Matrix Methods and Their Applications to Factor Analysis

Haruo Yanai

National Center for University Entrance Examinations, Tokyo, Japan

Yoshio Takane

Department of Psychology, McGill University, Montreal, Canada

Abstract

Since the introduction of the Spearman's two factor model in 1904, a number of books and articles on factor analysis theories have been published. During the same period, a number of matrix methods have also been developed, particularly in the theory of g-inverses and projection matrices. In this chapter, we integrated these two lines of developments, matrix methods and some important topics of factor analysis such as identifiability conditions, communality problems, analysis of image and anti-image variables, estimation of factor scores, and equivalence conditions on canonical factor analysis, thereby extending some of the earlier theories. In particular, we developed the conditions under which the SMC of a variable is equal to the communality of the variable, and some equivalent conditions under which the eigenvalues resulting from canonical factor analysis are either 1 or 0. We also introduced methods for estimating factor score matrices when the uniqueness variance matrix is singular.

1 Introduction

Over the past hundred years since the introduction of Spearman's two-factor model of intelligence (in 1904), a number of books and articles have been published on factor analysis theories. During the same period of time, there have been a number of interesting developments in matrix theory, particularly in the theory of g-inverses and projectors. In this chapter we attempt to integrate these two lines of developments, matrix methods and some important topics of factor analysis such as identifiability conditions, communality problems with special reference to squared multiple correlation (SMC), image and anti-image analysis, estimation of factor scores, equivalence conditions on canonical factor analysis, etc. Through this exercise, we also attempt to generalize some of the earlier theories. Throughout this chapter, we emphasize the use of g-inverse and projection matrices, which have been proven useful (Takeuchi, Yanai, & Mukherjee, 1982; see also Takane (2004)) in explicating some intricate concepts underlying factor analysis models as well as other multivariate data analysis techniques. All matrices considered in this paper are real matrices.

2 Fundamentals of matrix methods

2.1 General definitions of g-inverse matrices and orthogonal projectors

Let A be a matrix of order $n \times m$, and let X be a matrix of order $m \times n$. Consider the following four equations:

$$\text{(i) } AXA = A, \quad \text{(ii) } XAX = A, \quad \text{(iii) } (AX)' = AX, \quad \text{(iv) } (XA)' = XA. \tag{1}$$

Matrix X satisfying (i) is called g-inverse of A and is generally denoted as A^-, while X satisfying both (i) and (iii) is called least squares g-inverse of A, and X satisfying both (i) and (iv) is called minimum norm g-inverse. These three types of g-inverses are not uniquely determined. Matrix X satisfying all of the above four conditions is called Moore-Penrose (g)-inverse matrix and is generally denoted as A^+. The Moore-Penrose inverse is uniquely determined. (See (c) below.)

We give some basic properties of g-inverses and orthogonal projectors:

(a) Let $X = A_\ell^-$ be a least squares g-inverse of A. Then,

$$AX = AA_\ell^- = A(A'A)^- A' = P_A, \tag{2}$$

where P_A is the orthogonal projector onto $\text{Sp}(A)$, space spanned by the column vectors of A.

(b) Let $X = A_m^-$ be a minimum norm g-inverse of A. Then,

$$XA = A_m^- A = A'(AA')^- A = P_{A'}, \tag{3}$$

where $P_{A'}$ is the orthogonal projector onto $\text{Sp}(A')$, space spanned by the row vectors of A. Observe that P_A and $P_{A'}$ are symmetric and invariant over any choice of g-inverse of $A'A$ and AA', respectively, and for any choice of bases vectors spanning $\text{Sp}(A)$ and $\text{Sp}(A')$, respectively.

(c) Let X_1 and X_2 be two Moore-Penrose inverse matrices of A. Then,

$$X_1 = (X_1A)X_1 = (X_2A)X_1 = X_2(AX_1) = X_2(AX_2) = X_2 \tag{4}$$

due to the relationships given in (2) and (3). This shows the uniqueness of the Moore-Penrose inverse matrix.

2.2 Decompositions of the orthogonal projector

Let A and B be $n \times p$ and $n \times q$ matrices, respectively, and let $\text{Sp}(A)$ and $\text{Sp}(B)$ represent subspaces spanned by the column vectors of A and B. Let I_n be the identity matrix of order n. Then, $Q_A = I_n - P_A$ and $Q_B = I_n - P_B$ are the orthogonal projectors onto $\text{Sp}(A)^\perp$ and $\text{Sp}(B)^\perp$, respectively, where $\text{Sp}(A)^\perp$ and $\text{Sp}(B)^\perp$ are the ortho-complement subspaces of $\text{Sp}(A)$ and $\text{Sp}(B)$. Obviously, $P_AQ_A = Q_AP_A = P_BQ_B = Q_BP_B = O$.

We introduce three important properties of the orthogonal projectors:

Property 1 (Rao & Yanai, 1979). Let $\text{Sp}(A, B)$ represent the space spanned by column vectors of matrix $[A, B]$. Let $P_{[A,B]}$ be the orthogonal projector onto $\mathrm{Sp}(A, B)$. Then,

$$P_{[A,B]} = P_A + P_{Q_A B} = P_B + P_{Q_B A}. \tag{5}$$

Property 2 (Yanai & Puntanen, 1993). Let $Q_{[A,B]}$ be the orthogonal projector onto the ortho-complement subspace of $\mathrm{Sp}(A, B)$, that is, $\mathrm{Sp}(A, B)^{\perp}$. Then,

$$Q_{[A,B]} = Q_{Q_A B} Q_A = Q_{Q_B A} Q_B. \tag{6}$$

Property 3 (Baksalary, 1987). Let A and B be $n \times p$ and $n \times q$ matrices, respectively. Further, let P_A and P_B be orthogonal projectors onto $\mathrm{Sp}(A)$ and $\mathrm{Sp}(B)$. Then the following eight statements are equivalent:

1) $P_A P_B = P_B P_A$.	2) $A'B = A'P_B P_A B$.
3) $(P_A P_B)^2 = P_A P_B$.	4) $Q_B P_A B = O$.
5) $Q_A P_B A = O$.	6) $P_{[A,B]} = P_A + P_B - P_A P_B$.
7) $\mathrm{rank}(Q_B A) = \mathrm{rank}(A) - \mathrm{rank}(A'B)$.	8) $A'Q_B Q_A B = O$.

Bakasalary (1987, Theorem 1) provides thirty eight other equivalent conditions.

