PSYCHOMETRIKA—VOL. 56, NO. 1, 97–120
MARCH 1991

PRINCIPAL COMPONENT ANALYSIS WITH EXTERNAL INFORMATION ON BOTH SUBJECTS AND VARIABLES

YOSHIO TAKANE AND TADASHI SHIBAYAMA

MCGILL UNIVERSITY

A method for structural analysis of multivariate data is proposed that combines features of regression analysis and principal component analysis. In this method, the original data are first decomposed into several components according to external information. The components are then subjected to principal component analysis to explore structures within the components. It is shown that this requires the generalized singular value decomposition of a matrix with certain metric matrices. The numerical method based on the QR decomposition is described, which simplifies the computation considerably. The proposed method includes a number of interesting special cases, whose relations to existing methods are discussed. Examples are given to demonstrate practical uses of the method.

Key words: orthogonal projection operator, trace-orthogonality, generalized singular value decomposition (GSVD), QR decomposition, vector preference models, two-way CANDELINC, dual scaling, redundancy analysis, GMANOVA (growth curve models).

1. Introduction

Principal component analysis (PCA) is often used to explore structures in multivariate data. For example, a researcher may be interested in what attributes of stimuli (e.g., political candidates, commercial products, etc.) are important in determining preferences toward them. The researcher may collect preference judgments on a set of stimuli from a group of subjects, analyze how the preferences toward the different stimuli are related with each other, and find out what attributes of the stimuli are commonly preferred, or not preferred, by which subjects.

Simple PCA may not be the best method to apply, however, when additional information about variables and subjects is available. For example, in the preference judgment study, the stimuli may be presented in pairs to subjects, who are asked to indicate the degree to which they prefer one stimulus to the other. The fact that each judgment reflects a comparative process between two stimuli provides important structural information about the variable. Subjects' demographic information (e.g., sex, age, level of education, etc.) may also be available. The investigator may be interested in how the preference data are related to the subjects' demographic information. This type of information is particularly useful in identifying special subgroups of subjects who exhibit common patterns of individual differences in the preference judgments.

The external information can be used informally to aid subjective interpretations of analysis results. Alternatively, it can be directly incorporated in the formal analysis. The joint analysis of main data with auxiliary information can lead to more objective interpretations by enabling the assessment of how well structures supplied by the

The work reported in this paper was supported by grant A6394 from the Natural Sciences and Engineering Research Council of Canada to the first author. Thanks are due to Jim Ramsay, Haruo Yanai, Henk Kiers, and Shizuhiko Nishisato for their insightful comments on earlier versions of this paper. Jim Ramsay, in particular, suggested the use of the QR decomposition, which simplified the presentation of the paper considerably.

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

0033-3123/91/0300-89122 $00.75/0

external information can account for the data. This paper describes a method for a joint analysis of multivariate data with external information. In this method, the original data are first decomposed into several components (External Analysis), those that can be explained, and those that cannot be explained, by the external information. PCA is then applied to each component separately or to some of the components combined (Internal Analysis).

The proposed method combines regression analysis and PCA in a unified framework. Regression analysis decomposes the data according to known structures. Although regression analysis is effective when meaningful structures are known in advance, it obviously cannot be applied unless such structures are known. PCA, on the other hand, is useful when no obvious structures are known. The proposed method unifies the two methods, capturing advantages of both. It attempts to explain as much as possible of the data by known structures. At the same time it seeks to find unknown structures inside and outside the known structures.

The following section gives a detailed account of the proposed method. We first describe the basic model for the external analysis and the associated decomposition of the original data matrix (section 2.1). PCA of decomposed submatrices for the internal analysis is then discussed and some mathematical problems associated with it (section 2.2). In particular, it will be shown that the PCA of the decomposed submatrices involves the generalized singular value decomposition (GSVD) of a matrix with certain metric matrices. Some computational considerations are then given (section 2.3) that will simplify the computations. The proposed method subsumes a number of existing methods as special cases. In section 3, we discuss relations to several of these methods, including three vector models of preference (Takane & Shibayama, 1988a), two-way CANDELINC (Carroll, Pruzansky, & Kruskal, 1980), dual scaling (Nishisato, 1980a), and redundancy analysis (van den Wollenberg, 1977). In section 4, practical uses of the proposed method are illustrated using two empirical examples. In the final section, we consider some common problems in practical applications of PCA, such as stability assessment, missing data, and possible data transformations.

2. The Method

2.1. *External Analysis*

We denote an N-subject by n-variable data matrix by $\mathbf{Z}$. The data may consist of N subjects' preference ratings on n stimuli, profiles of N objects on n attributes, or any other multivariate observations. The data may be raw or preprocessed, for example, by standardizations or other transformations. The data may also contain dummy-coded discrete variables.

Assume there are an N by p ($\leq N$) subject information matrix, $\mathbf{G}$, and an n by q ($\leq n$) variable information matrix, $\mathbf{H}$. These matrices can take a variety of forms. For example, $\mathbf{G}$ may be an N-component vector of ones, a matrix of dummy variables, or a matrix of continuous variables characterizing the subjects. Similarly, $\mathbf{H}$ may be an n-component vector of ones, a design matrix for pair comparisons, or any other matrix of explanatory variables that capture relationships among columns of $\mathbf{Z}$. When no additional information is available on subjects or variables, we may simply set $\mathbf{G} = \mathbf{I}$ or $\mathbf{H} = \mathbf{I}$.

Consider the following model:

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

where $\mathbf{M}(p \times q)$, $\mathbf{B}(N \times q)$, and $\mathbf{C}(p \times n)$ are matrices of coefficients to be estimated, and $\mathbf{E}(N \times n)$ a matrix of error components. The four terms in (1) explain portions of the original data matrix, **Z**. The first term pertains to what can be explained by both **G** and **H**, the second term by **H**, the third term by **G**, and the fourth term by neither **G** nor **H**. There is redundancy in the model, however. Obviously, what can be explained by both **G** and **H** can also be explained by **H** alone or by **G** alone. Thus, the first term is completely subsumed under both the second term and the third term.

There are two possible strategies we may take to resolve the redundancy problem; simultaneous estimation and sequential estimation. In the former, identifiability constraints, $\mathbf{G'B} = \mathbf{0}$ and $\mathbf{CH} = \mathbf{0}$, are imposed. Least squares estimates of **M**, **B**, and **C** that minimize $SS(\mathbf{E}) = \text{tr}(\mathbf{E'E})$, are obtained simultaneously subject to these constraints. In the sequential estimation, we fit the first term first, as if there were no other terms in the model. We then fit the second and the third terms separately to the residual from the first term. What remains unexplained by the first three terms constitutes the fourth term in (1).

Although the two methods lead to identical results, we will follow the sequential estimation method more closely, since it is much easier to understand. Let $\mathbf{Z} = \mathbf{GMH'} + \mathbf{E}_1$, and consider the problem of estimating **M** so as to minimize $SS(\mathbf{E}_1) = \text{tr}(\mathbf{E}_1'\mathbf{E}_1)$. We obtain

$$\hat{\mathbf{M}} = (\mathbf{G'G})^{-}\mathbf{G'ZH}(\mathbf{H'H})^{-}, \tag{2}$$

where $(\mathbf{G'G})^{-}$ and $(\mathbf{H'H})^{-}$ are g-inverses of $\mathbf{G'G}$ and $\mathbf{H'H}$, respectively. The residual from the first term is now equal to

$$\hat{\mathbf{E}}_1 = \mathbf{Z} - \mathbf{G}\hat{\mathbf{M}}\mathbf{H'} = \mathbf{Z} - \mathbf{P}_G\mathbf{Z}\mathbf{P}_H,$$

where $\mathbf{P}_G = \mathbf{G}(\mathbf{G'G})^{-}\mathbf{G'}$ and $\mathbf{P}_H = \mathbf{H}(\mathbf{H'H})^{-}\mathbf{H'}$ are orthogonal projection operators (e.g., Yanai & Takeuchi, 1983) onto spaces spanned by the column vectors of **G** and **H**, respectively. It is well known that $\mathbf{P}_G$ and $\mathbf{P}_H$ are unique, even if $(\mathbf{G'G})^{-}$ and $(\mathbf{H'H})^{-}$ are nonunique. We now separately fit the second and the third terms to $\hat{\mathbf{E}}_1$.

$$\hat{\mathbf{E}}_1 = \mathbf{BH'} + \mathbf{E}_2;$$

$$\hat{\mathbf{E}}_1 = \mathbf{GC} + \mathbf{E}_3.$$

We obtain a least squares estimate of **B** that minimizes $SS(\mathbf{E}_2)$ by

$$\hat{\mathbf{B}} = \mathbf{Q}_G\mathbf{ZH}(\mathbf{H'H})^{-}, \tag{3}$$

where $\mathbf{Q}_G = \mathbf{I} - \mathbf{P}_G$. Similarly, we obtain

$$\hat{\mathbf{C}} = (\mathbf{G'G})^{-}\mathbf{G'Z}\mathbf{Q}_H, \tag{4}$$

that minimizes $SS(\mathbf{E}_3)$, where $\mathbf{Q}_H = \mathbf{I} - \mathbf{P}_H$. It is well known that $\mathbf{Q}_G$ and $\mathbf{Q}_H$ are both orthogonal projection operators that are orthogonal to $\mathbf{P}_G$ and $\mathbf{P}_H$, respectively. Now, the estimate of the fourth term is given by

$$\begin{aligned}\hat{\mathbf{E}} &= \mathbf{Z} - \mathbf{G}\hat{\mathbf{M}}\mathbf{H'} - \hat{\mathbf{B}}\mathbf{H'} - \mathbf{G}\hat{\mathbf{C}} \\ &= \mathbf{Z} - \mathbf{P}_G\mathbf{Z}\mathbf{P}_H - \mathbf{Q}_G\mathbf{Z}\mathbf{P}_H - \mathbf{P}_G\mathbf{Z}\mathbf{Q}_H \\ &= \mathbf{Q}_G\mathbf{Z}\mathbf{Q}_H.\end{aligned} \tag{5}$$

It may be pointed out that we could estimate **B** and **C** in a strictly sequential manner. That is, **GC** may be fitted to $\hat{\mathbf{E}}_2 = \hat{\mathbf{E}}_1 - \hat{\mathbf{B}}\mathbf{H'}$, where $\hat{\mathbf{B}}$ is given in (3). Alternatively, **BH'** may be fitted to $\hat{\mathbf{E}}_3 = \hat{\mathbf{E}}_1 - \mathbf{G}\hat{\mathbf{C}}$ after **GC** is fitted to $\hat{\mathbf{E}}_1$. These sequential estimations provide identical results as above. However, the second and the third terms cannot be fitted to $\hat{\mathbf{E}}_1$ simultaneously. Minimizing $SS(\hat{\mathbf{E}}_1 - \mathbf{BH}' - \mathbf{GC})$ with respect to both **B** and **C** will not lead to the same estimates of **B** and **C**, unless the same restrictions that are used in the simultaneous estimation (i.e., $\mathbf{G}'\mathbf{B} = \mathbf{0}$ and $\mathbf{CH} = \mathbf{0}$) are explicitly imposed. Note that in this estimation scheme, both column and row metrics are assumed to be identity matrices. Some generalizations will be given in Appendix section B.

By substituting the least squares estimates for the corresponding parameters in (1), we obtain the following decomposition of the data matrix, **Z**:

$$\mathbf{Z} = (\mathbf{P}_G + \mathbf{Q}_G)\mathbf{Z}(\mathbf{P}_H + \mathbf{Q}_H)$$

$$= \mathbf{P}_G\mathbf{Z}\mathbf{P}_H + \mathbf{Q}_G\mathbf{Z}\mathbf{P}_H + \mathbf{P}_G\mathbf{Z}\mathbf{Q}_H + \mathbf{Q}_G\mathbf{Z}\mathbf{Q}_H. \tag{6}$$

The four terms in (6) are the estimates of the corresponding four terms in (1). The decomposition is now unique, and a specific meaning can be attached to each term. Since the effect due to the first term was eliminated before the second term was fitted, the second term in (6) represents the portion of **Z** that can be accounted for by **H**, but not by **G**. The third term in (6) can be interpreted in a similar manner. Note that some of the terms in (6) may be zero. This is indeed the case when **G** or **H** is a square matrix of full rank (e.g., $\mathbf{G} = \mathbf{I}$ or $\mathbf{H} = \mathbf{I}$).

Not all the four terms in (6) are columnwise orthogonal, or rowwise orthogonal. They are, however, orthogonal in the following sense, which may be termed as trace-orthogonal. Two matrices of a same size, **X** and **Y**, are said to be trace-orthogonal when $\text{tr}\,(\mathbf{X}'\mathbf{Y}) = \text{tr}\,(\mathbf{XY}') = 0$. For example, the first and the third terms in (6) are trace-orthogonal, since $\text{tr}\,(\mathbf{P}_H\mathbf{Z}'\mathbf{P}_G\mathbf{P}_G\mathbf{Z}\mathbf{Q}_H) = \text{tr}\,(\mathbf{Q}_H\mathbf{P}_H\mathbf{Z}'\mathbf{P}_G\mathbf{Z}) = \text{tr}\,(\mathbf{0}) = 0$. Every pair of terms in (6) are either columnwise orthogonal, rowwise orthogonal or both, which in general implies the trace-orthogonality. The fact that the four terms in (6) are trace-orthogonal implies

$$SS(\mathbf{Z}) = SS(\mathbf{P}_G\mathbf{Z}\mathbf{P}_H) + SS(\mathbf{Q}_G\mathbf{Z}\mathbf{P}_H) + SS(\mathbf{P}_G\mathbf{Z}\mathbf{Q}_H) + SS(\mathbf{Q}_G\mathbf{Z}\mathbf{Q}_H), \tag{7}$$

where $SS(\mathbf{Z}) = \text{tr}\,(\mathbf{Z}'\mathbf{Z})$. That is, the sum of squares of elements in **Z** is decomposed into the sum of sums of squares corresponding to the four terms in (6).

