# CPCA: A COMPREHENSIVE THEORY[1]

**Yoshio Takane**
**Department of Psychology, McGill University**
**Montreal, Quebec H3A 1B1, CANADA**

## ABSTRACT

Constrained principal component analysis (CPCA) incorporates external information into principal component analysis (PCA). 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 (Takane & 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. We also discuss four special cases of CPCA; 1) CCA (canonical correspondence analysis) and CALC (canonical analysis with linear constraints), 2) GMANOVA, 3) Lagrange's theorem, and 4) CANO (canonical correlation analysis) and related methods.

## 1. DATA REQUIREMENT

PCA is often used for structural analysis of multivariate data. The data are, however, often accompanied by auxiliary information. CPCA incorporates such information in representing structures in the data. CPCA allows specifying metric matrices that modulate the effects of rows and columns of a data matrix. There are thus three important ingredients in CPCA; the main data, external information and metric matrices.

**1.1 The Main Data:** Let us denote an $N$ by $n$ data matrix by $\mathbf{Z}$. Rows of $\mathbf{Z}$ represent cases, while columns represent variables. The data can 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. The data may be preprocessed or not preprocessed. Preprocessing refers to centering, normalizing, standardizing, or any other prescribed data transformations. Results of PCA and CPCA are typically affected by what preprocessing is applied, so whatever the decision on the preprocessing must be made deliberately in the light of investigators' empirical interests.

**1.2 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 $\mathbf{G}$, and the latter by an $n$ by $q$ matrix $\mathbf{H}$. When there is no special row and/or column information, 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 $\mathbf{G}$. For example, we may take a matrix of dummy variables for $\mathbf{G}$ indicating subjects' group membership. We then analyze the differences among the groups.

When the columns of a data matrix represent stimuli, we may take a matrix of descriptor variables of the stimuli as $\mathbf{H}$. When the columns correspond to different within-subject experimental conditions, $\mathbf{H}$ could be a matrix of contrasts, or when the variables represent repeated observations, $\mathbf{H}$ could be a matrix of trend coefficients. There are several potential advantages of incorporating external information (Takane, Kiers & de Leeuw, 1995). By incorporating external information, we may obtain more interpretable solutions. We may also obtain more stable solutions by reducing the number of paramerters to be estimated. We may investigate the empirical validity of hypotheses incorporated as external constraints by evaluating the goodness of fit of the hypotheses. We may predict missing values via external constraints which serve as predictor variables. In some cases we can eliminate incidental parameters by reparameterizing them as linear combinations of a small number of external constraints.

**1.3 Metric Matrices:** There also are two kinds of metric matrices, one on the row side, $\mathbf{K}$, and the other on the column side, $\mathbf{L}$. Metric matrices are assumed to be *nnd*. They are closely related to the criteria employed in 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 $\mathbf{L} = \mathbf{I}$. This implies that we use the unweighted LS criterion. However, when variables in a data matrix are measured on uncomparable scales, a special nonidentity metric matrix is required, leading to a weighted LS criterion. It is common, when scales are uncomparable, 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 $\mathbf{L}$. A special metric is also necessary when rows of a data matrix are correlated. The rows of a data matrix are usually assumed statistically independent, which can be justified when they represent a random sample of subjects from a target population. They tend to be correlated, when they represent, for example, different time points in sigle-subject multivariate time series data. In such cases, a matrix of serial correlations has to be estimated, and its inverse be used as $\mathbf{K}$. When differences in importance and/or in reliability among the rows are suspected, a special diagonal matrix is used for $\mathbf{K}$ that has the effect of differentially weighting rows of a data matrix.

When columns of a data matrix are correlated, no special metric matrix is usually used. However, when the columns of the residual matrix are correlated

[1]In the *Proceedings of the 1997 IEEE-SMC International Conference*, pp. 35-40.

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 $\mathbf{L}$. This has the effect of improving the quality of parameter estimates by orthonormalizing the residuals in evaluating the overall goodness of fit of the model. Meredith & Millsap (1985) suggests to use reliability coefficients or inverses of variances of anti-images (Guttman, 1953) as a nonidentity $\mathbf{L}$. Although as typically used, PCA is not scale invariant, Rao (1964) has shown that specifying certain non-identity $\mathbf{L}$ matrices have the effect of attaining scale invariance.

## 2. BASIC THEORY

We present CPCA in its general form. The provision of nonidentity metric matrices widens the scope of CPCA. For example, it makes correspondence analysis of various kinds (Greenacre, 1984; Nishisato, 1980; Takane, Yanai & Mayekawa, 1991) a special case of CPCA.

**2.1 External Analysis:** 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 $\mathbf{H}$. We assume that metric matrices, $\mathbf{K}$ and $\mathbf{L}$ are both *nnd*, and that $\text{rank}(\mathbf{KG}) = \text{rank}(\mathbf{G})$, and $\text{rank}(\mathbf{LH}) = \text{rank}(\mathbf{H})$. These conditions are necessary for $\mathbf{P}_{G/K}$ and $\mathbf{P}_{H/L}$ (defined below) to be projectors. Model (1) is underidentified. To identify the model, it is convenient to impose the orthogonality constraints, $\mathbf{G}'\mathbf{KB} = \mathbf{0}$, and $\mathbf{H}'\mathbf{LC}' = \mathbf{0}$.

