Professor Yanai and Multivariate Analysis

Yoshio Takane[1,*]

[1]Department of Psychology, University of Victoria, British Columbia, Canada

Abstract Late Professor Yanai has contributed to many fields ranging from aptitude diagnostics, epidemiology, and nursing to psychometrics and statistics. This paper reviews some of his accomplishments in multivariate analysis through his collaborative work with the present author, along with some untold episodes for the inception of key ideas underlying the work. The various topics covered include constrained principal component analysis, extensions of Khatri's lemma, the Wedderburn-Guttman theorem, ridge operators, decompositions of the total association between two sets of variables, and ideal instruments. A common thread running through all of them is projectors and singular value decomposition (SVD), which are the main subject matters of a recent monograph by Yanai, Takeuchi, and Takane [35].

Keywords Projectors: Singular value decomposition (SVD); Constrained principal component analysis (CPCA); Khatri's lemma; The Wedderburn-Guttman theorem; Ridge operators; Generalized constrained canonical correlation analysis; Confounding variables; Propensity scores; Instrumental variables

1 Introduction

Professor Yanai passed away due to prostate cancer in December, 2013 at the age of 73. A quick glance at his home page reveals that his contributions extend over 7 broad categories, including aptitude diagnostics, test theories, educational psychology, epidemiology, nursing, linear algebra, statistics, and multivariate analysis (MVA). Here we focus on his contributions in the last category, namely multivariate analysis, through his collaborative works with me. Professor Yanai has been the most influential person in my career. In particular, if I had not met him when I was in the third year of college, I would not have been a statistician. We have 15 joint publications, including two books one in English [35] and one other in Japanese. The specific topics we cover today are:

(1) Constrained principal component analysis (CPCA)
(2) Khatri's lemma
(3) The Wedderburn-Guttman theorem
(4) Ridge operators
(5) Generalized constrained canonical correlation analysis
(6) Causal inference

Professor Yanai's idea about MVA can be most succinctly summarized as "partitioning the space of dimensionality n (the number of cases) into meaningful subspaces" identified by some external information or by some internal criterion (Takeuchi, Yanai, and Mukherjee [27]). Two major tools for partitioning are:

(1) Projectors

*Corresponding author. Email: yoshio.takane@mcgill.ca

(2) Singular value decomposition (SVD)

which are the main subject matters of a recent monograph by Yanai, Takeuchi, and Takane [35]. As is well known, projectors are used to partition the space of observation vectors on criterion variables into subspaces that can and cannot be explained by predictor variables, and SVD seeks to find the subspace most representative of the original subspace.

Before we begin, let us introduce some basic notations we use throughout this paper: Let $\text{Sp}(\mathbf{X})$ denote the space spanned by column vectors of $\mathbf{X}$, and let $\text{Ker}(\mathbf{X}')$ denote the orthogonal complement subspace to $\text{Sp}(\mathbf{X})$. Let

$$\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' \tag{1}$$

denote the orthogonal projector onto $\text{Sp}(\mathbf{X})$, and let

$$\mathbf{Q}_X = \mathbf{I} - \mathbf{P}_X \tag{2}$$

denote the orthogonal projectors onto $\text{Ker}(\mathbf{X}')$. Then,

$\mathbf{P}'_X = \mathbf{P}_X$, $\mathbf{Q}'_X = \mathbf{Q}_X$ (symmetric).
$\mathbf{P}^2_X = \mathbf{P}_X$, $\mathbf{Q}^2_X = \mathbf{Q}_X$ (idempotent).
$\mathbf{P}_X\mathbf{Q}_X = \mathbf{Q}_X\mathbf{P}_X = \mathbf{O}$ (orthogonal).

These projectors are useful in partitioning $\mathbf{y}$, the vector of observations on the dependent variable in regression analysis, into $\mathbf{P}_X\mathbf{y}$, the portions of $\mathbf{y}$ that can be accounted for by the predictor variables $\mathbf{X}$, and $\mathbf{Q}_X\mathbf{y}$, the portions of $\mathbf{y}$ that cannot be accounted for by $\mathbf{X}$.

Slight generalizations of the I-orthogonal projectors above lead to K-orthogonal projectors, which are useful in weighted least squares (LS) estimation in regression analysis: Let $\mathbf{K}$ be an *nnd* matrix such that $\text{rank}(\mathbf{KX}) = \text{rank}(\mathbf{X})$. Then,

$$\mathbf{P}_{X/K} = \mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{X})^{-}\mathbf{X}'\mathbf{K}, \tag{3}$$

and

$$\mathbf{Q}_{X/K} = \mathbf{I} - \mathbf{P}_{X/K} \tag{4}$$

are called K-orthogonal projectors onto $\text{Sp}(\mathbf{X}$ and $\text{Ker}(\mathbf{X}')$, respectively, with respect to the metric matrix $\mathbf{K}$.

These projectors have properties similar to those of the I-orthogonal projectors:

$(\mathbf{K}\mathbf{P}_{X/K})' = \mathbf{K}\mathbf{P}_{X/K}$, $(\mathbf{K}\mathbf{Q}_{X/K})' = \mathbf{K}\mathbf{Q}_{X/K}$ (K-symmetric).
$\mathbf{P}^2_{X/K} = \mathbf{P}_{X/K}$, $\mathbf{Q}^2_{X/K} = \mathbf{Q}_{X/K}$ (idempotent).
$\mathbf{P}'_{X/K}\mathbf{K}\mathbf{Q}_{X/K} = \mathbf{Q}'_{X/K}\mathbf{K}\mathbf{P}_{X/K} = \mathbf{O}$ (K-orthogonal).

