mathbf{G}$ are mutually orthogonal, and that Sp($\mathbf{ZQ}_{Z'G}$) is in Sp($\mathbf{Z}$).)

Guttman (1944, 1952; also, see Schönemann and Steiger, 1976) considered obtaining components which are given linear combinations of $\mathbf{Z}$, as, for example, in the group centroid method of factor analysis, and used Lagrange's theorem to successively obtain residual matrices. Let the weight matrix in the linear combinations be denoted by $\mathbf{W}$. Let $\mathbf{A} = \mathbf{ZW}$ and $\mathbf{B} = \mathbf{W}$ in (68). PCA of the part of data matrix $\mathbf{Z}$ that can be explained by $\mathbf{ZW}$ amounts to SVD of $\mathbf{P}_{ZW}\mathbf{Z} = \mathbf{ZP}_{W/Z'Z}$ and that of residual matrices to SVD of $\mathbf{Q}_{ZW}\mathbf{Z} = \mathbf{ZQ}'_{W/Z'Z}$. Both are special cases of CPCA (PCA of $\mathbf{P}_G\mathbf{Z} = \mathbf{ZP}_{H/L}$ and that of $\mathbf{Q}_G\mathbf{Z} = \mathbf{ZQ}_{H/L}$) with $\mathbf{G} = \mathbf{ZW}$ or with $\mathbf{H} = \mathbf{W}$ and $\mathbf{L} = \mathbf{Z}'\mathbf{Z}$.

A major difference between CPCA and the methods discussed in this section is that in the former, components are often constructed outside Sp($\mathbf{Z}'$) or Sp($\mathbf{Z}$), whereas in the latter they are always formed within the spaces.

5.4 Relationships among CPCA, CANO and Related Methods

A number of methods have been proposed for relating two sets of variables with or without additional constraints. In this section we show relationships among some of them: CPCA, canonical correlation analysis (CANO), CANOLC (CANO with linear constraints; Yanai and Takane, 1990), CCA (ter Braak, 1986), and the usual (unconstrained) correspondence analysis (CA; Greenacre, 1984). A common thread running through these techniques is the generalized singular value decomposition (GSVD) described in Section 3.2.

We first briefly discuss each method in turn, and then establish specific relationships among the methods.

(i) CPCA: As has been seen, there are five matrices involved in CPCA, and it is more explicitly written as CPCA($\mathbf{Z}$, $\mathbf{G}$, $\mathbf{H}$, $\mathbf{K}$, $\mathbf{L}$), where $\mathbf{Z}$ is a data matrix, $\mathbf{G}$ and $\mathbf{H}$ are matrices of external constraints, and $\mathbf{K}$ and $\mathbf{L}$ metric matrices. Row and column representations, $\mathbf{U}$ and $\mathbf{V}$, of $\mathbf{Z}$ are sought under the restrictions that $\mathbf{U} = \mathbf{G}\mathbf{U}^*$ and $\mathbf{V} = \mathbf{H}\mathbf{V}^*$, where $\mathbf{U}^*$ and $\mathbf{V}^*$ are weight matrices. Matrices $\mathbf{U}^*$ and $\mathbf{V}^*$ are obtained by GSVD$((\mathbf{G}'\mathbf{K}\mathbf{G})^-\mathbf{G}'\mathbf{K}\mathbf{Z}\mathbf{L}\mathbf{H}(\mathbf{H}'\mathbf{L}\mathbf{H})^-)_{G'KG,H'LH}$.