The decomposition presented in (6) is a very basic one, and $\mathbf{P}_G$ and/or $\mathbf{P}_H$ may be subjected to further decompositions. For example, in a balanced two-way ANOVA design, $\mathbf{P}_G$ can be uniquely decomposed into the sum of $\mathbf{P}_N$ (the orthogonal projection operator pertaining to the grand mean), $\mathbf{P}_A$ (the main effect of Factor A), $\mathbf{P}_B$ (the main effect of Factor B), and $\mathbf{P}_{AB}$ (the interaction between A and B). In an unbalanced design, $\mathbf{P}_A$ and $\mathbf{P}_B$ are not mutually orthogonal. This reflects the fact that what can be explained by Factor A and by Factor B are not mutually exclusive. There are two possible decompositions of $\mathbf{P}_G$ in this case: $\mathbf{P}_G = \mathbf{P}_N + \mathbf{P}_A + \mathbf{P}_{B-A} + \mathbf{P}_{AB}$, and $\mathbf{P}_G = \mathbf{P}_N + \mathbf{P}_B + \mathbf{P}_{A-B} + \mathbf{P}_{AB}$, where $\mathbf{P}_{A-B}$ and $\mathbf{P}_{B-A}$ refer to the effect of A eliminating the effect of B and the effect of B eliminating the effect of A, respectively, while $\mathbf{P}_A$ and $\mathbf{P}_B$ refer to the effect of A and the effect of B, respectively, ignoring the effect of the other. (See section 3.3 for related material.)

An interesting situation is when **G** and/or **H** are products of two or more matrices. For example, stimuli in pair comparison judgments may be constructed by factorial combinations of basic factors. Then **H** is a product of the design matrix for pair comparisons, **A**, and the design matrix for the stimuli, **S**; that is, $\mathbf{H} = \mathbf{AS}$. In this case, we may first obtain $\mathbf{P}_A$ and $\mathbf{Q}_A = \mathbf{I} - \mathbf{P}_A$, and split $\mathbf{P}_A$ further into $\mathbf{P}_{AS} = \mathbf{AS}(\mathbf{S}'\mathbf{A}'\mathbf{AS})^-\mathbf{S}'\mathbf{A}'$ and $\mathbf{P}_A - \mathbf{P}_{AS}$. The orthogonality among $\mathbf{P}_{AS}$, $\mathbf{P}_A - \mathbf{P}_{AS}$, and $\mathbf{Q}_A$ can be easily verified.

With further decompositions of $\mathbf{P}_G$ and $\mathbf{P}_H$, (6) may be more generally written as

$$\mathbf{Z} = \left(\sum_i \mathbf{P}_{Gi}\right)\mathbf{Z}\left(\sum_j \mathbf{P}_{Hj}\right), \tag{8}$$

where $\Sigma_i \mathbf{P}_{Gi} = \mathbf{I}$ and $\Sigma_j \mathbf{P}_{Hj} = \mathbf{I}$. This expression is due to Nishisato and Lawrence (1989).

2.2. *Internal Analysis*

Once the data matrix is decomposed according to the external information, principal component analysis (PCA) may be applied to each component separately. For example, the first term in (6), $\mathbf{P}_G\mathbf{Z}\mathbf{P}_H$, may be subjected to PCA. In certain cases, however, some of the decomposed submatrices may be recombined for PCA. For example, the first and the second terms in (6) may be combined, amounting to PCA of $\mathbf{Z}\mathbf{P}_H$. Since the components analyzed are associated with specific meanings of their own, PCA of the components may be more readily interpretable than direct PCA of the original data matrix.

PCA extracts the most important dimensions of the components to be analyzed. It provides the best fixed-rank approximation of the matrix, in the least squares sense. Rao (1979, 1980) has shown, however, that its optimality is much more general; PCA provides the best fixed-rank approximation under any orthogonally invariant norm. (A matrix norm $\|\cdot\|$ is called an orthogonally invariant norm if $\|\mathbf{X}\| = \|\mathbf{AXB}'\|$ for any orthogonal matrices, $\mathbf{A}$ and $\mathbf{B}$.)

An interesting relationship exists between singular values of the original data matrix and those of the decomposed submatrices. Let $\mathbf{Z}_0$ denote the sum of any subset of the terms in (6). Then, by a standard separation theorem for singular values (Yanai & Takeuchi, 1983) reproduced in Appendix section A, $s_j(\mathbf{Z}) \geq s_j(\mathbf{Z}_0)$, where $s_j(\mathbf{Z})$ and $s_j(\mathbf{Z}_0)$ are the j-th largest singular values of matrices $\mathbf{Z}$ and $\mathbf{Z}_0$, respectively. That is, the singular values of the submatrices are never larger than the corresponding singular values of the original data matrix.

Computationally PCA amounts to the singular value decomposition (SVD) of a rectangular matrix. The ordinary SVD can be used if the PCA is to be applied to the entire submatrices in decomposition (6). In some cases, however, only a portion of the decomposed submatrices may be meaningfully analyzed. For example, in $\mathbf{P}_G\mathbf{Z}\mathbf{P}_H = \mathbf{G}\hat{\mathbf{M}}\mathbf{H}$ (the first term in (6)), only $\hat{\mathbf{M}}$ may be subjected to PCA. (For a concrete example, see section 3.1.) In this case, the ordinary SVD does not apply. Instead, the generalized SVD (GSVD) is required.

The problem posed here is that of finding $\mathbf{M}^*$ which minimizes $\psi = SS(\mathbf{G}(\hat{\mathbf{M}} - \mathbf{M}^*)\mathbf{H}')$. Although the ordinary SVD of $\hat{\mathbf{M}}$ will find $\mathbf{M}^*$ that minimizes $SS(\hat{\mathbf{M}} - \mathbf{M}^*)$, the same $\mathbf{M}^*$ does not minimize ψ unless both $\mathbf{G}$ and $\mathbf{H}$ are columnwise orthonormal. Finding $\mathbf{M}^*$ which minimizes ψ, in general requires the GSVD of $\hat{\mathbf{M}}$ with $\mathbf{G}'\mathbf{G}$ and $\mathbf{H}'\mathbf{H}$ as column and row metrics, respectively.

Let us define the GSVD and show how it solves the minimization problem. For simplicity, we temporarily assume that both $\mathbf{G}'\mathbf{G}$ and $\mathbf{H}'\mathbf{H}$ are nonsingular. We later extend our results to singular metric matrices. In the following derivation, we do not use any special features of $\hat{\mathbf{M}}$ (i.e., $\hat{\mathbf{M}} = (\mathbf{G}'\mathbf{G})^-\mathbf{G}'\mathbf{Z}\mathbf{H}(\mathbf{H}'\mathbf{H})^-$). The special form of $\hat{\mathbf{M}}$ simplifies the computation, but this will not be discussed until the next section. In the remaining part of this section, we use the symbol $\mathbf{M}$ for $\hat{\mathbf{M}}$ to indicate the independence of our argument from the special features of $\hat{\mathbf{M}}$.

A decomposition of a p by $q(\leq p)$ matrix $\mathbf{M}$ into a product of three matrices (i.e., $\mathbf{M} = \mathbf{U}_M\mathbf{D}_M\mathbf{V}_M'$), is called the generalized singular value decomposition (GSVD) of $\mathbf{M}$ with respect to the column metric matrix, $\mathbf{G}'\mathbf{G}(p \times p)$ and the row metric matrix,

$\mathbf{H'H}(q \times q)$, if (a) $\mathbf{U}_M$ is $p \times q$ and $\mathbf{U}_M'\mathbf{G'GU}_M = \mathbf{I}_q$, (b) $\mathbf{V}_M$ is $q \times q$ and $\mathbf{V}_M'\mathbf{H'HV}_M = \mathbf{I}_q$, and (c) $\mathbf{D}_M$ is diagonal $(q \times q)$ and nonnegative definite (Greenacre & Underhill, 1982; Ramsay, ten Berge, & Styan, 1984). Diagonal elements in $\mathbf{D}_M$ are usually arranged in a descending order of magnitude. When $p < q$, this definition applies to $\mathbf{M}'$. The corresponding representation for $\mathbf{M}$ is obtained by transposition.

The GSVD of a matrix always exists. It is unique (up to permutation and reflection), if the diagonal elements in $\mathbf{D}_M$ are all distinct. Let $\mathbf{G'G} = \mathbf{R}_G\mathbf{R}_G'$ and $\mathbf{H'H} = \mathbf{R}_H\mathbf{R}_H'$ be square root decompositions of $\mathbf{G'G}$ and $\mathbf{H'H}$. Any square root decompositions may be used, but it is convenient to use the Cholesky decomposition, or triangularization by the Householder transformation. These methods theoretically lead to the same $\mathbf{R}_G'$ and $\mathbf{R}_H'$ that are upper triangular. Let $\mathbf{U}_J\mathbf{D}_J\mathbf{V}_J'$ be the ordinary SVD of $\mathbf{R}_G'\mathbf{MR}_H \equiv \mathbf{J}$, which is uniquely determined if all the singular values in $\mathbf{D}_J$ are distinct. We then obtain $\mathbf{U}_M$, $\mathbf{V}_M$, and $\mathbf{D}_M$ by $\mathbf{U}_M = (\mathbf{R}_G')^{-1}\mathbf{U}_J$, $\mathbf{V}_M = (\mathbf{R}_H')^{-1}\mathbf{V}_J$, and $\mathbf{D}_M = \mathbf{D}_J$ with the required properties.

How does the GSVD of $\mathbf{M}$ with metrics $\mathbf{G'G}$ and $\mathbf{H'H}$ relate to the minimization of ψ? We may rewrite

$$\psi = SS(\mathbf{G}(\mathbf{M} - \mathbf{M}^*)\mathbf{H}')$$
$$= SS(\mathbf{R}_G'(\mathbf{M} - \mathbf{M}^*)\mathbf{R}_H). \tag{9}$$

Let $\mathbf{U}_J\mathbf{D}_J\mathbf{V}_J'$ be the usual SVD of $\mathbf{R}_G'\mathbf{MR}_H = \mathbf{J}$. We then have $\mathbf{M} = (\mathbf{R}_G')^{-1}\mathbf{U}_J\mathbf{D}_J\mathbf{V}_J'\mathbf{R}_H^{-1} = \mathbf{U}_M\mathbf{D}_M\mathbf{V}_M'$, which is the GSVD of $\mathbf{M}$ with metrics $\mathbf{G'G}$ and $\mathbf{H'H}$. Let $\mathbf{U}_J^*\mathbf{D}_J^*\mathbf{V}_J^{*\prime}$ be the best rank-r approximation of $\mathbf{J} = \mathbf{R}_G'\mathbf{MR}_H$, obtained from $\mathbf{U}_J\mathbf{D}_J\mathbf{V}_J'$ by discarding the last $\min(p - r, q - r)$ columns of $\mathbf{U}_J$ and $\mathbf{V}_J$, and the same number of rows and columns of $\mathbf{D}_J$. However, $\mathbf{U}_J^*\mathbf{D}_J^*\mathbf{V}_J^{*\prime}$ has to be equal to $\mathbf{R}_G'\mathbf{M}^*\mathbf{R}_H$, and therefore, $\mathbf{M}^* = (\mathbf{R}_G')^{-1}\mathbf{U}_J^*\mathbf{D}_J^*\mathbf{V}_J^*\mathbf{R}_H^{-1} = \mathbf{U}_M^*\mathbf{D}_M^*\mathbf{V}_M^{*\prime}$. The $\mathbf{U}_M^*$, $\mathbf{D}_M^*$, and $\mathbf{V}_M^{*\prime}$ are obtained by discarding appropriate portions of $\mathbf{U}_M$, $\mathbf{D}_M$, and $\mathbf{V}_M$, respectively, which are obtained by the GSVD of $\mathbf{M}$ with metrics $\mathbf{G'G}$ and $\mathbf{H'H}$. The $\mathbf{M}^* = \mathbf{U}_M^*\mathbf{D}_M^*\mathbf{V}_M^{*\prime}$ gives the best rank-r approximation of $\mathbf{M}$ in the metrics of $\mathbf{G'G}$ and $\mathbf{H'H}$.

The necessity of the GSVD for the minimization of ψ may be intuitively understood as follows. Typically, $\mathbf{M}$ is not scale invariant, since the elements of $\mathbf{M}$ depend on diag $(\mathbf{G'G})$ and diag $(\mathbf{H'H})$. Thus, they are not directly comparable unless the columns of $\mathbf{G}$ and those of $\mathbf{H}$ have comparable scales. The metric matrices have the effect of recovering the comparability. (This situation is analogous to multiple regression in which regression coefficients, elements of $\mathbf{b}$, are usually not comparable, while the elements of the prediction vector, $\mathbf{y}^* = \mathbf{Xb}$, are. In this case, $\mathbf{X'X}$ serves as the metric for $\mathbf{b}$.)