These projectors are useful in weighted LS (WLS) estimation in regression analysis. When $\mathbf{K}$ is set to $\mathbf{K} = \mathbf{P}_Z$, the K-orthogonal projectors effect instrumental variable estimation. See Yanai [33] for other types of projectors.

2 Constrained Principal Component Analysis (CPCA)

In as early as 1970, Professor Yanai (Yanai [32]) proposed so-called partial principal component analysis (PPCA) to extract components unrelated to certain prescribed effects such as differences in gender, age, levels of education, etc., which amounts to SVD of $\mathbf{Q}_G\mathbf{Y}$ (where $\mathbf{Y}$ is the matrix of criterion variables, and $\mathbf{G}$ the matrix of predictor variables whose effects are to be eliminated). This process consists of two phases, decomposing $\mathbf{Y}$ int $\mathbf{P}_G\mathbf{Y}$ and $\mathbf{Q}_G\mathbf{Y}$, and applying SVD to the latter. While Yanai himself did not explicitly suggest the SVD of $\mathbf{P}_G\mathbf{Y}$, it was known as redundancy analysis, a special case of reduced-rank regression. The two phases may be called External and Internal Analyses.

Similarly, CPCA consists of two major phases: External Analysis and Internal Analysis. External Analysis decomposes the main data matrix according to the external information about the row and columns of a data matrix, which amounts to projections. Internal Analysis further analyses the decomposed matrices into components, which is equivalent to SVD (singular value decomposition).

In CPCA, we consider not only the row-side constraints, $\mathbf{G}$, but also the column-side constraints $\mathbf{H}$, analogously to growth curve models (Potthoff and Roy [12]). This leads to a four-way decomposition of the main data matrix $\mathbf{Y}$ (Takane and Shibayama [20]):

$$\mathbf{Y} = \mathbf{P}_G\mathbf{Y}\mathbf{P}_H + \mathbf{Q}_G\mathbf{Y}\mathbf{P}_H + \mathbf{P}_G\mathbf{Y}\mathbf{Q}_H + \mathbf{Q}_G\mathbf{Y}\mathbf{Q}_H. \tag{5}$$

A similar decomposition is also possible with K-orthogonal projectors.

The decomposition above is a very basic one. When $\mathbf{G}$ and/or $\mathbf{H}$ consist of more than one set of variables, finer decompositions of $\mathbf{Y}$ are possible, corresponding to analogous decompositions of $\mathbf{P}_G$ and.or $\mathbf{P}_H$ (e.g., Takane [17]; Takane and Yanai [21]):

Let $\mathbf{G} = [\mathbf{M}, \mathbf{N}]$, for example. Then,

(1) $\mathbf{P}_G = \mathbf{P}_M + \mathbf{P}_N \Leftrightarrow \mathbf{M}'\mathbf{N} = \mathbf{O}$. ($\mathbf{M}$ and $\mathbf{N}$ are mutually orthogonal.)

(2) $\mathbf{P}_G = \mathbf{P}_M + \mathbf{P}_N - \mathbf{P}_M\mathbf{P}_N \Leftrightarrow \mathbf{P}_M\mathbf{P}_N = \mathbf{P}_N\mathbf{P}_M$. ($\mathbf{M}$ and $\mathbf{N}$ are mutually orthogonal, except their common space, e.g., ANOVA w/o interactions). (3) $\mathbf{P}_G = \mathbf{P}_M + \mathbf{P}_{Q_M N} = \mathbf{P}_N + \mathbf{P}_{Q_N M}$. (Fit one first and the other to the residuals from the fist).

(4) $\mathbf{P}_G = \mathbf{P}_{M/Q_N} + \mathbf{P}_{N/Q_M} \Leftrightarrow \text{rank}(\mathbf{G}) = \text{rank}(\mathbf{M}) + \text{rank}(\mathbf{N})$. ($\mathbf{M}$ and $\mathbf{N}$ are disjoint. Fit both simultaneously).

(5) $\mathbf{P}_G = \mathbf{P}_{GA} + \mathbf{P}_{G(G'G)^-B} \Leftrightarrow \mathbf{A}'\mathbf{B} = \mathbf{O}$, and $\text{Sp}(\mathbf{A}) \oplus \text{Sp}(\mathbf{B}) = \text{Sp}(\mathbf{G}')$. (A matrix of regression coefficients $\mathbf{C}$ constrained by $\mathbf{C} = \mathbf{A}\mathbf{C}^*$ or by $\mathbf{B}'\mathbf{C} = \mathbf{O}$).

The first four decompositions above were noted in Rao and Yanai [13], while (5) is due to Yanai and Takane [34]. Analogous decompositions are possible for $\mathbf{P}_H$, $\mathbf{P}_{G/K}$, and $\mathbf{P}_{H/L}$.