Property 4 (Baksalary & Styan, 1990). Let A, B, P_A, P_B, Q_A and Q_B be matrices as defined in Properties 1, 2 and 3. Then,

$$\mathrm{rank}(A'B) = \mathrm{rank}(A'Q_B Q_A B) + \mathrm{rank}(A) + \mathrm{rank}(B) - \mathrm{rank}(A, B). \tag{7}$$

A straightforward proof of the equivalence between 7) and 8) in Property 3 can be given by combining Property 4 and the following rank formula:

$$\mathrm{rank}(A, B) = \mathrm{rank}(A) + \mathrm{rank}(Q_A B) = \mathrm{rank}(B) + \mathrm{rank}(Q_B A). \tag{8}$$

2.3 Image and anti-image vectors

Let $X = [x_1, \cdots, x_p]$ be a column-wise centered data matrix of order $n \times p$. Further, let $X_{(j)} = [x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_p]$ be the n by $p - 1$ matrix excluding the j-th column vector, x_j, from X. Then, using an orthogonal projector we can write the squared multiple correlation, SMC(j), obtained by regressing x_j onto $X_{(j)}$ as

$$\mathrm{SMC}(j) \equiv R^2_{j/(j)} = ||P_{X_{(j)}} x_j||^2/||x_j||^2. \tag{9}$$

We also have

$$x_j = P_{X_{(j)}} x_j + Q_{X_{(j)}} x_j \tag{10}$$

for $i = 1, \cdots, p$, where $P_{X_{(j)}} x_j$ and $Q_{X_{(j)}} x_j$ are called image vector of x_j on $\mathrm{Sp}(X_{(j)})$ and anti-image vector of x_j on $\mathrm{Sp}(X_{(j)})^{\perp}$, respectively. Note that the image and anti-image vectors of x_j are orthogonal.

Observe that (9) ensures that SMC(j) can be computed even if $X'_{(j)} X_{(j)}$ is singular. Let

$$X_I = [P_{X_{(1)}} x_1, \cdots, P_{X_{(p)}} x_p], \tag{11}$$

and

$$X_A = [Q_{X_{(1)}} x_1, \cdots, Q_{X_{(p)}} x_p]. \tag{12}$$

Then, it follows from (10) that $X = X_I + X_A$. Assume that X is columnwise standardized. Then, $R = (1/n)X'X$, where R is the correlation matrix.

Property 5 (Yanai & Mukherjee, 1987, Theorem 1).

$$\frac{1}{n} X'_A X = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_p \end{bmatrix} \equiv D, \tag{13}$$

where $1 - d_j = R^2_{j/(j)}$ (the latter having been defined in (9)), and

$$X_A = P_X X_A = X R^- D, \tag{14}$$

$$X'_A X_A = D R^- D, \tag{15}$$

$$X'_I X_I = (R - D) R^- (R - D), \tag{16}$$

and

$$X'_A X_I = D - DR^- D. \tag{17}$$

Proof. (13) follows immediately by noting that $x'_j Q_{X_{(j)}} x_i = 0$ because $x_i \in \mathrm{Sp}(X_{(j)})$ for $i \neq j$. (14) follows, since $\mathrm{Sp}(X_A) \subset \mathrm{Sp}(X)$ and $X_A = P_X X_A = X((1/n)X'X)^-(1/n)X'X_A = XR^-D$. (15), (16) and (17) are direct consequences of (13) and (14). To prove (15), note that $X'_A X_A = X(X'X)^- X_A = X'_A X R^- D = DR^- D$, observing that $X = X_A + X_I$. □

The above results are extensions of Kaiser (1976) in that R^+ (the Moore-Penrose inverse of R) is replaced by a weaker g-inverse R^-.

Now, from X one can construct an $n \times j$ matrix of the form $X_j = [x_1, \cdots, x_j]$. Further, let $P_{Y|X} = P_{Q_X Y}$. Then, Properties 1 and 2 can be extended to the following lemmas.

Lemma 1 (Rao & Yanai, 1979). Let $V_i = \mathrm{Sp}(X_i)$, and let $V_{i|i-1} = \{x | x = Q_{X_{i-1}} y,\ y \in V_i\}$. Then,

$$P_X = P_{X_1} + P_{X_2|X_1} + \cdots + P_{X_j|X_{j-1}} + \cdots + P_{X_p|X_{p-1}}, \tag{18}$$

where $P_{X_i|X_{i-1}}$ is the orthogonal projector onto $V_{i|i-1}$.

Lemma 2.

$$Q_X = Q_{X_1} Q_{X_2|X_1} \cdots Q_{X_j|X_{j-1}} \cdots Q_{X_p|X_{p-1}}. \tag{19}$$

2.4 Matrix inequalities

Property 6 (Beckenbach & Bellman, 1961). If A and B are positive-semidefinite (PSD) matrices of order p, such that $A - B$ is also PSD, then

$$\rho_j(A) \geq \rho_j(B) \tag{20}$$

for $1 \leq j \leq p$, where $\rho_j(A)$ is the j-th largest eigenvalue of A.

Property 7 (Poincare Separation Theorem). Let A be a symmetric matrix of order p, and let B be a $p \times m$ matrix such that $B'B = I_m$. Then,

$$\rho_{p-m+i}(A) \leq \rho_i(B'AB) \leq \rho_i(A) \tag{21}$$

for $i = 1, \cdots, p$.

Property 8 (Anderson & Gupta, 1963). If A and B are symmetric matrices of order p,

$$\rho_i(A+B) \leq \rho_j(A) + \rho_k(B) \tag{22}$$

for $j + k \leq i + 1$.

2.5 Miscellaneous properties of matrix and its rank

Property 9 (Yanai, 1990). Let A and B be matrices of order $p \times m$ and $q \times m$, respectively. Then, $\text{rank}(AB') = \text{rank}(A)$ is necessary and sufficient for

$$B'(AB')^{-}AB' = B'. \tag{23}$$

Further, let $\text{rank}(AB') = \text{rank}(A) = \text{rank}(B)$. Then, $B'(AB')^{-}A$ is the projector onto $\text{Sp}(B')$ along $\text{Ker}(A)$.

Property 10 (Kristof, 1970). Let T_j $(j = 1, \cdots, m)$ denote orthognal matrices of order p, and let D_j $(j = 1, \cdots, m)$ denote diagonal matrices of order p with nonnegative diagonal elements. Then, $\text{tr}(\prod_{j=1}^{m} T_j D_j) \leq \text{tr}(\prod_{j=1}^{m} D_j)$.

When $m = 1$, Property 10 reduces to the following.

Property 11 (ten Berge, 1993). Let T and X be $n \times p$ matrices. Let $T'T = I_p$, and let the singular value decomposition of X be given by $X = V\Delta U'$, where $V'V = U'U = UU' = I_p$. Then, $\text{tr}(T'X) \leq \text{tr}(\Delta)$, and the equality is attained when

$$T = VU' = X(X'X)^{-1/2}. \tag{24}$$

Proof. Let $H = U'T'V$ denote a square matrix of order p. Then, $H'H = V'TUU'T'V = V'P_TV \leq V'V = I_p$, where $V'P_TV \leq V'V$ indicate $V'V - V'P_T$ is PSD. Since $\sum_{k=1}^{p} h_{jk}^2 \leq 1$ implies $h_{jj}^2 \leq 1$, where $H = (h_{jk})$, we obtain $0 \leq h_{jj} \leq 1$, establishing

$$\text{tr}(T'X) = \text{tr}(T'V\Delta U') = \text{tr}((U'T'V)\Delta) = \text{tr}(H\Delta) = \sum_{j=1}^{p} h_{jj}\delta_j \leq \sum_{j=1}^{p} \delta_j = \text{tr}(\Delta),$$

where δ_j is the j-th diagonal element of Δ. The equality holds when $H = I_p$ yielding $U'T'V = I_p$, which implies $T = VU'$. $\square$

3 Applications of matrix methods to factor analysis

3.1 Lower bounds for communalities

Let X denote an $n \times p$ columnwise standardized data matrix, and consider the following traditional factor analysis model:

$$X = F\Lambda' + E\Psi^{1/2}, \tag{25}$$

where $F = [f_1, \cdots, f_m]$ is the $n \times m$ common factor score matrix, Λ is the $p \times m$ factor loading matrix, $E = [e_1, \cdots, e_p]$ is the $n \times p$ unique factor score matrix, and Ψ is the positive-definite diagonal matrix of order p of uniqueness variances. We typically assume $(1/n)E'E = I_p$, and $F'E = O$. In an orthogonal solution, we additionally assume $(1/n)F'F = I_m$, so that

$$R = \Lambda\Lambda' + \Psi, \tag{26}$$

where $R = (1/n)X'X$ is a correlation matrix, and Ψ is a diagonal matrix whose j-th diagonal element, ψ_j, is the uniqueness variance of the j-th variable, x_j. Let h_j^2 denote the communality of this variable satisfying $h_j^2 + \psi_j = 1$ for $j = 1, \cdots, p$.