When $\mathbf{G'G}$ and/or $\mathbf{H'H}$ are singular, we may replace the regular inverses of $\mathbf{R}_G'$ and $\mathbf{R}_H$ by their Moore-Penrose inverses (de Leeuw, 1984). The $\mathbf{R}_G'$ and $\mathbf{R}_H'$ are now upper trapezoidal (incomplete triangular) rather than complete triangular. The Moore-Penrose inverses of $\mathbf{R}_G'$ and $\mathbf{R}_H$ are given by

$$(\mathbf{R}_G')^+ = (\mathbf{R}_G\mathbf{R}_G')^+\mathbf{R}_G = \mathbf{R}_G(\mathbf{R}_G'\mathbf{R}_G)^{-1}, \tag{10}$$

and

$$\mathbf{R}_H^+ = \mathbf{R}_H'(\mathbf{R}_H\mathbf{R}_H')^+ = (\mathbf{R}_H'\mathbf{R}_H)^{-1}\mathbf{R}_H'.$$

The second equality in (10) follows from

$$(\mathbf{R}_G\mathbf{R}_G')^+ = \mathbf{R}_G(\mathbf{R}_G'\mathbf{R}_G)^{-2}\mathbf{R}_G'. \tag{11}$$

(The same for $(\mathbf{R}_H\mathbf{R}_H')^+$.) Ramsay (1980, Appendix) describes an elaborate method to calculate $(\mathbf{R}_G\mathbf{R}_G')^+$.

The use of the Moore-Penrose inverses maintains the uniqueness of the GSVD. However, with singular metric matrices, **M** may not be fully recovered by the GSVD. Instead, we have

$$\mathbf{P}_{R(G)}\mathbf{M}\mathbf{P}_{R(H)} = \mathbf{U}_M\mathbf{D}_M\mathbf{V}'_M = (\mathbf{R}'_G)^+\mathbf{U}_J\mathbf{D}_J\mathbf{V}'_J\mathbf{R}_H^+,$$

where $\mathbf{P}_{R(G)} = \mathbf{R}_G(\mathbf{R}'_G\mathbf{R}_G)^{-1}\mathbf{R}'_G$ and $\mathbf{P}_{R(H)} = \mathbf{R}_H(\mathbf{R}'_H\mathbf{R}_H)^{-1}\mathbf{R}'_H$ are orthogonal projection operators onto spaces spanned by column vectors of $\mathbf{R}_G$ and $\mathbf{R}_H$, respectively. (If $\mathbf{G}'\mathbf{G}$ and $\mathbf{H}'\mathbf{H}$ are nonsingular, $\mathbf{P}_{R(G)} = \mathbf{I}$ and $\mathbf{P}_{R(H)} = \mathbf{I}$.) Note that $\mathbf{P}_{R(G)}\mathbf{M}\mathbf{P}_{R(H)} = \mathbf{M}$, if and only if **M** is in the column spaces of $\mathbf{R}_G$ and $\mathbf{R}_H$. This condition is equivalent to rank ($\mathbf{GMH}'$) = rank (**M**). We have $\mathbf{P}_{R(G)} = \mathbf{G}'(\mathbf{GG}')^-\mathbf{G} = \mathbf{P}_{G'}$ and $\mathbf{P}_{R(H)} = \mathbf{H}'(\mathbf{HH}')^-\mathbf{H} = \mathbf{P}_{H'}$. Thus, **M** is projected onto spaces spanned by row vectors of **G** and **H**. For an arbitrary **M**, only the projected **M** can be recovered by the GSVD.

2.3. Some Computational Considerations

The preceding sections presented the basic analytic tools (orthogonal projection and GSVD) for the proposed method. At least equally important are considerations that should be taken into account for efficient computations. Both the external and the internal analyses can be facilitated by the special form of **M**.

Let $\mathbf{G} = \mathbf{F}_G\mathbf{R}'_G$ and $\mathbf{H} = \mathbf{F}_H\mathbf{R}'_H$ such that $\mathbf{F}'_G\mathbf{F}_G = \mathbf{I}_s$ and $\mathbf{F}'_H\mathbf{F}_H = \mathbf{I}_t$, where s = rank (**G**) and t = rank (**H**). In principle, any decompositions of the above form may be used (e.g., SVD, the Gram-Schmidt orthogonalization, or the QR (or QL) decomposition); however, the QR decomposition using the Householder transformation (e.g., Wilkinson, 1965) seems to be the best choice since it is computationally most efficient and numerically most stable. The $\mathbf{R}'_G$ and $\mathbf{R}'_H$ are upper triangular, if **G** and **H** are nonsingular, although they reduce to upper trapezoidal forms if $s < p$ and $t < q$. Since $\mathbf{G}'\mathbf{G} = \mathbf{R}_G\mathbf{R}'_G$ and $\mathbf{H}'\mathbf{H} = \mathbf{R}_H\mathbf{R}'_H$, $\mathbf{R}'_G$ and $\mathbf{R}'_H$ are the same square root factors of $\mathbf{G}'\mathbf{G}$ and $\mathbf{H}'\mathbf{H}$ introduced in the previous section. Note that

$$\mathbf{R}'_G(\mathbf{R}_G\mathbf{R}'_G)^-\mathbf{R}_G = \mathbf{R}'_G(\mathbf{R}_G\mathbf{R}'_G)^+\mathbf{R}_G = \mathbf{I}_s. \tag{12}$$

This follows from (11). A similar relationship holds for $\mathbf{R}_H$.

Using the QR decompositions of **G** and **H** with (11) and (12), we can simplify various quantities defined in the previous sections. First, we obtain

$$\hat{\mathbf{M}} = (\mathbf{R}_G\mathbf{R}'_G)^-\mathbf{R}_G\mathbf{F}'_G\mathbf{Z}\mathbf{F}_H\mathbf{R}'_H(\mathbf{R}_H\mathbf{R}'_H)^-, \tag{13}$$

and

$$\mathbf{J} = \mathbf{R}'_G\hat{\mathbf{M}}\mathbf{R}_H = \mathbf{F}'_G\mathbf{Z}\mathbf{F}_H. \tag{14}$$

As before, the GSVD of $\hat{\mathbf{M}}$ with metrics $\mathbf{G}'\mathbf{G} = \mathbf{R}_G\mathbf{R}'_G$ and $\mathbf{H}'\mathbf{H} = \mathbf{R}_H\mathbf{R}'_H$ can be derived from the regular SVD of **J**. The important point is that using (14), **J** can be calculated without explicitly calculating $\hat{\mathbf{M}}$.

Furthermore, $\mathbf{P}_G$ can be re-expressed as

$$\mathbf{P}_G = \mathbf{F}_G\mathbf{R}'_G(\mathbf{R}_G\mathbf{R}'_G)^-\mathbf{R}_G\mathbf{F}'_G$$
$$= \mathbf{F}_G\mathbf{F}'_G,$$

so that

$$\mathbf{K} = \mathbf{P}_G\mathbf{Z}\mathbf{P}_H = \mathbf{F}_G\mathbf{F}'_G\mathbf{Z}\mathbf{F}_H\mathbf{F}'_H$$
$$= \mathbf{F}_G\mathbf{J}\mathbf{F}'_H,$$

and

$$\mathbf{J} = \mathbf{F}_G'\mathbf{K}\mathbf{F}_H.$$

Let $\mathbf{K} = \mathbf{U}_K\mathbf{D}_K\mathbf{V}_K'$ be the regular SVD of $\mathbf{K}$ and let $\mathbf{J} = \mathbf{U}_J\mathbf{D}_J\mathbf{V}_J'$ be the GSVD of $\mathbf{J}$ with metrics $\mathbf{F}_G'\mathbf{F}_G = \mathbf{I}$ and $\mathbf{F}_H'\mathbf{F}_H = \mathbf{I}$, which is nothing but the regular SVD of $\mathbf{J}$. Then, $\mathbf{U}_K = \mathbf{F}_G\mathbf{U}_J$ (or $\mathbf{U}_J = \mathbf{F}_G'\mathbf{U}_K$), $\mathbf{V}_K = \mathbf{F}_H\mathbf{V}_J$ (or $\mathbf{V}_J = \mathbf{F}_H'\mathbf{V}_K$) and $\mathbf{D}_K = \mathbf{D}_J$. Note also, $\mathbf{U}_K = \mathbf{P}_G\mathbf{U}_K$ and $\mathbf{V}_K = \mathbf{P}_H\mathbf{V}_K$. This relationship provides an efficient way of obtaining the SVD of $\mathbf{K}$, and indicates the relationship between the SVD of a whole, $\mathbf{K} = \mathbf{P}_G\mathbf{Z}\mathbf{P}_H = \mathbf{G}\hat{\mathbf{M}}\mathbf{H}'$, and the GSVD of its part, $\hat{\mathbf{M}}$, with metrics $\mathbf{G}'\mathbf{G}$ and $\mathbf{H}'\mathbf{H}$. That is, $\mathbf{U}_K = \mathbf{F}_G\mathbf{R}_G'\mathbf{U}_M = \mathbf{G}\mathbf{U}_M$, $\mathbf{V}_K = \mathbf{F}_H\mathbf{R}_H'\mathbf{V}_M = \mathbf{H}\mathbf{V}_M$, and $\mathbf{D}_K = \mathbf{D}_M$, or $\mathbf{U}_M = (\mathbf{R}_G')^+\mathbf{F}_G'\mathbf{U}_K$, $\mathbf{V}_M = (\mathbf{R}_H')^+\mathbf{F}_H'\mathbf{V}_K$, and $\mathbf{D}_M = \mathbf{D}_K$.

When $\mathbf{G}'\mathbf{G}$ and/or $\mathbf{H}'\mathbf{H}$ are singular, $\hat{\mathbf{M}}$ is in general not unique. If the Moore-Penrose inverses of $\mathbf{G}'\mathbf{G}$ and $\mathbf{H}'\mathbf{H}$ are used to calculate $\hat{\mathbf{M}}$, $\hat{\mathbf{M}}$ will be unique. At the same time, $\mathbf{P}_{R(G)}\hat{\mathbf{M}}\mathbf{P}_{R(H)} = \hat{\mathbf{M}}$, and rank $(\mathbf{G}\hat{\mathbf{M}}\mathbf{H}')$ = rank $(\hat{\mathbf{M}})$ so that $\hat{\mathbf{M}}$ can be fully recovered by the GSVD with metrics $\mathbf{G}'\mathbf{G}$ and $\mathbf{H}'\mathbf{H}$. If, on the other hand, other g-inverses of $\mathbf{G}'\mathbf{G}$ and $\mathbf{H}'\mathbf{H}$ are used to define $\hat{\mathbf{M}}$, $\hat{\mathbf{M}}$ is no longer unique. Its GSVD with metrics $\mathbf{G}'\mathbf{G}$ and $\mathbf{H}'\mathbf{H}$ is also nonunique. However, since

$$(\mathbf{R}_G')^-\mathbf{R}_G'\hat{\mathbf{M}}\mathbf{R}_H\mathbf{R}_H^- = (\mathbf{R}_G\mathbf{R}_G')^-\mathbf{R}_G\mathbf{R}_G'(\mathbf{R}_G\mathbf{R}_G')^-\mathbf{R}_G\mathbf{J}\mathbf{R}_H'(\mathbf{R}_H\mathbf{R}_H')^-\mathbf{R}_H\mathbf{R}_H'(\mathbf{R}_H\mathbf{R}_H')^-$$

$$= (\mathbf{R}_G\mathbf{R}_G')^-\mathbf{R}_G\mathbf{J}\mathbf{R}_H'(\mathbf{R}_H\mathbf{R}_H')^- = \hat{\mathbf{M}},$$

$\hat{\mathbf{M}}$ is fully recovered by the GSVD, $(\mathbf{R}_G')^-\mathbf{U}_J\mathbf{D}_J\mathbf{V}_J'\mathbf{R}_H^-$, provided that the same g-inverses of $\mathbf{G}'\mathbf{G} = \mathbf{R}_G\mathbf{R}_G'$ and $\mathbf{H}'\mathbf{H} = \mathbf{R}_H\mathbf{R}_H'$ are used in calculating both $\hat{\mathbf{M}}$ and $(\mathbf{R}_G')^- = (\mathbf{R}_G\mathbf{R}_G')^-\mathbf{R}_G$ and $\mathbf{R}_H^- = \mathbf{R}_H'(\mathbf{R}_H\mathbf{R}_H')^-$.

3. Relations to Other Methods

In the previous sections we presented a method of decomposing the data matrix according to the external information and the internal criterion. This method is quite general and subsumes a number of existing methods as special cases. In this section we first briefly review the literature on related work, and then discuss some of the interesting special cases in some detail.