(ii) CANOLC: Four matrices are involved in CANOLC, and hence it is written as CANOLC($\mathbf{X}$, $\mathbf{Y}$, $\mathbf{G}$, $\mathbf{H}$). Canonical correlation analysis between $\mathbf{X}$ and $\mathbf{Y}$ is performed under the restrictions that canonical variates, $\mathbf{U}$ and $\mathbf{V}$, are linear functions of $\mathbf{G}$ and $\mathbf{H}$, respectively. That is, $\mathbf{U} = \mathbf{G}\mathbf{U}^*$ and $\mathbf{V} = \mathbf{H}\mathbf{V}^*$, where $\mathbf{U}^*$ and $\mathbf{V}^*$ are weight matrices obtained by GSVD$((\mathbf{G}'\mathbf{X}'\mathbf{X}\mathbf{G})^-\mathbf{G}'\mathbf{X}'\mathbf{Y}\mathbf{H}(\mathbf{H}'\mathbf{Y}'\mathbf{Y}\mathbf{H})^-)_{G'X'XG,H'Y'YH}$. Note that there is a symmetry between a pair of matrices, $\mathbf{X}$ and $\mathbf{Y}$, and the other pair of matrices, $\mathbf{G}$ and $\mathbf{H}$, so that their roles can be exchanged. We then have CANOLC($\mathbf{G}$, $\mathbf{H}$, $\mathbf{X}$, $\mathbf{Y}$).

(iii) CCA: When there are constraints on both rows and columns of a contingency table, five matrices are involved in CCA, and it is more explicitly written as CCA($\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$, $\mathbf{D}_R$, $\mathbf{D}_C$), where $\mathbf{F}$ is a two-way contingency table, $\mathbf{G}$ and $\mathbf{H}$ are matrices of external constraints, and $\mathbf{D}_R$ and $\mathbf{D}_C$ diagonal matrices of row and column totals of $\mathbf{F}$, respectively. Row and column representations, $\mathbf{U}$ and $\mathbf{V}$, of $\mathbf{F}$ are obtained under the restrictions that $\mathbf{U} = \mathbf{G}\mathbf{U}^*$ and $\mathbf{V} = \mathbf{H}\mathbf{V}^*$, where $\mathbf{U}^*$ and $\mathbf{V}^*$ are weight matrices. Matrices $\mathbf{U}^*$ and $\mathbf{V}^*$ are obtained by GSVD$((\mathbf{G}'\mathbf{D}_R\mathbf{G})^-\mathbf{G}'\mathbf{F}\mathbf{H}(\mathbf{H}'\mathbf{D}_C\mathbf{H})^-)_{G'D_RG,H'D_CH}$. CCA discussed in the main text of this paper (Section 5.1) is a simplified version, where $\mathbf{H} = \mathbf{I}$ is assumed.

(iv) CANO: Canonical correlation analysis between $\mathbf{G}$ and $\mathbf{H}$ denoted as CANO ($\mathbf{G}$, $\mathbf{H}$) amounts to GSVD$((\mathbf{G}'\mathbf{G})^-\mathbf{G}'\mathbf{H}(\mathbf{H}'\mathbf{H})^-)_{G'G,H'H}$.

(v) CA: The usual (unconstrained) correspondence analysis of a two-way contingency table, $\mathbf{F}$, is written as CA($\mathbf{F}$, $\mathbf{D}_R$, $\mathbf{D}_C$), where $\mathbf{D}_R$ and $\mathbf{D}_C$ are, as before, diagonal matrices of row and column totals of $\mathbf{F}$, respectively. CA($\mathbf{F}$, $\mathbf{D}_R$, $\mathbf{D}_C$) reduces to GSVD$(\mathbf{D}_R^-\mathbf{F}\mathbf{D}_C^-)_{D_R,D_C}$.

Specific relationships among these methods are depicted in Fig. 2. In the figure, methods placed higher are more general. By specializing some of the matrices involved in more general methods, more specialized methods result:

