



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Linear Algebra and its Applications 388 (2004) 341–361

LINEAR ALGEBRA
AND ITS
APPLICATIONSwww.elsevier.com/locate/laa

Matrices with special reference to applications in psychometrics[☆]

Yoshio Takane

*Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montréal,
Que., Canada H3A 2K6*

Received 24 February 2002; accepted 18 February 2003

Submitted by H.J. Werner

Abstract

Multidimensional scaling, item response theory, and factor analysis may be considered three major contributions of psychometricians to statistics. Matrix theory played an important role in early developments of these techniques. Unfortunately, nonlinear models are currently very prevalent in these areas. Still, one can identify several areas of psychometrics where matrix algebra plays a prominent role. They include analysis of asymmetric square tables, multiway data analysis, reduced-rank regression analysis, and multiple-set (T -set) canonical correlation analysis among others. In this article we review some of the important matrix results in these areas and suggest future studies.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Multidimensional scaling; Singular value decomposition; Reduced-rank regression; Constrained principal component analysis; Different constraints on different dimensions; Multiple-set canonical correlation analysis; Wedderburn–Guttman theorem

1. Introduction

There were days when matrix algebraists and psychometricians were much more closely related. Mathematicians/statisticians used to publish their papers more often in substantive journals. Harold Hotelling, for example, published his papers on principal component analysis (PCA) in *Journal of Educational Psychology*. Alston

[☆] Invited lecture given at the Ninth International Workshop on Matrices and Statistics in Celebration of C.R. Rao's 80th Birthday held in Hyderabad, India in December 2000. The work reported in this paper has been supported by grant A6394 from the Natural Sciences and Engineering Research Council of Canada.
E-mail address: takane@takane2.psych.mcgill.ca (Y. Takane).

Householder published three papers in *Psychometrika* in 1937 alone (10 in total), although part of this could be due to the fact that he was at the University of Chicago around that period with more substantive interests. The University of Chicago was one of the central sites in psychometrics with L.L. Thurstone, founder of psychometrics, who had just started the Psychometric Society and its journal *Psychometrika*.

The following episode tells a close tie between psychometricians and mathematicians at the University of Chicago in early days of psychometrics. One day Thurstone was having lunch with a mathematician in school cafeteria, talking about factor analysis (FA). The mathematician told him that it was a matrix in form. Thurstone immediately realized the importance of matrix algebra in his work and started studying it, which later culminated in his book entitled “*Vectors of Mind*” (The University of Chicago Press, 1935). Some mathematicians at the university, including Carl Eckart, Gale Young, and Alston Householder, also got interested in psychometric research and published some of their papers in *Psychometrika*. Other prominent statisticians (not necessarily at the University of Chicago) such as T.W. Anderson, Quinn McNemar, Frederick Mosteller, C.R. Rao, John Tukey, and S.S. Wilks have also contributed one or more papers to *Psychometrika*.

In the meantime, matrix theory has sufficiently pervaded virtually every aspect of psychometrics and its knowledge has become common sense among psychometricians. Somewhat ironically, however, with the advancement of specializations and the publications of many more new journals specialized in one specific area, those days are long gone when mathematicians and statisticians looked for journals outside their own disciplines to publish their work. This is a bit unfortunate state of affairs because it is getting increasingly more difficult to keep track of more recent developments in matrix algebra useful in psychometrics. The Psychometric Society is making every effort, however, to provide its members the opportunity to keep their knowledge in matrix algebra up to date. For example, the Society invited Ingram Olkin to its 1996 annual meeting to deliver an insightful lecture on “Interface between multivariate analysis and matrix theory”.

This article has a somewhat reversed role (to that of Olkin’s). I am a psychometrician addressing to mathematicians and statisticians, reviewing some of the areas in psychometrics where matrix algebra plays an essential role. Along the way, I would like to suggest some of the interesting matrix algebra problems yet to be solved and encourage further studies. Areas of psychometrics to be discussed in this article include multidimensional scaling (MDS) (with special emphasis on analysis of asymmetric square tables and multiway data analysis), various extensions of reduced-rank regression analysis, multiple-set (T -set) canonical correlation analysis (CANO), and the Wedderburn–Guttman theorem.

2. Multidimensional scaling

MDS is a data analysis technique to locate a set of points in a multidimensional space in such a way that points corresponding to similar stimuli are located close

together, while those corresponding to dissimilar stimuli are located far apart. Many road maps, for example, have a matrix of intercity distances. Put simply, MDS recovers a map based on the intercity distances. Given a map it is relatively straightforward to measure the distances between cities. However, the reverse operation, that of recovering a map from a given set of distances is not as straightforward. MDS is a method to perform this reverse operation [65].

A variety of MDS procedures have been developed, depending on the kind of similarity data analyzed, the form of functional relationships assumed between the observed data and the distance model, the type of fitting criteria used, etc. During the past 40 years or so, however, a form of MDS called nonmetric MDS [42,43,63] has been very popular because of its flexibility. In this paper, however, we focus on foundational aspects of MDS that can be easily seen through simple matrix manipulations.

2.1. MDS for a single square symmetric table

The reverse operation mentioned above is particularly simple when the set of error-free Euclidean distances are given between stimuli. Let x_{ir} denote the coordinate of point i ($i = 1, \dots, n$) on dimension r ($r = 1, \dots, p$). The squared Euclidean distance between points i and j is then given by $d_{ij}^2 = \sum_{r=1}^p (x_{ir} - x_{jr})^2$. Let \mathbf{X} denote an n by p ($\leq n - 1$) matrix of stimulus coordinates, assumed nonsingular. Then, the matrix of squared Euclidean distances, $\mathbf{D}^{(2)}(\mathbf{X})$, between stimuli can be expressed as

$$\mathbf{D}^{(2)}(\mathbf{X}) = \mathbf{1}_n \mathbf{1}_n' \text{diag}(\mathbf{X}\mathbf{X}') - 2\mathbf{X}\mathbf{X}' + \text{diag}(\mathbf{X}\mathbf{X}')\mathbf{1}_n \mathbf{1}_n', \tag{1}$$

where $\mathbf{1}_n$ is an n -element vector of ones. Define $\mathbf{S} = (-1/2)\mathbf{J}_n \mathbf{D}^{(2)}(\mathbf{X}) \mathbf{J}_n$, where $\mathbf{J}_n = (\mathbf{I}_n - \mathbf{1}_n \mathbf{1}_n' / n)$. Then, $\mathbf{S} = \mathbf{J}_n \mathbf{X}\mathbf{X}' \mathbf{J}_n = \mathbf{X}\mathbf{X}'$, where it is assumed that $\mathbf{J}_n \mathbf{X} = \mathbf{X}$. (The origin of the space is placed at its centroid.) The matrix of stimulus coordinates (\mathbf{X}) can be obtained by a square root decomposition of \mathbf{S} . Note that $\text{rank}(\mathbf{S}) = \text{rank}(\mathbf{X}) = p$.

The above procedure suggests the following theorem known as the Young–Householder theorem [60,96].

Theorem 2.1. *A set of dissimilarities, $\{\delta_{ij}\}$, defined on the set of pairs of n stimuli can be embedded in the irreducible p -dimensional Euclidean space if and only if $\widehat{\mathbf{S}} = (-1/2)\mathbf{J}_n \Delta^{(2)} \mathbf{J}_n$ is positive semi-definite (PSD) of rank p , where $\Delta^{(2)}$ is the matrix of δ_{ij}^2 .*

More generally, let $\widehat{\mathbf{S}}$ denote a matrix of observed “similarities” between n stimuli. The stimuli can be embedded in the p -dimensional (but not less than p -dimensional) Euclidean space if and only if $\widehat{\mathbf{S}}$ is PSD of rank p [21]. Let $\widehat{\mathbf{S}} = \mathbf{X}\mathbf{X}'$ be a square root decomposition of $\widehat{\mathbf{S}}$, where \mathbf{X} is n by p , and nonsingular. Then, the matrix of squared Euclidean distances between the stimuli are given by (1).

The exact reverse operation presented above strictly applies to only error-free data. However, a similar procedure can be used in fallible cases as well [82]. This method obtains the best rank p approximation (in the least squares (LSs) sense) of $\widehat{\mathbf{S}}$ by the eigenvalue-vector decomposition of $\widehat{\mathbf{S}}$. This method is now known as classical MDS. More recently, this approach has been extended to nonmetric MDS by Trosset [83] (see [11] for a more comprehensive review of literature in MDS).

Various extensions of the above scheme of MDS are possible. In the following subsections, three such extensions are considered: (1) MDS for a rectangular table, (2) MDS for $m (\geq 1)$ square symmetric tables, and (3) MDS for a square asymmetric table.

2.2. MDS for a rectangular table

In the above discussion, there is only one set of objects (stimuli) between which dissimilarities are observed. In some cases, however, dissimilarities are defined between objects that belong to two distinct sets. Such data often arise when a group of m subjects make preference judgments on a set of n stimuli, and the preference data are assumed inversely related to the distances between subjects' ideal stimuli and the actual stimuli. MDS designed for such situations is called unfolding analysis [10].

Let y_{kr} denote the coordinate of subject k 's ideal point on dimension r , and let x_{ir} denote the coordinate of stimulus i on dimension r . Then, the squared Euclidean distance between them is given by $d_{ki}^2 = \sum_{r=1}^p (y_{kr} - x_{ir})^2$. Let \mathbf{Y} ($m \times p$) and \mathbf{X} ($n \times p$) denote the matrices of y_{kr} and x_{ir} , respectively. Then, similarly to (1), the m by n matrix of squared Euclidean distances, $\mathbf{D}^{(2)}(\mathbf{Y}, \mathbf{X})$, between subjects' ideal points and stimulus points can be expressed as

$$\mathbf{D}^{(2)}(\mathbf{Y}, \mathbf{X}) = \mathbf{1}_m \mathbf{1}_n' \text{diag}(\mathbf{X}\mathbf{X}') - 2\mathbf{Y}\mathbf{X}' + \text{diag}(\mathbf{Y}\mathbf{Y}') \mathbf{1}_m \mathbf{1}_n', \quad (2)$$

where $\mathbf{1}_m$ is an m -element vector of ones.