In Internal Analysis, on the other hand, we apply PCA to terms obtained by the external analysis, e.g., $\mathbf{P}_G\mathbf{Y}\mathbf{P}_H$, which amounts to $\text{SVD}(\mathbf{P}_G\mathbf{Y}\mathbf{P}_H)$, whose computation time can be economized considerably by the following procedure:

A theorem on $\text{SVD}(\mathbf{P}_G\mathbf{Y}\mathbf{P}_H)$ (Takane and Hunter [18]): Let $\mathbf{F}_G$ and $\mathbf{F}_H$ be columnwise orthogonal matrices such that $\text{Sp}(\mathbf{G}) = \text{Sp}(\mathbf{F}_G)$ and $\text{Sp}(\mathbf{H}) = \text{Sp}(\mathbf{F}_H)$. Then, $\mathbf{P}_G\mathbf{Y}\mathbf{P}_H = \mathbf{F}_G\mathbf{F}'_G\mathbf{Y}\mathbf{F}_H\mathbf{F}'_H$. Let $\text{SVD}(\mathbf{F}'_G\mathbf{Y}\mathbf{F}_H)$ be denoted as $\mathbf{U}\mathbf{D}\mathbf{V}'$, and let $\text{SVD}(\mathbf{F}_G\mathbf{F}'_G\mathbf{Y}\mathbf{F}_H\mathbf{F}'_H)$ be denoted as $\mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*'}$. Then, $\mathbf{U}^* = \mathbf{F}_G\mathbf{U}$, $\mathbf{V}^* = \mathbf{F}_H\mathbf{V}$, and $\mathbf{D}^* = \mathbf{D}$.

3 Khatri's Lemma

Toward the end of 1980's, I was interested in the relationships among various methods of constrained correspondence analysis (CCA), a special case of CPCA. When I looked through the literature on CCA, I found that there were two ways of incorporating the constraints. Let $\mathbf{U}$ denote the row representation matrix. (For explanation, we consider only the row side constraints.) Two equivalent ways of constraining $\mathbf{U}$ are: (1) $\mathbf{U} = \mathbf{A}\mathbf{U}^*$ (e.g., ter Braak [28]), and (2) $\mathbf{B}'\mathbf{U} = \mathbf{O}$ (e.g., Böckenholt and Böckenholt [1]), where $\mathbf{A}$ and $\mathbf{B}$ are mutually orthogonal, and jointly span the entire row space of a contingency table. The relationship is rather trivial, i.e.,

$$\mathbf{P}_A = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}' = \mathbf{I} - \mathbf{B}(\mathbf{B}'\mathbf{B})^-\mathbf{B}' = \mathbf{Q}_B, \tag{6}$$

if the identity metric is used. I was not sure what would happen if non-identity metric $\mathbf{K}$ is used. Khatri's lemma states the exact relationship for this case (Takane, Yanai, and Mayekawa [26]):

Let $\mathbf{A}$ $(p \times r)$ and $\mathbf{B}$ $(p \times (p-r))$ be matrices such that $\text{rank}(\mathbf{A}) = r$, $\text{rank}(\mathbf{B}) = p - r$, and $\mathbf{A}'\mathbf{B} = \mathbf{O}$. Then (Khatri [9]),

$$\mathbf{I} = \mathbf{A}(\mathbf{A}'\mathbf{K}\mathbf{A})^{-1}\mathbf{A}\mathbf{K} + \mathbf{K}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{K}^{-1}\mathbf{B})^{-1}\mathbf{B}', \tag{7}$$

where $\mathbf{K}$ is a symmetric *pd* (positive definite) matrix.

Several remarks are in order on Khtri's original lemma given above. Khatri's lemma may sometimes be expressed in an alternative form:

$$\mathbf{K} = \mathbf{K}\mathbf{A}(\mathbf{A}'\mathbf{K}\mathbf{A})^{-1}\mathbf{A}\mathbf{K} + \mathbf{B}(\mathbf{B}'\mathbf{K}^{-1}\mathbf{B})^{-1}\mathbf{B}'. \tag{8}$$

Note also that $\mathbf{K}$ and $\mathbf{K}^{-1}$ are interchangeable. Khatri's lemma is useful for rewriting P-type projectors into Q-type projectors (LaMotte [11]; Shapiro [15]; Seber [14]; Takane and Zhou [24]; Verbyla [30]). Khatri's lemma has been generalized in various ways, e.g., let $\mathbf{K}$ be square, but not necessarily symmetric or nonsingular, but $\text{Sp}(\mathbf{B}) \subset \text{Sp}(\mathbf{K})$ and $\text{Sp}(\mathbf{B}) \subset \text{Sp}(\mathbf{K}')$. Then (Khatri [10]),

$$\mathbf{K} = \mathbf{K}\mathbf{A}(\mathbf{A}'\mathbf{K}\mathbf{A})^{-}\mathbf{A}'\mathbf{K} + \mathbf{B}(\mathbf{B}'\mathbf{K}^{-}\mathbf{B})^{-}\mathbf{B}'. \tag{9}$$

Professor Yanai (Yanai and Takane [34]) further extended Khatri's lemma as follows; Let $\mathbf{A}$ $(p \times r)$ and $\mathbf{B}$ $(p \times (p-r))$ be matrices such that $\text{rank}(\mathbf{A}) = r$ and $\text{rank}(\mathbf{B}) = p - r$, and let $\mathbf{M}$ and $\mathbf{N}$ be *nnd* matrices such that
(i) $\mathbf{A}'\mathbf{M}\mathbf{N}\mathbf{B} = \mathbf{O}$,
(ii) $\text{rank}(\mathbf{M}\mathbf{A}) = \text{rank}(\mathbf{A})$,
(iii) $\text{rank}(\mathbf{N}\mathbf{B}) = \text{rank}(\mathbf{B})$.
Then,

$$\mathbf{I} = \mathbf{A}(\mathbf{A}'\mathbf{M}\mathbf{A})^{-}\mathbf{A}'\mathbf{M} + \mathbf{N}\mathbf{B}(\mathbf{B}'\mathbf{N}\mathbf{B})^{-}\mathbf{B}'. \tag{10}$$

This reduces to the original lemma when $\mathbf{M} = \mathbf{K}$ and $\mathbf{N} = \mathbf{K}^{-1}$. Takane [17] further extends it to a rectangular $\mathbf{K}$.