Special cases of decomposition (6) have been worked out by many authors. Corsten and Van Eijnsbergen (1972; see also Corsten, 1976), Gabriel (1978), and Rao (1980) proposed a decomposition in which either the first and the second terms or the first and the third terms in (6) were not separated. That is, $\mathbf{Z} = \mathbf{Z}\mathbf{P}_H + \mathbf{P}_G\mathbf{Z}\mathbf{Q}_H + \mathbf{Q}_G\mathbf{Z}\mathbf{Q}_H$, $\mathbf{Z} = \mathbf{P}_G\mathbf{Z} + \mathbf{Q}_G\mathbf{Z}\mathbf{P}_H + \mathbf{Q}_G\mathbf{Z}\mathbf{Q}_H$ or $\mathbf{Z} = \mathbf{Z}\mathbf{P}_H + \mathbf{P}_G\mathbf{Z} - \mathbf{P}_G\mathbf{Z}\mathbf{P}_H + \mathbf{Q}_G\mathbf{Z}\mathbf{Q}_H$. Common to all these decompositions is the last term, $\mathbf{Q}_G\mathbf{Z}\mathbf{Q}_H$, which is minimal under any orthogonally invariant norm (Rao, 1980), which includes the euclidean norm (SS) as a special case. Gollob (1968) in his FANOVA model used $\mathbf{G} = \mathbf{1}_N$ (an N-component vector of ones) and $\mathbf{H} = \mathbf{1}_n$ (an n-component vector of ones). Yanai (1970) proposed PCA with external criteria with $\mathbf{G}$ taken to be a matrix of dummy variables indicating categories (e.g., male or female) of subjects. In the context of PCA of functional data, Besse and Ramsay (1986) proposed to "filter out" the effect of $\mathbf{H}$, taken to be a matrix of regular seasonal variations in average monthly temperatures of various French cities (also see Winsberg, 1988).

In all the above proposals, PCA is applied to the residual term (the last term in (6)). Okamoto (1972) also set $\mathbf{G} = \mathbf{1}_N$ and $\mathbf{H} = \mathbf{1}_n$ (as in Gollob) and proposed four techniques of PCA, PCA of $\mathbf{Z}$, $\mathbf{Q}_G\mathbf{Z}$, $\mathbf{Z}\mathbf{Q}_H$, and $\mathbf{Q}_G\mathbf{Z}\mathbf{Q}_H$. Recently, Takane and Shibayama (1988a) showed that three representative vector models for pairwise preference data were special cases of (6) with PCA applied to the second term or the first and the second terms combined (see section 3.1). GMANOVA or the growth curve models (Khatri, 1966; Potthoff & Roy, 1964; Rao, 1965) and two-way CANDELINC (Carroll, et al., 1980) analyze the first term in (6). (See Appendix Section C for GMANOVA and section 3.2 for two-way CANDELINC.) Nishisato (1980a, 1988; Nishisato & Lawrence, 1981, 1989) also proposed a similar approach to "ANOVA" of multiple-choice data in the framework of dual scaling (section 3.3). Also, see Escoufier and Holmes (1988), and Sabatier, Lebreton, and Chessel (1989). Böckenholt and Bökenholt (1990) proposed linearly constrained correspondence analysis that eliminates certain effects in the analysis of contingency tables (also see Takane, Yanai, & Mayekawa, in press; van der Heijden, de Falguerolles, & de Leeuw, 1989). This method amounts to PCA of the fourth term in our decomposition, (6), except for the use of special metric matrices peculiar to correspondence analysis (see Appendix Section C).

3.1. *Three Vector Models for Pairwise Preference Data*

Vector models are often used to represent individual differences in preference (Bechtel, 1976; Carroll, 1972; Slater, 1960; Tucker, 1959). In these models, stimuli are represented as points in a multidimensional space and subjects as vectors emanating from the origin in various directions. Preferences of the individual subjects are supposedly obtained by projections of the stimulus points onto the subject vectors. The relative length of a subject vector indicates how well the individual's preference is represented in the space. Similar models have been used for pairwise preference judgments described earlier. In this case, the differences in the projections of two stimuli onto the subject vectors are directly related to the observed judgments.

Suppose that N subjects make all possible pairwise judgments of m stimuli. Denote the N by $n = m(m-1)/2$ data matrix by $\mathbf{Z}$. Bechtel, Tucker, and Chang (1971) proposed

$$\mathbf{Z} = \mathbf{Y}^*\mathbf{X}'\mathbf{A}' + \mathbf{1}_N\mathbf{c}' + \mathbf{E}, \qquad \text{(BTC model)} \tag{15}$$

where $\mathbf{A}$ is an n by m design matrix for pair comparisons, $\mathbf{1}_N$ the N-component vector of ones, $\mathbf{Y}^*$ the N by r matrix of subject vectors, $\mathbf{X}$ the m by r matrix of stimulus coordinates, $\mathbf{c}$ the n-component vector of pairwise unscalability, and $\mathbf{E}$ the matrix of error components. Let $\hat{\mathbf{M}} = \mathbf{ZA}(\mathbf{A}'\mathbf{A})^+ = \mathbf{ZA}/m$, where $(\mathbf{A}'\mathbf{A})^+$ is the Moore-Penrose inverse of $\mathbf{A}'\mathbf{A}$, and $\mathbf{Q}_A = \mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^+\mathbf{A}'$. Bechtel, Tucker, and Chang split $\mathbf{E}$ into two parts:

$$\mathbf{E} = (\hat{\mathbf{M}} - \mathbf{Y}^*\mathbf{X}')\mathbf{A}' + (\mathbf{ZQ}_A - \mathbf{1}_N\mathbf{c}'), \tag{16}$$

and obtain $\mathbf{Y}^*$ and $\mathbf{X}$ that minimize the SS of the first term, and $\mathbf{c}$ that minimizes the SS of the second term. Such a $\mathbf{Y}^*$ and an $\mathbf{X}$ is obtained by the SVD of $\hat{\mathbf{M}}$. (The reason for the ordinary SVD and not the GSVD, in this particular instance, will be explained later.) The estimate of $\mathbf{c}'$ is obtained by $\hat{\mathbf{c}}' = \mathbf{1}_N'\mathbf{ZQ}_A/N$.

Two similar models have since been proposed. One is called the THL model (Heiser & de Leeuw, 1981; Takane, 1980, 1987), and the other the wandering vector model (WVM; De Soete & Carroll, 1983). Although both of these models were originally intended to capture stochastic components in the data, they can be easily translated into the form analogous to (15):

$$\mathbf{Z} = (\mathbf{1}_N\mathbf{m}' + \mathbf{YX}')\mathbf{A}' + \mathbf{E}, \qquad \text{(THL model)} \tag{17}$$

and

$$\mathbf{Z} = (\mathbf{1}_N\mathbf{v}' + \mathbf{Y})\mathbf{X}'\mathbf{A}' + \mathbf{E}, \qquad \text{(WVM model)} \tag{18}$$

where $\mathbf{m}$ is the m-component mean preference vector, and $\mathbf{v}$ the r-component mean subject vector. Notice that (18) can be obtained from (17) by setting $\mathbf{m} = \mathbf{X}\mathbf{v}$. On the assumption that various terms in (17) and (18) are accounted for in a specific order, least squares (LS) estimates of parameters are given by: For the THL model: $\hat{\mathbf{m}}' = \mathbf{1}_N'\hat{\mathbf{M}}/N$, where $\hat{\mathbf{M}} = \mathbf{Z}\mathbf{A}/m$ as before (m is the number of stimuli), and $\mathbf{Y}$ and $\mathbf{X}$ are obtained by the SVD of $\mathbf{Q}_N\hat{\mathbf{M}}$, where $\mathbf{Q}_N = \mathbf{I} - \mathbf{1}_N\mathbf{1}_N'/N$. For the WVM: $\hat{\mathbf{v}}' = \mathbf{1}_N'\hat{\mathbf{M}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}/N = \mathbf{1}_N'\mathbf{Y}^*/N$, and $\mathbf{Y} = \mathbf{Q}_N\mathbf{Y}^*$, where $\mathbf{Y}^*$ and $\mathbf{X}$ are obtained by the SVD of $\hat{\mathbf{M}}$.

The three vector preference models described above can be regarded as special cases of our general model. First, we set $\mathbf{G} = \mathbf{1}_N$ and $\mathbf{H} = \mathbf{A}$. Decomposition (6) then becomes

$$\mathbf{Z} = \mathbf{P}_N\mathbf{Z}\mathbf{P}_A + \mathbf{Q}_N\mathbf{Z}\mathbf{P}_A + \mathbf{P}_N\mathbf{Z}\mathbf{Q}_A + \mathbf{Q}_N\mathbf{Z}\mathbf{Q}_A, \tag{19}$$

where

$$\mathbf{P}_N\mathbf{Z}\mathbf{P}_A = \mathbf{1}_N(\mathbf{1}_N'\hat{\mathbf{M}}/N)\mathbf{A}' = \mathbf{1}_N\hat{\mathbf{m}}'\mathbf{A}',$$

$$\mathbf{Q}_N\mathbf{Z}\mathbf{P}_A = \mathbf{Q}_N\hat{\mathbf{M}}\mathbf{A}',$$

and

$$\mathbf{P}_N\mathbf{Z}\mathbf{Q}_A = \mathbf{1}_N(\mathbf{1}_N'\mathbf{Z}\mathbf{Q}_A/N) = \mathbf{1}_N\hat{\mathbf{c}}'.$$

In the above derivation, the Moore-Penrose inverse of $\mathbf{A}'\mathbf{A}$ (i.e., $(\mathbf{A}'\mathbf{A})^+$), was used for a g-inverse of $\mathbf{A}'\mathbf{A}$. In the BTC model, the first and the second terms of (19) are not separated, and the SVD of $\hat{\mathbf{M}}$ in $\mathbf{Z}\mathbf{P}_A = \hat{\mathbf{M}}\mathbf{A}'$ is obtained. In the THL model, the SVD of $\mathbf{Q}_N\hat{\mathbf{M}}$ in the second term is obtained. The third and the fourth terms are not separated. In the WVM, the SVD of $\hat{\mathbf{M}}$ is obtained as in the BTC model, and then $\mathbf{Y}^*\mathbf{X}'$ is decomposed into $\mathbf{P}_N\mathbf{Y}^*\mathbf{X}' = \mathbf{1}_N\hat{\mathbf{v}}'\mathbf{X}' = \mathbf{P}_N\hat{\mathbf{M}}\mathbf{P}_x$ and $\mathbf{Q}_N\mathbf{Y}^*\mathbf{X}' = \mathbf{Y}\mathbf{X}'$. Again, the third and the fourth terms are not separated.

Note that in the above cases, the minimization of $\psi_1 = SS((\hat{\mathbf{M}} - \mathbf{M}^*)\mathbf{A}')$ or $\psi_2 = SS((\mathbf{Q}_N\hat{\mathbf{M}} - \mathbf{M}^*)\mathbf{A}')$ did not require the GSVD of $\hat{\mathbf{M}}$ or $\mathbf{Q}_N\hat{\mathbf{M}}$ with row metric $\mathbf{A}'\mathbf{A}$, but only the ordinary SVD. This is because $\mathbf{A}'\mathbf{A} = m\mathbf{Q}_m$, $\hat{\mathbf{M}}\mathbf{Q}_m = \hat{\mathbf{M}}$ and $\mathbf{M}^*\mathbf{Q}_m = \mathbf{M}^*$, and so we can rewrite $\psi_1 = mSS(\hat{\mathbf{M}} - \mathbf{M}^*)$ and $\psi_2 = mSS(\mathbf{Q}_N\hat{\mathbf{M}} - \mathbf{M}^*)$. Thus, no special treatment (GSVD) was necessary in this case. Note that $\mathbf{A}'\mathbf{A}$ is singular, and thus, were it not for the simplification in the minimization criterion, this case would have required the general treatment for singular metric matrices by de Leeuw (1984). An example will be given in section 4.1 to illustrate practical uses of the vector preference models.

3.2. *Two-way CANDELINC*

Carroll, Pruzansky, and Kruskal (1980) considered the minimization of $SS(\mathbf{Z} - \mathbf{G}\mathbf{M}^*\mathbf{H}')$ over a fixed-rank matrix, $\mathbf{M}^*$. The model, $\mathbf{G}\mathbf{M}^*\mathbf{H}'$, is called the two-way CANDELINC (canonical decomposition under linear constraints). For the special case in which $\mathbf{G}$ and $\mathbf{H}$ are both orthonormal, Carroll et al. obtained:

$$SS(\mathbf{Z} - \mathbf{G}\mathbf{M}^*\mathbf{H}') = SS(\hat{\mathbf{M}} - \mathbf{M}^*) + SS(\mathbf{Z}) - SS(\hat{\mathbf{M}}), \tag{20}$$

where $\hat{\mathbf{M}} = \mathbf{G}'\mathbf{Z}\mathbf{H}$. Since $SS(\mathbf{Z}) - SS(\hat{\mathbf{M}})$ is a constant for a given $\mathbf{Z}$, $\mathbf{G}$, and $\mathbf{H}$, the minimum of $SS(\mathbf{Z} - \mathbf{G}\mathbf{M}^*\mathbf{H}')$ is attained at the minimum of $SS(\hat{\mathbf{M}} - \mathbf{M}^*)$. The $\mathbf{M}^*$ minimizing $SS(\hat{\mathbf{M}} - \mathbf{M}^*)$ is obtained by the SVD of $\hat{\mathbf{M}}$. When $\mathbf{G}$ and $\mathbf{H}$ are nonorthonormal, they are "orthonormalized" and the decomposition (20) is applied. As before, let $\mathbf{G} = \mathbf{F}_G\mathbf{R}_G'$ and $\mathbf{H} = \mathbf{F}_H\mathbf{R}_H'$ be the QR decompositions of $\mathbf{G}$ and $\mathbf{H}$. Then,

$$SS(\mathbf{Z} - \mathbf{G}\mathbf{M}^*\mathbf{H}') = SS(\mathbf{J} - \mathbf{J}^*) + SS(\mathbf{Z}) - SS(\mathbf{J}), \tag{21}$$

where $\mathbf{J} = \mathbf{F}'_G\mathbf{Z}\mathbf{F}_H$ and $\mathbf{J}^* = \mathbf{R}'_G\mathbf{M}^*\mathbf{R}_H$. Again, $SS(\mathbf{Z}) - SS(\mathbf{J})$ is a constant, and $SS(\mathbf{Z} - \mathbf{G}\mathbf{M}^*\mathbf{H}')$ is minimized by minimizing $SS(\mathbf{J} - \mathbf{J}^*)$ with respect to $\mathbf{M}^*$. To obtain such an $\mathbf{M}^*$, the SVD of $\mathbf{J}$, $\mathbf{J} = \mathbf{U}_J\mathbf{D}_J\mathbf{V}'_J$, is first obtained. Then, define $\mathbf{U}_M = (\mathbf{R}'_G)^-\mathbf{U}_J$, $\mathbf{V}_M = (\mathbf{R}'_H)^-\mathbf{V}_J$, and $\mathbf{D}_M = \mathbf{D}_J$. By discarding appropriate portions of $\mathbf{U}_M$, $\mathbf{V}_M$, and $\mathbf{D}_M$, the terms $\mathbf{U}^*_M$, $\mathbf{V}^*_M$, and $\mathbf{D}^*_M$ are obtained. Finally, $\mathbf{M}^* = \mathbf{U}^*_M\mathbf{D}^*_M\mathbf{V}^{*\prime}_M$ is formed that gives the desired $\mathbf{M}^*$. However, from sections 2.2 and 2.3 we already know that $\mathbf{U}_M\mathbf{D}_M\mathbf{V}'_M$ is nothing but the GSVD of $\hat{\mathbf{M}} = (\mathbf{G}'\mathbf{G})^-\mathbf{G}'\mathbf{Z}\mathbf{H}(\mathbf{H}'\mathbf{H})^- = (\mathbf{R}'_G)^-\mathbf{J}\mathbf{R}_H^-$, with metrics $\mathbf{G}'\mathbf{G} = \mathbf{R}_G\mathbf{R}'_G$ and $\mathbf{H}'\mathbf{H} = \mathbf{R}_H\mathbf{R}'_H$. This indicates a close relationship between two-way CANDELINC and the GSVD.