CPCA ⟶ CANOLC:	Set $\mathbf{Z} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-}$, $\mathbf{K} = \mathbf{X}'\mathbf{X}$, and $\mathbf{L} = \mathbf{Y}'\mathbf{Y}$.
CPCA ⟶ CCA:	Set $\mathbf{Z} = \mathbf{D}_R^{-}\mathbf{F}\mathbf{D}_C^{-}$, $\mathbf{K} = \mathbf{D}_R$, and $\mathbf{L} = \mathbf{D}_C$.
CPCA ⟶ CANO:	Set $\mathbf{Z} = \mathbf{I}$, $\mathbf{K} = \mathbf{I}$, and $\mathbf{L} = \mathbf{I}$.
CPCA ⟶ CA:	Set $\mathbf{Z} = \mathbf{D}_R^{-}\mathbf{F}\mathbf{D}_C^{-}$, $\mathbf{G} = \mathbf{I}$, $\mathbf{H} = \mathbf{I}$, $\mathbf{K} = \mathbf{D}_R$, and $\mathbf{L} = \mathbf{D}_C$.
CANOLC ⟶ CCA:	Set $\mathbf{X}'\mathbf{Y} = \mathbf{F}$, $\mathbf{X}'\mathbf{X} = \mathbf{D}_R$, and $\mathbf{Y}'\mathbf{Y} = \mathbf{D}_C$.
CANOLC ⟶ CANO:	Set $\mathbf{X} = \mathbf{I}$, and $\mathbf{Y} = \mathbf{I}$.
CANOLC ⟶ CA:	Set $\mathbf{G}'\mathbf{H} = \mathbf{F}$, $\mathbf{X} = \mathbf{I}$, $\mathbf{Y} = \mathbf{I}$, $\mathbf{G}'\mathbf{G} = \mathbf{D}_R$, and $\mathbf{H}'\mathbf{H} = \mathbf{D}_C$.
CCA ⟶ CANO:	Set $\mathbf{F} = \mathbf{I}$, $\mathbf{D}_R = \mathbf{I}$, and $\mathbf{D}_C = \mathbf{I}$.
CCA ⟶ CA:	Set $\mathbf{G} = \mathbf{I}$, and $\mathbf{H} = \mathbf{I}$.
CANO ⟶ CA:	Set $\mathbf{G}'\mathbf{H} = \mathbf{F}$, $\mathbf{G}'\mathbf{G} = \mathbf{D}_R$, and $\mathbf{H}'\mathbf{H} = \mathbf{D}_C$.

Note that the relationship between CPCA and CANO implies relationships between CPCA and MANOVA and between CPCA and canonical discriminant analysis, as both MANOVA and canonical discriminant analysis are special cases of CANO.

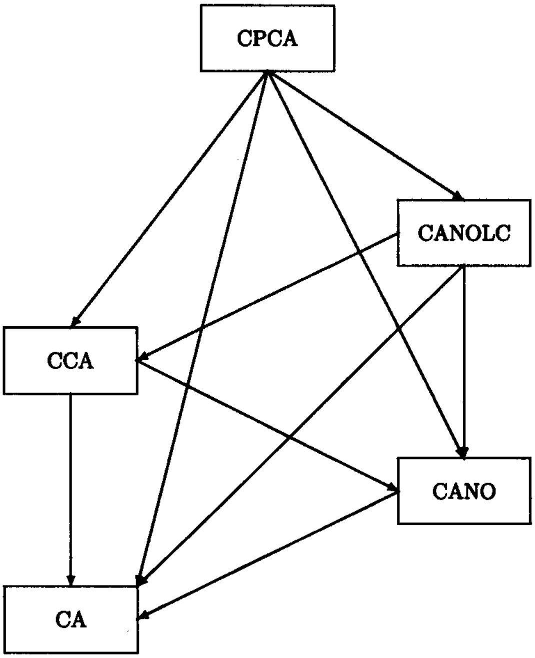

Fig. 2. Relationships among CPCA, CANOLC, CCA, CANO and CA

5.5 Historical Remarks on CPCA

Special cases of partition (20) have been proposed by many authors (Gabriel, 1978; Rao, 1980). These authors proposed models in which in addition to $\mathbf{K} = \mathbf{I}$ and $\mathbf{L} = \mathbf{I}$, either the first and the second terms, or the first and the third terms in (20) are not separated. These models are written as