There is an exact reverse operation applicable to this case, which is similar to the one discussed above. The method "recovers" \mathbf{Y} and \mathbf{X} from $\mathbf{D}^{(2)}(\mathbf{Y}, \mathbf{X})$ [61]. Let

$$\mathbf{Z} = (-1/2)\mathbf{J}_m \mathbf{D}^{(2)}(\mathbf{Y}, \mathbf{X}) \mathbf{J}_n = \mathbf{J}_m \mathbf{Y}\mathbf{X}' \mathbf{J}_n = \mathbf{Y}^* \mathbf{X}^*,$$

where $\mathbf{J}_m = \mathbf{I}_m - \mathbf{1}_m \mathbf{1}_m' / m$, $\mathbf{Y}^* = \mathbf{J}_m \mathbf{Y}$, and $\mathbf{X}^* = \mathbf{J}_n \mathbf{X}$. By rank factorization, $\mathbf{Z} = \widetilde{\mathbf{Y}} \widetilde{\mathbf{X}}'$, where $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{X}}$ are m by p and n by p nonsingular matrices. The origin of the space may be set at the centroid of \mathbf{X}^* , and the origin of \mathbf{Y}^* is to be adjusted accordingly. For a square nonsingular matrix \mathbf{T} of order p and a p -element translation vector \mathbf{y}_0 ,

$$\mathbf{X} = \mathbf{X}^* = \widetilde{\mathbf{X}} \mathbf{T},$$

and

$$\mathbf{Y} = \mathbf{Y}^* + \mathbf{1}_m \mathbf{y}_0' = \widetilde{\mathbf{Y}} \mathbf{T}^{-1} + \mathbf{1}_m \mathbf{y}_0'.$$

Matrix \mathbf{T} and vector \mathbf{y}_0 can be found by putting the above expressions of \mathbf{X} and \mathbf{Y} into (2). Again, this method can only be applied to error-free data. Due to the

additional steps needed, the procedure is rather sensitive to errors. This is in contrast to the similar procedure for a square symmetric table discussed earlier. Some remedial measures have been suggested by Gold [20] to make the method more robust against errors. See Heiser and Meulman [30] for more recent developments and issues surrounding the unfolding analysis.

2.3. MDS for m square symmetric tables

So far, it is assumed that there is a single set of dissimilarity data, either square or rectangular. In some cases, however, there are m sets of square symmetric data obtained from, say, m individuals. MDS applicable to such data is called individual differences MDS. One useful technique for individual differences MDS represents both commonality and uniqueness in such data sets by the weighted Euclidean distance model [5].

Let x_{ir} denote the coordinate of stimulus i on dimension r , and let w_{kr} denote the weight individual k attaches to dimension r . Then, the squared weighted Euclidean distance between stimuli i and j for individual k is given by $d_{ijk}^2 = \sum_{r=1}^p w_{kr}(x_{ir} - x_{jr})^2$. The matrix of d_{ijk}^2 can be expressed, analogously to (1), as

$$\mathbf{D}_k^{(2)}(\mathbf{X}, \mathbf{W}_k) = \mathbf{1}_n \mathbf{1}'_n \text{diag}(\mathbf{X} \mathbf{W}_k \mathbf{X}') - 2 \mathbf{X} \mathbf{W}_k \mathbf{X}' + \text{diag}(\mathbf{X} \mathbf{W}_k \mathbf{X}') \mathbf{1}_n \mathbf{1}'_n \quad (3)$$

for $k = 1, \dots, m$, where \mathbf{X} ($n \times p$) is the matrix of x_{ir} , and \mathbf{W}_k ($p \times p$) is the diagonal matrix of w_{kr} , assumed to be nonnegative definite (NND). For identification, it is convenient to require that $\text{diag}(\mathbf{X}' \mathbf{X}) = \mathbf{I}_p$. This model attempts to explain differences among the sets of dissimilarities defined on a same set of stimuli by differential weighting of dimensions by different individuals.

Again, there is an exact reverse operation for this model [62]. Let $\mathbf{S}_k = (-1/2) \mathbf{J}_n \mathbf{D}_k^{(2)}(\mathbf{X}, \mathbf{W}_k) \mathbf{J}_n$. Then, $\mathbf{S}_k = \mathbf{X} \mathbf{W}_k \mathbf{X}'$, where $\mathbf{J}_n \mathbf{X} = \mathbf{X}$ is assumed. Let

$$\bar{\mathbf{S}} = \sum_{k=1}^m \mathbf{S}_k / m = \mathbf{X} \left(\sum_{k=1}^m \mathbf{W}_k / m \right) \mathbf{X}' = \mathbf{X} \mathbf{X}'$$

where it is temporarily assumed that $\sum_{k=1}^m \mathbf{W}_k / m = \mathbf{I}_p$. (This can be done without loss of generality; it amounts to using $\sum_{k=1}^m \tilde{\mathbf{W}}_k / m = \mathbf{I}_p$ as the identification restriction.) By a square root decomposition, $\bar{\mathbf{S}} = \tilde{\mathbf{X}} \tilde{\mathbf{X}}'$. Then, for some p by p orthogonal matrix \mathbf{T} , $\mathbf{X} = \tilde{\mathbf{X}} \mathbf{T}$. Let

$$\bar{\mathbf{S}} = \sum_{k=1}^m \mathbf{S}_k e_k$$

denote a linear combination of \mathbf{S}_k . Then,

$$\bar{\mathbf{S}} = \mathbf{X} \left(\sum_{k=1}^m \mathbf{W}_k e_k \right) \mathbf{X}' = \mathbf{X} \bar{\mathbf{W}} \mathbf{X}'$$

where $\bar{\mathbf{W}} = \sum_{k=1}^m \mathbf{W}_k e_k$. Then,

$$\bar{\mathbf{W}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\bar{\mathbf{S}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{T}'(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}'\bar{\mathbf{S}}\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \mathbf{T},$$

since $\mathbf{T}^{-1} = \mathbf{T}'$ and $\mathbf{T}'\mathbf{T} = \mathbf{I}_p$. That is, $\mathbf{C} = \mathbf{T}\bar{\mathbf{W}}\mathbf{T}'$, where $\mathbf{C} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}'\bar{\mathbf{S}}\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}$. Matrices \mathbf{T} and $\bar{\mathbf{W}}$ is given by the eigenvalue-vector decomposition of \mathbf{C} . For this decomposition to be unique there must be at least one linear combination of \mathbf{W}_k such that the diagonal elements of $\bar{\mathbf{W}}$ are all distinct. Again, this procedure can only be applied to infallible data. Iterative procedures for fallible data have been developed by Carroll and Chang [5] and de Leeuw and Pruzansky [12].

The data analyzed by individual differences MDS are three-way (stimuli by stimuli by individuals). Psychometrics has a long tradition in dealing with multiway data, starting from Tucker's [84] three-mode FA, Harshman's [26] PARAFAC (parallel FA), etc. The latter is a kind of three-way component analysis postulating $\mathbf{Z}_k = \mathbf{Y}\mathbf{W}_k\mathbf{X}'$ ($k = 1, \dots, m$) for a rectangular matrix \mathbf{Z}_k . An iterative parameter estimation procedure has been developed for PARAFAC [41,58] as well as for Tucker's three-mode FA. Interesting results are also due to psychometricians on some algebraic properties of multiway tables (e.g., ranks of multiway tables), see [44,45,79,80], and references therein for further details.

In statistics, a model similar to $\mathbf{S}_k = \mathbf{X}\mathbf{W}_k\mathbf{X}'$ has been proposed by Flury [17] with the additional restriction that $\mathbf{X}'\mathbf{X} = \mathbf{I}$, and is called "Common Principal Component Analysis". Carroll and Chang also proposed a model called IDIOSCAL (individual differences in orientation scaling), where the NND diagonal matrix \mathbf{W}_k in the weighted Euclidean distance model was replaced by an NND matrix \mathbf{C}_k .

2.4. MDS for a square asymmetric table

Relationships between stimuli are often asymmetric. For example, the degree to which person A likes B is not necessarily the same as the degree to which person B likes A. Such examples of asymmetric relationships abound in psychology and elsewhere, for example, mobility tables, stimulus identification data, brand switching data, journal citation data, husband's and wife's occupations in two-earner families, etc.

A variety of models that capture asymmetries in the data have been proposed, some of which will be briefly discussed below:

1. *DEDICOM (DEcomposing Directional COMponent)* [27]. Let \mathbf{A} denote an n by n asymmetric table. DEDICOM postulates $\mathbf{A} = \mathbf{X}\mathbf{R}\mathbf{X}' + \mathbf{E}$, where \mathbf{X} is an n by p ($< n$) matrix, \mathbf{R} is a square asymmetric matrix of order p , and \mathbf{E} is a matrix of residuals. This model attempts to explain asymmetric relationships between pairs of n objects by a smaller number (p) of asymmetric relationships (represented by \mathbf{R}), and by their relations to the objects (represented by \mathbf{X}). In the infallible case ($\mathbf{E} = \mathbf{0}$), the model implies that $\text{Sp}(\mathbf{A}) = \text{Sp}(\mathbf{A}')$, which always holds for $p = n$. For $\text{rank}(\mathbf{A}) = p < n$, this condition characterizes the falsifiable DEDICOM model.

A closed form solution exists in this case [39]. An iterative algorithm has also been developed for fallible data [39].

2. *Generalized GIPSCAL* [38]. An asymmetric square matrix \mathbf{A} can be generally expressed as the sum of symmetric and skew-symmetric parts: $\mathbf{A} = \mathbf{S}_s + \mathbf{S}_{sk}$, where $\mathbf{S}_s = (\mathbf{A} + \mathbf{A}')/2$, and $\mathbf{S}_{sk} = (\mathbf{A} - \mathbf{A}')/2$ with $\mathbf{S}'_s = \mathbf{S}_s$, and $\mathbf{S}'_{sk} = -\mathbf{S}_{sk}$. In the DEDICOM model, $\mathbf{X}\mathbf{R}\mathbf{X}'$ can also be decomposed in a similar way: $\mathbf{X}\mathbf{R}\mathbf{X}' = \mathbf{X}\mathbf{R}_s\mathbf{X}' + \mathbf{X}\mathbf{R}_{sk}\mathbf{X}'$, where $\mathbf{R}_s = (\mathbf{R} + \mathbf{R}')/2$, and $\mathbf{R}_{sk} = (\mathbf{R} - \mathbf{R}')/2$. It can be further rewritten as $\mathbf{X}\mathbf{R}\mathbf{X}' = \tilde{\mathbf{X}}(\mathbf{I}_p + \mathbf{K})\tilde{\mathbf{X}}'$ if and only if \mathbf{R}_s is positive definite (PD), where \mathbf{K} contains 2 by 2 blocks of the form $\begin{pmatrix} 0 & k_l \\ -k_l & 0 \end{pmatrix}$ along the diagonal when p is even. There is an additional 0 diagonal entry when n is odd. This model is called generalized GIPSCAL.

3. *CASK (Canonical Analysis of SKew-symmetric Data)* [22]. As a method for analyzing square asymmetric tables, CASK precedes all other methods discussed in this section. However, it analyzes only skew-symmetric data (or that part of data). The SVD of \mathbf{S}_{sk} yields $\mathbf{S}_{sk} = \mathbf{P}\mathbf{D}\mathbf{Q}'$. Singular values of a skew-symmetric matrix come in pairs (except for the one extra zero singular value obtained when n is odd), and $\mathbf{P}\mathbf{D}\mathbf{Q}'$ can be further rewritten as $\mathbf{P}\mathbf{D}\mathbf{L}\mathbf{P}' = \mathbf{X}\mathbf{K}\mathbf{X}'$, where $\mathbf{X} = \mathbf{P} = \mathbf{Q}\mathbf{L}$, $\mathbf{K} = \mathbf{D}\mathbf{L}$, and \mathbf{L} is similar in form to \mathbf{K} above except that all k_l 's are unities.