Although (21) is derived as a special case of (20), there is a more general expression for decomposition (21) that more directly shows the close relationship between two-way CANDELINC and the GSVD. The expression is:

$$\begin{aligned} SS(\mathbf{Z} - \mathbf{G}\mathbf{M}^*\mathbf{H}') &= SS(\mathbf{P}_G\mathbf{Z}\mathbf{P}_H - \mathbf{G}\mathbf{M}^*\mathbf{H}') + SS(\mathbf{Z} - \mathbf{P}_G\mathbf{Z}\mathbf{P}_H) \\ &= SS(\mathbf{G}(\hat{\mathbf{M}} - \mathbf{M}^*)\mathbf{H}') + SS(\mathbf{Z}) - SS(\mathbf{G}\hat{\mathbf{M}}\mathbf{H}'), \end{aligned} \tag{22}$$

where $\hat{\mathbf{M}} = (\mathbf{G}'\mathbf{G})^-\mathbf{G}'\mathbf{Z}\mathbf{H}(\mathbf{H}'\mathbf{H})^-$. Decomposition (22) reduces to (20) when $\mathbf{G}$ and $\mathbf{H}$ are columnwise orthonormal. Notice that the first term on the right hand side of (22) is nothing but the ψ minimized in section 2.2. Since $SS(\mathbf{Z} - \mathbf{P}_G\mathbf{Z}\mathbf{P}_H) = SS(\mathbf{Z}) - SS(\mathbf{G}\hat{\mathbf{M}}\mathbf{H}')$ is a constant, the left hand side of (22) is minimized by minimizing ψ. As we have seen in section 2.2, ψ is minimized by the GSVD of $\hat{\mathbf{M}}$ with metrics $\mathbf{G}'\mathbf{G}$ and $\mathbf{H}'\mathbf{H}$.

Decomposition (22) points to an interesting observation. Two-way CANDELINC is a special case of our general model, in which only the first term in (6) is separated from the rest. The $SS(\mathbf{Z} - \mathbf{P}_G\mathbf{Z}\mathbf{P}_H)$ represents the portion of the sum of squares of $\mathbf{Z}$ left unexplained by the first term. PCA is applied to the first term and the SS left unaccounted for by the r principal components corresponds with the first term in (22).

3.3. *Dual Scaling with External Criteria*

Nishisato (1978, 1980a, 1980b, 1988; Nishisato & Lawrence, 1981, 1989) proposed dual scaling of pair comparison data, successive categories data, and ANOVA of multiple-choice categorical data. His approach is similar to two-way CANDELINC, and consequently, may be recast in the light of our general model (Takane & Shibayama, 1988b).

For pair comparison data, let a specially coded data matrix be denoted by $\mathbf{Z}$. As before (section 3.1), the design matrix for pair comparisons is denoted by $\mathbf{A}$. Our approach invokes the GSVD of $\mathbf{Z}\mathbf{A}(\mathbf{A}'\mathbf{A})^+ = \mathbf{Z}\mathbf{A}/m$ with row metric $\mathbf{A}'\mathbf{A} = m\mathbf{Q}_m$. However, as noted in section 3.1, this metric has no substantial effect since $\mathbf{A}(\mathbf{A}'\mathbf{A}) = m\mathbf{A}$. The GSVD problem thus reduces to the ordinary SVD of $\mathbf{Z}\mathbf{A}$, which is identical to Nishisato's (1978) solution.

For successive categories data, data matrix $\mathbf{Z}$ is again specially coded (see Nishisato, 1980b, for details). The design matrix (denoted by $\mathbf{R}$) in this case represents comparisons between stimuli and category boundaries. Let s represent the number of category boundaries (which is equal to the number of observation categories minus one). Then $\mathbf{R}$ has the form:

$$\mathbf{R} = \left.\begin{bmatrix} \mathbf{I}_s & -\mathbf{1}_s & \cdot & \cdot & \cdot & \mathbf{0} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ \mathbf{I}_s & \mathbf{0} & \cdot & \cdot & \cdot & -\mathbf{1}_s \end{bmatrix}\right\} \quad m \text{ times}.$$

Our method again invokes the GSVD of $\mathbf{ZR}(\mathbf{R}'\mathbf{R})^+$ with row metric $\mathbf{R}'\mathbf{R}$. However, unlike the pair comparison case, this does not reduce to the ordinary SVD, since

$$\mathbf{R}'\mathbf{R} = \begin{bmatrix} m\mathbf{I}_s & -\mathbf{1}_s\mathbf{1}_m' \\ -\mathbf{1}_m\mathbf{1}_s' & s\mathbf{I}_m \end{bmatrix},$$

and

$$(\mathbf{R}'\mathbf{R})^+ = (\mathbf{R}'\mathbf{R} + \mathbf{1}_t\mathbf{1}_t')^{-1} - \mathbf{1}_t\mathbf{1}_t'/t^2,$$

where $t = m + s$.

Nishisato and Sheu (1984), on the other hand, proposed a different solution. They augmented the data matrix and the design matrix by including implied relationships among stimuli and among category boundaries. This has the effect of making $\mathbf{R}'\mathbf{R} = t\mathbf{Q}_t$. Consequently, the problem reduced to the ordinary SVD as in the pair comparison case. In section 4.2, we will demonstrate how strikingly similar results our procedure can obtain in relation to those obtained by Nishisato and Sheu's method without artificially augmenting the data.

The pair comparison method may be used in conjunction with the successive categories method (Sjöberg, 1967). In this case, the design matrix will take the multiplicative form, $\mathbf{H} = \mathbf{RA}$, and so the decomposition suggested toward the end of section 2.1 seems more appropriate. In this decomposition, the row space is split by orthogonal projection operators, $\mathbf{P}_{RA}$, $\mathbf{P}_R - \mathbf{P}_{RA}$, and $\mathbf{Q}_R$.

The model underlying Nishisato's ANOVA of multiple-choice categorical data is very similar to that of two-way CANDELINC, although in the former emphasis is placed on finding out how much variability in the data is accounted for by which effects, while in the latter, emphasis is placed on graphical representations of the data. For the two-way ANOVA case, for example, Nishisato's procedure decomposes the data matrix into:

$$\mathbf{Z} - \mathbf{P}_N\mathbf{Z} = \mathbf{Q}_N\mathbf{Z} = \mathbf{P}_A\mathbf{Q}_N\mathbf{Z} + \mathbf{P}_B\mathbf{Q}_N\mathbf{Z} + \mathbf{P}_{AB}\mathbf{Q}_A\mathbf{Q}_B\mathbf{Q}_N\mathbf{Z} + \mathbf{Q}_{AB}\mathbf{Q}_A\mathbf{Q}_B\mathbf{Q}_N\mathbf{Z}, \tag{25}$$

where

$$\mathbf{P}_N = \mathbf{1}_N\mathbf{1}_N'/N, \qquad \mathbf{Q}_N = \mathbf{I} - \mathbf{P}_N;$$

$$\mathbf{P}_A = \mathbf{G}_A(\mathbf{G}_A'\mathbf{G}_A)^{-1}\mathbf{G}_A', \qquad \mathbf{Q}_A = \mathbf{I} - \mathbf{P}_A;$$

$$\mathbf{P}_B = \mathbf{G}_B(\mathbf{G}_B'\mathbf{G}_B)^{-1}\mathbf{G}_B', \qquad \mathbf{Q}_B = \mathbf{I} - \mathbf{P}_B;$$

$$\mathbf{P}_{AB} = \mathbf{G}_{AB}(\mathbf{G}_{AB}'\mathbf{G}_{AB})^{-1}\mathbf{G}_{AB}', \qquad \mathbf{Q}_{AB} = \mathbf{I} - \mathbf{P}_{AB},$$

and $\mathbf{G}_A$, $\mathbf{G}_B$, and $\mathbf{G}_{AB}$ are matrices of dummy variables pertaining to the A main effect, the B main effect, and the interaction between A and B, respectively. Note that $\mathbf{P}_A\mathbf{Q}_N = \mathbf{P}_A - \mathbf{P}_N$, $\mathbf{P}_B\mathbf{Q}_N = \mathbf{P}_B - \mathbf{P}_N$, $\mathbf{P}_A\mathbf{P}_B = \mathbf{P}_N$ (assuming that the design is balanced), $\mathbf{P}_{AB}\mathbf{Q}_A\mathbf{Q}_B\mathbf{Q}_N = \mathbf{P}_{AB} - \mathbf{P}_A - \mathbf{P}_B + \mathbf{P}_N$, and $\mathbf{Q}_{AB}\mathbf{Q}_A\mathbf{Q}_B\mathbf{Q}_N = \mathbf{Q}_{AB}$ (Yanai & Takeuchi, 1983). PCA is applied to subsets of the terms on the right hand side of (25), either separately or combined. The PCA may involve the SVD of relatively large matrices. However, the results in section 2.3 suggest that these SVD can be derived from the SVD of much smaller matrices using the QR decompositions of $\mathbf{G}$ and $\mathbf{H}$. This is particularly advantageous, when the QR decompositions of $\mathbf{G}$ and $\mathbf{H}$ can easily be obtained.

3.4. Redundancy Analysis

Several methods have been developed in multivariate analysis literature to relate two sets of variables. Canonical correlation analysis and canonical discriminant analysis are two representative examples. A question that naturally arises is how our development relates to these methods. For a direct comparison we temporarily assume that only the subject information, **G**, is available, which amounts to assuming $\mathbf{H} = \mathbf{I}$.

One interesting method in this context is the second type of quantification method (Q2) developed by Hayashi (1952). This method is similar to Fisher's (1948) method of additive scoring, and may be viewed as a special case of canonical discriminant analysis, where not only the criterion but also the predictor variables are discrete. Canonical discriminant analysis is, in turn, viewed as a special case of canonical correlation analysis. Canonical correspondence analysis (ter Braak, 1986) is also equivalent to Q2 (Takane et al., 1989), and consequently, is distinct from our method.

Our approach and consequently Nishisato's ANOVA of multiple-choice categorical data are distinct from canonical correlation analysis, and consequently, are also distinct from canonical discriminant analysis and Q2. Canonical discriminant analysis amounts to obtaining the GSVD of $(\mathbf{G}'\mathbf{G})^{-}\mathbf{G}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-}$ in the column and row metrics of $\mathbf{G}'\mathbf{G}$ and $\mathbf{Z}'\mathbf{Z}$, respectively. Our approach, on the other hand, obtains the GSVD of $(\mathbf{G}'\mathbf{G})^{-}\mathbf{G}'\mathbf{Z}$ with metrics $\mathbf{G}'\mathbf{G}$ and **I**. It is interesting to note that this is equivalent to redundancy analysis (Israëls, 1984; van den Wollenberg, 1977) of **Z**-variables given **G**-variables. Redundancy analysis is variously called "principal components of instrumental variables" (Rao, 1964), "reduced rank regression" (Rao, 1979, 1980), and so on. In the usual redundancy analysis, however, only the subject information matrix, **G**, can be incorporated. In our approach, on the other hand, both column and row structures (**G** and **H**) can be quite naturally incorporated. On the basis of this observation, Henk Kiers (personal communication, November 9, 1988) called our method "double redundancy analysis".