We first give a property that allows to represent the communality in terms of orthogonal projector onto $\text{Sp}(F)$, the space spanned by column vectors of F.

Property 12. Let h_j^2 denote the communality of the j-th observed variable, x_j, and let P_F denote the orthogonal projector onto $\text{Sp}(F)$. Then,

$$h_j^2 = ||P_F x_j||^2/||x_j||^2. \tag{27}$$

The relationships among x_j, $\text{Sp}(F)$, $P_F x_j$, and $h_j = ||P_F x_j||$ are depicted in Figure 1.

Insert Figure 1 about here

Observe that $0 \leq h_j^2 \leq 1$, where the first equality holds if $x_j \in \mathrm{Sp}(F)^\perp$ and the second equality if $x_j \in \mathrm{Sp}(F)$. The uniqueness variance, ψ_j, is obtained by

$$\psi_j = ||Q_F x_j||^2/||x_j||^2. \tag{28}$$

We next give a well-known property on the relationship between communality and SMC.

Property 13 (Roff, 1936). For $1 \leq j \leq p$, SMC(j) is a lower bound to communality h_j^2, i.e.,

$$\mathrm{SMC}(j) \leq h_j^2. \tag{29}$$

Using Property 1, we are in a position to give a straightforward proof of Property 13 and look into the conditions under which the equality in (29) holds.

Theorem 1. Let $X_{(j)} = [x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_p]$ be a columnwise standardized n by $p-1$ matrix obtained by eliminating x_j from X. Then, (29) holds, and the equality in (29) holds in the following two cases:

$$\mathrm{SMC}(j) = 1 \quad \text{(Case 1)}, \tag{30}$$

and

$$\mathrm{SMC}(j) \neq 1, \text{ and } r^{ji}\psi_i = 0 \text{ for any } i \neq j \quad \text{(Case 2)}, \tag{31}$$

where r^{ji} is the (j,i)-th element of R^-.

Proof. By Property 1, we have

$$P_F + P_{Q_F X_{(j)}} = P_{X_{(j)}} + P_{Q_{X_{(j)}} F}. \tag{32}$$

By pre- and post-multiplying the above equation by x_j' and x_j, respectively, we obtain from (26) and $(1/n)F'F = I_m$ that

$$(1/n)x_j' Q_F X_{(j)} = (1/n)x_j' X_{(j)} - x_j' F F' P_{X_{(j)}} = r_{j/(j)}' - \lambda_j' \Lambda_{(j)} = 0, \tag{33}$$

where $\Lambda_{(j)}$ is the factor loading matrix with $p-1$ variables excluding x_j, λ_j is the vector of factor loadings of the j-th variable, and $r_{j/(j)}$ is the vector of correlation coefficients between x_j and the remaining $p-1$ variables. It follows from (27) and (33) that

$$h_j^2 = \mathrm{SMC}(j) + x_j' P_{Q_{X_{(j)}} F} x_j, \tag{34}$$

which implies (29), since $x_j' P_{Q_{X_{(j)}} F} x_j$ is nonnegative.

We now look for conditions under which the equality holds in (29). If (33) is true, then $x_j' P_{Q_{X_{(j)}} F} x_j = 0$, leading to $x_j' Q_{X_{(j)}} F = 0'$, which implies that anti-image vector $Q_{X_{(j)}} x_j$ $(j = 1, \cdots, p)$ is orthogonal to $\mathrm{Sp}(F)$. Noting that $\mathrm{Sp}(Q_{X_{(j)}}) \subset \mathrm{Sp}(X)$, we obtain

$$(P_X Q_{X_{(j)}} x_j)' F = x_j' Q_{X_{(j)}} X R^- \Lambda = O. \tag{35}$$

By postmultiplying (35) by Λ', and using (26), we obtain from (13) that

$$(Q_{X_{(j)}} x_j)' X((1/n)X'X)^- \{(1/n)X'X - \Psi\} =$$
$$(Q_{X_{(j)}} x_j)' X(I_p - ((1/n)X'X)^- \Psi) = (0, \cdots, 0, 1 - \mathrm{SMC}(j), 0, \cdots, 0)(I_p - R^- \Psi) = 0',$$

which implies

$$(1 - \mathrm{SMC}(j)) r^{ji} \psi_i = 0 \ (i \neq j), \ \text{ and } \ (1 - \mathrm{SMC}(j))(1 - r^{jj} \psi_j) = 0. \tag{36}$$

This completes the proof of Theorem 1. □

We provide an example of Theorem 1, assuming that $r^{ji} \neq 0$, which implies $\psi_i = 0$ $(i \neq j)$, and $h_i^2 = 1$ $(i \neq j)$. We will discuss the case in which $r^{ji} = 0$ $(i \neq j)$ later.

Example 1. Suppose that the correlation matrix among four variables, x_1, x_2, x_3, and x_4, is given by

$$R = \begin{bmatrix} 1 & 0 & a & a \\ 0 & 1 & a & -a \\ a & a & 1 & 0 \\ a & -a & 0 & 1 \end{bmatrix},$$

where $2a^2 \leq 1$. The SMC of x_1 can be computed as

$$1 - \det(R)/\det \begin{bmatrix} 1 & a & -a \\ a & 1 & 0 \\ -a & 0 & 1 \end{bmatrix} = 1 - \frac{(1-2a^2)^2}{(1-2a^2)} = 2a^2,$$

provided that $2a^2 \neq 1$. Similarly, it can be shown that the SMC's of all the four variables are equal to $2a^2$.

The following factor loading matrix Λ and the uniqueness variance matrix Ψ,

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ a & a \\ a & -a \end{bmatrix}, \text{ and } \Psi = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1-2a^2 & 0 \\ 0 & 0 & 0 & 1-2a^2 \end{bmatrix},$$

on the other hand, satisfy the factor analysis model, (26). The communalities of the four variables can be computed as $(1, 1, 2a^2, 2a^2)$. Thus, the SMC's are equal to the communalities for variables 3 and 4, while the SMC's are smaller than (or equal to) the communalities for variables 1 and 2. Since the communalities of variables 1 and 2 are unity, factors f_1 and f_2 can be rotated to coincide with them. Since the SMC's are equal to the squared length of the projection of x_3 and x_4 onto the factor space which are now spanned by x_1 and x_2 , it can be easily seen that the SMC's of variables x_3 and x_4 coincide with their communalities. In terms of the factor analysis model, we can write

$$\psi_1 = \psi_2 = 0, \quad \text{and} \quad \psi_3 = \psi_4 = 1 - 2a^2. \tag{37}$$

This result covers Case 2 in (31). It also covers Theorem 3 of Roff (1936, p. 5), which states that SMC(j) is equal to the communality of variable j, if variables contain m ($m < p$) statistically independent variables each with unit communality (where p is the number of variables and m is the number of common factors).