4 The Wedderburn-Guttman (WG) Theorem

The Wedderburn-Guttman (WG) theorem is stated as follows: Let $\mathbf{Y}$ $(n \times p)$ be of rank r, and let $\mathbf{A}$ $(n \times s)$ and $\mathbf{B}$ $(p \times s)$ be such that $\mathbf{A}'\mathbf{Y}\mathbf{B}$ is invertible. Then,

$$\text{rank}(\mathbf{Y}_1) = \text{rank}(\mathbf{Y}) - \text{rank}(\mathbf{Y}\mathbf{B}(\mathbf{A}'\mathbf{Y}\mathbf{B})^{-1}\mathbf{A}'\mathbf{Y}) \tag{11}$$

$$= \text{rank}(\mathbf{Y}) - \text{rank}(\mathbf{A}'\mathbf{Y}\mathbf{B}) = r - s, \tag{12}$$

where

$$\mathbf{Y}_1 = \mathbf{Y} - \mathbf{Y}\mathbf{B}(\mathbf{A}'\mathbf{Y}\mathbf{B})^{-1}\mathbf{A}'\mathbf{Y}. \tag{13}$$

Wedderburn [31] first proved the theorem for $s = 1$. Guttman [5] extended it for $s > 1$. Guttman [6] further proved the reverse, i.e., $\mathbf{Y}_1$ must be of the above form to satisfy the rank condition stated above.

Guttman [5] used the matrix rank method for a proof of the above theorem. In this method, we apply a series of elementary block matrix operations to a matrix to derive a rank formula. We apply another series of elementary block matrix operations to the same matrix to derive another rank formula. Neither operations change the rank of the original matrix, so the two must be equal. Guttman's proof is given in the appendix. Yongge Tian (many papers) derived many interesting rank formula based on this method. It is intriguing to find that Guttman [5] already used the method in 1944 (cf. Khatri [8]).

My initial interest in this theorem stemmed from Hubert's talk (Hubert, Meulman, and Heiser [7]) at the 1989 Meeting of the Psychometric Society at Illinois. This talk was to criticise the ignorance of numerical analysts (e.g., Chu, Funderlic, and Golub [2]) about Guttman's contributions (Guttman [5, 6]) in the WG theorem. When the talk was over, I asked a question: When $\mathbf{A}'\mathbf{Y}\mathbf{B}$ is not invertible, can we replace it by a generalized inverse? I had a feeling that it was possible, while Hubert said it was probably impossible. It has turned out that both of us are only half correct. The answer is yes, but it requires a condition. I initially thought this was purely a rank additivity (subtractivity) problem. That is, we are to prove that

$$\begin{aligned}&\operatorname{rank}(\mathbf{Y} - \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}) \\ &\quad = \operatorname{rank}(\mathbf{Y}) - \operatorname{rank}(\mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}). \qquad (14)\end{aligned}$$

This supposition also included that

$$\operatorname{rank}(\mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}) = \operatorname{rank}(\mathbf{A}'\mathbf{YB}) \qquad (15)$$

always holds. However, Tian and Styan [29] showed the following always holds:

$$\operatorname{rank}(\mathbf{Y} - \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}) = \operatorname{rank}(\mathbf{Y}) - \operatorname{rank}(\mathbf{A}'\mathbf{YB}). \qquad (16)$$

This implies that (15) requires a condition, as does (14), and that the two conditions are equivalent.

The necessary and sufficient (*ns*) condition is stated as follows (Takane and Yanai [22]): Let $\mathbf{C} = \mathbf{B}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'$. Then, the *ns* condition for (14) and (15) to hold is:

$$\mathbf{YCYCY} = \mathbf{YCY}. \qquad (17)$$

There are a number of equivalent conditions, e.g., $(\mathbf{YCYY}^{-})^2 = \mathbf{YCYY}^{-} \Leftrightarrow (\mathbf{Y}^{-}\mathbf{YCY})^2 = \mathbf{Y}^{-}\mathbf{YCY}$. There are also a number of interesting sufficient (but not necessary) conditions, e.g., $(\mathbf{YC})^2 = \mathbf{YC}$ or $(\mathbf{CY})^2 = \mathbf{CY}$, and $\mathbf{CYC} = \mathbf{C}$ (Cline, Funderlic, and Golub [3]; Galantai [4]). The latter is even stronger than the idempotency of $\mathbf{YC}$ or $\mathbf{CY}$.