$$\begin{aligned} \mathbf{Z} &= \mathbf{P}_G\mathbf{Z} + \mathbf{Q}_G\mathbf{Z}\mathbf{P}_H + \mathbf{Q}_G\mathbf{Z}\mathbf{Q}_H \\ &= \mathbf{Z}\mathbf{P}_H + \mathbf{P}_G\mathbf{Z}\mathbf{Q}_H + \mathbf{Q}_G\mathbf{Z}\mathbf{Q}_H, \end{aligned} \quad (71)$$

where $\mathbf{P}_G = \mathbf{G}(\mathbf{G}'\mathbf{G})^-\mathbf{G}'$, $\mathbf{Q}_G = \mathbf{I} - \mathbf{P}_G$, $\mathbf{P}_H = \mathbf{H}(\mathbf{H}'\mathbf{H})^-\mathbf{H}'$ and $\mathbf{Q}_G = \mathbf{I} - \mathbf{P}_G$ are I-orthogonal projectors. Gollob's (1968) FANOVA is a special case of CPCA in which $\mathbf{G} = \mathbf{1}_N$ and $\mathbf{H} = \mathbf{1}_n$. Yanai (1970) proposed PCA with external criteria, where $\mathbf{G}$ represented a matrix of dummy variables indicating subjects' group membership. Okamoto (1972) set $\mathbf{G} = \mathbf{1}_N$ and $\mathbf{H} = \mathbf{1}_n$ as in Gollob, and proposed PCA's of four matrices, $\mathbf{Z}$, $\mathbf{Q}_G\mathbf{Z}$, $\mathbf{Z}\mathbf{Q}_H$ and $\mathbf{Q}_G\mathbf{Z}\mathbf{Q}_H$.

In all the above proposals, PCA's of residual terms are recommended. Several lines of development in PCA of the structural parts have also taken place. Rao (1964) gave a solution to a constrained generalized eigenvalue problem, which is closely related to GSVD. He also proposed PCA of instrumental variables, also known as reduced-rank regression (Anderson, 1951) and redundancy analysis (van den Wollenberg, 1977). This method amounts to SVD($\mathbf{P}_G\mathbf{Z}$) or GSVD($(\mathbf{G}'\mathbf{G})^-\mathbf{G}'\mathbf{Z})_{G'G,I}$. Golub (1973) gave a solution to the problem of maximizing a bilinear form $\mathbf{x}'\mathbf{A}\mathbf{y}/\|\mathbf{x}\| \cdot \|\mathbf{y}\|$ subject to linear restrictions of the form, $\mathbf{C}'\mathbf{x} = \mathbf{0}$ and $\mathbf{R}'\mathbf{y} = \mathbf{0}$. Ter Braak (1986; CCA) and Böckenholt and Böckenholt (1990; CALC) proposed similar methods for analysis of contingency tables (see Section 5.1). Nishisato and his collaborators (Nishisato, 1980; Nishisato and Lawrence, 1989) also proposed similar methods called ANOVA of categorical data. Carroll et al. (1980) two-way CANDELINC applies PCA to only the first term in model (1). GMANOVA also fits only the first term in model (1), and optionally applies PCA to residuals (see Section 5.2).

We also should not forget many interesting contributions by French data analysts in related areas (e.g., Bonifas et al., 1984; Durand, 1993; Sabatier, Lebreton and Chessel, 1989). The use of the term GSVD in this paper follows their tradition (Cailliez and Pages, 1976; Escoufier, 1987; Greenacre, 1984). Among North American numerical analysts, however, the same terminology has been used to refer to a related, but different, procedure (Van Loan, 1976), which is a technique to solve the generalized eigenvalue problem of the form $(\mathbf{A}'\mathbf{A} - \lambda\mathbf{B}'\mathbf{B})\mathbf{x} = 0$ without explicitly forming $\mathbf{A}'\mathbf{A}$ and $\mathbf{B}'\mathbf{B}$. De Moor and Golub (1991) recently proposed to call it QSVD (Quotient SVD) instead of GSVD. QSVD has been extended to RSVD (Restricted SVD) which involves not two, but three, rectangular matrices simultaneously.