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

$$f = SS(\mathbf{E})_{K,L} = \text{tr}(\mathbf{E}'\mathbf{KEL}). \tag{2}$$

This leads to the following LS estimates of $\mathbf{M}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{E}$:

$$\hat{\mathbf{M}} = (\mathbf{G}'\mathbf{KG})^{-}\mathbf{G}'\mathbf{KZLH}(\mathbf{H}'\mathbf{LH})^{-}, \tag{3}$$

$$\hat{\mathbf{B}} = \mathbf{K}^{-}\mathbf{KQ}_{G/K}\mathbf{ZLH}(\mathbf{H}'\mathbf{LH})^{-}, \tag{4}$$

$$\hat{\mathbf{C}} = (\mathbf{G}'\mathbf{KG})^{-}\mathbf{G}'\mathbf{KZQ}'_{H/L}\mathbf{LL}^{-}, \tag{5}$$

and

$$\hat{\mathbf{E}} = \mathbf{Z} - \mathbf{P}_{G/K}\mathbf{ZP}'_{H/L} - \mathbf{K}^{-}\mathbf{KQ}_{G/K}\mathbf{ZP}'_{H/L} - \mathbf{P}_{G/K}\mathbf{ZQ}'_{H/L}\mathbf{LL}^{-}, \tag{6}$$

where superscript "–" indicates a g-inverse of a matrix, $\mathbf{P}_{G/K} = \mathbf{G}(\mathbf{G}'\mathbf{KG})^{-}\mathbf{G}'\mathbf{K}$, $\mathbf{Q}_{G/K} = \mathbf{I} - \mathbf{P}_{G/K}$, $\mathbf{P}_{H/L} = \mathbf{H}(\mathbf{H}'\mathbf{LH})^{-}\mathbf{H}'\mathbf{L}$, $\mathbf{Q}_{H/L} = \mathbf{I} - \mathbf{P}_{H/L}$. Matrices $\mathbf{P}_{G/K}$, and $\mathbf{Q}_{G/K}$ are projectors such that $\mathbf{P}^2_{G/K} = \mathbf{P}_{G/K}$, $\mathbf{Q}^2_{G/K} = \mathbf{Q}_{G/K}$, $\mathbf{P}_{G/K}\mathbf{Q}_{G/K} = \mathbf{Q}_{G/K}\mathbf{P}_{G/K} = \mathbf{0}$, $\mathbf{P}'_{G/K}\mathbf{KP}_{G/K} = \mathbf{P}'_{G/K}\mathbf{K} = \mathbf{KP}_{G/K}$, and $\mathbf{Q}'_{G/K}\mathbf{KQ}_{G/K} = \mathbf{Q}'_{G/K}\mathbf{K} = \mathbf{KQ}_{G/K}$. Similar properties hold for $\mathbf{P}_{H/L}$ and $\mathbf{Q}_{H/L}$. They reduce to the usual ($I$-orthogonal) projectors when $\mathbf{K} = \mathbf{I}$ and $\mathbf{L} = \mathbf{I}$.

Putting the LS estimates of $\mathbf{M}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{E}$ above in model (1) yields the following decomposition of the data matrix, $\mathbf{Z}$:

$$\mathbf{Z} = \mathbf{P}_{G/K}\mathbf{ZP}'_{H/L} + \mathbf{K}^{-}\mathbf{KQ}_{G/K}\mathbf{ZP}'_{H/L} + \mathbf{P}_{G/K}\mathbf{ZQ}'_{H/L}\mathbf{LL}^{-} + (\mathbf{Z} - \mathbf{P}_{G/K}\mathbf{ZP}'_{H/L} - \mathbf{K}^{-}\mathbf{KQ}_{G/K}\mathbf{ZP}'_{H/L} - \mathbf{P}_{G/K}\mathbf{ZQ}'_{H/L}\mathbf{LL}^{-}). \tag{7}$$

The four terms in (7) are mutually orthogonal with respect to $\mathbf{K}$ and $\mathbf{L}$, so that

$$SS(\mathbf{Z})_{K,L} = SS(\mathbf{G}\hat{\mathbf{M}}\mathbf{H}')_{K,L} + SS(\hat{\mathbf{B}}\mathbf{H}')_{K,L} + SS(\mathbf{G}\hat{\mathbf{C}})_{K,L} + SS(\hat{\mathbf{E}})_{K,L}. \tag{8}$$

That is, sum of squares of $\mathbf{Z}$ is decomposed into the sum of sums of squares of the four terms in (7).