When $\mathbf{Z} = \mathbf{I}$, $\mathbf{P}_G\mathbf{Z}\mathbf{P}_H$ reduces to $\mathbf{P}_G\mathbf{P}_H$, whose SVD is equivalent to canonical correlation analysis between **G** and **H**. Thus, one way to characterize canonical correlation analysis is: Find **U** and **V** that minimize

$$SS(\mathbf{I} - \mathbf{G}\mathbf{U}\mathbf{D}\mathbf{V}'\mathbf{H}'),$$

such that $\mathbf{U}'\mathbf{G}'\mathbf{G}\mathbf{U} = \mathbf{I}$, $\mathbf{V}'\mathbf{H}'\mathbf{H}\mathbf{V} = \mathbf{I}$ and **D** is diagonal and positive definite. The $\text{tr}\,(\mathbf{P}_G\mathbf{P}_H) = SS(\mathbf{P}_G\mathbf{P}_H)$ is called the generalized coefficient of determination (GCD) by Yanai (1974).

4. Examples of Application

In this section we present two examples. The first concerns pairwise preference rating data, and the second a contrived data set used by Nishisato and Sheu (1984) to demonstrate the feasibility of their method for successive categories data. Both of our examples are somewhat specialized in that they need specific subject and/or variable design matrices (and hence, our approach) for meaningful analyses. Our method is also useful when this information is not so essential. For example, ordinary multivariate data may be analyzed by simple PCA, then portions of the data explained by the main effect of A may be analyzed, etc., and the results may be compared. The series of analyses emphasize and reveal different aspects of the data. Examples of such analyses can be found in Escoufier and Holmes (1988), and Sabatier, et al. (1989) for continuous multivariate data, and in Nishisato (1980a, 1982) and van der Heijden, et al. (1989) for discrete multivariate data.

4.1. Pairwise Preference Rating Data

Pairwise preference data were collected on nine stimuli using 25-point rating scales. The stimulus set was constructed in a manner similar to Rumelhart and Greeno (1971). It consisted of three distinct groups of people, three politicians, three athletes, and three popular singers: (a) Brian Mulroney (Prime Minister of Canada), (b) Ronald Reagan (President of the United States), (c) Margaret Thatcher (Prime Minister of United Kingdom), (d) Jacqueline Gareau (twice winner of the Boston marathon in the woman's division), (e) Wayne Gretzky (NHL hockey player), (f) Steve Podborski (former champion of the World Cup downhill ski race), (g) Paul Anka (male vocalist), (h) Tommy Hunter (country song singer), and (i) Ann Murray (female vocalist). The stimuli were presented in pairs. The subjects were asked to indicate with whom and the degree to which they would rather spend an hour, if they were given a chance. One member of a pair is placed on one end and the other on the other end of a rating scale. The rating scale was marked by integers from -12 to $+12$, a large negative value indicating more preference toward a member placed on the left end and a large positive value indicating just opposite; zero indicates a midpoint and neutrality.

The data were initially collected from 501 subjects, each responding to all $9 \times 8/2 = 36$ possible pairs of stimuli. The data analyzed are a subset of the data pertaining to 100 subjects for whom some demographic information was available. The subjects were mostly university students and some high school students living in the Montreal area. Approximately 3/5 of the 100 subjects were female students. Also, 55 subjects were anglophones (English speaking), 17 were francophones (French speaking), and the remaining 28 had other linguistic backgrounds (e.g., Italian, Greek, etc.).

Our method was applied first with $\mathbf{G} = \mathbf{1}_N$ and $\mathbf{H} = \mathbf{A}$ (design matrix for pair comparisons). The four terms in (6) accounted for 11%, 59%, less than 1%, and 29% of the total sum of squares ($SS(\mathbf{Z})$). (In this particular context, the first and the third terms pertain to the mean pairwise preference judgments that, respectively, can and cannot be explained by differences between mean stimulus preference values. The second and the fourth terms, on the other hand, pertain to the covariances among the judgments that can and cannot be explained by stimuluswise covariances.) This implies that the pairwise unscalability in the BTC model (the third term in (6)) accounted for only a negligible portion of the total SS. The mean preference vector, $\mathbf{m}$, in the THL model, on the other hand, accounted for 11% of the total SS. Estimated mean preference values were, in the order of the most preferred to the least preferred: Gretzky (2.13), Podborski (1.98), Reagan (1.13), Mulroney (.53), Thatcher (.17), Murray ($-.24$), Anka ($-.56$), Gareau (-2.31), and Hunter (-2.83). There seems to be a strong correlation between the mean preference values and the occupational categories of the stimuli. Athletes tend to be preferred most, politicians next, and entertainers least. The only exception was Jacqueline Gareau, who happened to be relatively unknown.

The THL model applies PCA to $\mathbf{B}$ in the second term of (6). The first two principal components accounted for 73% of the SS in the second term. Figure 1 displays the two dimensional stimulus configuration (the plot of $\mathbf{X}$). The nine stimuli formed three clusters roughly corresponding with the three occupational categories from which the stimuli were sampled. Jacqueline Gareau (Number 4) is again an exception, being placed closer to the singer/entertainer group than to the athlete group. Stimuli that are located close together are generally more similar to each other, and consequently, are more comparable. They also tend to have higher covariances (Takane, 1987).

The mean preference vector, $\mathbf{m}$, seems highly correlated with a certain direction in the space. This direction is indicated by the vector, $\mathbf{w}$, labeled $\mathbf{W}$ in the figure. The correlation between $\mathbf{m}$ and $\mathbf{Xw}$ was .776. This phenomenon may appear somewhat

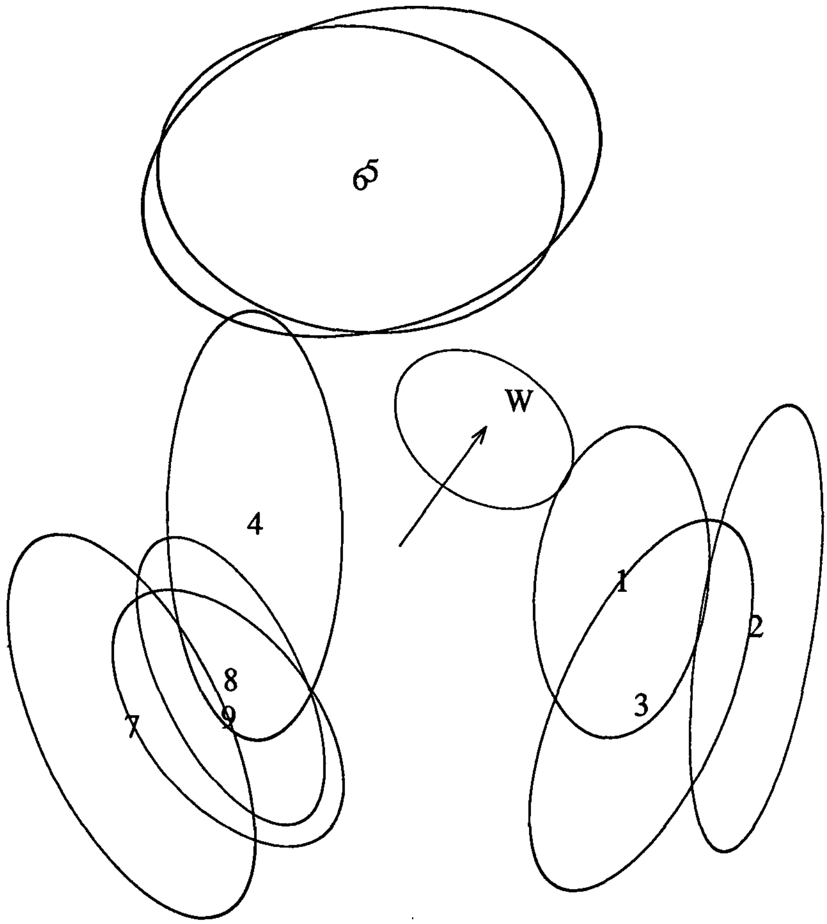

FIGURE 1
Stimulus configuration and 95% Bootstrap confidence regions for the preference data derived from the THL model.

surprising, since the information pertaining to **m** is explicitly eliminated from the second term of (6). It happens quite frequently, however, when categories of stimuli are closely related to their preference values, as in the present case.

Ellipses surrounding the estimated stimulus points are 95% confidence regions (Ramsay, 1978) obtained through the Bootstrap method (Efron, 1979; Weinberg, Carroll, & Cohen, 1984). These confidence regions are based on repeated analyses of 100 samples of 100 observations each resampled with replacement from the original data set. The derived confidence regions indicate that the estimated stimulus points are fairly variable with just a hundred observations. Still, the current situation is a favorable one in that each subject responded to all possible pairs of stimuli, so that on average, there were four judgments per stimulus obtained from each subject.

The BTC model combines the first and the second terms in (6), and applies PCA to $\mathbf{1}_N\hat{\mathbf{M}} + \hat{\mathbf{B}}$. The first two principal components accounted for 64% of the *SS* of the combined term that in turn accounted for 70% of the total *SS*. The derived two dimensional stimulus configuration is presented in Figure 2 along with ten subjects' preference vectors indicated by pointed arrows. This stimulus configuration is strikingly similar to the one derived from the THL model (Figure 1). This similarity between the two configurations corresponds with the fact that the mean preference vector, **m**, in the THL model is highly correlated with a particular direction in Figure 1.

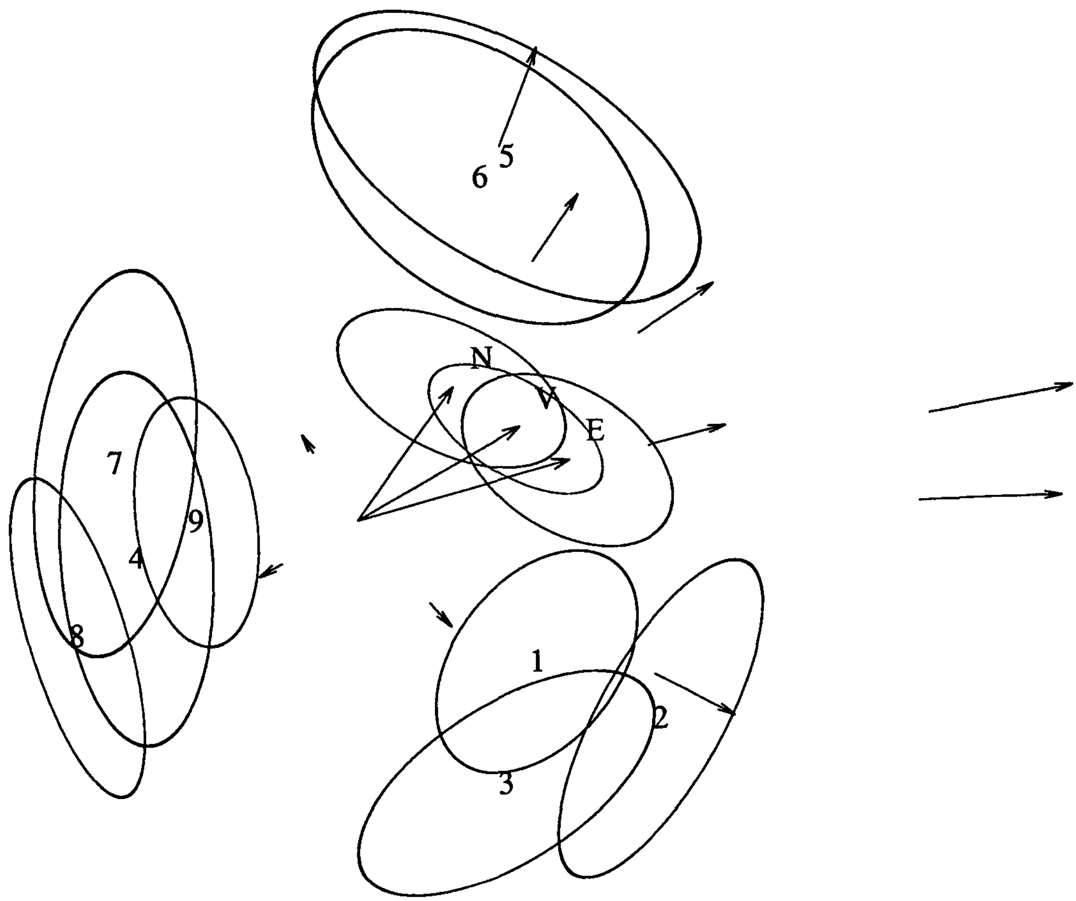

FIGURE 2
Stimuls configuration and 95% Bootstrap confidence regions for the preference data derived from the BTC model.

The mean subject vector, **v**, in the WVM model was calculated and superposed in Figure 2 (the vector labeled **V**). The correlation between **m** and **Xv** is extremely high (.951). This is quite natural, since the WVM model can be derived by setting **m** = **Xv** in the THL model.

It is interesting to see where mean vectors of various subgroups of subjects would be located in Figure 2 in a manner similar to **V**. For this, we are in effect fitting a generalized version of the WVM model,

$$\mathbf{Z} = (\mathbf{GW} + \mathbf{Y})\mathbf{X}'\mathbf{A}' + \mathbf{E}, \tag{26}$$

where **G** is a matrix of dummy variables indicating categories of subjects (e.g., male, female), and **W** a matrix of weights analogous to the **v** vector. When the model was applied, only a slight difference was found between males and females. A substantial difference was found between the anglophone group and the nonanglophone groups, but only a small difference between francophones and other nonanglophones. The latter two groups were combined into one; two mean vectors, one for the anglophones (labeled **E**) and the other for all the nonanglophones combined (labeled **N**) are depicted in Figure 2. It appears that the anglophones are more inclined toward the politicians (who are all English speaking) in comparison with the nonanglophones.