6 Discussion

CPCA is a versatile technique for structural analysis of multivariate data. It is widely applicable and subsumes a number of existing methods as special cases. Technically, CPCA amounts to two major analytic techniques, projection and GSVD, both of which can be obtained non-iteratively. The computation involved is simple, efficient, and free from dangers of suboptimal solutions. Component scores are uniquely defined (unlike in factor analysis, there is no factorial indeterminacy problem), and solutions are nested in the sense that lower dimensions are retained in higher dimensional solutions.

No distributional assumptions were deliberately made on the data so as not to limit the applicability of CPCA. It may be argued, however, that this has a negative impact on statistical model evaluation. Goodness of fit evaluation and dimensionality selection are undoubtedly more difficult, although various cross-validation approaches (Eastment and Krzanowski, 1982; Geisser, 1975; Stone, 1974) are feasible. For example, the bootstrap method (Efron, 1979) can easily be used to assess the degree of stability of the analysis results. There are also some attempts to develop analytic distribution theories in some special cases of CPCA (e.g., Denis, 1987; Rao, 1985).

It may also be argued that in contrast to ACOVS (e.g., Jöreskog, 1970), CPCA does not take into account measurement errors. Although it is true that the treatment of measurement errors is totally different in the two methods, CPCA has its mechanism to reduce the amount of measurement errors in the solution. Discarding components associated with smaller singular values in the internal analysis has the effect of eliminating measurement errors (Gleason and Staelin, 1973). Furthermore, information concerning reliability of measurement can be incorporated into CPCA via metric matrices (see Section 2.3).

PCA and CPCA are generally considered scale variant, in contrast to ACOVS which is scale invariant (e.g., Bollen, 1989) if the maximum likelihood or the generalized least squares method is used for estimation. This statement is only half true. While PCA and CPCA are not scale invariant with $\mathbf{L} = \mathbf{I}$, they can be made scale invariant by specifying an appropriate non-identity $\mathbf{L}$, as has been discussed in Section 2.3. A crucial question is how to choose an appropriate $\mathbf{L}$. This seems to be a long neglected area of research that requires further investigations (but see Meredith and Millsap, 1985).

One limitation of CPCA is that it cannot fit different sets of constraints imposed on different dimensions, unless they are mutually orthogonal or orthogonalized *a priori*. A separate method (DCDD) has been developed specifically to deal with this kind of constraints in PCA-like settings (Takane et al., 1995).

Development of CPCA is still under progress. It will be interesting to extend CPCA to cover structural equation models, multilevel analysis, time series analysis, dynamical systems, etc. Extensions of CPCA into structural equation models may make CPCA similar to the PLS (Lohmöller, 1989) approach to structural equation models. In both methods, models are fitted to data matrices rather than covariance matrices. However, solutions are analytic in CPCA, while they are iterative in the latter. In view of the nature of solutions, PLS is in fact more similar to DCDD, which is also iterative. Takane et al. (1995) discussed similarities and distinctions between PLS and DCDD.

There are a few problems left undiscussed or only briefly discussed in this paper. They include, among others, optimal data transformations, graphic displays, missing observations, and robust estimations. These, however, have to await separate publications. Also, no illustrative examples are given in this paper. They are given in a companion paper (Hunter and Takane, 2000).

Acknowledgments. We are grateful to Henk Kiers, Shizuhiko Nishisato, Jim Ramsay, Cajo ter Braak, and Haruo Yanai for their insightful comments on earlier drafts of this paper.