In (31) it is important to consider the case in which $\psi_i = 0$ ($i \neq j$) does not hold. In such a case, $r^{ji} = 0$ ($i \neq j$) should be true. Since r^{ji} is the (j, i)-th element of R^-, it follows from 8) of Property 3, that

$$(Q_{X_{(j)}} x_j)'(Q_{X_{(i)}} x_i) = 0 \quad (i \neq j), \tag{38}$$

provided that $R_{j/(j)} \neq 1$, which implies that the diagonal matrix D as defined by (13) is nonsingular. (38) implies that the anti-image of variable j is uncorrelated with that of variable i. It is interesting to note that (38) is closely related to Theorem 4 of Guttman (1953), which states that if a common-factor space of dimensionality m is determinate for an infinitely large universe of content, then there is no other determinate common factor space. In this case, the communalities are uniquely determined and are equal to the corresponding total norms, and in addition the common-factor scores are the total image scores, and the unique factor scores are the total ant-images. If (38) holds for any combination of i and j, then anti-image variable $Q_{X_{(j)}} x_j$ behaves like the unique factor, e_j, corresponding to the j-th variable, x_j.

Note 1. If Sp(F) is a subspace of Sp($X_{(j)}$), then $P_{X_{(j)}} F = F$, leading to $X'_{(j)} Q_{X_{(j)}} F = O$. In case of orthogonal factor analysis, F is columnwise orthogonal. Then, if m vectors in $X_{(j)}$ are orthogonal, Sp(F) can be embedded in a subspace of Sp($X_{(j)}$). Thus, the equality in (29) holds.

Note 2. Let $X_1 = [x_1, \cdots, x_k]$ and $X_2 = [x_{k+1}, \cdots, x_p]$, which satisfies the following factor analysis model:

$$[X_1, X_2] = F[\Lambda_1', \Lambda_2'] + E \begin{bmatrix} \Psi_1^{1/2} & O \\ O & \Psi_2^{1/2} \end{bmatrix},$$

where F is the $n \times m$ matrix of common factor scores, Λ_1 and Λ_2 are $k \times m$ and $(p-k) \times m$ factor loading matrices corresponding to X_1 and X_2, respectively, and Ψ_1 and Ψ_2 are diagonal matrices of uniqueness variances of order k and $p-k$, corresponding to X_1 and X_2, respectively. Then,

$$\Lambda_j \Lambda_j' \geq X_j' P_{X_i} X_j, \ (j, i = 1, 2, j \neq i) \tag{39}$$

where $\Lambda_j \Lambda_j' \geq X_j' P_{X_i} X_j$ means $\Lambda_j \Lambda_j' - X_j' P_{X_i} X_j$ is PSD.

From Property 1, we have

$$P_{[F,X_i]} = P_F + P_{Q_F X_i} = P_{X_i} + P_{Q_{X_i} F}. \tag{40}$$

Premultiplying (40) by P_{X_j} and noting that $(1/n)X_j' Q_F X_i = (1/n)X_j' X_i - \Lambda_j \Lambda_i' = O$, we obtain $X_j' P_F X_j = X_j' P_{X_i} X_j + X_j' P_{Q_{X_i} F} X_j$, establishing (39). The term on the left side of (39) represents generalized forms of communalities for variables X_j, and the term on the right may be called generalized SMC's.

3.2 Stronger upper and lower bounds for communalities

In this section, we consider the random model of factor analysis as opposed to the traditional model of factor analysis introduced earlier in (25). The random model of factor analysis is written as

$$x = \Lambda f + e \tag{41}$$

with $\mathrm{E}(f) = 0$, $\mathrm{E}(e) = 0$, $\mathrm{Cov}(f, e) = \mathrm{E}(f'e) = O$, $\mathrm{V}(f) = \Phi$, and $\mathrm{V}(e) = \Psi$, where E, V, and Cov are expectation, variance, and covariance operators, respectively. The corresponding representation of the factor analysis model in terms of a correlation matrix can be expressed as

$$\Sigma = \Lambda \Phi \Lambda' + \Psi, \tag{42}$$

where Σ is the population correlation matrix. We have

Property 14 (Yanai & Ichikawa, 1990).

(a) Let $h^2_{(j)}$ denote the j-th largest communality among the p variables. Then, for $1 \leq j \leq p$,

$$h^2_{(j)} \geq 1 - \rho_{p+1-j}(\Sigma), \tag{43}$$

where $\rho_{p+1-j}(\Sigma)$ is the $(p+1-j)$-th largest eigenvalue of Σ.

(b) For any positive definite correlation matrix Σ with distinct eigenvalues, we have

$$1 - \rho_p(\Sigma) \geq \mathrm{SMC}(j) \tag{44}$$

for $1 \geq j \geq p$.

(c) For any $1 \geq j \geq p$,

$$h^2_{(j)} \leq 1 - \rho_{(p+m+1-j)}(\Sigma). \tag{45}$$

These results can be proved by the matrix inequalities given by (20), (21), and (22).

Example 2. Suppose we have the following factor loading matrix, Λ, and the uniqueness variance matrix, Ψ:

$$\Lambda = \begin{bmatrix} .6 & .5 & .4 \\ .6 & .4 & .4 \\ .2 & .6 & .5 \\ .2 & .5 & .6 \\ .4 & .6 & .2 \\ .5 & .4 & .2 \end{bmatrix} \quad \text{and} \quad \Psi = \text{diag}\begin{pmatrix} .23 \\ .32 \\ .35 \\ .35 \\ .44 \\ .55 \end{pmatrix},$$

which yields

$$\Sigma = \Lambda\Lambda' + \Psi = \begin{bmatrix} 1 & .72 & .62 & .61 & .62 & .58 \\ .72 & 1 & .56 & .56 & .56 & .54 \\ .62 & .56 & 1 & .64 & .54 & .44 \\ .61 & .56 & .64 & 1 & .50 & .42 \\ .62 & .56 & .54 & .50 & 1 & .48 \\ .58 & .54 & .44 & .42 & .48 & 1 \end{bmatrix}.$$

With some calculations, the eigenvalues of Σ are found to be $\rho_1 = 3.810$, $\rho_2 = .648$, $\rho_3 = .490$, $\rho_4 = .432$, $\rho_5 = .355$, and $\rho_6 = .265$, from which we obtain the new lower bounds (NLB), and the upper bounds (UB) summarized in Table 1. For comparison we also give SMC's in the last column of the table. The NLB's for variables 1, 2, 3, and 4 improve upon SMC's used as lower bounds of communalities. It seems that there are generally more than one variable in which the NLB is better than the SMC.