TABLE 1

Analysis of Nishisato and Sheu's Data

a. Data Matrix:

-1	-1	-1	1	1	1
-1	1	-1	1	-1	1
-1	1	-1	1	1	1
-1	-1	-1	1	1	1
-1	-1	1	1	1	1
-1	-1	-1	-1	1	1
-1	1	-1	1	-1	1
-1	-1	-1	-1	-1	1
-1	1	-1	1	1	1
-1	-1	1	1	1	1

b. Design Matrix:

1	0	-1	0	0
0	1	-1	0	0
1	0	0	-1	0
0	1	0	-1	0
1	0	0	0	-1
0	1	0	0	-1

c. Coordinates of Stimulus Points and Category Boundaries:

t1	0.230	-0.329
t2	-0.273	0.287
m1	-0.367	-0.359
m2	-0.004	0.163
m3	0.414	0.238

d. Matrix of Subject Vectors:

s1	-2.067	-0.577
s2	-1.511	1.850
s3	-1.878	0.716
s4	-2.067	-0.577
s5	-1.598	-1.560
s6	-1.528	-0.826
s7	-1.511	1.850
s8	-1.160	0.308
s9	-1.878	0.716
s10	-1.598	-1.560

Again, ellipses indicate 95% Bootstrap confidence regions for the estimated points and vectors. Note that there is an overlap between the confidence region for vector E and that for vector N. This roughly indicates that there is no significant difference between the two vectors at $\alpha = .05$. (To be more precise, we also have to take into account covariances among the estimated points.)

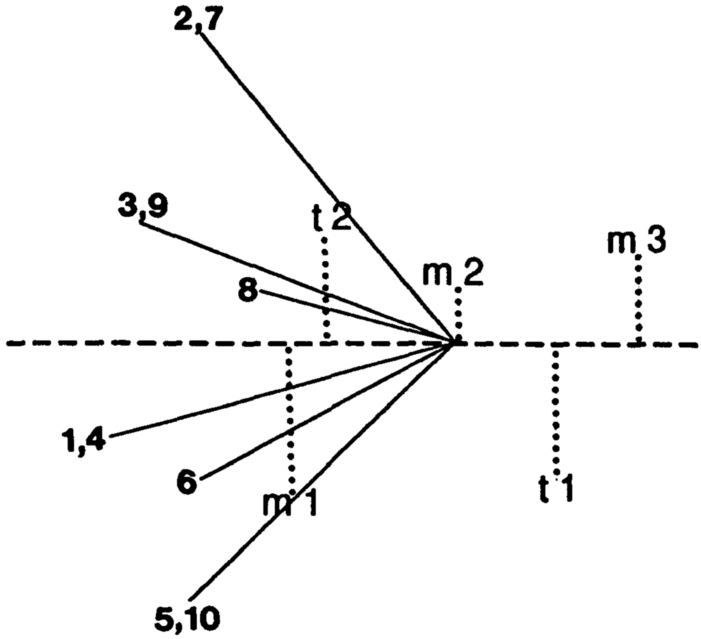

FIGURE 3A
Two dimensional configuration for Nishisato and Sheu's data derived by our procedure.

4.2. Successive Categories Data

Nishisato and Sheu (1984) analyzed a hypothetical data set to illustrate their proposed solution for successive categories data. In generating the data set, it was supposed that ten subjects rated three stimuli on three-point rating scales. The original data were coded into a 10 by 6 matrix of +1 and −1's according to Nishisato's (1980b) coding scheme (see Table 1a).

Our method was applied with **G** = **I** and **H** = **R** (design matrix given in Table 1b for successive categories data). Since **G** = **I**, the data matrix was decomposed into two parts, one that could be explained and the other that could not be explained by **R**. As it turned out, nearly 90% of the total *SS* could be explained by **R**. PCA was applied to the portion explained by **R**. This involved the GSVD of $\mathbf{ZR}(\mathbf{R}'\mathbf{R})^+$ with row metric $\mathbf{R}'\mathbf{R}$. The first two principal components accounted for 81% (54% by the first component and 27% by the second) of the *SS* explained by **R**. The derived two dimensional solution is presented numerically in Table 1c (coordinates of stimulus points and category boundaries) and Table 1d (subject vectors), and is displayed graphically in Figure 3a. When the figure was drawn, the subject vectors were multiplied by some constant, since only their relative lengths were meaningful. In this figure stimulus points are indicated by m1, m2, and m3, and category boundaries by t1 and t2. Subject vectors that account for individual differences in response patterns are indicated by line segments numbered from 1 to 10. There are only six subject vectors in the figure, because some had identical response patterns.

The corresponding unidimensional solution is obtained by projecting points and vectors onto the horizontal axis indicated by a broken line. The points lie on the unidimensional continuum in the order of m1, t2, m2, t1, and m3 from left to right.

Figure 3b displays the two dimensional configuration derived from Nishisato and Sheu's (1984) procedure, which involves the ordinary SVD of $\mathbf{Z}^*\mathbf{R}^*/t$, where $\mathbf{Z}^*$ and $\mathbf{R}^*$ are augmented data and augmented design matrices, respectively, and t is the number of stimuli plus the number of category boundaries. (In Figure 3b, the stimulus config-

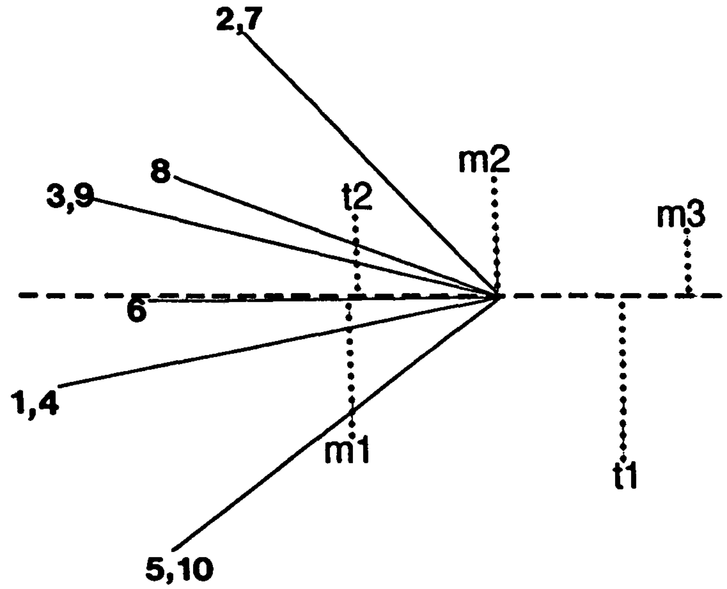

FIGURE 3B
Two dimensional configuration for Nishisato and Sheu's data derived by Nishisato and Sheu's procedure.

uration and subject vectors were independently adjusted in size to make them comparable to Figure 3a.) Notice that Figure 3b is strikingly similar to Figure 3a. The similarity is even more remarkable in the unidimensional case, indicating that our procedure is capable of obtaining results similar to those obtained by Nishisato and Sheu's procedure without artificially augmenting the data.

5. Discussion

We discussed PCA of multivariate data with external information on both subjects and variables. The idea of incorporating external information in scaling procedures is by no means new. A host of examples can be found in multidimensional scaling (e.g., Bloxom, 1978; Heiser & Meulman, 1983b), in unfolding analysis (DeSarbo & Rao, 1984; Heiser & Meulman, 1983a), in conjoint analysis (DeSarbo, Carroll, Lehmann, & O'Shaughnessy, 1982), and so on. In this paper we presented a general methodology for PCA that encompassed a wide range of special cases. In this final section, we briefly discuss three important practical considerations: assessment of stability, the problem of missing observations, and possible data transformations. For a wider range of problems in PCA, see, for example, Jolliffe (1986).

Assessing the reliability of derived solutions is an important aspect of any data analysis. Although our proposed method itself is largely descriptive, the reliability of the solutions obtained by our procedure can be easily assessed by a Bootstrap method (Efron, 1979) or similar resampling methods, as demonstrated in section 4.1. The simplicity of computation in our method is an asset in this process, since all the resampling methods require repeated solutions of many data sets. Sensitivity analysis (Critchley, 1985; Tanaka, 1988) is also feasible, since solutions can be obtained analytically in our method. The sensitivity analysis identifies influential observations.

Certain types of hypothesis testing are also possible with the Bootstrap method (for an example, see section 4.1). Eastment and Krzanowski (1982) proposed a cross-validational procedure for choosing the number of components in PCA. Monte Carlo techniques may also be used to obtain benchmarks for the number of significant singular values. Some attempts have also been made, in a special case of our method, to mathematically develop a goodness-of-fit test and a test of additional constraints on model parameters. Appendix Section C gives some examples.

Missing observations raise a serious problem in PCA as well as other related techniques (e.g., Nishisato, 1980a; Meulman, 1982). The simplest way to deal with the problem is to fill in some "neutral" values (e.g., zero, means, etc.) that have "minimal" effects on derived solutions. Once this is done, the analysis can proceed as if there were no missing data. A potential problem with this strategy is that in some cases, it is difficult to decide a priori which values have the "minimal" effects on the solutions. To avoid this difficulty, we may iteratively re-estimate "optimal" values for missing observations (e.g., Gabriel & Zamir, 1979; Gifi, 1981). Model values are first estimated with temporary estimates of missing observations, which are then used as estimates. This iterative process is repeated until no significant changes occur in the estimates. This process, however, may be time consuming. Shibayama (1988) proposed a closed form solution for PCA with missing observations, but unfortunately, his method only applies to columnwise standardized data.

Data transformation is another important consideration, since it makes the data more in line with the model, goodness-of-fit will be improved, and as a result, a more parsimonious representation of the data may be possible. A specific form of the transformation may also be of interest in its own sake; it may reflect some empirically important process. For example, in rating data the form of the transformation may indicate subjects' response styles, and this may be an interesting aspect of individual differences. Kruskal's (1964) least squares monotonic transformation and the monotone spline transformation (Ramsay, 1989; Winsberg & Ramsay, 1983) are excellent candidates for possible data transformations and can easily be incorporated in our method by alternating the model estimation phase and the optimal data transformation phase until convergence is reached. We have tried Kruskal's least squares monotonic transformation in fitting the BTC model to the pairwise preference data described in section 4.1. In this particular instance, however, we failed to obtain results substantially more interesting than in the untransformed case.

Appendix

In this appendix, we present (a) a separation theorem for singular values used in section 2.1, (b) some results on the use of metric matrices in decomposition (6), and (c) some results on GMANOVA (or the growth curve models).

A. Separation Theorem (Yanai & Takeuchi, 1983, p. 128)

Let $\mathbf{Z}$ be an N by n matrix. Let $\mathbf{P}_G$ (N by N) and $\mathbf{P}_H$ (n by n) be orthogonal projection matrices of rank p and rank q, respectively. Then

$$s_{j+t}(\mathbf{Z}) \le s_j(\mathbf{P}_G\mathbf{Z}\mathbf{P}_H) \le s_j(\mathbf{Z}), \tag{27}$$

where $s_j(\mathbf{Z})$ is the jth largest singular value of $\mathbf{Z}$, and $t = N + n - (p + q)$.

The $\mathbf{P}_G$ and $\mathbf{P}_H$ can be decomposed into $\mathbf{P}_G = \mathbf{T}_G\mathbf{T}_G'$ and $\mathbf{P}_H = \mathbf{T}_H\mathbf{T}_H'$ such that $\mathbf{T}_G'\mathbf{T}_G = \mathbf{I}_p$ and $\mathbf{T}_H'\mathbf{T}_H = \mathbf{I}_q$. The generalized Poincare separation theorem (e.g., Rao, 1980, p. 10), on the other hand, states that $s_{j+t}(\mathbf{Z}) \le s_j(\mathbf{T}_G'\mathbf{Z}\mathbf{T}_H) \le s_j(\mathbf{Z})$. However, from the results of section 2.3, $s_j(\mathbf{T}_G'\mathbf{Z}\mathbf{T}_H) = s_j(\mathbf{T}_G\mathbf{T}_G'\mathbf{Z}\mathbf{T}_H\mathbf{T}_H')$.

B. Metric Matrices in External Analysis

Let $\mathbf{G}$ ($N \times p$) and $\mathbf{H}$ ($n \times q$) be subject and variable information matrices, respectively, and let $\mathbf{K}$ and $\mathbf{L}$ be column and row metric matrices, respectively, in the least squares problem for estimating parameters in model (1). Define

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

and

$$\mathbf{P}_{H/L} = \mathbf{H}(\mathbf{H}'\mathbf{L}\mathbf{H})^{-}\mathbf{H}'\mathbf{L},$$

and $\mathbf{Q}_{G/K} = \mathbf{I} - \mathbf{P}_{G/K}$ and $\mathbf{Q}_{H/L} = \mathbf{I} - \mathbf{P}_{H/L}$, where we assume rank $(\mathbf{KG})$ = rank $(\mathbf{G})$ and rank $(\mathbf{LH})$ = rank $(\mathbf{H})$ (Yanai, 1990). The $\mathbf{P}_{G/K}$, $\mathbf{P}_{H/L}$, $\mathbf{Q}_{G/K}$, and $\mathbf{Q}_{H/L}$ are oblique projection operators (i.e., they are idempotent but not symmetric). Note that $\mathbf{P}_{G/K}\mathbf{Q}_{G/K} = \mathbf{Q}_{G/K}\mathbf{P}_{G/K} = \mathbf{0}$ and $\mathbf{P}_{H/L}\mathbf{Q}_{H/L} = \mathbf{Q}_{H/L}\mathbf{P}_{H/L} = \mathbf{0}$. Then, analogous to (6) we obtain

$$\mathbf{Z} = \mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L} + \mathbf{Q}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L} + \mathbf{P}_{G/K}\mathbf{Z}\mathbf{Q}'_{H/L} + \mathbf{Q}_{G/K}\mathbf{Z}\mathbf{Q}'_{H/L}. \tag{28}$$

The notion of trace-orthogonality should also be generalized. Two matrices, $\mathbf{X}$ and $\mathbf{Y}$, are said to be trace-orthogonal in the metrics of $\mathbf{K}$ and $\mathbf{L}$ when tr $(\mathbf{X}'\mathbf{KYL})$ = tr $(\mathbf{KXLY}') = 0$. The four terms on the right hand side of (28) are mutually orthogonal in this sense.