4. *HCM (Hermitian Canonical Model)* [15]. Form an hermitian matrix by $\mathbf{H} = \mathbf{S}_s + i\mathbf{S}_{sk}$. HCM obtains the eigenvalue decomposition of \mathbf{H} . Assume that \mathbf{H} is NND. Then,

$$\mathbf{H} = \tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{U}}^* = \mathbf{U}\mathbf{U}^*, \tag{4}$$

where $*$ indicates a conjugate transpose, and $\mathbf{U} = \tilde{\mathbf{U}}\mathbf{D}^{1/2}$. Let $\mathbf{U} = \mathbf{X} + i\mathbf{Y}$. Then, $\mathbf{H} = (\mathbf{X}\mathbf{X}' + \mathbf{Y}\mathbf{Y}') + i(\mathbf{Y}\mathbf{X}' - \mathbf{X}\mathbf{Y}')$, where $\mathbf{X}\mathbf{X}' + \mathbf{Y}\mathbf{Y}' = \mathbf{S}_s$, and $\mathbf{Y}\mathbf{X}' - \mathbf{X}\mathbf{Y}' = \mathbf{S}_{sk}$ with $\mathbf{X}'\mathbf{X} + \mathbf{Y}'\mathbf{Y} = \mathbf{I}$, and $\mathbf{Y}'\mathbf{X} = \mathbf{X}'\mathbf{Y}$ (symmetric). The following theorem due to Chino and Shiraiwa [7] is an extension of the Young–Householder theorem (Theorem 2.1) to a finite dimensional complex Hilbert space.

Theorem 2.2. Let $\mathbf{V} = [\mathbf{X}, \mathbf{Y}]$, and $\bar{\mathbf{V}} = \mathbf{V}\bar{\mathbf{L}}$, where $\bar{\mathbf{L}} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{pmatrix}$, and \mathbf{X} and \mathbf{Y} are as defined above. Matrix $\bar{\mathbf{L}}$ has the effect of rotating \mathbf{V} counter clockwise by 90° . (Note that $\overline{\mathbf{V}\mathbf{V}'} = \mathbf{V}\mathbf{V}'$.) Let

$$\begin{aligned} \mathbf{D}^{(2)} &= \mathbf{D}^{(2)}(\mathbf{V}) = \mathbf{1}_n\mathbf{1}'_n \text{diag}(\mathbf{X}\mathbf{X}' + \mathbf{Y}\mathbf{Y}') - 2(\mathbf{X}\mathbf{X}' + \mathbf{Y}\mathbf{Y}') \\ &\quad + \text{diag}(\mathbf{X}\mathbf{X}' + \mathbf{Y}\mathbf{Y}')\mathbf{1}_n\mathbf{1}'_n \\ &= \mathbf{1}_n\mathbf{1}'_n \text{diag}(\mathbf{S}_s) - 2\mathbf{S}_s + \text{diag}(\mathbf{S}_s)\mathbf{1}_n\mathbf{1}'_n, \end{aligned} \tag{5}$$

and

$$\begin{aligned} \bar{\mathbf{D}}^{(2)} &= \mathbf{D}^{(2)}(\mathbf{V}, \bar{\mathbf{V}}) = \mathbf{1}_n\mathbf{1}'_n \text{diag}(\mathbf{X}\mathbf{X}' + \mathbf{Y}\mathbf{Y}') - 2(\mathbf{Y}\mathbf{X}' - \mathbf{X}\mathbf{Y}') \\ &\quad + \text{diag}(\mathbf{X}\mathbf{X}' + \mathbf{Y}\mathbf{Y}')\mathbf{1}_n\mathbf{1}'_n \\ &= \mathbf{1}_n\mathbf{1}'_n \text{diag}(\mathbf{S}_s) - 2\mathbf{S}_{sk} + \text{diag}(\mathbf{S}_s)\mathbf{1}_n\mathbf{1}'_n. \end{aligned} \tag{6}$$

Then, $\mathbf{H} = (-1/2)\mathbf{J}_n(\mathbf{D}^{(2)} + i\overline{\mathbf{D}}^{(2)})\mathbf{J}_n = (\mathbf{X}\mathbf{X}' + \mathbf{Y}\mathbf{Y}') + i(\mathbf{Y}\mathbf{X}' - \mathbf{X}\mathbf{Y}') = \mathbf{S}_s + i\mathbf{S}_{sk}$ (this is an instance of the polar identity in the finite dimensional complex Hilbert space) is PSD of rank p .

Conversely, if \mathbf{H} formed from \mathbf{A} by $\mathbf{H} = \mathbf{S}_s + i\mathbf{S}_{sk}$ where $\mathbf{A} = \mathbf{S}_s + \mathbf{S}_{sk}$ is PSD of rank p , a set of stimuli whose “similarities” are defined by \mathbf{A} can be represented as points in the irreducible p -dimensional complex Hilbert space, where interpoint distances are given by (5).

Let θ_{ij} denote the angle (in radian) between points i and j with respect to the origin (0) of the space for a particular (real–imaginary) pair of corresponding dimensions. Then, $h_{ij} = d_{i0}d_{j0}(\cos \theta_{ij} + i \sin \theta_{ij}) = d_{i0}d_{j0} \exp(-i\theta_{ij})$. Thus, $\theta_{ij} = (-1/2)i \log(h_{ij}/h_{ji})$.

MDS of asymmetric data began only 20 years ago or so and is still in a maturing stage. There are still a lot of things left to be done, including developments of fitting procedures under various distributional assumptions on observed data, under various measurement characteristics of the data, etc.

3. Singular value decomposition

Singular value decomposition (SVD) [2,14,37,47,59] continues to play a central role in many multivariate data analysis techniques used in psychometrics. Although there are many proofs of optimalities of SVD (e.g., [54]), we will briefly discuss ten Berge’s [78] proofs based on the notion of *sub-orthogonal* (s.o.) matrices.

There are two major uses of SVD: (1) finding the best reduced-rank approximation to a matrix and (2) finding the best orthogonal approximation to a matrix. The first kind of optimality has been widely recognized in statistics, while the second kind is not well known outside the psychometric community. ten Berge [78] proves both kinds of optimality quite elegantly.

Kristof [40] obtained an upper bound of the following function:

$$f(\mathbf{B}_1, \dots, \mathbf{B}_T) = \text{tr} \left(\prod_{j=1}^T \mathbf{B}_j \mathbf{C}_j \right), \quad (7)$$

where \mathbf{C}_j and \mathbf{B}_j ($j = 1, \dots, T$) are a diagonal matrix and an orthogonal matrix of order n , respectively. Kristof’s result is a generalization of von Neumann’s [89] theorem for $T = 1$ and $T = 2$. Only the case of $T = 1$ is necessary for the present purpose. For $T = 1$, Neumann’s theorem can be stated as follows:

Let $\mathbf{C}_1 = \mathbf{C}$ be NND, and write $\mathbf{B}_1 = \mathbf{B}$. Then, $f(\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{C}) \leq \text{tr}(\mathbf{C})$ because $\text{tr}(\mathbf{B}\mathbf{C}) = \sum_{i=1}^n b_{ii}c_i \leq \sum_{i=1}^n c_i = \text{tr}(\mathbf{C})$. Note that $b_{ii} \leq 1$ for all i since \mathbf{B} is orthogonal.

ten Berge [77] generalized this theorem to a s.o. matrix \mathbf{B} .

Definition 3.1. A matrix is s.o. if it can be completed to an orthogonal matrix by appending rows or columns, or both. Every s.o. matrix can be viewed as a submatrix of some orthogonal matrix.

Property 3.1. Every columnwise or rowwise orthogonal matrix is s.o. This is because it can readily be completed to be orthogonal.

Property 3.2. The product of any two s.o. matrices is also s.o.

Theorem 3.1 [77]. If \mathbf{B} is an $n \times n$ s.o. matrix of rank $p \leq n$, and \mathbf{C} is diagonal, with diagonal elements $c_1 \geq c_2 \geq \dots \geq c_n \geq 0$, then

$$f(\mathbf{B}) = \text{tr}(\mathbf{BC}) \leq c_1 + \dots + c_p, \tag{8}$$

which is the sum of the p largest elements in \mathbf{C} . This upper bound is attained for the s.o. matrix $\mathbf{B} = \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$.

ten Berge [78] used the above theorem to show the following two results:

(1) Let \mathbf{Z} denote an m by n matrix of rank p , and consider the problem of approximating \mathbf{Z} by another matrix \mathbf{Z}_0 of the same size but of a lower rank. That is, finding \mathbf{Z}_0 such that $f(\mathbf{Z}_0) = \text{SS}(\mathbf{Z} - \mathbf{Z}_0)$ is minimized subject to $\text{rank}(\mathbf{Z}_0) = q \leq p$. Since $\text{rank}(\mathbf{Z}_0) = q$, it can be written as $\mathbf{Z}_0 = \mathbf{FA}'$, where \mathbf{F} (m by q) is columnwise orthogonal, and \mathbf{A} is an n by q matrix of rank q . Minimizing $f(\mathbf{Z}_0)$ with respect to \mathbf{A} for fixed \mathbf{F} leads to $\mathbf{A} = \mathbf{F}'\mathbf{Z}$. Then,

$$f^*(\mathbf{F}) = \min_{\mathbf{A}|\mathbf{F}} f(\mathbf{Z}_0) = \text{SS}(\mathbf{Z} - \mathbf{FF}'\mathbf{Z}). \tag{9}$$

The minimum of $f(\mathbf{Z}_0)$ can be obtained by minimizing $f^*(\mathbf{F})$ with respect to \mathbf{F} . This is equivalent to maximizing $\text{tr}(\mathbf{F}'\mathbf{ZZ}'\mathbf{F})$ with respect to \mathbf{F} . Let the (incomplete) SVD of \mathbf{Z} be denoted by $\mathbf{Z} = \mathbf{UDV}'$. Then, $\mathbf{ZZ}' = \mathbf{UD}^2\mathbf{U}'$ and $\text{tr}(\mathbf{F}'\mathbf{ZZ}'\mathbf{F}) = \text{tr}(\mathbf{F}'\mathbf{UD}^2\mathbf{U}'\mathbf{F}) = \text{tr}(\mathbf{U}'\mathbf{FF}'\mathbf{UD}^2) \leq \text{tr}(\mathbf{D}^2)$ by ten Berge's theorem. The maximum is attained when $\mathbf{F} = \mathbf{U}_q\mathbf{T}$, where \mathbf{U}_q is the portion of \mathbf{U} pertaining to the q largest singular values of \mathbf{Z} , and \mathbf{T} is an arbitrary orthogonal matrix of order q . Note that

$$\text{when } \mathbf{F} = \mathbf{U}_q\mathbf{T}, \mathbf{U}'\mathbf{FF}'\mathbf{U} = \mathbf{U}'\mathbf{U}_q\mathbf{TT}'\mathbf{U}_q\mathbf{U} = \begin{pmatrix} \mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

(2) Let \mathbf{Z} denote an m by n nonsingular matrix, and consider the problem of approximating \mathbf{Z} by a columnwise orthogonal matrix \mathbf{Z}_1 of the same order. That is, finding \mathbf{Z}_1 such that $f(\mathbf{Z}_1) = \text{SS}(\mathbf{Z} - \mathbf{Z}_1)$ is minimized subject to $\mathbf{Z}'_1\mathbf{Z}_1 = \mathbf{I}$. First, note that minimizing $f(\mathbf{Z}_1)$ is equivalent to maximizing $\text{tr}(\mathbf{Z}'\mathbf{Z}_1)$. Let the (incomplete) SVD of \mathbf{Z} be denoted by $\mathbf{Z} = \mathbf{UDV}'$. Then, $\text{tr}(\mathbf{Z}'\mathbf{Z}_1) = \text{tr}(\mathbf{VDU}'\mathbf{Z}_1) = \text{tr}(\mathbf{U}'\mathbf{Z}_1\mathbf{VD}) \leq \text{tr}(\mathbf{D})$. The maximum is attained by $\mathbf{Z}_1 = \mathbf{UV}' = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1/2}$.