The WG theorem states the rank condition for the residual matrix. However, from a data analytic viewpoint, the decomposition of the data matrix $\mathbf{Y}$ the theorem implies is even more interesting:

$$\mathbf{Y} = \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y} + (\mathbf{Y} - \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}). \qquad (18)$$

Takane and Hunter [19] developed a new family of CPCA almost exclusively based on this decomposition. The second term of the above decomposition involves a Q-type projector, but it can be replaced by a P-type projector as follows (Takane [17]): Let $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$ be matrices such that

(i) $\operatorname{Sp}(\tilde{\mathbf{A}}) \subset \operatorname{Sp}(\mathbf{Y})$,
(ii) $\operatorname{Sp}(\tilde{\mathbf{B}}) \subset \operatorname{Sp}(\mathbf{Y}')$,
(iii) $\operatorname{rank}(\mathbf{A}'\mathbf{YB}) + \operatorname{rank}(\tilde{\mathbf{B}}'\mathbf{Y}^{-}\tilde{\mathbf{A}}) = \operatorname{rank}(\mathbf{Y})$,
(iv) $\mathbf{A}'\mathbf{YY}^{-}\tilde{\mathbf{A}} = \mathbf{A}'\tilde{\mathbf{A}} = \mathbf{O}$,
(v) $\tilde{\mathbf{B}}'\mathbf{Y}^{-}\mathbf{YB} = \tilde{\mathbf{B}}'\mathbf{B} = \mathbf{O}$.

Then,

$$\mathbf{Y} = \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y} + \tilde{\mathbf{A}}(\tilde{\mathbf{B}}'\mathbf{Y}^{-}\tilde{\mathbf{A}})^{-}\tilde{\mathbf{B}}'. \qquad (19)$$

5 Ridge Operators

In the mid 2000's, I was interested in extending the ridge-type of regularized least squares (RLS) estimation to various multivariate (MV) techniques. These extensions were rather straightforward, and I wrote most of the papers on them with my graduate students. I did not have to bother Professor Yanai. However, as I applied the RLS to so many MV procedures, I thought it would be important to write a paper on ridge operators, which was a common thread running through all of them (Takane [16, 23]).

The simplest form of ridge operators is defined as:

$$\mathbf{R}_X(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda\mathbf{P}_{X'})^{-}\mathbf{X}', \tag{20}$$

where $\mathbf{P}_{X'} = \mathbf{X}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{X}$ is the orthogonal projector onto $\mathrm{Sp}(\mathbf{X}')$. ($\mathbf{P}_{X'} = \mathbf{I}$ if $\mathbf{X}$ is columnwise nonsingular.) This operator arises in the RLS estimation $\min_{\mathbf{c}} = \phi_\lambda(\mathbf{c})$ in regression analysis, where $\phi_\lambda(\mathbf{c}) = \mathrm{SS}(\mathbf{e}) + \lambda \mathrm{SS}(\mathbf{c})_{P_{X'}}$ and $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{c}$. (We assume, w/o loss of generality, that $\mathrm{Sp}(\mathbf{c}) \subset \mathrm{Sp}(\mathbf{X}')$.)

The $\mathbf{R}_X(\lambda)$ and $\mathbf{S}_X(\lambda)$ have properties similar to those of $\mathbf{P}_X$ and $\mathbf{Q}_X$, where $\mathbf{S}_X(\lambda) = \mathbf{I} - \mathbf{R}_X(\lambda)$. For example:

$\mathbf{R}_X(\lambda)$ and $\mathbf{S}_X(\lambda)$ are symmetric and invariant over the choice of a g-inverse of $(\mathbf{X}'\mathbf{X} + \lambda\mathbf{P}_{X'})$.

$\mathbf{R}_X(\lambda)\mathbf{K}_X(\lambda)\mathbf{R}_X(\lambda) = \mathbf{R}_X(\lambda)$ (i.e., $\mathbf{K}_X(\lambda) = \mathbf{R}_X(\lambda)^{+}$.).

$\mathbf{R}_X(\lambda) - \mathbf{R}_X(\lambda)^2 = \mathbf{R}_X(\lambda)\mathbf{S}_X(\lambda) = \mathbf{S}_X(\lambda)\mathbf{R}_X(\lambda) \geq \mathbf{O}$.

$\mathbf{R}_X(\lambda)\mathbf{K}_X(\lambda) = \mathbf{P}_X$, etc.

Similar decompositions of $\mathbf{R}_X(\lambda)$ to those of $\mathbf{P}_X$ are also possible.

The ridge operators defined above can be rewritten as follows using a ridge metric matrix defined below: Let

$$\mathbf{K}_X(\lambda) = \mathbf{P}_X + \lambda(\mathbf{X}\mathbf{X}')^{+} \quad \text{(Ridge Metric Matrix)}. \tag{21}$$

Then, $\mathbf{R}_X(\lambda)$ can be rewritten as:

$$\mathbf{R}_X(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{K}_X(\lambda)\mathbf{X})^{-}\mathbf{X}'. \tag{22}$$

The simple ridge operators introduced above can be generalized into generalized ridge operators:

$$\mathbf{R}_X^{(W,L)}(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X} + \lambda\mathbf{L})^{-}\mathbf{X}'\mathbf{W}, \tag{23}$$

where $\mathbf{L}$ is an *nnd* matrix such that $\mathrm{Sp}(\mathbf{L}) \subset \mathrm{Sp}(\mathbf{X}')$, and $\mathbf{W}$ is an *nnd* matrix such that $\mathrm{rank}(\mathbf{W}\mathbf{X}) = \mathrm{rank}(\mathbf{X})$. As before, the generalized ridge operators can be rewritten as follows using a generalized ridge metric matrix defined below: Let

$$\mathbf{K}_X^{(W,L)}(\lambda) = \mathbf{P}_X + \lambda\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{L}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}. \tag{24}$$

Then,

$$\mathbf{R}_X^{(W,L)}(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{K}_X^{(W,L)}(\lambda)\mathbf{X})^{-}\mathbf{X}'\mathbf{W}. \tag{25}$$

6 Generalized Constrained Canonical Correlation Analysis

In the external analysis of CPCA, a data matrix is decomposed into several components by external information. I initially thought we could do the same in generalized constrained canonical correlation analysis (CANO). We decompose $\mathbf{X}$ and $\mathbf{Y}$ (the matrix of observations on the two sets of variables) separately into several orthogonal components, and then choose one term from each decomposition, and apply CANO to the pair, which amounts to SVD of the product of the orthogonal projectors. It has turned out that this strategy will not do.