When $\mathbf{K}$ and $\mathbf{L}$ are both nonsingular, $\mathbf{K}^{-}\mathbf{K} = \mathbf{I}$ and $\mathbf{L}^{-}\mathbf{L} = \mathbf{I}$, so that decomposition (7) reduces to,

$$\mathbf{Z} = \mathbf{P}_{G/K}\mathbf{ZP}'_{H/L} + \mathbf{Q}_{G/K}\mathbf{ZP}'_{H/L} + \mathbf{P}_{G/K}\mathbf{ZQ}'_{H/L} + \mathbf{Q}_{G/K}\mathbf{ZQ}'_{H/L}, \tag{9}$$

and (8) to

$$SS(\mathbf{Z})_{K,L} = SS(\mathbf{P}_{G/K}\mathbf{ZP}'_{H/L})_{K,L} + SS(\mathbf{Q}_{G/K}\mathbf{ZP}'_{H/L})_{K,L} + SS(\mathbf{P}_{G/K}\mathbf{ZQ}'_{H/L})_{K,L} + SS(\mathbf{Q}_{G/K}\mathbf{ZQ}'_{H/L})_{K,L}. \tag{10}$$

**2.2 Internal Analysis:** In the internal analysis, the decomposed matrices in (7) or (9) 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 (7) 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 (Rao, 1980; Yanai, 1970).

PCA with nonindentity 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 *nnd* matrices. Let $\mathbf{A}$ be an $N$ by $n$ matrix of rank $r$. Then,

$$\mathbf{R}'_K\mathbf{AR}_L = \mathbf{R}'_K\mathbf{UDV}'\mathbf{R}_L \tag{11}$$

is called GSVD of $\mathbf{A}$ under metrics $\mathbf{K}$ and $\mathbf{L}$, and is written as $\text{GSVD}(\mathbf{A})_{K,L}$, where $\mathbf{R}_K$ and $\mathbf{R}_L$ are such that $\mathbf{K} = \mathbf{R}_K\mathbf{R}'_K$ and $\mathbf{L} = \mathbf{R}_L\mathbf{R}'_L$, $\mathbf{U}$ ($N$ by $r$) is such that $\mathbf{U'KU} = \mathbf{I}$, $\mathbf{V}$ ($n$ by $r$) is such that $\mathbf{V'LV} = \mathbf{I}$, and $\mathbf{D}$ ($r$ by $r$) is diagonal and *pd*. When $\mathbf{K}$ and $\mathbf{L}$ are nonsingular, (11) reduces to

$$\mathbf{A} = \mathbf{UDV'}, \tag{12}$$

where $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{D}$ have the same properties as above.

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

$$\mathbf{R}'_K\mathbf{A}\mathbf{R}_L = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime}. \tag{13}$$

Then, $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{D}$ in GSVD$(\mathbf{A})_{K,L}$ are obtained by $\mathbf{U} = (\mathbf{R}'_K)^-\mathbf{U}^*$, $\mathbf{V} = (\mathbf{R}'_L)^-\mathbf{V}^*$, and $\mathbf{D} = \mathbf{D}^*$. It can easily be verified that these $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{D}$ satisfy the required properties of GSVD.

GSVD plays an important role in CPCA. The following two theorems (given without proof) 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., $\mathbf{T'T} = \mathbf{I}$ and $\mathbf{W'W} = \mathbf{I}$. 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 $\mathbf{TAW'}$ by $\mathbf{TAW'} = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime}$. Then, $\mathbf{U}^* = \mathbf{TU}_A$ ($\mathbf{U}_A = \mathbf{T'U}^*$), $\mathbf{V}^* = \mathbf{WV}_A$ ($\mathbf{V}_A = \mathbf{W'V}^*$), and $\mathbf{D}_A = \mathbf{D}^*$.

**Theorem 2.** Let $\mathbf{T}$ and $\mathbf{W}$ 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 $\mathbf{U}_A = (\mathbf{T'KT})^-\mathbf{T'KU}$, $\mathbf{V}_A = (\mathbf{W'LW})^-\mathbf{W'LV}$ and $\mathbf{D}_A = \mathbf{D}$.

In some cases, GSVD$(\hat{\mathbf{M}})_{G'KG,H'LH}$, where $\hat{\mathbf{M}}$ is given in (3), may be of direct interest. For example, Takane & Shibayama (1991) discussed vector preference models, in which $\mathbf{K} = \mathbf{I}$, $\mathbf{L} = \mathbf{I}$, $\mathbf{G} = \mathbf{I}$, and $\mathbf{H}$ is a design matrix for pair comparisons. In those models $\mathbf{M}$ contains scale values of stimuli, and consequently GSVD$(\hat{\mathbf{M}})_{I,H'H}$ is of direct interest. GSVD$(\hat{\mathbf{M}})_{G'KG,H'LH}$ may be calculated directly, or indirectly from the related GSVD discussed above. In particular, if $\hat{\mathbf{M}} = \mathbf{U}_M\mathbf{D}_M\mathbf{V}'_M$ represents GSVD $(\hat{\mathbf{M}})_{G'KG,H'LH}$, then because of Theorem 2, $\mathbf{U}_M = (\mathbf{G'KG})^-\mathbf{G'KU}$, $\mathbf{V}_M = (\mathbf{H'LH})^-\mathbf{H'LV}$ 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 $\mathbf{V} = \mathbf{HV}_M$, 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 $((\mathbf{G'G})^-\mathbf{G'H}(\mathbf{H'H})^-)_{G'G,H'H}$, whereas canonical variates are directly obtained by SVD$(\mathbf{P}_G\mathbf{P}_H)$, where $\mathbf{P}_G = \mathbf{P}_{G/I}$ and $\mathbf{P}_H = \mathbf{P}_{H/I}$.