This second use of SVD is popular in psychometrics where the best orthogonal transformation (rotation) of a stimulus configuration and of a factor loading matrix is looked for that facilitates interpretation.

4. Reduced-rank regression models

SVD plays one of the two most crucial roles in constrained PCA (CPCA) proposed by Takane and Shibayama [72] (see also [35,67]). This technique incorporates external information into PCA by first decomposing the data matrix according to the external information, and then applying PCA to decomposed matrices. The former amounts to orthogonal projections of the data matrix onto the spaces spanned by the matrices of external information (often called design matrices, constraint matrices, etc.), while the latter involves SVD or generalized SVD (GSVD). CPCA subsumes a number of existing techniques as its special cases.

Let \mathbf{Z} be an m by n data matrix, and let \mathbf{G} and \mathbf{H} be m by u and n by v matrices of external information on rows and columns of the data matrix, respectively. CPCA postulates the following model for \mathbf{Z} ,

$$\mathbf{Z} = \mathbf{GMH}' + \mathbf{BH}' + \mathbf{GC} + \mathbf{E}, \quad (10)$$

where \mathbf{M} (u by v), \mathbf{B} (m by v), and \mathbf{C} (u by n) are matrices of unknown parameters, and \mathbf{E} (m by n) a matrix of residuals. To identify the model, it is convenient to require

$$\mathbf{G}'\mathbf{KB} = \mathbf{0}, \quad (11)$$

and

$$\mathbf{CLH} = \mathbf{0}, \quad (12)$$

where \mathbf{K} and \mathbf{L} denote the row and column metric (weight) matrices, respectively. For simplicity, it is assumed that both \mathbf{K} and \mathbf{L} are symmetric PD. (Takane and Hunter [67] discuss the more general case in which \mathbf{K} and \mathbf{L} are possibly singular.) The first term in model (10) pertains to the portions of the data matrix that 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 by \mathbf{H} .

Model parameters are estimated in such a way that the following extended (weighted) LSs criterion is minimized:

$$f = \text{SS}(\mathbf{E})_{\mathbf{K},\mathbf{L}} = \text{tr}(\mathbf{E}'\mathbf{KEL}). \quad (13)$$

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

$$\hat{\mathbf{M}} = (\mathbf{G}'\mathbf{KG})^{-1}\mathbf{G}'\mathbf{KZLH}(\mathbf{H}'\mathbf{LH})^{-1}, \quad (14)$$

$$\hat{\mathbf{B}} = \mathbf{Q}_{\mathbf{G}/\mathbf{K}}\mathbf{ZLH}(\mathbf{H}'\mathbf{LH})^{-1}, \quad (15)$$

$$\hat{\mathbf{C}} = (\mathbf{G}'\mathbf{KG})^{-1}\mathbf{G}'\mathbf{KZQ}'_{\mathbf{H}/\mathbf{L}}, \quad (16)$$

$$\hat{\mathbf{E}} = \mathbf{Q}_{\mathbf{G}/\mathbf{K}}\mathbf{ZQ}'_{\mathbf{H}/\mathbf{L}}, \quad (17)$$

where $\mathbf{Q}_{\mathbf{G}/\mathbf{K}} = \mathbf{I} - \mathbf{P}_{\mathbf{G}/\mathbf{K}}$ and $\mathbf{P}_{\mathbf{G}/\mathbf{K}} = \mathbf{G}(\mathbf{G}'\mathbf{KG})^{-1}\mathbf{G}'\mathbf{K}$ are orthogonal projectors onto $\text{Ker}(\mathbf{G}')$ (the null space of \mathbf{G}') and $\text{Sp}(\mathbf{G})$ (the range space of \mathbf{G}), respectively, in

the metric of \mathbf{K} , and $\mathbf{Q}_{H/L} = \mathbf{I} - \mathbf{P}_{H/L}$ and $\mathbf{P}_{H/L} = \mathbf{H}(\mathbf{H}'\mathbf{L}\mathbf{H})^{-1}\mathbf{H}\mathbf{L}$ are orthogonal projectors onto $\text{Ker}(\mathbf{H}')$ (the null space of \mathbf{H}') and $\text{Sp}(\mathbf{H})$ (the range space of \mathbf{H}), respectively, in the metric of \mathbf{L} [56,94]. Putting the above expressions in (10) leads to the following decomposition of the data matrix, \mathbf{Z} :

$$\mathbf{Z} = \mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L} + \mathbf{Q}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L} + \mathbf{P}_{G/K}\mathbf{Z}\mathbf{Q}'_{H/L} + \mathbf{Q}_{G/K}\mathbf{Z}\mathbf{Q}'_{H/L}. \quad (18)$$

The four terms on the right hand side of (18) correspond to the four terms in model (10). Because of the trace-orthogonality of the four terms in (18), the total SS in \mathbf{Z} is uniquely decomposed into the sum of component sums of squares, namely

$$\begin{aligned} \text{SS}(\mathbf{Z})_{K,L} &= \text{SS}(\mathbf{P}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L})_{K,L} + \text{SS}(\mathbf{Q}_{G/K}\mathbf{Z}\mathbf{P}'_{H/L})_{K,L} \\ &+ \text{SS}(\mathbf{P}_{G/K}\mathbf{Z}\mathbf{Q}'_{H/L})_{K,L} + \text{SS}(\mathbf{Q}_{G/K}\mathbf{Z}\mathbf{Q}'_{H/L})_{K,L}. \end{aligned} \quad (19)$$

Two matrices, \mathbf{X} and \mathbf{Y} , are said to be trace-orthogonal, when $\text{tr}(\mathbf{X}'\mathbf{K}\mathbf{Y}\mathbf{L}) = 0$ for given metric matrices, \mathbf{K} and \mathbf{L} .

The decomposed matrices in (18) are subjected to PCA either separately or jointly (i.e., recombining some of them together). This leads to the (GSVD/SVD) of a matrix with certain metric matrices.

Definition 4.1 (GSVD). Let \mathbf{K} and \mathbf{L} denote PD matrices of orders m and n , respectively. Let \mathbf{A} be an m by n matrix of rank p . Then,

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}' \quad (20)$$

is called the (incomplete) GSVD of \mathbf{A} under the metric matrices \mathbf{K} and \mathbf{L} and is written as $\text{GSVD}(\mathbf{A})_{K,L}$, where $\mathbf{U}'\mathbf{K}\mathbf{U} = \mathbf{I}_p = \mathbf{V}'\mathbf{L}\mathbf{V}$, and \mathbf{D} is diagonal and PD.

$\text{GSVD}(\mathbf{A})_{K,L}$ can be obtained as follows. Let $\mathbf{K} = \mathbf{R}_K\mathbf{R}'_K$ and $\mathbf{L} = \mathbf{R}_L\mathbf{R}'_L$ be any square root decompositions of \mathbf{K} and \mathbf{L} . Let the usual SVD of $\mathbf{R}'_K\mathbf{A}\mathbf{R}_L$ (i.e., $\text{GSVD}(\mathbf{R}'_K\mathbf{A}\mathbf{R}_L)_{I_m, I_n}$) be denoted by

$$\mathbf{R}'_K\mathbf{A}\mathbf{R}_L = \tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{V}}'. \quad (21)$$

Then, \mathbf{U} , \mathbf{D} , and \mathbf{V} in $\text{GSVD}(\mathbf{A})_{K,L}$ can be obtained by $\mathbf{U} = (\mathbf{R}'_K)^{-1}\tilde{\mathbf{U}}$, $\mathbf{V} = (\mathbf{R}'_L)^{-1}\tilde{\mathbf{V}}$, and $\mathbf{D} = \tilde{\mathbf{D}}$.

The following theorems (given without proofs) are very useful in facilitating computations of SVD and GSVD in CPCA.

Theorem 4.1. Let \mathbf{T} (m by u ; $m \geq u$) and \mathbf{W} (n by v ; $n \geq v$) be columnwise orthogonal matrices. Let the usual SVD of \mathbf{A} (u by v) be denoted by $\mathbf{A} = \mathbf{U}_A\mathbf{D}_A\mathbf{V}'_A$, and that of \mathbf{TAW}' by $\mathbf{TAW}' = \tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{V}}'$. Then, $\tilde{\mathbf{U}} = \mathbf{T}\mathbf{U}_A$ (or $\mathbf{U}_A = \mathbf{T}'\tilde{\mathbf{U}}$), $\tilde{\mathbf{V}} = \mathbf{W}\mathbf{V}_A$ (or $\mathbf{V}_A = \mathbf{W}'\tilde{\mathbf{V}}$), and $\tilde{\mathbf{D}} = \mathbf{D}$.

Theorem 4.2. Let \mathbf{T} and \mathbf{W} be matrices of orders specified above but not necessarily orthogonal. Let $\text{GSVD}(\mathbf{TAW}')_{K,L}$ be denoted by $\mathbf{TAW}' = \mathbf{U}\mathbf{D}\mathbf{V}'$, and let

GSVD(\mathbf{A}) $_{T'KT, W'LV}$ be denoted by $\mathbf{A} = \mathbf{U}_A \mathbf{D}_A \mathbf{V}'_A$. Then, $\mathbf{U} = \mathbf{T} \mathbf{U}_A$ (or $\mathbf{U}_A = (\mathbf{T}' \mathbf{K} \mathbf{T})^{-1} \mathbf{T}' \mathbf{K} \mathbf{U}$), $\mathbf{V} = \mathbf{W} \mathbf{U}_A$ (or $\mathbf{V}_A = (\mathbf{W}' \mathbf{L} \mathbf{W})^{-1} \mathbf{W}' \mathbf{L} \mathbf{V}$) and $\mathbf{D}_A = \mathbf{D}$.