Insert Table 1 about here

Example 3. Let Σ_k ($k = p-1$ and p) be a correlation matrix of order k with all the correlation coefficients being equal to a ($0 < a < 1$). The SMC's of all p variables are computed by

$$\begin{aligned} \mathrm{SMC}(j) &= 1 - \det(\Sigma_p)/\det(\Sigma_{p-1}) \\ &= 1 - \{(1+(p-1)a)(1-a)^p\}/\{1+(p-2)a)\}(1-a)^{p-1} \\ &= a^2/(a+(1-a)/(p-1)), \end{aligned} \tag{46}$$

and the eigenvalues of Σ_p are:

$$\rho_1(\Sigma_p) = 1 + (p-1)a, \;\text{ and }\; \rho_j(\Sigma_p) = 1 - a \;\text{ for }\; 2 \le j \le p.$$

Note that the last $p-1$ eigenvalues are equal to a. Furthermore, from (a) of Property 14, a gives a stronger lower bound of communality for each of the $p-1$ variables, since

$$a - a^2/(a+(1-a)/(p-1)) = a(1-a)/((1-a)+a(p-1)) > 0.$$

It is interesting to note that as p approaches infinity, SMC(j) computed by (46) approaches a which coincides with the communalities of the p variables. This is consistent with the suggestion first made by Roff (1936) and later proved by Guttman (1940).

3.3 Variable selection in factor analysis

It is recommended that some rotation methods be applied to the factor loading matrix derived by some initial factor extraction method to construct some psychological scales such as personality, vocational interest, and so on. In some cases, a number of items load highly on some factors, while smaller numbers of items load highly on other factors. In such cases, it is important to check whether a particular variable is a suitable indicator of a factor extracted. We present a stepwise variable selection method in factor analysis, following Yanai (1980).

Let $X = [x_1, \cdots, x_p]$ denote a standardized data matrix, and assume that the factor score matrix, $F = [f_1, \cdots, f_m]$, consists of m orthogonal factors. Let R^2_{X/f_j} denote the squared multiple correlation obtained by regressing f_j onto X. Then, from Lemma 1 we obtain

$$\begin{aligned} s &= R^2_{X/f_1} + \cdots + R^2_{X/f_m} \\ &= \sum_{j=1}^{m} (f_j' P_X f_j)/(f_j' f_j) = \mathrm{tr}(P_X \sum_{j=1}^{m} P_{f_j}) = \mathrm{tr}(P_X P_F), \end{aligned} \quad (47)$$

using the relationship, $P_F = P_{f_1} + \cdots + P_{f_m}$, which follows from Lemma 1 and the orthogonality of F. Note that s defined in (47) is the sum of the squared canonical correlation coefficients between F and X representing the relationship between the extracted factors and observed data. We propose a forward inclusion method for stepwise selection of variables in factor analysis by employing the following decomposition of $\mathrm{tr}(P_X P_F)$:

$$\begin{aligned} s = \mathrm{tr}(P_X P_F) &= \mathrm{tr}(P_{x_1} P_F) + \mathrm{tr}(P_{x_2|x_1} P_F) + \cdots + \mathrm{tr}(P_{x_p|X_{[p-1]}} P_F) \\ &= s_1 + s_2 + \cdots + s_p. \end{aligned} \quad (48)$$

Then, the proposed procedure of stepwise selection can be described as:

Step 1: Select a variable x_j with the largest communality h_j^2, since $\mathrm{tr}(P_{x_j} P_F) = ||P_F x_j||^2/||x_j||^2 = h_j^2$ follows from (27).

Step 2: Suppose that variable x_j is selected. Then, select variable x_k $(k \neq j)$ with the largest value of

$$s_k = \mathrm{tr}(P_{x_k|x_j} P_F) = \mathrm{tr}(P_{Q_{x_j} x_k} P_F) = (h_j^2 r_{jk}^2 + h_k^2 - 2r_{jk}\sqrt{h_j^2 h_k^2})/(1 - r_{jk}^2), \quad (49)$$

where r_{jk} is the correlation coefficient between x_j and x_k.

Step 3: In earlier $j-1$ steps, suppose, for simplicity, that $j-1$ variables $X_{j-1} = [x_1, \cdots, x_{j-1}]$

are selected. (This is just for notational convenience.) Then, select a variable x_k $(k \geq j)$ with the largest value of

$$s_k = \text{tr}(P_{x_k|X_{j-1}} P_F) = ||b_k||^2/(1 - R^2_{X_{j-1}/x_k}), \tag{50}$$

where $b_k = \lambda_k - \Lambda'_{j-1} R^-_{j-1,j-1} r_{(j-1)/k}$.

Example 4. We first performed principal factor analysis and extracted four common factors from the data with twelve scales in Yatabe-Guilford Personality Inventory. (This is the most popular personality inventory in Japan.) We show the result of stepwise selection of the variables in Table 2, in which scales are arranged in descending order of s_j for $j = 1, \cdots, p$. (The list of the twelve scales is given in Table 3.) We then computed the communalities for the twelve scales. It turned out that the Depression scale (Scale D for short) had the largest communality of .728 among the twelve scales. In the second step, Scale A with the s_k value (defined in (49)) of .698 was selected. Interestingly, Scale D had the highest factor loading on the first factor, while Scale A had the highest factor loading on the second factor. Continuing this way, we came to the final step where Scale I was selected with the s_k value (defined in (50)) of only .020. In reference to the values of $s_{(j)} = s_1 + \cdots + s_j$ given in the last column of Table 2, we may say that only five or six scales are sufficient for explaining the information contained in the four common factors. As an alternative method of stepwise selection in factor analysis, Kano and Harada (2000) developed SEFA (Stepwise variable selection in Exploratory Factor Analysis) by employing several goodness-of-fit measures used in structural equation modeling.

Insert Tables 2 and 3 about here

3.4 Representation of SMC when the correlation matrix may be singular

Let R denote a correlation matrix with three variables, x_1, x_2, and x_3, with correlation coefficients $r_{x_1x_2} = r_{x_1x_2} = a$ $(a \neq \pm 1)$ and $r_{x_2x_3} = 1$. We write R, and a g-inverse of R, denoted by $P = R^-$, as

$$R = \begin{bmatrix} 1 & a & a \\ a & 1 & 1 \\ a & 1 & 1 \end{bmatrix}, \text{ and } P = (1/(1-a^2)) \begin{bmatrix} 1 & -wa & -(1-w)a \\ -xa & t_1 & t_2 \\ -(1-x)a & t_3 & 1-t_1-t_2-t_3 \end{bmatrix},$$

where $-1 \leq a \leq 1$, and t_1, t_2, t_3, w, and x are arbitrary. Let $I_3 - RP = [g_1, g_2, g_3]$. Then with some computations, we obtain $g_1 = 0$, and $g_j \neq 0$ $(j = 2, 3)$. According to Theorem 1 of Khatri (1976), SMC(1) $= a^2$, and SMC(j) $= 1$ for $j = 2, 3$. Thus, SMC's can be computed for all variables, even if R is singular.