## 3. SOME EXTENSIONS

Within the basic framework of CPCA, various extensions are possible. We discuss two of them here.

### 3.1 Decompositions into Finer Components:

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. 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. Suppose $\mathbf{G}$ consists of two constraint sets, $\mathbf{X}$ and $\mathbf{Y}$; i.e., $\mathbf{G} = [\mathbf{X}|\mathbf{Y}]$. A variety of decompositions are possible (Rao & Yanai, 1979), depending on the relationship between $\mathbf{X}$ and $\mathbf{Y}$.

When $\mathbf{X}$ and $\mathbf{Y}$ are mutually orthogonal (with respect to $\mathbf{K}$), we have

$$\mathbf{P}_{G/K} = \mathbf{P}_{X/K} + \mathbf{P}_{Y/K}. \tag{14}$$

This simply partitions the joint effect of $\mathbf{X}$ and $\mathbf{Y}$ into the sum of the independent effects of $\mathbf{X}$ and $\mathbf{Y}$.

When $\mathbf{X}$ and $\mathbf{Y}$ are orthogonal except in their intersection space, $\mathbf{P}_{X/K}$ and $\mathbf{P}_{Y/K}$ are still commutative, and

$$\mathbf{P}_{G/K} = \mathbf{P}_{X/K} + \mathbf{P}_{Y/K} - \mathbf{P}_{X/K}\mathbf{P}_{Y/K}. \tag{15}$$

This decomposition plays an important role in ANOV A for factorial designs.

When $\mathbf{X}$ and $\mathbf{Y}$ are not mutually orthogonal in any way, two decompositions are possible:

$$\begin{aligned}\mathbf{P}_{G/K} &= \mathbf{P}_{X/K} + \mathbf{P}_{Q_{Y/K}X/K} \\ &= \mathbf{P}_{Y/K} + \mathbf{P}_{Q_{X/K}Y/K},\end{aligned} \tag{16}$$

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 $\mathbf{X}$ that is unaccounted for by $\mathbf{Y}$) and $\mathbf{Q}_{X/K}\mathbf{Y}$, respectively. The above decompositions are useful when one of $\mathbf{X}$ and $\mathbf{Y}$ is fitted first and the other is fitted to the residuals.

When $Sp(\mathbf{X})$ and $Sp(\mathbf{Y})$ are not mutually orthogonal but disjoint, we may use

$$\begin{aligned}\mathbf{P}_{G/K} = &\ \mathbf{X}(\mathbf{X'KQ}_{Y/K}\mathbf{X})^-\mathbf{X'KQ}_{Y/K} \\ &+ \mathbf{Y}(\mathbf{Y'KQ}_{X/K}\mathbf{Y})^-\mathbf{Y'KQ}_{X/K}.\end{aligned} \tag{17}$$

Note that $\mathbf{KQ}_{Y/K}$ and $\mathbf{KQ}_{X/K}$ are both symmetric. This decomposition is useful when $\mathbf{X}$ and $\mathbf{Y}$ are fitted simultaneously. Note that unlike all other decompositions discussed in this section, the two terms in this decomposition are not mutually orthogonal.

When additional information is given as constraints on the weight matrix $\mathbf{U}_A$ applied to $\mathbf{G}$, i.e., $\mathbf{U}_G = \mathbf{AU}_A$ for a given matrix, $\mathbf{A}$. Then,

$$\mathbf{P}_{G/K} = \mathbf{P}_{GA/K} + \mathbf{P}_{G(G'KG)^-B/K}, \tag{18}$$

where $\mathbf{A'B} = 0$, $Sp(\mathbf{A}) \oplus Sp(\mathbf{B}) = Sp(\mathbf{G}')$, and $\mathbf{B} = \mathbf{G'KW}$ for some $\mathbf{W}$ (Yanai & Takane, 1992). Since $\mathbf{B}'(\mathbf{G'KG})^-\mathbf{G'KGU}_A = 0$ for $\mathbf{B}$ such that $\mathbf{B} = \mathbf{G'KW}$, the constraint $\mathbf{U}_G = \mathbf{AU}_A$ can also be expressed as $\mathbf{B'U}_G = 0$. This decomposition is an example of higher-order structures to be discussed in the next section.

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.

**3.2 Higher-Order Structures:** External information other than $\mathbf{G}$ or $\mathbf{H}$ can also be incorporated. This information is often provided in the form of a hypothesis about the parameters in the model. In such cases we may be interested in obtaining an estimate of the parameters under that hypothesis. For example, a model similar to (1) for $\mathbf{Z}$ may also be assumed for $\mathbf{M}$. 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 $\mathbf{S}$ denote the design matrix for the stimuli. It may be assumed that $\mathbf{M} = \mathbf{WS}' + \mathbf{E}^*$, where $\mathbf{W}$ is a matrix of weights applied to $\mathbf{S}'$. The entire model may then be written as

$$\begin{aligned} \mathbf{Z} &= \mathbf{G}(\mathbf{WS}' + \mathbf{E}^*)\mathbf{A}' + \mathbf{E} \\ &= \mathbf{GWS}'\mathbf{A}' + \mathbf{GE}^*\mathbf{A}' + \mathbf{E}. \end{aligned} \quad (19)$$

This model partitions $\mathbf{Z}$ into three parts: what can be explained by $\mathbf{G}$ and $\mathbf{AS}$, what can be explained by $\mathbf{G}$ and $\mathbf{A}$ but not by $\mathbf{AS}$, and the residual.