PCA of the first term in (18) amounts to GSVD($\mathbf{P}_{G/K} \mathbf{Z} \mathbf{P}'_{H/L}$) $_{K,L}$. This can be computed as follows: Notice that $\mathbf{R}'_K \mathbf{P}_{G/K} \mathbf{Z} \mathbf{P}'_{H/L} \mathbf{R}_L = \mathbf{P}'_{\tilde{G}} \tilde{\mathbf{Z}} \tilde{\mathbf{P}}_{\tilde{H}}$, where $\tilde{\mathbf{Z}} = \mathbf{R}'_K \mathbf{Z} \mathbf{R}_L$, and $\mathbf{P}_{\tilde{G}} = \tilde{\mathbf{G}} (\tilde{\mathbf{G}}' \tilde{\mathbf{G}})^{-1} \tilde{\mathbf{G}}'$ with $\tilde{\mathbf{G}} = \mathbf{R}'_K \mathbf{G}$ and $\mathbf{P}_{\tilde{H}} = \tilde{\mathbf{H}} (\tilde{\mathbf{H}}' \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}'$ with $\tilde{\mathbf{H}} = \mathbf{R}'_L \mathbf{H}$ are orthogonal projectors. Since orthogonal projectors can be written as products of a columnwise orthogonal matrix and its transpose (i.e., $\mathbf{P}_{\tilde{G}} = \mathbf{F}_G \mathbf{F}'_G$ and $\mathbf{P}_{\tilde{H}} = \mathbf{F}_H \mathbf{F}'_H$, where $\mathbf{F}'_G \mathbf{F}_G = \mathbf{I}$ and $\mathbf{F}'_H \mathbf{F}_H = \mathbf{I}$), $\mathbf{R}'_K \mathbf{P}_{G/K} \mathbf{Z} \mathbf{P}'_{H/L} \mathbf{R}_L = \mathbf{F}_G \mathbf{F}'_G \tilde{\mathbf{Z}} \mathbf{F}_H \mathbf{F}'_H$, whose SVD can be easily derived from SVD of $\mathbf{F}'_G \tilde{\mathbf{Z}} \mathbf{F}_H$ which is much smaller in size than $\mathbf{R}'_K \mathbf{P}_{G/K} \mathbf{Z} \mathbf{P}'_{H/L} \mathbf{R}_L$.

In some cases, GSVD($\hat{\mathbf{M}}$) $_{G'KG, H' LH}$ may be of direct interest [72]. Let $\mathbf{P}_{G/K} \mathbf{Z} \mathbf{P}'_{H/L} = \mathbf{U} \mathbf{D} \mathbf{V}'$ and $\hat{\mathbf{M}} = \mathbf{U}_M \mathbf{D}_M \mathbf{V}'_M$ denote GSVD($\mathbf{P}_{G/K} \mathbf{Z} \mathbf{P}'_{H/L}$) $_{K,L}$, and GSVD($\hat{\mathbf{M}}$) $_{G'KG, H' LH}$, respectively. Then, \mathbf{U} , \mathbf{V} and \mathbf{D} , and \mathbf{U}_M , \mathbf{V}_M and \mathbf{D}_M are related by $\mathbf{U} = \mathbf{G} \mathbf{U}_M$ (or $\mathbf{U}_M = (\mathbf{G}' \mathbf{K} \mathbf{G})^{-1} \mathbf{G}' \mathbf{K} \mathbf{U}$), $\mathbf{V} = \mathbf{H} \mathbf{V}_M$ (or $\mathbf{V}_M = (\mathbf{H}' \mathbf{L} \mathbf{H})^{-1} \mathbf{H}' \mathbf{L} \mathbf{V}$), and $\mathbf{D} = \mathbf{D}_M$.

It is of interest to further explore relationships among various kinds of SVD [66] including OSVD (ordinary SVD), GSVD, PSVD (product SVD) [16], QSVD (quotient SVD) [86], and RSVD (restricted SVD) [13,97].

CPCA subsumes a number of interesting techniques as its special cases:

(1) When $\mathbf{B} = \mathbf{0}$, $\mathbf{C} = \mathbf{0}$, and no rank restrictions are imposed on $\hat{\mathbf{M}}$, CPCA reduces to the growth curve models [50], where some additional linear constraints such as $\mathbf{R}' \mathbf{M} \mathbf{T} = \mathbf{0}$ may be imposed [55]. If in addition $\mathbf{H} = \mathbf{I}$, the ordinary multivariate multiple regression analysis model results.

(2) When $\mathbf{B} = \mathbf{0}$, $\mathbf{C} = \mathbf{0}$, $\mathbf{H} = \mathbf{I}$, and $\text{rank}(\mathbf{M}) = p (< \text{rank}(\mathbf{P}_{G/K} \mathbf{Z}))$, CPCA reduces to the reduced-rank regression analysis model [1], which is variously called PCA of instrumental variables [52] and redundancy analysis [85]. Yanai [93] proposed FA with external criteria, which analyses the residual term from redundancy analysis. When $\mathbf{H} \neq \mathbf{I}$, CPCA specializes into two-way CANDELINC [6] or the reduced-rank growth curve models [57]. This case involves the minimization of $SS(\mathbf{Z} - \mathbf{G} \mathbf{M} \mathbf{H}')_{K,L}$, which can be decomposed into

$$\begin{aligned} SS(\mathbf{Z} - \mathbf{G} \mathbf{M} \mathbf{H}')_{K,L} &= SS(\mathbf{P}_{G/K} \mathbf{Z} \mathbf{P}'_{H/L} - \mathbf{G} \mathbf{M} \mathbf{H}')_{K,L} \\ &\quad + SS(\mathbf{Z} - \mathbf{P}_{G/K} \mathbf{Z} \mathbf{P}'_{H/L})_{K,L} \\ &= SS(\mathbf{G}(\hat{\mathbf{M}} - \mathbf{M}) \mathbf{H}')_{K,L} \\ &\quad + SS(\mathbf{Z})_{K,L} + SS(\mathbf{G} \hat{\mathbf{M}} \mathbf{H}')_{K,L}. \end{aligned} \tag{22}$$

It can be minimized by minimizing the first term, which can be done by GSVD ($\hat{\mathbf{M}}$) $_{G'KG, H' LH}$. When additionally $\mathbf{Z} = \mathbf{I}$, this case reduces to CANO between \mathbf{G} and \mathbf{H} .

(3) When $\mathbf{B} = \mathbf{0}$, $\mathbf{C} = \mathbf{0}$, $\mathbf{G} = \mathbf{I}$, and $\mathbf{H} = \mathbf{I}$, CPCA reduces to (unconstrained) correspondence analysis (CA). Set $\mathbf{Z} = \mathbf{D}_R^{-1}\mathbf{F}\mathbf{D}_C^{-1}$, $\mathbf{K} = \mathbf{D}_R$, and $\mathbf{L} = \mathbf{D}_C$, where \mathbf{F} is a two-way contingency table, \mathbf{D}_R and \mathbf{D}_C are diagonal matrices of row and column sums of \mathbf{F} , respectively. (This case can also be obtained by CANO of two matrices, \mathbf{G} and \mathbf{H} , of dummy variables, where $\mathbf{F} = \mathbf{G}'\mathbf{H}$, $\mathbf{D}_R = \mathbf{G}'\mathbf{G}$, and $\mathbf{D}_C = \mathbf{H}'\mathbf{H}$. As mentioned in (2) above, CANO is also realized by setting $\mathbf{Z} = \mathbf{I}$ in CPCA.) When \mathbf{G} and/or \mathbf{H} are nonidentity matrices, this case leads to canonical correspondence analysis [81] which amounts to GSVD $(\mathbf{G}(\mathbf{G}'\mathbf{D}_R\mathbf{G})^{-1}\mathbf{G}'\mathbf{F}\mathbf{H}(\mathbf{H}'\mathbf{D}_C\mathbf{H})^{-1}\mathbf{H}')_{D_R, D_C}$, and canonical analysis of contingency tables with linear constraints (CALC) [3] which amounts to GSVD $(\mathbf{D}_R^{-1}\mathbf{Q}_{S/D_R^{-1}}\mathbf{F}\mathbf{Q}'_{T/D_C^{-1}}\mathbf{D}_C^{-1})_{D_R, D_C}$, where \mathbf{S} and \mathbf{T} are such that $\text{Ker}(\mathbf{S}') = \text{Sp}(\mathbf{X})$ and $\text{Ker}(\mathbf{T}') = \text{Sp}(\mathbf{H})$ [76]. (Recall that $\text{Ker}(\mathbf{A})$ and $\text{Sp}(\mathbf{A})$ indicate the null and range spaces of matrix \mathbf{A} , respectively.)

The decomposition of the data matrix given in (18) is a very basic one. When \mathbf{G} and/or \mathbf{H} consist of more than one distinct set of variables, $\mathbf{P}_{G/K}$ and/or $\mathbf{P}_{H/L}$ may be further decomposed in various ways, depending on how the subsets of \mathbf{G} and/or those of \mathbf{H} are related with each other. Takane and Yanai [73] (see also [56]) present a variety of such decompositions. With decompositions into finer and finer components, model (18) can be ultimately written as

$$\mathbf{Z} = \left(\sum_i \mathbf{P}_{G_i/K} \right) \mathbf{Z} \left(\sum_j \mathbf{P}_{H_j/L} \right)', \tag{23}$$

where $\sum_i \mathbf{P}_{G_i/K} = \mathbf{I}_m$, and $\sum_j \mathbf{P}_{H_j/L} = \mathbf{I}_n$. Matrices \mathbf{G}_i and \mathbf{H}_j are subsets of \mathbf{G} and \mathbf{H} , respectively.

A closely related technique for structured component analysis has been proposed by Takane et al. [70]. Their method is called DCDD (Different Constraints on Different Dimensions). Let \mathbf{G}_i and \mathbf{H}_i be T sets of row and column information (constraint) matrices, not necessarily mutually orthogonal. Consider approximating the data matrix \mathbf{Z} by the sum of $\mathbf{G}_i\mathbf{M}_i\mathbf{H}_i'$ ($i = 1, \dots, T$), where it is assumed that $\text{rank}(\mathbf{M}_i) = q_i$. (In most cases, it is assumed that $q_i = 1$ for all i .) This leads to the minimization problem of

$$f = \text{SS} \left(\mathbf{Z} - \sum_{i=1}^T \mathbf{G}_i\mathbf{M}_i\mathbf{H}_i' \right)_{K,L} \tag{24}$$

with respect to \mathbf{M}_i ($i = 1, \dots, T$) subject to $\text{rank}(\mathbf{M}_i) = q_i$, where \mathbf{K} and \mathbf{L} are metric matrices (assumed to be PD). Unfortunately, this minimization problem has no closed-form solutions except for a few special cases. Efficient algorithms have been developed, although they are iterative.

Verbyla and Venables [88] proposed a model similar to DCDD but without the rank restrictions on \mathbf{M}_i . They also developed an iterative algorithm for parameter estimation. von Rosen [90,91] proposed a model similar to that by Verbyla and Venables as an extension to the growth curve models. He derived a closed-form solution

for the maximum likelihood estimators under the normality assumption on \mathbf{Z} , but under the additional assumption that \mathbf{H}_i 's had special nested structures. More recently, Fujikoshi et al. [18] developed some inferential procedures for von Rosen's model. Velu [87] (see also [57]) considered a special case of DCDD, where $T = 2$, $\mathbf{K} = \mathbf{I}$, and $\mathbf{L} = \mathbf{I}$. Hwang and Takane [36] have recently extended DCDD to fit structural equation models (SEM) specifying a set of assumed relationships between observed and latent variables. SEM has been traditionally (and predominantly) fitted via ACOVS (Analysis of Covariance Structures). Hwang and Takane's approach, on the other hand, relies on the structured reduced-rank regression analysis approach.

Ramsay and Silverman [51] proposed structured analysis of functional data. Many of the mathematical tools used in functional data analysis such as reproducing kernels and Green's function, etc. have clear analogues in matrix algebra. It is of interest to explore further correspondence between terminologies used in functional data analysis and multivariate data analysis.