CANO analyzes total association between $\mathbf{X}$ and $\mathbf{Y}$, i.e., $\mathrm{tr}(\mathbf{P}_X\mathbf{P}_Y)$. However, $\mathbf{X} = \mathbf{M} + \mathbf{N}$, where $\mathbf{M}'\mathbf{N} = \mathbf{O}$ does not guarantee $\mathbf{P}_X = \mathbf{P}_M + \mathbf{P}_N$. This may be contrasted with a similar situation in which $\mathbf{X} = [\mathbf{M}, \mathbf{N}]$, where $\mathbf{M}'\mathbf{N} = \mathbf{O}$, in which case we indeed have $\mathbf{P}_X = \mathbf{P}_M + \mathbf{P}_N$. This suggests that we need orthogonal decompositions of orthogonal projectors to derive additive decompositions of the total association.

Takane, Yanai, and Hwang [25] derived the following two orthogonal decompositions of $\mathbf{P}_{[X,G]}$ by combining two orthogonal decompositions ((3) and (5)) of the orthogonal projector given in the CPCA section:

(1) Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{W}$ be matrices such that $\mathrm{Sp}(\mathbf{A}) = \mathrm{Ker}(\mathbf{H}'\mathbf{X}'\mathbf{P}_G\mathbf{X})$, $\mathrm{Sp}(\mathbf{B}) = \mathrm{Ker}(\mathbf{H}'\mathbf{X}'\mathbf{Q}_G\mathbf{X})$, and $\mathrm{Sp}(\mathbf{W}) = \mathrm{Ker}(\mathbf{X}'\mathbf{G})$. Then,

$$\mathbf{P}_{[X,G]} = \mathbf{P}_{P_GXH} + \mathbf{P}_{P_GXA} + \mathbf{P}_{Q_GXH} + \mathbf{P}_{Q_GXB} + \mathbf{P}_{GW}. \tag{26}$$

(2) Let $\mathbf{K}$, $\mathbf{U}$, and $\mathbf{V}$ be matrices such that $\mathrm{Sp}(\mathbf{K}) = \mathrm{Ker}(\mathbf{H}'\mathbf{X}'\mathbf{X})$, $\mathrm{Sp}(\mathbf{U}) = \mathrm{Ker}(\mathbf{G}'\mathbf{X}\mathbf{H})$, and $\mathrm{Sp}(\mathbf{V}) = \mathrm{Ker}(\mathbf{G}'\mathbf{X}\mathbf{K})$. Then,

$$\mathbf{P}_{[X,G]} = \mathbf{P}_{P_{XH}G} + \mathbf{P}_{XHU} + \mathbf{P}_{P_{XK}G} + \mathbf{P}_{XKV} + \mathbf{P}_{Q_XG}. \tag{27}$$

We can derive similar decompositions of $\mathbf{P}_{[Y,G_Y]}$ (The subscript Y is put on G to indicate that this is a $\mathbf{G}$ for $\mathbf{Y}$.) We take one term each from a decomposition of $\mathbf{P}_{[X,G_X]}$ and that of $\mathbf{P}_{[Y,G_Y]}$, and apply SVD to the product of the two, e.g.,

$$\mathrm{SVD}(\mathbf{P}_{Q_{G_X}XH_X}\mathbf{P}_{YH_YU_Y}). \tag{28}$$

7 Causal Inference

Causal inference is one of the most important roles of statistics. This was the topic of our conversation when I met him last in the fall of 2013. When randomization is unavailable, there are a lot of pitfalls in establishing causal relationships based on correlational relationships alone. One crucial aspect of the problem is how to eliminate the effects of confounding variables.

The easiest way is to include the effects of the confounding variables in regression analysis along with the predictor variable of interest, although this is easier said than done. Identifying the set of confounding variables is not so easy, although here we assume that they are known. Let $\mathbf{y}$ denote the criterion variable, let $\mathbf{x}$: denote the predictor variable of interest, and let $\mathbf{U}$ denote the matrix of confounding variables. The suggested regression model can be written as:

$$\mathbf{y} = \mathbf{x}a_1 + \mathbf{U}\mathbf{c} + \mathbf{e}_1. \tag{29}$$

The ordinary least squares (OLS) estimate of $\mathbf{x}a_1$ is given by

$$\mathbf{x}\hat{a}_1 = \mathbf{P}_{x/Q_u}\mathbf{y} \tag{30}$$

.