Alternatively, $\mathbf{M}$ may be subjected to PCA first, and then some hypothesized structure may be imposed on its row representation, $\mathbf{U}_M$, or on $\mathbf{U} = \mathbf{GU}_M$. In the latter case, we may have

$$\begin{aligned} \mathbf{Z} &= \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime} + \mathbf{GE}^*\mathbf{H}' + \mathbf{E} \\ &= (\mathbf{TW} + \tilde{\mathbf{E}})\mathbf{D}^*\mathbf{V}^{*\prime} + \mathbf{GE}^*\mathbf{H}' + \mathbf{E}, \end{aligned} \quad (20)$$

where $\mathbf{T}$ is an additional row information matrix. An LS estimate of $\mathbf{W}$ in this model, given the estimate of $\mathbf{U}^*$, is obtained by $\hat{\mathbf{W}} = (\mathbf{T}'\mathbf{KT})^-\mathbf{T}'\mathbf{KU}^*$. This $\hat{\mathbf{W}}$ can also be obtained directly by GSVD $(\mathbf{P}_{GT/K}\mathbf{ZP}'_{H/L})_{K,L}$.

## 4. SPECIAL CASES

CPCA subsumes a number of interesting special cases. In this paper we focuss on those which have not been discussed previously (Takane & Shibayama, 1991). Specifically, we discuss four groups of methods; canonical correspondence analysis (CCA; ter Braak, 1986) and canonical analysis with linear constraints (CALC; Böckenholt & Böckenholt, 1990); GMANOVA (Potthoff & Roy, 1964) and its extensions (Khatri, 1966); CPCA with components within row and column spaces of data matrices (Guttman, 1944; Rao, 1964); and relationships among CPCA, canonical correlation analysis (CANO) and related methods.

**4.1 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. Let $\mathbf{F}$ denote a two-way contingency table. Correspondence analysis (CA) of $\mathbf{F}$ obtains an "optimal" row and column representation of $\mathbf{F}$, which technically amounts to obtaining GSVD$(\mathbf{D}_R^-\mathbf{FD}_C^-)_{D_R,D_C}$, where $\mathbf{D}_R$ and $\mathbf{D}_C$ are diagonal matrices of row and column totals of $\mathbf{F}$, respectively. Let $\mathbf{UDV}'$ denote the GSVD. Then, the row and column representations of $\mathbf{F}$ are obtained by simple rescaling of $\mathbf{U}$ and $\mathbf{VD}$, respectively.

Let $\mathbf{X}$ denote the row constraint matrix. CCA obtains $\mathbf{U}$ under the restriction that $\mathbf{U} = \mathbf{XU}^*$, where $\mathbf{U}^*$ is a matrix of weights. This amounts to GSVD $((\mathbf{X}'\mathbf{D}_R\mathbf{X})^-\mathbf{X}'\mathbf{FD}_C^-)_{X'D_RX,D_C}$ from which $\mathbf{U}^*$ is obtained (and then, $\mathbf{U}$ is derived by $\mathbf{U} = \mathbf{XU}^*$), orto GSVD$(\mathbf{X}(\mathbf{X}'\mathbf{D}_R\mathbf{X})^-\mathbf{X}'\mathbf{FD}_C^-)_{D_R,D_C}$ from which $\mathbf{U}$ is directly obtained (Takane, Yanai, & Mayekawa, 1991). 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($\mathbf{F}$, $\mathbf{X}$) = GSVD$(\mathbf{X}(\mathbf{X}'\mathbf{D}_R\mathbf{X})^-\mathbf{X}'\mathbf{FD}_C^-)_{D_R,D_C}$.

CALC 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 GSVD$(\mathbf{D}_R^-(\mathbf{I} - \mathbf{R}(\mathbf{R}'\mathbf{D}_R^-\mathbf{R})^-\mathbf{R}'\mathbf{D}_R^-)\mathbf{FD}_C^-)_{D_R,D_C}$, which will be denoted as CALC($\mathbf{F}$, $\mathbf{R}$), or simply CALC($\mathbf{R}$).