References

1. Anderson, T. W.: Estimating linear restrictions on regression coefficients for multivariate normal distributions. Annals Math. Stat. **22**, 327–351 (1951)
2. Bechtel, G. G., Tucker, L. R., Chang, W.: A scalar product model for the multidimensional scaling of choice. Psychometrika **36**, 369–387 (1971)
3. Böckenholt, U., Böckenholt, I.: Canonical analysis of contingency tables with linear constraints. Psychometrika **55**, 633–639 (1990)
4. Bollen, K. A.: Structural Equations with Latent Variables. New York: Wiley (1989)
5. Bonifas, L., et al.: Choix de variables en analyse composantes principales. Revue de Statistique Appliqueée **32**(2), 5–15 (1984)
6. Cailliez, F., Pages, J. P.: Introduction à l'Analyse des Données. Paris: Societe de Mathematique Appliquees et de Sciences Humaines (1976)
7. Carroll, J. D., Pruzansky, S., Kruskal, J. B.: CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. Psychometrika **45**, 3–24 (1980)
8. De Moor, B. L. R, Golub, G. H.: The restricted singular value decomposition: properties and applications. SIAM Journal: Matr. Anal. Appl. **12**, 401–425 (1991)
9. Denis, J. B.: Two way analysis using covariates. Statistics **19**, 123–132 (1988)
10. De Soete, G., Carroll, J. D.: A maximum likelihood method for fitting the wandering vector model. Psychometrika **48**, 553–566 (1983)
11. Durand, J. F.: Generalized principal component analysis with respect to instrumental variables via univariate spline transformations. Comput. Stat. Data Anal. **16**, 423–440 (1993)
12. Eastment, H. T., Krzanowski, W. J.: Cross-validatory choice of the number of components from a principal component analysis. Technometrics **24**, 73–77 (1982)
13. Efron, B.: Bootstrap methods: another look at the Jackknife. Annals Stat. **7**, 1–26 (1979)
14. Escoufier, Y.: The duality diagram: a means for better practical applications. In: Legenre, P., Legendre, L. (eds.) Development in numerical ecology, pp. 139–156. Berlin: Springer (1987)
15. Faddeev, D. K., Faddeeva, V. N.: Computational Methods of Linear Algebra. San Francisco: Freeman (1963)
16. Gabriel, K. R.: Least squares approximation of matrices by additive and multiplicative models. J. Royal Statistical Soc., Series B **40**, 186–196 (1978)
17. Geisser, S.: The predictive sample reuse method with applications. J. Am. Stat. Assoc. **70**, 320–328 (1975)

18. Gollob, H. F.: A statistical model which combines features of factor analytic and analysis of variance technique. Psychometrika **33**, 73–115 (1968)
19. Golub, G. H.: Some modified eigenvalue problems. SIAM Journal: Review **15**, 318–335 (1973)
20. Golub, G. H., Van Loan, C. F.: Matrix computations. 2nd edn. Baltimore: Johns Hopkins University Press (1989)
21. Greenacre, M. J.: Theory and Applications of Correspondence Analysis. London: Academic Press (1984)
22. Guttman, L.: General theory and methods for matric factoring. Psychometrika **9**, 1–16 (1944)
23. Guttman, L.: Multiple group methods for common-factor analysis: their basis, computation and interpretation. Psychometrika **17**, 209–222 (1952)
24. Guttman, L.: Image theory for the structure of quantitative variables. Psychometrika **9**, 277–296 (1953)
25. Heiser, W. J., de Leeuw, J.: Multidimensional mapping of preference data. Mathematiqué et sciences humaines **19**, 39–96 (1981)
26. Hunter, M. A., Takane, Y.: Constrained principal component analysis: applications. Submitted to J. Edu. Behav. Stat. (2000)
27. Ihara, M., Kano, Y.: A new estimator of the uniqueness in factor analysis. Psychometrika **51**, 563–566 (1986)
28. Johnston, J.: Econometric methods. 3rd edn. New York: McGraw Hill (1984)
29. Jöreskog, K. G.: A general method for analysis of covariance structures. Biometrika **57**, 239–251 (1970)
30. Khatri, C. G.: A note on a MANOVA model applied to problems in growth curves. Annals Inst. Stat. Math. **18**, 75–86 (1966)
31. Kiers, H. A. L.: Simple structure in component analysis techniques for mixtures of qualitative and quantitative variables. Psychometrika **56**, 197–212 (1991)
32. Lohmöller, J.: Latent Variable Path Modeling with Partial Least Squares. Heidelberg: Physica Verlag (1989)
33. McDonald, R. P.: A simple comprehensive model for the analysis of covariance structures. Br. J. Math. Stat. Psychol. **31**, 59–72 (1978)
34. Meredith, W., Millsap, R. E.: On component analysis. Psychometrika **50**, 495–507 (1985)
35. Nishisato, S.: Analysis of Categorical Data: Dual Scaling and its Applications. Toronto: University of Toronto Press (1980)
36. Nishisato, S., Lawrence, D. R.: Dual scaling of multiway data matrices: several variants. In: Coppi, R., Bolasco, S. (eds.) Multiway data analysis, pp. 317–326. Amsterdam: North-Holland (1989)
37. Okamoto, M.: Four techniques of principal component analysis. J. Jap. Stat. Soc. **2**, 63–69 (1972)
38. Potthoff, R. F., Roy, S. N.: A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika **51**, 313–326 (1964)
39. Rao, C. R.: The use and interpretation of principal component analysis in applied research. Sankhyã A **26**, 329–358 (1964)
40. Rao, C. R.: The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. Biometrika **52**, 447–458 (1965)
41. Rao, C. R.: Linear Statistical Inference and its Application. New York: Wiley (1973)
42. Rao, C. R.: Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In: Krishnaiah P. R. (ed.) Multivariate analysis V, pp. 3–22. Amsterdam: North-Holland (1980)
43. Rao, C. R.: Tests for dimensionality and interaction of mean vectors under general and reducible covariance structures. J. Multivariate Anal. **16**, 173–184 (1985)
44. Rao, C. R., Yanai, H.: General definition and decomposition of projectors and some applications to statistical problems. J. Stat. Infer. Planning **3**, 1–17 (1979)