Furthermore, let $W = [X, Y]$ be a matrix of order $n \times (p+q)$, where $X = [x_1, \cdots, x_p]$ and $Y = [y_1, \cdots, y_q]$ are matrices of orders $n \times p$ and $n \times q$, respectively. Let $R = R_{WW}$ be the correlation matrix, and let P $= R^-$ denote a g-inverse of R. Let R and P be partitioned analogously:

$$R = \begin{bmatrix} R_{XX} & R_{XY} \\ R_{YX} & R_{YY} \end{bmatrix}, \text{ and } P = \begin{bmatrix} P^{XX} & P^{XY} \\ P^{YX} & P^{YY} \end{bmatrix}. \tag{51}$$

Then, with some computations, we obtain

$$R_{XX.Y} P^{XX} R_{XX.Y} = R_{XX.Y}, \text{ where } R_{XX.Y} = R_{XX} - R_{XY} R_{YY}^- R_{YX}. \tag{52}$$

Let

$$\text{SMC}(X|Y) = \begin{bmatrix} R^2_{x_1|Y} & R_{x_1x_2|Y} & \cdots & R_{x_1x_p|Y} \\ R_{x_2x_1|Y} & R^2_{x_2|Y} & \cdots & R_{x_2x_p|Y} \\ \vdots & \vdots & \ddots & \vdots \\ R_{x_px_1|Y} & R_{x_px_2|Y} & \cdots & R^2_{x_p|Y} \end{bmatrix}$$

a square matrix of order p, where

$$R^2_{x_j|Y} = ||P_Y x_j||^2, \text{ and } R_{x_ix_j|Y} = x_i' P_Y x_j.$$

Then, we have the following lemma.

Lemma 3. Let R and P be matrices defined in (51). Further, let

$$H = RP = \begin{bmatrix} H_{XX} & H_{XY} \\ H_{YX} & H_{YY} \end{bmatrix}, \text{ and } B = \begin{bmatrix} H_{XX} - I_p \\ H_{XY} \end{bmatrix}$$

be matrices of orders $p+q$ and $(p+q)\times p$, respectively. If $B=O$, then

$$\mathrm{SMC}(X|Y) = R_{XY}R_{YY}^{-}R_{YX} = R_{XX} - (P^{XX})^{-1}, \tag{53}$$

and if $\mathrm{rank}(B)=p$, then $\mathrm{SMC}(X|Y)=R_{XX}$.

Proof. It follows from Lemma 4 of Khatri (1976) that $\mathrm{rank}(B) = p - \mathrm{rank}(X'Q_YX)$. If $B = O$, then $\mathrm{rank}(X'Q_YX) = p$, indicating $R_{XY}R_{YY}^{-}R_{YX}$ is nonsingular. Then, $R_{XX.Y}P^{XX} = I_p$ follows from (52), establishing (53). If $\mathrm{rank}(B) = p$, then $X'Q_YX = O$, which implies $\mathrm{SMC}(X|Y) = X'P_YX = X'X$. □

The term $\mathrm{SMC}(X|Y)$ defined in Lemma 3 coincides with the right hand side of (39), which we call generalized SMC, since it reduces to the communality of a variable when X consists of a single vector. The above result represents an extension of Khatri (1976).

3.5 Interpretation of communalities from a regularization perspective

A major difference between principal factor analysis (PFA) and principal component analysis (PCA) is that the former obtains the eigen-decomposition of $R-\Psi$ (assuming that Ψ is tentatively known), whereas the latter obtains that of R. The analysis of $R-\Psi$ may be justified from a regularization perspective. In the ridge type of regularization (Hoerl & Kennard, 1970) estimates of regression coefficients in linear regression models are obtained by

$$\tilde{b} = (X'X + \kappa I_p)^{-1}X'y, \tag{54}$$

where X is an $n\times p$ matrix of predictor variables, y is an n-component vector of criterion variable, and κ is a ridge parameter, which typically takes a small positive value. The prediction vector is obtained by

$$X\tilde{b} = P(\kappa)y, \tag{55}$$

where $P(\kappa) = X(X'X+\kappa I_p)^{-1}X'$ is called ridge operator. The ridge estimates of regression coefficients are usually biased, but are associated with smaller MSE (mean square errors; Hoerl & Kennard, 1970).

Takane and Yanai (2005) recently introduced the following metric,

$$M(\kappa) = I_n + \kappa(XX')^+, \tag{56}$$

in an effort to generalize the ridge type of regularization to other techniques of multivariate analysis. Using $M(\kappa)$, they could rewrite $P(\kappa)$ as $P(\kappa) = X(X'M(\kappa)X)^- X'$, where $X'M(\kappa)X = X'X + \kappa P_{X'}$, and $P_{X'}$ is the orthogonal projector onto $\mathrm{Sp}(X')$, which reduces to I_p when X is columnwise nonsingular. Note that $P(\kappa)$ is invariant over the choice of g-inverse of $X'M(\kappa)X$, and that $(X'X + \kappa I_p)^{-1} \in \{(X'M(\kappa)X)^-\}$. Takane and Yanai (2005) further extended the metric matrix to:

$$M^{(L)}(\kappa) = I_n + \kappa(XL^- X')^+, \tag{57}$$

where L is PSD with $\mathrm{Sp}(L) = \mathrm{Sp}(X')$. With this generalized metric matrix, we obtain $X'M^{(L)}(\kappa)X = X'X + \kappa L$.

The ridge estimation generally has the effect of shrinking the estimates toward zero by adding $\kappa P_{X'}$ or κL to $X'X$ on the predictor side. Presumably, a similar shrinkage effect can be obtained by subtracting the same from the criterion side. Let

$$M^{(\Psi)}(-1) = I_n - (X\Psi^{-1}X')^+. \tag{58}$$

Then,

$$(1/n)X'M^{(\Psi)}(-1)X = R - \Psi, \tag{59}$$

which, as noted earlier, is the matrix whose eigen-decomposition is taken in PFA. (59) may be seen from $(X\Psi^{-1}X')^+ = X\Psi^{-1/2}((\Psi^{-1/2}X'X\Psi^{-1/2})^+)^2\Psi^{-1/2}X' = X(X'X\Psi^{-1}X'X)^+X'$. This indicates that in PFA we are sort of obtaining shrinkage estimates (of factor loadings) relative to PCA loadings. This leads to the idea that the estimate of Ψ is chosen in such a way that it reproduces an R that cross-validates best.

3.6 Methods of estimating factor scores

It is useful to estimate factor scores of individual subjects. A number of methods of estimating factor scores have been proposed so far.

The first estimator starts from the parametric model of factor analysis, $x = \Lambda f + e$, where $\mathrm{E}(e) = 0$, and $\mathrm{V}(e) = \Psi$ is a nonsingular diagonal matrix of uniqueness variances. It is assumed that the factor loading matrix, Λ, and the uniqueness variance matrix, Ψ, are known, and only e is a vector of random variables analogous to the disturbance terms in linear regression models. The generalized least squares estimate of f, which we denote by f_1, minimizing

$$(x - \Lambda f)'\Psi^{-1}(x - \Lambda f) \tag{60}$$

is given by

$$f_1 = (\Lambda'\Psi^{-1}\Lambda)^{-1}\Lambda'\Psi^{-1}x. \tag{61}$$

It can be easily verified that this estimator is unbiased and its covariance matrix is given by $\mathrm{V}(f_1) = (\Lambda'\Psi^{-1}\Lambda)^{-1}$. This is called Bartlett estimator (Bartlett, 1937).