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($\mathbf{X}$) = CALC($\mathbf{R}$) if $\mathbf{X}$ and $\mathbf{R}$ are mutually orthogonal, and together they span the entire column space of $\mathbf{F}$. For a given $\mathbf{R}$, such an $\mathbf{X}$ can be obtained by $\mathbf{X}$ such that $\mathbf{I} - \mathbf{R}(\mathbf{R}'\mathbf{R})^-\mathbf{R}' = \mathbf{XX}'$. Similarly, an $\mathbf{R}$ can be obtained from a given $\mathbf{X}$ by $\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}' = \mathbf{RR}'$. It can also be shown that CCA and CALC are both special cases of CPCA. When $\mathbf{H} = \mathbf{I}$, decomposition (7) reduces to

$$\mathbf{Z} = \mathbf{P}_{G/K}\mathbf{Z} + \mathbf{Q}_{G/K}\mathbf{Z}. \quad (21)$$

The first term in (21) can be rewritten as

$$\mathbf{P}_{G/K}\mathbf{Z} = \mathbf{G}(\mathbf{G}'\mathbf{KG})^-\mathbf{G}'(\mathbf{KZL})\mathbf{L}^-, \quad (22)$$

which is equal to $\mathbf{X}(\mathbf{X}'\mathbf{D}_R\mathbf{X})^-\mathbf{X}'\mathbf{FD}_C^-$, if $\mathbf{G} = \mathbf{X}$, $\mathbf{K} = \mathbf{D}_R$, $\mathbf{L} = \mathbf{D}_C$, and $\mathbf{Z} = \mathbf{D}_R^-\mathbf{FD}_C^-$. That is, GSVD$(\mathbf{P}_{G/K}\mathbf{Z})_{K,L}$ = CCA($\mathbf{F}$, $\mathbf{X}$).

The second term in (21) can be rewritten as

$$\begin{aligned} \mathbf{Q}_{G/K}\mathbf{Z} &= (\mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{KG})^-\mathbf{G}'\mathbf{K})\mathbf{Z} \\ &= \mathbf{K}^-(\mathbf{I} - \mathbf{KG}(\mathbf{G}'\mathbf{KK}^-\mathbf{KG})^- \\ &\quad \mathbf{G}'\mathbf{KK}^-)(\mathbf{KZL})\mathbf{L}^-, \end{aligned} \quad (23)$$

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

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

**4.2 GMANOVA:** GMANOVA (Potthoff & Roy, 1964) postulates

$$\mathbf{Z} = \mathbf{GMH}' + \mathbf{E}. \quad (24)$$

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 $\mathbf{E}$ are *iid* multivariate normal, a maximum likelihood estimate of $\mathbf{M}$ is obtained by $\hat{\mathbf{M}} = (\mathbf{G}'\mathbf{G})^-\mathbf{G}'\mathbf{Z}\mathbf{S}^{-1}\mathbf{H}(\mathbf{H}'\mathbf{S}^{-1}\mathbf{H})^-$ (Khatri, 1966), where $\mathbf{S} = \mathbf{Z}'(\mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{G})^-\mathbf{G}')\mathbf{Z}$ assumed to be nonsingular. This estimate of $\mathbf{M}$ is equivalent to an LS estimate of $\mathbf{M}$ in (3) with $\mathbf{K} = \mathbf{I}$ and $\mathbf{L} = \mathbf{S}^{-1}$.

In GMANOVA, tests of hypotheses about $\mathbf{M}$ are typically of interest: $\mathbf{R}'\mathbf{M}\mathbf{C} = \mathbf{0}$, 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 $\mathbf{W}_C$. 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 $Sp[\mathbf{R}|\mathbf{X}] = Sp(\mathbf{G}')$, and $\mathbf{C}'\mathbf{Y} = \mathbf{0}$ and $Sp[\mathbf{C}|\mathbf{Y}] = Sp(\mathbf{H}')$. Then, $\mathbf{M}$ can be reparameterized as

$$\mathbf{M} = \mathbf{X}\mathbf{M}_{XY}\mathbf{Y}' + \mathbf{M}_Y\mathbf{Y}' + \mathbf{X}\mathbf{M}_X, \qquad (25)$$

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 $\mathbf{X}'\mathbf{G}'\mathbf{K}\mathbf{G}\mathbf{M}_Y = \mathbf{0}$, and $\mathbf{Y}'\mathbf{H}'\mathbf{L}\mathbf{H}\mathbf{M}_X' = \mathbf{0}$. Putting (25) into (24), we obtain

$$\begin{aligned}\mathbf{Z} = \mathbf{G}\mathbf{X}\mathbf{M}_{XY}\mathbf{Y}'\mathbf{H}' + \mathbf{G}\mathbf{M}_Y\mathbf{Y}'\mathbf{H}' \\ + \mathbf{G}\mathbf{X}\mathbf{M}_X\mathbf{H}' + \mathbf{E}.\end{aligned} \qquad (26)$$

LS estimates of $\mathbf{M}_{XY}$, $\mathbf{M}_Y$, and $\mathbf{M}_X$ can be obtained in a manner similar to (3) through (6). Putting them into (26) leads to the following decomposition:

$$\begin{aligned}\mathbf{Z} = \mathbf{P}_{GX/K}\mathbf{Z}\mathbf{P}'_{HY/L} + \mathbf{P}_{G(G'KG)^-R/K}\mathbf{Z}\mathbf{P}'_{HY/L} \\ + \mathbf{P}_{GX/K}\mathbf{Z}\mathbf{P}'_{H(H'LH)^-C/L} + \hat{\mathbf{E}},\end{aligned} \qquad (27)$$

where $\hat{\mathbf{E}}$ is defined as $\mathbf{Z}$ minus the sum of the first three terms in (27). In the above decpomposition $Sp(\mathbf{Z})$ is split into four mutually orthogonal subspaces.