5. Multiple-set (T -set) canonical correlation analysis

CANO is used to explore linear relationships between two sets of multivariate data. Multiple-set CANO, on the other hand, explores relationships among T (≥ 2) sets of multivariate data. A number of techniques have been proposed for multiple-set CANO. Only one of them due to Horst [32] will be discussed here because it is the only technique with a closed form solution (see also [4]). See Gifi [19] for more comprehensive review of multiple-set CANO and related technique.

Let \mathbf{Z}_k ($k = 1, \dots, T$) denote the set of T data matrices. Consider minimizing

$$f = \sum_{k=1}^T \text{SS}(\mathbf{Y} - \mathbf{Z}_k \mathbf{W}_k) \quad (25)$$

with respect to \mathbf{W}_k ($k = 1, \dots, T$) and \mathbf{Y} subject to $\mathbf{Y}'\mathbf{Y} = \mathbf{I}$, where \mathbf{W}_k is the matrix of weights applied to \mathbf{Z}_k , and \mathbf{Y} is the matrix of hypothetical variables closely related to canonical variates. The above criterion is called homogeneity criterion [19]; it attempts to make $\mathbf{Z}_k \mathbf{W}_k$ ($k = 1, \dots, T$) as homogeneous as possible among themselves by making each of them as close as possible to \mathbf{Y} . Assume temporarily that \mathbf{Y} is known, and minimize f with respect to \mathbf{W}_k . Then,

$$\mathbf{W}_k = (\mathbf{Z}'_k \mathbf{Z}_k)^{-1} \mathbf{Z}'_k \mathbf{Y} \quad (26)$$

for $k = 1, \dots, T$. Putting this estimate of \mathbf{W}_k in the above criterion leads to

$$f^* = \min_{\mathbf{Y}} f = \sum_{k=1}^K \text{SS}(\mathbf{Y} - \mathbf{P}_{\mathbf{Z}_k} \mathbf{Y}) = \text{tr} \left(\mathbf{Y}' \left(\sum_{k=1}^K \mathbf{Q}_{\mathbf{Z}_k} \right) \mathbf{Y} \right), \quad (27)$$

where $\mathbf{P}_{\mathbf{Z}_k} = \mathbf{Z}_k (\mathbf{Z}'_k \mathbf{Z}_k)^{-1} \mathbf{Z}'_k$, and $\mathbf{Q}_{\mathbf{Z}_k} = \mathbf{I} - \mathbf{P}_{\mathbf{Z}_k}$. Minimizing f^* with respect to \mathbf{Y} subject to $\mathbf{Y}'\mathbf{Y} = \mathbf{I}$ is equivalent to maximizing

$$g = \text{tr} \left(\mathbf{Y}' \left(\sum_{k=1}^K \mathbf{P}_{Z_k} \right) \mathbf{Y} \right) \tag{28}$$

with respect to \mathbf{Y} subject to the same restriction. This amounts to obtaining the eigenvalue decomposition of $\mathbf{R} = \sum_{k=1}^K \mathbf{P}_{Z_k}$, or equivalently obtaining the SVD of $\tilde{\mathbf{Z}} = [\tilde{\mathbf{Z}}_1, \dots, \tilde{\mathbf{Z}}_T]$, where $\tilde{\mathbf{Z}}_k = \mathbf{Z}_k (\mathbf{Z}'_k \mathbf{Z}_k)^{-1/2}$ ($k = 1, \dots, T$).

Multiple-set CANO is interesting partly because it subsumes a number of existing techniques as its special cases:

(1) *PCA*: When each \mathbf{Z}_k consists of a single continuous variable, say \mathbf{z}_k , generalized CANO reduces to PCA. In this case $\mathbf{R} = \sum_{k=1}^T \mathbf{z}_k (\mathbf{z}'_k \mathbf{z}_k)^{-1} \mathbf{z}'_k = \tilde{\mathbf{Z}} \tilde{\mathbf{Z}}'$, where $\tilde{\mathbf{Z}}$ is the standardized data matrix (i.e., $\tilde{\mathbf{Z}} = [\mathbf{z}_1 (\mathbf{z}'_1 \mathbf{z}_1)^{-1/2}, \dots, \mathbf{z}_T (\mathbf{z}'_T \mathbf{z}_T)^{-1/2}]$). The eigenvalue decomposition of matrix \mathbf{R} is equivalent to the SVD of $\tilde{\mathbf{Z}}$.

(2) *Multiple correspondence analysis (MCA)* [23]: When each \mathbf{Z}_k denotes a matrix of dummy variables, multiple-set CANO specializes into MCA (e.g., [23]), variously known as the quantification method of the third kind [28], dual scaling [48], etc. In this case $\mathbf{R} = \mathbf{Z} (\mathbf{D}'_Z \mathbf{D}_Z)^{-1} \mathbf{Z}'$, where $\mathbf{Z} = [\mathbf{Z}_1, \dots, \mathbf{Z}_T]$, and \mathbf{D}_Z is a block diagonal matrix with \mathbf{Z}_k ($k = 1, \dots, T$) as the k th diagonal block.

(3) *CANO, DISC, and MANOVA*: Multiple-set CANO reduces to the usual 2-set CANO when $T = 2$ (and the two sets of variables both consist of sets of continuous variables), which in turn specializes into canonical discriminant analysis (DISC) and MANOVA when one of the two sets of variables consists of dummy variables and the other a set of continuous variables. When $T = 2$, the eigenvalue decomposition of \mathbf{R} reduces to

$$(\mathbf{P}_{Z_1} + \mathbf{P}_{Z_2}) \mathbf{Y} = \mathbf{Y} \mathbf{\Delta}, \tag{29}$$

where \mathbf{Y} is the matrix of eigenvectors, and $\mathbf{\Delta}$ is the diagonal matrix of eigenvalues of $\mathbf{R} = \mathbf{P}_{Z_1} + \mathbf{P}_{Z_2}$. Premultiplying both sides of (29) by \mathbf{P}_{Z_2} leads to

$$\mathbf{V}_2 = \mathbf{P}_{Z_2} \mathbf{V}_1 (\mathbf{\Delta} - \mathbf{I})^{-1}, \tag{30}$$

where $\mathbf{V}_1 = \mathbf{P}_{Z_1} \mathbf{Y}$, and $\mathbf{V}_2 = \mathbf{P}_{Z_2} \mathbf{Y}$. Similarly, premultiplying both sides of (29) by \mathbf{P}_{Z_1} leads to

$$\mathbf{P}_{Z_1} \mathbf{V}_2 = \mathbf{V}_1 (\mathbf{\Delta} - \mathbf{I}). \tag{31}$$

Substituting \mathbf{V}_2 in (30) for \mathbf{V}_2 in (31) leads to

$$(\mathbf{P}_{Z_1} \mathbf{P}_{Z_2}) \mathbf{V}_1 = \mathbf{V}_1 (\mathbf{\Delta} - \mathbf{I})^2. \tag{32}$$

This is essentially the same eigenequation encountered in the two-set CANO. Matrices \mathbf{V}_1 and \mathbf{V}_2 should be normalized to obtain the *canonical scores* obtained in the two-set CANO.

(4) *CA*: When $T = 2$, and both data sets consist of dummy variable matrices, simple CA of single two-way contingency tables results, which amounts to the GSVD of $\mathbf{D}_R^{-1} \mathbf{F} \mathbf{D}_C^{-1}$ with metrics \mathbf{D}_R and \mathbf{D}_C , where $\mathbf{D}_R = \mathbf{Z}'_1 \mathbf{Z}_1$, and $\mathbf{D}_C = \mathbf{Z}'_2 \mathbf{Z}_2$ are diagonal matrices of row and column sums of $\mathbf{F} = \mathbf{Z}'_1 \mathbf{Z}_2$.

It is interesting to see that CA can also be derived from unfolding analysis (Section 2.2) [29,64]. Let \mathbf{U} and \mathbf{V} denote matrices of coordinates of row and column points in a two-way contingency table, \mathbf{F} . Then, the matrix of squared Euclidean distances between row and column points is obtained by (2). Minimize

$$\begin{aligned} f &= \text{tr}(\mathbf{F}'\mathbf{D}^2(\mathbf{U}, \mathbf{V})) = \text{tr}(\mathbf{F}'\mathbf{1}_m\mathbf{1}'_n \text{diag}(\mathbf{V}\mathbf{V}') - 2\mathbf{F}'\mathbf{U}\mathbf{V}' + \mathbf{F}' \text{diag}(\mathbf{U}\mathbf{U}')\mathbf{1}_m\mathbf{1}'_n) \\ &= \text{tr}(\mathbf{V}'\mathbf{D}_C\mathbf{V}) - 2\text{tr}(\mathbf{V}'\mathbf{F}'\mathbf{U}) + \text{tr}(\mathbf{U}'\mathbf{D}_R\mathbf{U}), \end{aligned} \quad (33)$$

with respect to \mathbf{U} and \mathbf{V} subject to $\mathbf{U}'\mathbf{D}_R\mathbf{U} = \mathbf{I}$, where m and n are the numbers of rows and columns of the contingency table, respectively. Minimizing f with respect to \mathbf{V} for fixed \mathbf{U} yields

$$\mathbf{V} = \mathbf{D}_C^{-1}\mathbf{F}\mathbf{U}. \quad (34)$$

Putting this estimate of \mathbf{V} into the above criterion leads to

$$f^* = \min_{\mathbf{U}} f = -\text{tr}(\mathbf{U}'\mathbf{F}\mathbf{D}_C^{-1}\mathbf{F}'\mathbf{U}) + \text{tr}(\mathbf{U}'\mathbf{D}_R\mathbf{U}). \quad (35)$$

Minimizing this criterion with respect to \mathbf{U} subject to $\mathbf{U}'\mathbf{D}_R\mathbf{U} = \mathbf{I}$ is equivalent to maximizing $\text{tr}(\mathbf{U}'\mathbf{F}\mathbf{D}_C^{-1}\mathbf{F}'\mathbf{U})$ under the same restriction. This can be obtained by the generalized eigenvalue decomposition of $\mathbf{F}\mathbf{D}_C^{-1}\mathbf{F}'$ with respect to \mathbf{D}_R , or equivalently by $\text{GSVD}(\mathbf{D}_R^{-1}\mathbf{F}\mathbf{D}_C^{-1})_{D_R, D_C}$.

Let \mathbf{G} and \mathbf{H} be matrices of dummy variables such that $\mathbf{F} = \mathbf{G}'\mathbf{H}$. It is interesting to note that linear transformations of \mathbf{G} and \mathbf{H} into canonical variates by $\mathbf{G}\mathbf{U}$ and $\mathbf{H}\mathbf{V}$ provides best nonlinear transformations of arbitrarily quantified categories of \mathbf{G} and \mathbf{H} . This was shown by Otsu [49] who used a variational method to find optimal nonlinear transformations of the predictor variables in discriminant analysis. As it has turned out, Otsu's results are closely related to the Bayesian decision rule for classification that requires classifying subjects (cases) into the group associated with the maximum posterior probabilities (see also [19]).

Multiple-set CANO can also be viewed as a method for information integration from T concurrent sources. Takane and Oshima-Takane [71] proposed a nonlinear extension of multiple-set CANO using multilayered feed-forward neural network models. Takane et al. [69] proposed another extension of multiple-set CANO based on a kernel method. See Herbrich [31] for a general account of kernel methods. This technique allows nonlinear multivariate analyses by a series of linear matrix operations.