The separation theorem for singular values given in Appendix section A can also be generalized:

$$s_{j+t}(\mathbf{R}'_K\mathbf{Z}\mathbf{R}_L) \le s_j(\mathbf{R}'_K\mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L}\mathbf{R}_L) \le s_j(\mathbf{R}'_K\mathbf{Z}\mathbf{R}_L), \tag{29}$$

where $\mathbf{R}_K$ and $\mathbf{R}_L$ are such that $\mathbf{K} = \mathbf{R}_K\mathbf{R}'_K$ and $\mathbf{L} = \mathbf{R}_L\mathbf{R}'_L$. This directly follows from (27), since

$$\mathbf{R}'_K\mathbf{P}_{G/K}\mathbf{P}'_{H/L}\mathbf{R}_L = \mathbf{P}^*_G(\mathbf{R}'_K\mathbf{Z}\mathbf{R}_L)\mathbf{P}^*_H,$$

where $\mathbf{P}^*_G$ and $\mathbf{P}^*_H$ are orthogonal projection operators defined by $\mathbf{G}^* = \mathbf{R}'_K\mathbf{G}$ and $\mathbf{H}^* = \mathbf{R}'_L\mathbf{H}$, respectively.

C. GMANOVA (The Growth Curve Models)

When we fit only the first term in model (1) to $\mathbf{Z}$ with $\mathbf{K} = \mathbf{I}$ and $\mathbf{L} = \mathbf{S}^{-1}$, where $\mathbf{S} = \mathbf{Z}'\mathbf{Q}_G\mathbf{Z}$,

$$\hat{\mathbf{M}} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{Z}\mathbf{S}^{-1}\mathbf{H}(\mathbf{H}'\mathbf{S}^{-1}\mathbf{H})^{-1}, \tag{30}$$

which is equal to the maximum likelihood estimate of $\mathbf{M}$ in the GMANOVA model under the iid multivariate normal distribution (Khatri, 1966; Rao, 1965). A goodness-of-fit test of the GMANOVA model to $\mathbf{Z}$ may be performed by first calculating

$$\mathbf{E} = \mathbf{H}^{*\prime}\mathbf{S}\mathbf{H}^*,$$

and

$$\mathbf{F} = \mathbf{H}^{*\prime}\mathbf{Z}'\mathbf{P}_G\mathbf{Z}\mathbf{H}^*,$$

where $\mathbf{H}^*$ is any n by $n - q$ matrix of rank $n - q$ orthogonal to $\mathbf{H}$ (Grizzle & Allen, 1969). Eigenvalues of $\mathbf{FE}^{-1}$ are then used in one of Wilks' lambda criterion, Roy's max root criterion, Lawley-Hotelling's trace criterion, and Bartlett-Nanda-Pillai criterion for the hypothesis testing. The test of a hypothesis such as $\mathbf{AMB} = \mathbf{C}$ is also possible, where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are known matrices. Let

$$\mathbf{E} = \mathbf{B}'(\mathbf{H}'\mathbf{S}^{-1}\mathbf{H})^{-1}\mathbf{B},$$

and

$$\mathbf{F} = (\mathbf{A}\hat{\mathbf{M}}\mathbf{B} - \mathbf{C})'(\mathbf{ARA}')^{-1}(\mathbf{A}\hat{\mathbf{M}}\mathbf{B} - \mathbf{C}),$$

where $\mathbf{R} = (\mathbf{G}'\mathbf{G})^{-1} + (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{Z}\mathbf{S}^{-1}\mathbf{Q}_{H/S^{-1}}\mathbf{Z}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}$. $\mathbf{E}$ and $\mathbf{F}$ are used in the same way as above for the hypothesis testing. For more detail, see, for example, Siotani, Hayakawa, and Fujikoshi (1985).

References

Bechtel, G. G. (1976). *Multidimensional preference scaling*. The Hague: Mouton.

Bechtel, G. G., Tucker, L. R., & Chang, W. (1971). A scalar product model for the multidimensional scaling of choice. *Psychometrika*, *36*, 369–387.

Besse, P., & Ramsay, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika*, *51*, 285–311.

Bloxom, B. (1978). Constrained multidimensional scaling in *N* spaces. *Psychometrika*, *43*, 397–408.

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

Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling, Vol. I* (pp. 105–155). New York: Seminar Press.

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

Corsten, L. C. A. (1976). *Matrix approximation, a key to application of multivariate methods*. Invited paper presented at the 9th Biometric Conference, Boston.

Corsten, L. C. A., & Van Eijnsbergen, A. C. (1972). Multiplicative effects in two-way analysis of variance. *Statistica Neelandica*, *26*, 61–68.

Critchley, F. (1985). Influence in principal component analysis. *Biometrika*, *72*, 627–636.

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

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

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

De Soete, G., & Carroll, J. D. (1983). A maximum likelihood method for fitting the wandering vector model. *Psychometrika*, *48*, 553–566.

Eastment, H. T., & Krzanowski, W. J. (1982). Cross-validatory choice of the number of components from a principal component analysis. *Technometrics*, *24*, 73–77.

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

Escoufier, Y., & Holmes, S. (1988). *Décomposition de la variabilité dans les analyses exploratoires: Un exemple d'analyse en composantes principles en presence de variables qualitatives concomittantes* [Descomposition of variabilities in exploratory data analysis: An example of principal component analysis in the presence of qualitative concomitant variables.] Unpublished manuscript, ENSAM-INRA, Montpellier, France.

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

Gabriel, K. R. (1978). Least squares approximation of matrices by additive and multiplicative models. *Journal of Royal Statistical Society, Series B*, *40*, 186–196.

Gabriel, K. R., & Zamir, S. (1979). Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics*, *21*, 489–498.

Gifi, A. (1981). *Non-linear multivariate analysis*. Leiden: University of Leiden, Department of Data Theory.

Gollob, H. F. (1968). A statistical model which combines features of factor analytic and analysis of variance technique. *Psychometrika*, *33*, 73–115.

Greenacre, M. J., & Underhill, L. G. (1982). Scaling a data matrix in a low-dimensional euclidean space. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 183–268). Cambridge: Cambridge University Press.

Grizzle, J. E., & Allen, D. M. (1969). Analysis of growth and dose response curves. *Biometrics*, *25*, 357–381.

Hayashi, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics, 2*, 69–98.
Heiser, W. J., & de Leeuw, J. (1981). Multidimensional mapping of preference data. *Mathematiqué et sciences humaines, 19*, 39–96.
Heiser, W. J., & Meulman, J. (1983a). Analyzing rectangular tables by joint and constrained multidimensional scaling. *Journal of Econometrics, 22*, 139–167.
Heiser, W. J., & Meulman, J. (1983b). Constrained multidimensional scaling, including confirmation. *Applied Psychological Measurement, 7*, 381–404.
Israëls, A. Z. (1984). Redundancy analysis for qualitative variables. *Psychometrika, 49*, 331–346.
Jolliffe, I. T. (1986). *Principal component analysis*. Berlin: Springer Verlag.
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.
Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika, 29*, 115–129.
Meulman, J. (1982). *Homogeneity analysis of incomplete data*. Leiden: DSWO Press.
Nishisato, S. (1978). Optimal scaling of paired comparison and rank order data: An alternative to Guttman's formulation. *Psychometrika, 43*, 263–271.
Nishisato, S. (1980a). *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.
Nishisato, S. (1980b). Dual scaling of successive categories data. *Japanese Psychological Research, 22*, 134–143.
Nishisato, S. (1982). *Quantification of qualitative data*. Tokyo: Asakurashoten. (in Japanese)
Nishisato, S. (1988). Dual scaling: Its development and comparisons with other quantification methods. In H. D. Pressmar, K. E. Jager, H. Krallmann, H. Schellhaas, & L. Streitferdt (Eds.), *Deutsche Geselleschaft für operations research proceedings* (pp. 376–389). Berlin: Springer.
Nishisato, S., & Lawrence, D. R. (1981, May). *Dual scaling of multidimensional tables, a comparative study*. Paper presented at the annual meeting of the Psychometric Society, Chapel Hill, NC.
Nishisato, S., & Lawrence, D. R. (1989). Dual scaling of multiway data matrices: Several variants. In R. Coppi & S. Bolasco (Eds.) *Multiway data analysis* (pp. 317–326). Amsterdam: North Holland.
Nishisato, S., & Sheu, W. (1984). A note on dual scaling of successive categories data. *Psychometrika, 49*, 493–500.
Okamoto, M. (1972). Four techniques of principal component analysis. *Journal of Japanese Statistical Society, 2*, 63–69.
Potthoff, R. F., & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika, 51*, 313–326.
Ramsay, J. O. (1978). Confidence regions for multidimensional scaling analysis. *Psychometrika, 43*, 145–160.
Ramsay, J. O. (1980). Joint analysis of direct ratings, pairwise preferences and dissimilarities. *Psychometrika, 45*, 149–165.
Ramsay, J. O. (1989). Monotone regression splines in actions. *Statistical Science, 4*, 425–441.
Ramsay, J. O., ten Berge, J., & Styan, G. P. H. (1984). Matrix correlation. *Psychometrika, 49*, 403–423.
Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya A, 26*, 329–358.
Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika, 52*, 447–458.
Rao, C. R. (1979). Separation theorems for singular values of matrices and their applications in multivariate analysis. *Journal of Multivariate Analysis, 9*, 362–377.
Rao, C. R. (1980). Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (pp. 3–22). Amsterdam: North Holland.
Rumelhart, D. L., & Greeno, J. G. (1971). Similarity between stimuli: An experimental test of the Luce and Restle choice models. *Journal of Mathematical Psychology, 8*, 370–381.
Sabatier, R., Lebreton, J. D., & Chessel, D. (1989). Multivariate analysis of composition data accompanied by qualitative variables describing a structure. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 341–352). Amsterdam: North-Holland.
Shibayama, T. (1988). *Multivariate analysis of test scores with missing values*. Unpublished Doctoral Dissertation, University of Tokyo. (in Japanese)
Siotani, M., Hayakawa, T., & Fujikoshi, Y. (1985). *Modern multivariate statistical analysis: A graduate course handbook*. Columbus, OH: American Sciences Press.
Slater, P. (1960). The analysis of personal preferences. *The British Journal of Statistical Psychology, 13*, 119–135.
Sjöberg, L. (1967). Successive categories scaling of paired comparisons. *Psychometrika, 32*, 297–308.

Takane, Y. (1980). Maximum likelihood estimation in the generalized case of Thurstone's model of comparative judgment. *Japanese Psychological Research*, *22*, 188–196.

Takane, Y. (1987). Analysis of covariance structures and binary choice data. *Communication and Cognition*, *20*, 45–62.

Takane, Y., & Shibayama, T. (1988a). Three vector models of pairwise preference ratings and their generalizations. In S. Kashiwagi (Ed.), *Proceedings of the 16th Annual Meeting of the Behaviormetric Society* (pp. 131–132). Tokyo: Behaviormetric Society of Japan.

Takane, Y., & Shibayama, T. (1988b). Dual scaling with external criteria reconsidered. In S. Kashiwagi (Ed.), *Proceedings of the 16th Annual Meeting of the Behaviormetric Society* (pp. 133–134). Tokyo: Behaviormetric Society of Japan.

Takane, Y., Yanai, H., & Mayekawa, S. (in press). *Relationships among several methods of linearly constrained correspondence analysis. Psychometrika.*

Tanaka, Y. (1988). Sensitivity analysis in principal component analysis: Influence on the subspace spanned by principal components. *Communications in Statistics—Theory and Methods*, *17*, 3157–3175.

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

Tucker, L. R. (1959). Intra-individual and inter-individual multidimensionality. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling* (pp. 155–167). New York: Wiley.

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

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

Weinberg, S. L., Carroll, J. D., & Cohen, H. S. (1984). Confidence regions for INDSCAL using the jackknife and bootstrap techniques. *Psychometrika*, *49*, 475–491.

Wilkinson, J. H. (1965). *The algebraic eigenvalue problem*. Oxford: Oxford University Press.

Winsberg, S. (1988). Two techniques: Monotone spline transformations for dimension reduction in PCA and easy-to-generate metrics for PCA of sampled functions. In J. L. A. van Rijckevorsel & J. de Leeuw (Eds.), *Component and correspondence analysis* (pp. 115–135). New York: Wiley.

Winsberg, S., & Ramsay, J. O. (1983). Monotone spline transformations for dimension reduction. *Psychometrika*, *48*, 575–595.

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

Yanai, H. (1974). Unification of various techniques of multivariate analysis by means of generalized coefficients of determination. (G.C.D.) *Journal of Behaviormetrics*, *1*, 45–54. (in Japanese)

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

Manuscript received 5/9/89
Final version received 2/8/90