**4.3 Lagrange's Theorem:** It is well known (e.g., Yanai, 1990) that $(\mathbf{A}')^-_{ZB} = \mathbf{Z}\mathbf{B}(\mathbf{A}'\mathbf{Z}\mathbf{B})^-$, and $\mathbf{B}^-_{A'Z} = (\mathbf{A}'\mathbf{Z}\mathbf{B})^-\mathbf{A}'\mathbf{Z}$ are reflexive g-inverses of $\mathbf{A}'$ and $\mathbf{B}$, respectively, under rank$(\mathbf{A}'\mathbf{Z}\mathbf{B})$ = rank$(\mathbf{A}')$ and rank $(\mathbf{A}'\mathbf{Z}\mathbf{B})$ = rank$(\mathbf{B})$, respectively. Define $\mathbf{Q}_{ZB,A} = \mathbf{I} - (\mathbf{A}')^-_{ZB}\mathbf{A}'$, and $\mathbf{Q}_{Z'A,B} = \mathbf{I} - \mathbf{B}\mathbf{B}^-_{A'Z}$. Then, $\mathbf{Q}_{ZB,A}$ is the projector onto $Ker(\mathbf{A}')$ along $Sp(\mathbf{Z}\mathbf{B})$, and $\mathbf{Q}_{Z'A,B}$ onto $Ker(\mathbf{A}'\mathbf{Z})$ along $Sp(\mathbf{B})$. Define

$$\mathbf{Z}_1 = \mathbf{Q}_{ZB,A}\mathbf{Z} = \mathbf{Z}\mathbf{Q}_{Z'A,B}. \qquad (28)$$

Then, rank$(\mathbf{Z}_1)$ = rank$(\mathbf{Z})$ − rank$(\mathbf{A}'\mathbf{Z}\mathbf{B})$. This is called Lagrange's theorem (Rao, 1973, p. 69).

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

$$\begin{aligned}\mathbf{Z}\mathbf{Q}_{Z'G} &= \mathbf{Z}(\mathbf{I} - \mathbf{Z}'\mathbf{G}(\mathbf{G}'\mathbf{Z}\mathbf{Z}'\mathbf{G})^-\mathbf{G}'\mathbf{Z}) \\ &= (\mathbf{I} - \mathbf{Z}\mathbf{Z}'\mathbf{G}(\mathbf{G}'\mathbf{Z}\mathbf{Z}'\mathbf{G})^-\mathbf{G}')\mathbf{Z} \\ &= \mathbf{Q}_{G/ZZ'}\mathbf{Z}.\end{aligned} \qquad (29)$$

This reduces to $\mathbf{Z}_1$ in (28) by setting $\mathbf{A} = \mathbf{G}$ and $\mathbf{B} = \mathbf{Z}'\mathbf{G}$. It is obvious that this is also a special case of $\mathbf{Z}\mathbf{Q}_H$ with $\mathbf{H} = \mathbf{Z}'\mathbf{G}$, and of $\mathbf{Q}_{G/K}\mathbf{Z}$ with $\mathbf{K} = \mathbf{Z}\mathbf{Z}'$.

Guttman (1944) considered obtaining components which are given linear combinations of $\mathbf{Z}$, and used Lagrange's theorem to successively obtain residual matrices. Let the weight matrix in the linear combinations be denoted by $\mathbf{W}$. Let $\mathbf{A} = \mathbf{Z}\mathbf{W}$ and $\mathbf{B} = \mathbf{W}$ in (28). PCA of the part of data matrix $\mathbf{Z}$ that can be explained by $\mathbf{Z}\mathbf{W}$ amounts to the SVD of $\mathbf{P}_{ZW}\mathbf{Z} = \mathbf{Z}\mathbf{P}_{W/Z'Z}$ and that of residual matrices to the SVD of $\mathbf{Q}_{ZW}\mathbf{Z} = \mathbf{Z}\mathbf{Q}'_{W/Z'Z}$. Both are special cases of CPCA.

**4.4 Relationships among CPCA, CANO and Related Methods:** A number of methods have been proposed for relating two sets of variables. In this section we show relationships among some of them: CPCA, canonical correlation analysis (CANO), CANOLC (CANO with linear constraints; Yanai & 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.

**(1)** CPCA: There are five matrices involved. It is more explicitly written as CPCA($\mathbf{Z}$, $\mathbf{G}$, $\mathbf{H}$, $\mathbf{K}$, $\mathbf{L}$).