Note 3. If we neglect Ψ, the minimization of $(x - \Lambda f)'(x - \Lambda f)$ with respect to f yields

$$f_2 = (\Lambda'\Lambda)^{-1}\Lambda' x, \tag{62}$$

which was first derived by Horst (1965). Note that f_2 defined above is also unbiased.

We now consider an estimation of f when Ψ is possibly singular.

Lemma 4 (Rao & Yanai, 1979). Under the Gauss-Markov model, $(y, X\beta, \alpha^2 G)$ where G may be singular, the BLUE (the best linear unbiased estimator) of $X\beta$ can be expressed as

$$Xb = Py \tag{63}$$

where P satisfies both (i) $PX = X$, and (ii) $PGZ = O$, where $Z = Q_X$ is the orthogonal projector onto $\mathrm{Sp}(X)^{\perp}$. If $\mathrm{Sp}(X)$ and $\mathrm{Sp}(GZ)$ cover the entire space of E^n, P is the projector onto $\mathrm{Sp}(X)$ along $\mathrm{Sp}(GZ)$, and it can be expressed in the following three forms:

1) $X(X'Q_{GZ}X)^{-}X'Q_{GZ}$. 2) $I_n - GZ(ZGZ)^{-}Z$.

3) $X(X'T^{-1}X)^{-}X'T^{-1}$, where $T = XUX' + G$ and $\mathrm{rank}(T) = \mathrm{rank}(X, G)$.

We attempt to minimize (60) when Ψ is singular. To deal with this problem, Bentler and Yuan (1997) minimized $(x - \Lambda f)'\Psi^{+}(x - \Lambda f)$. Our solution, on the other hand, is based on Lemma 4.

Lemma 5. If $E^n = \mathrm{Sp}(\Lambda) + \mathrm{Sp}(\Psi)$, and $W = Q_\Lambda$ is the orthogonal projector onto $\mathrm{Sp}(\Lambda)^\perp$, the BLUE of f can be expressed in the following three equivalent forms:

1) $\Lambda(\Lambda' Q_{\Psi W}\Lambda)^- \Lambda' Q_{\Psi W} x$. 2) $(I_n - \Psi W(W\Psi W)^- W)x$.

3) $\Lambda(\Lambda' T^{-1}\Lambda)^- \Lambda' T^{-1} x$, where $T = \Lambda U \Lambda' + \Psi$ and $\mathrm{rank}(T) = \mathrm{rank}(\Lambda, \Psi)$.

Note that in the parametric model of factor analysis, a factor score vector and a raw data vector can be defined for each of n individual subjects. Let $f_{(j)}$ and $x_{(j)}$ denote these vectors for the j-th subject. These vector may be represented collectively by matrices $F' = [f_{(1)}, \cdots, f_{(n)}]$ and $X' = [x_{(1)}, \cdots, x_{(n)}]$. Anderson and Rubin (1956) obtained an estimate of F which minimizes

$$(1/n)\sum_{j=1}^{n}(x_{(j)} - \Lambda f_{(j)})'\Psi^{-1}(x_{(j)} - \Lambda f_{(j)}) \tag{64}$$

subject to $(1/n)F'F = (1/n)\sum_{j=1}^n f_{(j)}f'_{(j)} = \Phi$, where Φ is the matrix of correlations among m factors and thus is positive-definite (PD). Observe that $(1/n)\sum_{j=1}^n \mathrm{tr}(f'_{(j)}\Lambda'\Psi^{-1}\Lambda f_{(j)})$ $= \mathrm{tr}(\Lambda'\Psi^{-1}\Lambda((1/n)\sum_{j=1}^n f_{(j)}f'_{(j)})) = \mathrm{tr}(\Lambda'\Psi^{-1}\Lambda\Phi)$. Thus, the minimization of (64) is equivalent to maximizing $\sum_{j=1}^n f'_{(j)}\Lambda'\Psi^{-1}x_{(j)} = \mathrm{tr}(F\Lambda'\Psi^{-1}X') = \mathrm{tr}(F\Phi^{-1/2}(X\Psi^{-1}\Lambda\Phi^{1/2})')$ subject to $(1/n)F'F = \Phi$. Note that $(1/n)F'F = \Phi$ is equivalent to $\Phi^{-1/2}(1/n)F'F\Phi^{-1/2} = I_m$. We obtain from Property 11 that

$$F = X\Psi^{-1}\Lambda\Phi^{1/2}(\Phi^{1/2}\Lambda'\Psi^{-1}X'X\Psi^{-1}\Lambda\Phi^{1/2})^{-1/2}\Phi^{1/2}, \tag{65}$$

which yields

$$f_{(j)} = \Phi^{1/2}(\Phi^{1/2}\Lambda'\Psi^{-1}X'X\Psi^{-1}\Lambda\Phi^{1/2})^{-1/2}\Phi^{1/2}\Lambda'\Psi^{-1}x_{(j)} \ (j = 1, \cdots, n). \tag{66}$$

Note that

$$\Phi^{1/2}\Lambda'\Psi^{-1}X'X\Psi^{-1}\Lambda\Phi^{1/2} =$$
$$\Phi^{1/2}\Lambda'\Psi^{-1}(\Lambda\Phi\Lambda' + \Psi)\Psi^{-1}\Lambda\Phi^{1/2} = (\Phi^{1/2}\Lambda'\Psi^{-1}\Lambda\Phi^{1/2})^2 + \Phi^{1/2}\Lambda'\Psi^{-1}\Lambda\Phi^{1/2}.$$

By denoting $L = \Phi^{1/2}\Lambda'\Psi^{-1}\Lambda\Phi^{1/2}$, we may rewrite (66) as

$$f_{(j)} = \Phi^{1/2}(L^2 + L)^{-1/2}\Psi^{1/2}\Lambda'\Psi^{-1}x_{(j)}. \tag{67}$$

We denote (67) by f_3 for any j. Obviously, $(1/n)F'F = \Phi$ holds. This estimator was further discussed by Rao (1979) and ten Berge (1999).

Next, let us consider a random effect model of the form, $x = \Lambda f + e$, where Λ is a factor loading matrix of order $p \times m$, x and e are p dimensional random vectors, the latter satisfying $\mathrm{E}(fe') = O$. Let P denote a square matrix of order p and define Px as an estimate of Λf where f is assumed to be random. Differentiating

$$\begin{aligned} g(P) &= \mathrm{tr}(\mathrm{E}(Px - \Lambda f)(Px - \Lambda f)') \\ &= \mathrm{tr}(\mathrm{E}(Pxx'P' - Pxf'\Lambda' - \Lambda fx'P' + \Lambda ff'\Lambda')) \\ &= \mathrm{tr}(P\Sigma P' - P\Lambda\Phi\Lambda' - \Lambda\Phi\Lambda'P' + \Lambda\Phi\Lambda') \end{aligned} \tag{68}$$

with respect to P, we obtain

$$P\Sigma = \Lambda\Phi\Lambda'. \tag{69}$$

If Σ is nonsingular, we have

$$\Lambda f_4 = Px = (\Lambda\Phi\Lambda'\Sigma^{-1})x = \Lambda\Phi\Lambda'(\Lambda\Phi\Lambda' + \Psi)^{-1}x,$$

yielding

$$f_4 = \Phi\Lambda'\Sigma^{-1}x = \Phi\Lambda'(\Lambda\Phi\Lambda' + \Psi)^{-1}x, \tag{70}$$

which coincides with the regression estimator of f on x first introduced by Thurstone (1935) and further discussed by Thomson (1946).