Yanai and Takane [95] proposed constrained CANO. Takane and Hwang [68] and Takane et al. [75] extended constrained CANO by incorporating constraints on both row and column sides of the two matrices to be related. The technique is called generalized constrained CANO (GCCANO).

There is no analytical procedures for investigating sampling characteristics of associations among T sets of variables even under the standard multivariate normal assumptions. This is something to be explored in the future.

6. The Wedderburn–Guttman theorem

Let \mathbf{Z} be a m by n matrix of $\text{rank}(\mathbf{Z}) = p$, and let \mathbf{M} and \mathbf{N} be m by r and n by r matrices, respectively, such that $\mathbf{M}'\mathbf{Z}\mathbf{N}$ is nonsingular (i.e., $\text{rank}(\mathbf{M}'\mathbf{Z}\mathbf{N}) = q = \text{rank}(\mathbf{Z}\mathbf{N}(\mathbf{M}'\mathbf{Z}\mathbf{N})^{-1}\mathbf{M}'\mathbf{Z})$). Then,

$$\text{rank}(\mathbf{Z} - \mathbf{Z}\mathbf{N}(\mathbf{M}'\mathbf{Z}\mathbf{N})^{-1}\mathbf{M}'\mathbf{Z}) = p - q. \tag{36}$$

This is called Wedderburn–Guttman theorem. It was originally established for $q = 1$ by Wedderburn [92, p. 69] but was later extended to $q > 1$ by Guttman [24]. (Guttman is a prominent psychometrician.) Guttman called the case in which $q = 1$ Lagrange’s theorem while referring to Wedderburn [92], and Rao [53, p. 69] also calls it Lagrange’s theorem. However, apparently there is no reference to Lagrange in Wedderburn [92] according to Hubert et al. [34]. It may thus be more appropriately called Wedderburn–Guttman theorem. Guttman [25] also showed the reverse of the theorem, that is, for (36) to hold the matrix to be subtracted from \mathbf{Z} must be of the form $\mathbf{Z}\mathbf{N}(\mathbf{M}'\mathbf{Z}\mathbf{N})^{-1}\mathbf{M}'\mathbf{Z}$. The theorem has been used extensively in psychometrics [33] and in computational linear algebra [8] as a basis for extracting components which are known linear combinations of observed variables.

Both necessity and sufficiency (NS) parts of the theorem follow immediately from Marsaglia and Styan’s [46] condition (7.9) of Theorem 17 [9], which states that the NS conditions for $\text{rank}(\mathbf{A} - \mathbf{B}) = \text{rank}(\mathbf{A}) - \text{rank}(\mathbf{B})$ are: (i) $\text{Sp}(\mathbf{B}) \subset \text{Sp}(\mathbf{A})$, (ii) $\text{Sp}(\mathbf{B}') \subset \text{Sp}(\mathbf{A}')$, and (iii) $\mathbf{B}\mathbf{A}^{-}\mathbf{B} = \mathbf{B}$ (i.e., $\mathbf{A}^{-} \in \{\mathbf{B}^{-}\}$). It is obvious that $\mathbf{A} = \mathbf{Z}$ and $\mathbf{B} = \mathbf{Z}\mathbf{N}(\mathbf{M}'\mathbf{Z}\mathbf{N})^{-1}\mathbf{M}'\mathbf{Z}$ satisfy these conditions. Conversely, to satisfy (i) and (ii), \mathbf{B} has to be of the form $\mathbf{B} = \mathbf{Z}\mathbf{N}\mathbf{R}\mathbf{M}'\mathbf{Z}$ for some \mathbf{N} , \mathbf{M} and \mathbf{R} . To satisfy (iii), \mathbf{R} must be of the form $\mathbf{R} = (\mathbf{M}'\mathbf{Z}\mathbf{N})^{-1}$ if $\mathbf{M}'\mathbf{Z}\mathbf{N}$ is nonsingular. Takane and Yanai [74] further discuss the condition under which $(\mathbf{M}'\mathbf{Z}\mathbf{N})^{-1}$ can be replaced by a g -inverse of $\mathbf{M}'\mathbf{Z}\mathbf{N}$ of some kind.

Two examples of application of the theorem are given:

(1) In the group centroid method of component analysis, components are defined as centroids of some subsets of observed variables. Suppose there are six observed variables, and the first three variables define the first component and the last three the second component. Define

$$\mathbf{N}' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix},$$

and $\mathbf{M} = \mathbf{Z}\mathbf{N}$. Then, $\mathbf{Q}_{\mathbf{Z}\mathbf{N}}\mathbf{Z} = \mathbf{Z}\mathbf{Q}_{\mathbf{N}'\mathbf{Z}'\mathbf{Z}}$, where $\mathbf{Q}_{\mathbf{Z}\mathbf{N}} = \mathbf{I} - \mathbf{Z}\mathbf{N}(\mathbf{N}'\mathbf{Z}'\mathbf{Z}\mathbf{N})^{-1}\mathbf{N}'\mathbf{Z}'$ and $\mathbf{Q}_{\mathbf{N}'\mathbf{Z}'\mathbf{Z}} = \mathbf{I} - \mathbf{N}(\mathbf{N}'\mathbf{Z}'\mathbf{Z}\mathbf{N})^{-1}\mathbf{N}'\mathbf{Z}'\mathbf{Z}$, gives the residual matrix.

(2) Rao [52] derived a method of component analysis in which components were required to be orthogonal to a given matrix \mathbf{G} in $\text{Sp}(\mathbf{Z})$. This amounts to setting $\mathbf{M} = \mathbf{G}$ and $\mathbf{N} = \mathbf{Z}'\mathbf{G}$, and obtaining the SVD of $\mathbf{Q}'_{\mathbf{G}'\mathbf{Z}\mathbf{Z}'}\mathbf{Z} = \mathbf{Z}\mathbf{Q}_{\mathbf{Z}'\mathbf{G}}$, where $\mathbf{Q}'_{\mathbf{G}'\mathbf{Z}\mathbf{Z}'} = \mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{Z}\mathbf{Z}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{Z}\mathbf{Z}'$ and $\mathbf{Q}_{\mathbf{Z}'\mathbf{G}} = \mathbf{I} - \mathbf{Z}'\mathbf{G}(\mathbf{G}'\mathbf{Z}\mathbf{Z}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{Z}$.

7. Concluding remarks

There are a wide variety of contexts in psychometrics where matrix theory plays a crucial role. Psychometrics today is requiring more and more advanced mathematical skills with matrix algebra becoming only one of them. This, however, by no means implies that knowledge in matrix algebra is becoming less important in psychometrics. On the contrary, its importance has never been greater before than it is today and perhaps in many years to come. I always keep telling my graduate students that matrix algebra is the single most important subject to learn if one is to pursue psychometrics as one's profession.

References

- [1] T.W. Anderson, Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Ann. Math. Stat.* 22 (1951) 327–351.
- [2] E. Bertrami, Sulle funzioni bilineari, *Giornale di Matematiche ud uso Degli Studenti Delle Università* 11 (1873) 98–106.
- [3] U. Böckenholt, I. Böckenholt, Canonical analysis of contingency tables with linear constraints, *Psychometrika* 55 (1990) 633–639.
- [4] J.D. Carroll, A generalization of canonical correlation analysis to three or more sets of variables, *Proceedings of the 76th Annual Convention of the American Psychological Association*, 1968, pp. 227–228.
- [5] J.D. Carroll, J.-J. Chang, Analysis of individual differences in multidimensional scaling via an N -way generalization of “Eckart-Young” decomposition, *Psychometrika* 35 (1970) 282–319.
- [6] J.D. Carroll, S. Pruzansky, J.B. Kruskal, CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters, *Psychometrika* 45 (1980) 3–24.
- [7] N. Chino, K. Shiraiwa, Geometric structures of some non-distance models for asymmetric MDS, *Behaviormetrika* 20 (1993) 35–47.
- [8] T. Chu, R.E. Funderlic, G.H. Golub, A rank-one reduction formula and its applications to matrix factorizations, *SIAM Rev.* 37 (1995) 512–530.
- [9] R.E. Cline, R.E. Funderlic, The rank of a difference of matrices and associated generalized inverses, *Linear Algebra Appl.* 24 (1979) 185–215.
- [10] C.H. Coombs, *A Theory of Data*, Wiley, New York, 1964.
- [11] J. de Leeuw, W.J. Heiser, Theory of multidimensional scaling, in: P.R. Krishnaiah, L. Kanal (Eds.), *Handbook of Statistics*, North-Holland, Amsterdam, 1982, pp. 285–316.
- [12] J. de Leeuw, S. Pruzansky, A new computational method to fit the weighted Euclidean distance model, *Psychometrika* 43 (1978) 479–490.
- [13] B.L.R. De Moor, G.H. Golub, The restricted singular value decomposition: Properties and applications, *SIAM J.: Matrix Anal. Appl.* 12 (1991) 401–425.
- [14] C. Eckart, G. Young, The approximation of one matrix by another of lower rank, *Psychometrika* 1 (1936) 211–218.
- [15] Y. Escoufier, A. Grorud, Analyse factorielle des matrices carrees non-symetriques, in: E. Diday et al. (Eds.), *Data Analysis and Informatics*, North-Holland, Amsterdam, 1980, pp. 263–276.
- [16] K.V. Fernando, S.J. Hammarling, A product induced singular value decomposition for two matrices and balanced realization, in: B.N. Datta et al. (Eds.), *Linear Algebra in Signal Processings*, SIAM, Philadelphia, 1988, pp. 128–140.
- [17] B. Flury, *Common Principal Components and Related Multivariate Models*, Wiley, New York, 1988.