**(2)** CANOLC: Four matrices are involved. 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 the weights to define canonical variates, $\mathbf{U}$ and $\mathbf{V}$, are linear functions of $\mathbf{G}$ and $\mathbf{H}$, respectively. That is, $\mathbf{U} = \mathbf{X}\mathbf{G}\mathbf{U}^*$ and $\mathbf{V} = \mathbf{Y}\mathbf{H}\mathbf{V}^*$, where $\mathbf{U}^*$ and $\mathbf{V}^*$ are weight matrices obtained by GSVD$((\mathbf{G}'\mathbf{X}'\mathbf{X}\mathbf{G})^-\mathbf{G}'\mathbf{X}'\mathbf{Y}\mathbf{H}(\mathbf{H}'\mathbf{Y}'\mathbf{Y}\mathbf{H})^-)_{G'X'XG,H'Y'YH}$.

**(3)** CCA: Five matrices are involved. 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. CCA amounts to GSVD$((\mathbf{G}'\mathbf{D}_R\mathbf{G})^-\mathbf{G}'\mathbf{F}\mathbf{H}(\mathbf{H}'\mathbf{D}_C\mathbf{H})^-)_{G'D_RG,H'D_CH}$

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

**(5)** 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 relations among these methods are as follows:

1) CPCA $\longrightarrow$ CANOLC: Set $\mathbf{Z} = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^-$, $\mathbf{K} = \mathbf{X}'\mathbf{X}$, and $\mathbf{L} = \mathbf{Y}'\mathbf{Y}$.

2) CPCA $\longrightarrow$ CCA: Set $\mathbf{Z} = \mathbf{D}_R^-\mathbf{F}\mathbf{D}_C^-$, $\mathbf{K} = \mathbf{D}_R$, and $\mathbf{L} = \mathbf{D}_C$.

3) CPCA $\longrightarrow$ CANO: Set $\mathbf{Z} = \mathbf{I}$, $\mathbf{K} = \mathbf{I}$, and $\mathbf{L} = \mathbf{I}$.

4) CPCA $\longrightarrow$ 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 $\mathbf{L} = \mathbf{D}_C$.

5) CANOLC $\longrightarrow$ CCA: Set $\mathbf{X}'\mathbf{Y} = \mathbf{F}$, $\mathbf{X}'\mathbf{X} = \mathbf{D}_R$, and $\mathbf{Y}'\mathbf{Y} = \mathbf{D}_C$.

6) CANOLC $\longrightarrow$ CANO: Set $\mathbf{X} = \mathbf{I}$, and $\mathbf{Y} = \mathbf{I}$.

7) 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 $\mathbf{H}'\mathbf{H} = \mathbf{D}_C$.

8) CCA ⟶ CANO: Set $\mathbf{F} = \mathbf{I}$, $\mathbf{D}_R = \mathbf{I}$, and $\mathbf{D}_C = \mathbf{I}$.

9) CCA ⟶ CA: Set $\mathbf{G} = \mathbf{I}$, and $\mathbf{H} = \mathbf{I}$.

10) 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 all special cases of CANO including MANOVA and canonical discriminant analysis.

## 5. 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, and solutions are nested in the sense that lower dimensions are retained in higher dimensional solutions.

No distributional assumptions were made on the data not to limit the applicability of CPCA. This may have some negative consequence in statistical model evaluation. Goodness of fit evaluation and dimensionality selection are undoubtedly more difficult, although various cross-validation approaches are feasible. For example, the bootstrap method (Efron, 1979) can easily be used to assess the degree of stability of the analysis results.

It may 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. Furthermore, information concerning reliability of measurement can be incorporated into CPCA via metric matrices.

PCA and CPCA are generally considered scale variant, in contrast to ACOVS which is scale invariant 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 nonidentity $\mathbf{L}$. A crucial question is how to choose an appropriate $\mathbf{L}$ (Meredith & 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 progess. It will be interesting to extend CPCA to cover structural equation models, multilevel analysis, time series analysis, dynamical systems, etc.

## REFERENCES


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

Efron, B. (1979). Bootstrap methods: another look at the Jackknife. *Annals of Statistics*, **7**, 1-26.

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

Guttman, L. (1944). General theory and methods for matric factoring. *Psychometrika*, **9**, 1-16.

Guttman, L. (1953). Image theory for the structure of quantitative variables. *Psychometrika*, **9**, 277-296.

Jöreskog, K.G. (1970). A general method for analysis of covariance structures. *Biometrika*, **57**, 239-251.

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

Meredith, W., & Millsap, R.E. (1985). On component analysis. *Psychometrika*,**50**, 495-507.

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

Potthoff, R.F., & Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313-326.

Rao, C.R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyã A*, **26**, 329-358.

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

Rao, C.R. (1980). Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In P.R. Krishnaiah (Ed.), *Multivariate analysis V* (pp. 3-22). Amsterdum: North Holland.

Rao, C.R. & Yanai, H. (1979). General definition and decomposition of projectors and some applications to statistical problems. *Journal of Statistical Inference and Planning*, **3**, 1-17.

Takane, Y., Kiers, H.A.L., & de Leeuw, J. (1995). Component analysis with different sets of constraints on different dimensions. *Psychometrika*, **60**, 259-280.

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

Takane, Y., Yanai, H., & Mayekawa, S. (1991). Relationships among several methods of linearly constrained correspondence analysis. *Psychometrika*, **56**, 667-684.

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

Yanai, H. (1970). Factor analysis with external criteria. *Japanese Psychological Research*, **12**, 143-153.

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

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