Consider next the regression of $\mathbf{x}$ onto $\mathbf{U}$, i.e.,

$$\mathbf{x} = \mathbf{U}\mathbf{d} + \mathbf{e}_2. \tag{31}$$

The OLS estimate of $\mathbf{Ud}$ is given by

$$\mathbf{U}\hat{\mathbf{d}} = \mathbf{P}_U\mathbf{x}. \tag{32}$$

We call $\mathbf{P}_U\mathbf{x}$ linear propensity scores. Residuals from the above regression $\mathbf{Q}_U\mathbf{x}$ represent the portions of $\mathbf{x}$ left unaccounted for by $\mathbf{U}$.

We now consider using $\mathbf{P}_U\mathbf{x}$ instead of $\mathbf{U}$ in the first regression, i.e.,

$$\mathbf{y} = \mathbf{x}a_2 + \mathbf{P}_U\mathbf{x}b + \mathbf{e}_3. \tag{33}$$

The OLS estimate of $\mathbf{x}a_2$ is given by

$$\mathbf{x}\hat{a}_2 = \mathbf{P}_{x/Q_{P_U x}}\mathbf{y}, \tag{34}$$

where $\mathbf{Q}_{P_U x} = \mathbf{I} - \mathbf{P}_U\mathbf{x}(\mathbf{x}'\mathbf{P}_U\mathbf{x})^{-1}\mathbf{x}'\mathbf{P}_U$.

Since

$$\mathbf{Q}_{P_U x}\mathbf{x} = \mathbf{x} - \mathbf{P}_U\mathbf{x}(\mathbf{x}'\mathbf{P}_U\mathbf{x})^{-1}\mathbf{x}'\mathbf{P}_U\mathbf{x} = \mathbf{Q}_U\mathbf{x}, \tag{35}$$

we obtain

$$\mathbf{P}_{x/Q_{P_U x}}\mathbf{y} = \mathbf{P}_{x/Q_U}\mathbf{y}. \tag{36}$$

This means (30) and (34) are equivalent. This gives the rationale for replacing $\mathbf{U}$ by $\mathbf{P}_U\mathbf{x}$. The latter is more convenient because it is a single variable, and matching on a single variable is much easier than matching on multiple variables.

More recently, methods of of causal inference based on instrumental variables are getting popular. An instrumental variable $\mathbf{z}$ has the following properties:

(1) $\mathbf{z}'\mathbf{U} = \mathbf{0}$ ($\mathbf{z}$ and $\mathbf{U}$ are uncorrelated),
(2) $\mathbf{z}'\mathbf{x} \neq 0$ ($\mathbf{z}$ and $\mathbf{x}$ are correlated),
(3) $\mathbf{z}'\mathbf{Q}_{[U,x]}\mathbf{y} = 0$ (*i.e.*, $\mathbf{z}$ has a predictive power on $\mathbf{Y}$ only through $\mathbf{x}$).

How is $\mathbf{z}$ related to $\mathbf{P}_U\mathbf{x}$ or $\mathbf{Q}_U\mathbf{x}$?

Assume $\mathbf{z} = c\mathbf{Q}_U\mathbf{x}$, where c is a normalization factor. This $\mathbf{z}$ satisfies (1) and (2) above. That it also satisfies (3) can be seen from:

$$(1/c)\mathbf{z}'\mathbf{Q}_{[U,x]}\mathbf{y} = \mathbf{x}'\mathbf{Q}_U\mathbf{Q}_{[U,x]}\mathbf{y} = \mathbf{x}'\mathbf{Q}_{[U,x]}\mathbf{y} = 0. \tag{37}$$

Consider the regression model:

$$\mathbf{y} = \mathbf{x}a_3 + \mathbf{e}_4. \tag{38}$$

The IV estimate of $\mathbf{x}a_3$ is given by

$$\mathbf{x}\hat{a}_3 = \mathbf{P}_{x/P_z}\mathbf{y} = \mathbf{P}_{x/Q_U}\mathbf{y}. \tag{39}$$

Since $\mathbf{P}_z = \mathbf{Q}_U\mathbf{x}(\mathbf{x}'\mathbf{Q}_U\mathbf{x})^{-1}\mathbf{x}'\mathbf{Q}_U$ and $\mathbf{x}'\mathbf{P}_z = \mathbf{x}'\mathbf{Q}_U$, this is identical to (30) and (34). This implies that the $\mathbf{z}$ defined above is an ideal IV.

8 Conclusions

This paper overviewed Professor Yanai's contributions to MV analysis. He adamantly emphasized linear algebraic aspects of MV analysis. His framework was grad, yet easy to understand. After almost half a century since I got to know him, I am still working within the framework of Professor Yanai.

9 Appendix: The Matrix Rank Method used by Guttman

The following is the proof of the original WG theorem by Guttman [5]. Let

$$\mathbf{C} = \begin{bmatrix} \mathbf{I}_s & (\mathbf{A}'\mathbf{Y}\mathbf{B})^{-1}\mathbf{A}'\mathbf{Y} \\ \mathbf{Y}\mathbf{B} & \mathbf{Y} \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ -\mathbf{Y}\mathbf{B} & \mathbf{I} \end{bmatrix},$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{I} & -(\mathbf{A}'\mathbf{Y}\mathbf{B})^{-1}\mathbf{A}'\mathbf{Y} \\ \mathbf{O} & \mathbf{I} \end{bmatrix}.$$

Then,

$$\mathbf{ECF} = \begin{bmatrix} \mathbf{I}_s & \mathbf{O} \\ \mathbf{O} & \mathbf{Y}_1 \end{bmatrix},$$

so that

$$\text{rank}(\mathbf{C}) = s + \text{rank}(\mathbf{Y}_1). \tag{40}$$

On the other hand, let

$$\mathbf{G} = \begin{bmatrix} \mathbf{I} & -(\mathbf{A}'\mathbf{Y}\mathbf{B})^{-1} \\ \mathbf{O} & \mathbf{I} \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ -\mathbf{B} & \mathbf{I} \end{bmatrix}.$$

Then,

$$\mathbf{GCH} = \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{Y} \end{bmatrix},$$

so that

$$\text{rank}(\mathbf{C}) = \text{rank}(\mathbf{Y}). \tag{41}$$

We obtain the WG theorem by combining (40) and (41).