- [18] Y. Fujikoshi, T. Kanda, M. Ohtaki, Growth curve model with hierarchical within-individuals design matrices, *Ann. Inst. Stat. Math.* 51 (1999) 707–721.
- [19] A. Gifi, *Nonlinear Multivariate Analysis*, Wiley, Chichester, 1990.
- [20] E.M. Gold, Metric unfolding: Data requirement for unique solution and clarification of Schönemann’s algorithm, *Psychometrika* 37 (1973) 555–570.
- [21] J.C. Gower, Some distance properties of latent root and vector methods used in multivariate analysis, *Biometrika* 53 (1966) 325–338.
- [22] J.C. Gower, The analysis of asymmetry and orthogonality, in: J.R. Barra, F. Brodeau, G. Romier, B. van Cutem (Eds.) *Recent Developments in Statistics*, North-Holland, Amsterdam, 1977, pp. 109–203.
- [23] M.J. Greenacre, *Theory and Applications of Correspondence Analysis*, Academic Press, London, 1984.
- [24] L. Guttman, General theory and methods of matrix factoring, *Psychometrika* 9 (1944) 1–16.
- [25] L. Guttman, A necessary and sufficient formula for matrix factoring, *Psychometrika* 22 (1957) 79–81.
- [26] R.A. Harshman, Foundations of the PARAFAC procedure: Models and conditions for an “exploratory” multi-modal factor analysis, *UCLA Working Papers in Phonetics* 16 (1970) 1–84.
- [27] R.A. Harshman, P.E. Green, Y. Wind, M.E. Lundy, A model for the analysis of asymmetric data in marketing research, *Market. Sci.* 1 (1982) 205–242.
- [28] C. Hayashi, On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view, *Ann. Inst. Stat. Math.* 3 (1952) 69–98.
- [29] W.J. Heiser, *Unfolding analysis of proximity data*, Unpublished doctoral dissertation, Leiden University, 1981.
- [30] W.J. Heiser, J. Meulman, Analyzing rectangular tables by joint and constrained multidimensional scaling, *J. Econometrics* 22 (1983) 139–167.
- [31] R. Herbrich, *Learning Kernel Classifiers: Theory and Algorithms*, MIT Press, Cambridge, MA, 2001.
- [32] P. Horst, Generalized canonical correlations and their applications to experimental data, *J. Clin. Psychol.* 17 (1961) 331–347.
- [33] P. Horst, *Matrix Algebra for Social Scientists*, Holt, Rinehart and Winston, New York, 1965.
- [34] L.J. Hubert, J. Meulman, W.J. Heiser, Two purposes for matrix factorization: A historical appraisal, *SIAM Rev.* 42 (2000) 68–82.
- [35] M.A. Hunter, Y. Takane, Constrained principal component analysis: Various applications, *J. Educ. Behav. Stat.* 27 (2002) 105–145.
- [36] H. Hwang, Y. Takane, *Structural equation modeling by extended redundancy analysis*, *Psychometrika*, submitted for publication.
- [37] C. Jordan, *Memoire sur les formes bilineaires*, *J. Math. Pures Appl., Deuxième Série* 19 (1874) 35–54.
- [38] H.A.L. Kiers, Y. Takane, A generalization of GIPSCAL for the analysis of nonsymmetric data, *J. Classif.* 11 (1994) 79–99.
- [39] H.A.L. Kiers, J.M.F. ten Berge, Y. Takane, J. de Leeuw, A generalization of Takane’s algorithm for DEDICOM, *Psychometrika* 55 (1990) 151–158.
- [40] W. Kristof, A theorem on the trace of certain matrix products and some applications, *J. Math. Psychol.* 7 (1970) 515–530.
- [41] P.M. Kroonenberg, J. de Leeuw, Principal component analysis of three-mode data by means of alternating least squares, *Psychometrika* 45 (1980) 69–97.
- [42] J.B. Kruskal, Nonmetric multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrika* 29 (1964) 1–27.
- [43] J.B. Kruskal, Nonmetric multidimensional scaling: A numerical method, *Psychometrika* 29 (1964) 28–42.
- [44] J.B. Kruskal, Three-way arrays: Rank and uniqueness of trilinear decompositions with applications to arithmetic complexity and statistics, *Linear Algebra Appl.* 18 (1977) 95–138.

- [45] J.B. Kruskal, Rank, decomposition, and uniqueness for 3-way, N -way arrays, in: R. Coppi, S. Bolasco (Eds.) *Multiway Data Analysis*, North-Holland, Amsterdam, 1989, pp. 7–18.
- [46] G. Marsaglia, G.P.H. Styan, Equalities and inequalities for ranks of matrices, *Linear and Multilinear Algebra* 2 (1974) 269–292.
- [47] L. Mirsky, Symmetric gage functions and unitarily invariant norms, *Quart. J. Math.* 11 (1960) 50–59.
- [48] S. Nishisato, *Analysis of Categorical Data: Dual Scaling and its Applications*, University of Toronto Press, Toronto, 1980.
- [49] N. Otsu, Nonlinear discriminant analysis as a natural extension of the linear case, *Behaviormetrika* 2 (1975) 45–59.
- [50] R.F. Potthoff, S.N. Roy, A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika* 51 (1964) 313–326.
- [51] J.O. Ramsay, B.W. Silverman, *Functional Data Analysis*, Springer, New York, 1997.
- [52] C.R. Rao, The use and interpretation of principal component analysis in applied research, *Sankhyā A* 26 (1964) 329–358.
- [53] C.R. Rao, *Linear Statistical Inference and its Applications*, Wiley, New York, 1973.
- [54] C.R. Rao, Matrix approximations and reduction of dimensionality in multivariate statistical analysis, in: P.R. Krishnaiah (Ed.), *Multivariate Analysis*, vol. V, North-Holland, Amsterdam, 1980, pp. 3–22.
- [55] C.R. Rao, Tests for dimensionality and interactions of mean vectors under general and reducible covariance structures, *J. Multivariate Anal.* 16 (1985) 173–184.
- [56] C.R. Rao, H. Yanai, General definition and decomposition of projectors and some applications to statistical problems, *J. Stat. Inference Plann.* 3 (1979) 1–17.
- [57] G.C. Reinsel, R.P. Velu, *Multivariate Reduced-Rank Regression*, Springer, New York, 1998.
- [58] R. Sands, F.W. Young, Component models for three-way data: An alternating least squares algorithm with optimal scaling features, *Psychometrika* 45 (1980) 39–67.
- [59] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Teil. Entwicklung willkürlichen Funktionen nach System vorgeschriebener, *Math. Ann.* 63 (1907) 433–476.
- [60] I.J. Schoenberg, Remarks to Maurice Fréchet’s article “Sur la définition axiomatique d’une classe d’espace distanciés vectoriellement applicable sur l’espace de Hilbert”, *Ann. Math.* 36 (1935) 724–732.
- [61] P.H. Schönemann, On metric multidimensional unfolding, *Psychometrika* 35 (1970) 349–366.
- [62] P.H. Schönemann, An algebraic solution for the class of subjective metrics models, *Psychometrika* 37 (1972) 441–451.
- [63] R.N. Shepard, The analysis of proximities: Multidimensional scaling with an unknown distance function, *Psychometrika* 27 (1962) 219–246.
- [64] Y. Takane, Analysis of categorizing behavior by a quantification method, *Behaviormetrika* 8 (1980) 57–67.
- [65] Y. Takane, Basic concepts and applications of multidimensional scaling, *Proceedings of the 1984 Japan Applied Statistical Society Meeting*, 1984, pp. 30–41.
- [66] Y. Takane, Relationships among various kinds of eigenvalue and singular value decompositions, in: H. Yanani et al. (Eds.), *New Developments in Psychometrics*, Springer, Tokyo, 2002, pp. 45–56.
- [67] Y. Takane, M.A. Hunter, Constrained principal component analysis: A comprehensive theory, *Appl. Algebra Engng., Commun. Comput.* 12 (2001) 391–419.
- [68] Y. Takane, H. Hwang, Generalized constrained canonical correlation analysis, *Multivariate Behav. Res.* 37 (2002) 163–195.
- [69] Y. Takane, H. Hwang, Y. Oshima-Takane, A kernel method for multiple-set canonical correlation analysis, *IEEE Trans. Neural Networks*, submitted for publication.
- [70] Y. Takane, H.A.L. Kiers, J. de Leeuw, Component analysis with different sets of constraints on different dimensions, *Psychometrika* 60 (1995) 259–280.
- [71] Y. Takane, Y. Oshima-Takane, Nonlinear generalized canonical correlation analysis by neural network models, in: S. Nishisato et al. (Eds.), *Measurement and Multivariate Analysis*, Springer, Tokyo, 2002, pp. 183–190.

- [72] Y. Takane, T. Shibayama, Principal component analysis with external information on both subjects and variables, *Psychometrika* 56 (1991) 97–120.
- [73] Y. Takane, H. Yanai, On oblique projectors, *Linear Algebra Appl.* 289 (1999) 297–310.
- [74] Y. Takane, H. Yanai, On the Wedderburn–Guttman theorem, *Linear Algebra Appl.*, submitted for publication.
- [75] Y. Takane, H. Yanai, H. Hwang, An improved method for generalized constrained canonical correlation analysis, *Appl. Algebra in Engng., Commun. Comput.*, submitted for publication.
- [76] Y. Takane, H. Yanai, S. Mayekawa, Relationships among several methods of linearly constrained correspondence analysis, *Psychometrika* 56 (1991) 667–684.
- [77] J.M.F. ten Berge, A generalization of Kristof’s theorem on the trace of certain matrix products, *Psychometrika* 48 (1983) 519–523.
- [78] J.M.F. ten Berge, *Least Squares Optimization in Multivariate Analysis*, DSWO Press, Leiden, 1993.
- [79] J.M.F. ten Berge, The typical rank of tall three-way arrays, *Psychometrika* 65 (2000) 525–532.
- [80] J.M.F. ten Berge, H.A.L. Kiers, Simplicity of core arrays in three-way principal component analysis and the typical rank of $p \times p \times 2$ arrays, *Linear Algebra Appl.* 294 (1999) 169–179.
- [81] C.J.F. ter Braak, Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis, *Ecology* 67 (1986) 1167–1179.
- [82] W.S. Torgerson, Multidimensional scaling: I. Theory and method, *Psychometrika* 17 (1952) 401–409.
- [83] M.W. Trosset, A new formulation of the nonmetric strain problem in multidimensional scaling, *J. Classif.* 15 (1998) 15–35.
- [84] L.R. Tucker, The extension of factor analysis to three-dimensional matrices, in: N. Frederiksen, H. Gulliksen (Eds.) *Contributions to Mathematical Psychology*, Holt, Rinehart and Winston, New York, 1964, pp. 110–127.
- [85] A.L. van den Wollenberg, Redundancy analysis: An alternative for canonical correlation analysis, *Psychometrika* 42 (1977) 207–219.
- [86] C.F. Van Loan, Generalizing the singular value decomposition, *SIAM J. Numer. Anal.* 13 (1976) 76–83.
- [87] R.P. Velu, Reduced rank models with two sets of regressions, *Appl. Stat.* 40 (1991) 159–170.
- [88] A.P. Verbyla, W.N. Venables, An extension of the growth curve model, *Biometrika* 75 (1988) 129–138.
- [89] J. von Neumann, Some matrix inequalities and metrization of matrix space, *Tomask University Review* 1 (1937) 286–300 Reprinted in A.H. Taub (Ed.), *John von Neumann Collected Works*, vol. IV, Pergamon Press, New York, 1962.
- [90] D. von Rosen, Maximum likelihood estimators in multivariate linear normal models, *J. Multivariate Anal.* 31 (1989) 187–200.
- [91] D. von Rosen, The growth curve model: A review, *Commun. Stat.—Theory Meth.* 20 (1991) 2791–2822.
- [92] J.H.M. Wedderburn, *Lectures on Matrices*, Colloquium Publications, American Mathematical Society, Providence, 1934 Dover, New York, 1964.
- [93] H. Yanai, Factor analysis with external criteria, *Jpn. Psychol. Res.* 12 (1970) 143–153.
- [94] H. Yanai, Some generalized forms of least squares g -inverse minimum norm g -inverse and Moore–Penrose inverse matrices, *Comput. Stat. Data Anal.* 10 (1990) 251–260.
- [95] H. Yanai, Y. Takane, Canonical correlation analysis with linear constraints, *Linear Algebra Appl.* 176 (1992) 75–89.
- [96] G. Young, A.S. Householder, Discussion of a set of points in terms of their mutual distances, *Psychometrika* 3 (1938) 19–22.
- [97] H. Zha, The restricted singular value decomposition of matrix triplets, *SIAM J. Matrix Anal. Appl.* 12 (1991) 172–194.