45. Sabatier, R., Lebreton, J. D., Chessel, D.: Principal component analysis with instrumental variables as a tool for modelling composition data. In: Coppi, R., Bolasco, S. (eds.) Multiway data analysis, pp. 341–352. Amsterdam: North Holland (1989)
46. Schönemann, P. H., Steiger, J. H.: Regression component analysis. Br. J. Stat. Math. Psychol. **29**, 175–189 (1976)
47. Stone, M.: Cross-validatory choice and assessment of statistical prediction (with discussion) J. Royal Stat. Soc., Series B **36**, 111–147 (1974)
48. Takane, Y.: Maximum likelihood estimation in the generalized case of Thurstone's model of comparative judgment. Jap. Psychol. Res. **22**, 188–196 (1980)
49. Takane, Y., Kiers, H. A. L., de Leeuw, J.: Component analysis with different sets of constraints on different dimensions. Psychometrika **60**, 259–280 (1995)
50. Takane, Y., Shibayama, T.: Principal component analysis with external information on both subjects and variables. Psychometrika **56**, 97–120 (1991)
51. Takane, Y., Yanai, H.: On oblique projectors. Linear Algebra Appl. **289**, 297–310 (1999)
52. Takane, Y., Yanai, H., Mayekawa, S.: Relationships among several methods of linearly constrained correspondence analysis. Psychometrika **56**, 667–684 (1991)
53. ter Braak, C. J. F.: Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology **67**, 1167–1179 (1986)
54. van den Wollenberg, A. L.: Redundancy analysis: an alternative for canonical correlation analysis. Psychometrika **42**, 207–219 (1977)
55. Van Loan, C. F.: Generalizing the singular value decomposition. SIAM J. Num. Anal. **13**, 76–83 (1976)
56. Yanai, H.: Factor analysis with external criteria. Jap. Psychol. Res. **12**, 143–153 (1970)
57. Yanai, H.: Some generalized forms of least squares g-inverse, minimum norm g-inverse and Moore-Penrose inverse matrices. Comput. Stat. Data Anal. **10**, 251–260 (1990)
58. Yanai, H., Takane, Y.: Canonical correlation analysis with linear constraints. Linear Algebra Appl. **176**, 75–89 (1992)