Acknowledgement

The work reported in this talk has been supported by the NSERC (Natural Sciences and Engineering Research Council of Canada) Discovery Grant (# 10630) awarded to the author.

References

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

[2] Chu, M. T., Funderlic, R. E., and Golub, G. H. (1995). A rank-one reduction formula and its applications to matrix factorizations. *SIAM Review*, 37, 512-530.

[3] Cline, R. E., and Funderlic, R. E. (1979). The rank of a difference of matrices and associated generalized inverses. *Linear Algebra and its Applications*, 24, 185-215.

[4] Galantai, A. (2007). A note on generalized rank reduction. *Act Math Hung*, 116, 239-246.

[5] Guttman, L. (1944). General theory and methos for matric factoring. *Psychometrika*, 9, 1-16.

[6] Guttman, L. (1957). A necessary and sufficient formula for matric factoring. *Psychometrika*, 22, 79-81.

[7] Hubert, L., Meulman, J., and Heiser, W. J. (2000). Two purposes of matrix factorization: A historical appraisal. *SIAM Review*, 42, 68-82.

[8] Khatri, C. G. (1961). A simplified approach to the derivation of the theorems on the rank of a matrix. *Journal of the Maharaja Sayajirao University of Baroda*, 10, 1-5.

[9] Khatri, C. G. (1966). A note on a MANOVA model applied to problems in growth curve. *Ann I Stat Math*, 18, 75-86.

[10] Khatri, C. G. (1990). Some properties of BLUE in a linear model and canonical correlations associated with linear transformations. *J Multivariate Anal*, 34, 211-226.

[11] LaMotte, L. R. (2007). A direct derivation of the REML likelihood function. *Stat Pap*, 48, 321-327.

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

[13] Rao, C. R., and Yanai, H. (1979). General definition and decomposition of projectors and some applications to statistical problems. *J Stat Plan Infer*, 3, 1-17.

[14] Seber, G. F. A. (1984). *Multivariate Observations*. New York: Wiley.

[15] Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *J Am Stat Assoc*, 81, 142-149.

[16] Takane, Y. (2008). More on regularization and (generalized) ridge operators. In K. Shigemasu, A. Okada, T. Imaizumi, and T. Hoshino (Eds.), *New Trends in Psychometrics*, (pp. 443-452). Tokyo: University Academic Press.

[17] Takane, Y. (2013). *Constrained Principal Component Analysis and Related Techniques*. Boca Raton, FL: Chapman and Hall/CRC Press.

[18] Takane, Y., and Hunter, M. A. (2001). Constrained principal component analysis: A comprehensive theory. *Appl Algebr Eng Comm*, 12, 391-419.

[19] Takane, Y., and Hunter, M. A. (2011). New family of constrained principal component analysis (CPCA). *Linear Algebra Appl*, 434, 2539-2555.

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

[21] Takane, Y., and Yanai, H. (1999). On oblique projectors. *Linear Algebra Appl*, 289, 297-310.

[22] Takane, Y., and Yanai, H. (2005). On the Wedderburn-Guttman theorem. *Linear Algebra Appl*, 410, 267-278.

[23] Takane, Y., and Yanai, H. (2008). On ridge operators. *Linear Algebra Appl*, 428, 1778-1790.

[24] Takane, Y., and Zhou, L. (2012). On two expressions of the MLE for a special case of the extended growth curve models. *Linear Algebra Appl*, 436, 2567-2577.

[25] Takane, Y., Yanai, H., and Hwang, H. (2006). An improved method for generalized constrained canonical correlation analysis. *Comp Stat Data An*, 50, 221-241.

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

[27] Takeuchi, K., Yanai, H., and Mukherjee, B. N. (1982). *The Foundation of Multivariate Analysis*. New Delhi: Wiley Eastern and New York: Halsted Press.

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

[29] Tian, Y., and Styan, G. P. H. (2009). On some matrix equalities for generalized inverese with applications. *Linear Algebra Appl*, 430, 2716-2733.

[30] Verbyla, A. P. (1990). A conditional derivation of residual maximum likelihood. *Aust J Stat*, 32, 227-230.

[31] Wedderburn, J. H. M. (1934). *Lectures on Matrices*. Colloquium Publication, Vol. 17, Providence: American Mathematical Society.

[32] Yanai, H. (1970). Factor analysis with external criteria. *Jpn Psychol Res*, 12, 143-153.

[33] Yanai, H. (1990). Some generalized forms of least squares g-inverse, minimum norm g-inverse and Moore-Penrose inverse matrices. *Comp Stat Data An*, 10, 251-260.

[34] Yanai, H., and Takane, Y. (1992). Canonical correlation analysis with linear constraints. *Linear Algebra Appl*, 176, 75-82.

[35] Yanai, H., Takeuchi, K., Takane, Y. (2011). *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition*. New York: Springer.