If Σ is singular, we obtain from (69) that

$$P = \Lambda\Phi\Lambda'\Sigma^{-} + Z(I - \Sigma\Sigma^{-}), \tag{71}$$

where Z is arbitrary. Let $S = \Sigma\Sigma^{-}x - x$. Then, $\mathrm{E}(SS') = O$, which implies $\Sigma\Sigma^{-}x = x$. Postmultiplying (71) by x, we obtain $\Lambda f_4 = Px = \Lambda\Phi\Lambda'\Sigma^{-}x$, yielding

$$f_4 = \Phi\Lambda'\Sigma^{-}x. \tag{72}$$

The relationships among the four methods of estimating factor scores were discussed in McDonald and Burr (1967).

Now we consider the relationship between f_1 and f_4. With some derivations, it follows that

$$\begin{aligned} f_4 &= \Phi\Lambda'\Sigma^{-1}x = \Phi\Lambda'(\Psi + \Lambda'\Phi\Lambda)^{-1}x \\ &= \Phi\Lambda'(\Psi^{-1} - \Psi^{-1}\Lambda(\Lambda'\Psi^{-1}\Lambda + \Phi^{-1})^{-1}\Lambda'\Psi^{-1})x \\ &= \Phi(I_m - \Lambda'\Psi^{-1}\Lambda(\Lambda\Psi^{-1}\Lambda + \Phi^{-1})^{-1})\Lambda'\Psi^{-1}x \\ &= (\Lambda'\Psi^{-1}\Lambda + \Phi^{-1})^{-1}\Lambda'\Psi^{-1}x \qquad (73) \\ &= (I + \Phi^{-1}(\Lambda'\Psi^{-1}\Lambda)^{-1})^{-1}(\Lambda'\Psi^{-1}\Lambda)^{-1}\Lambda'\Psi^{-1}x \\ &= (I + \Phi^{-1}(\Lambda'\Psi^{-1}\Lambda)^{-1})^{-1}f_1. \end{aligned}$$

Assume that $\Phi = I_m$. Then, it follows from Anderson (2003, Section 14.7) that the mean square errors of f_4 given by

$$\mathrm{E}[(f_4 - f)(f_4 - f)'] = (I_m + \Lambda'\Psi^{-1}\Lambda)^{-1}$$

are smaller than the variances of the unbiased estimator, f_1, given by $\mathrm{V}(f_1) = (\Lambda'\Psi^{-1}\Lambda)^{-1}$. The above result indicates that f_4 is a linear combination of f_1.

3.7 Application of Property 3 to canonical factor analysis

Let F denote a matrix of common factor scores, and let X denote a standardized data matrix of p variables. Further, let $f = Xw$ denote a linear composite score vector. Then, maximizing

$$||P_F f||^2/||f||^2 \qquad (74)$$

with respect to w yields

$$(X'P_FX)w = \lambda(X'X)w, \qquad (75)$$

leading to

$$Rw = \overline{\lambda}\Psi w, \quad \text{where} \quad \overline{\lambda} = 1/(1+\lambda)$$

in view of $\Lambda\Lambda' = R - \Psi$. (75) is the eigen-equation resulting from canonical factor analysis introduced by Rao (1955). Note that the sum of eigenvalues obtained from (75) coincides with $\text{tr}(P_X P_F)$, as defined by (47).

Using Property 3, we can establish:

Theorem 2. Let Λ denote a factor loading matrix of order $p \times m$. Then, the following seven statements are all equivalent:

1) $\lambda_j = 1$ or 0 for $j = 1, \cdots, m$.

2) $P_X P_F = P_F P_X$.

3) $\Lambda\Lambda'(X'X)^- \Lambda = \Lambda$.

4) $\Psi(X'X)^- \Lambda = O$.

5) $((X'X)^- \Lambda\Lambda')^2 = (X'X)^- \Lambda\Lambda'$.

6) $\text{rank}(X) = \text{rank}(\Lambda) + \text{rank}(\Psi)$.

7) $(X - F\Lambda')' Q_X F = O$.

Proof. Equivalence between 1) and 2) is well known. To show 2) implies 3), we note $\Lambda = (1/n)F'X$. To show 3) implies 4), we note $\Lambda\Lambda'(X'X)^-\Lambda = \Lambda$, which implies $(R - \Psi)R^-\Lambda = RR^-\Lambda - \Psi R^-\Lambda = \Lambda$. This establishes the desired result, since $RR^-\Lambda = (1/n)(X'X)(X'X)^-X'F = (1/n)X'F = \Lambda$. To show 4) implies 3), we have $(X'X - \Lambda\Lambda')$ $(X'X)^-X'F = \Lambda - \Lambda\Lambda'(X'X)^-\Lambda = O$. To show 7) implies 3), $\Lambda F' Q_X F == \Lambda F'(I_n - X(X'X)^-X')F = \Lambda - \Lambda\Lambda'(X'X)^-\Lambda = O$. To show 4) implies 7), observe that $\text{Sp}(\Lambda\Lambda' + \Psi) = \text{Sp}(\Lambda, \Psi^{1/2})$. Further, suppose that $\Lambda\Lambda'\alpha + \Psi\beta = 0$. By premultiplying both sides by $\Lambda'(X'X)^-$, we obtain $\Lambda\Lambda'\alpha = 0$. Thus, $\text{Sp}(\Lambda)$ and $\text{Sp}(\Psi)$ are disjoint, establishing $\text{rank}(X'X) = \text{rank}(X) = \text{rank}(\Lambda) + \text{rank}(\Psi)$.

3.8 Some extension of the identifiability condition

It is well known that a sufficient condition for the matrix of uniqueness variances to be uniquely determined is that there exists at least two disjoint square matrices both nonsingular and of rank m in the factor loading matrix Λ when any one row vector is deleted from Λ (Anderson & Rubin, 1956). Ihara and Kano (1986), and Kano (1989) gave some extensions to Anderson and Rubin's result. We give an alternative extension using Property

9.

Lemma 6. Suppose that the factor loading matrix, Λ, of order $(p_1 + p_2 + r) \times m$, where $p_1 \geq m$, $p_2 \geq m$, and $r \leq \min(p_1, p_2)$, is partitioned as $\Lambda' = [\Lambda_1', \Lambda_2', \Lambda_3']$, where Λ_1, Λ_2, and Λ_3 are of orders $p_1 \times m$, $p_2 \times m$, and $r \times m$, respectively. The population correlation (covariance) matrix, Σ, is then expressed in a partitioned form as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix} = \begin{bmatrix} \Lambda_1\Lambda_1' + \Psi_1 & \Lambda_1\Lambda_2' & \Lambda_1\Lambda_3' \\ \Lambda_2\Lambda_1' & \Lambda_2\Lambda_2' + \Psi_2 & \Lambda_2\Lambda_3' \\ \Lambda_3\Lambda_1' & \Lambda_3\Lambda_2' & \Lambda_3\Lambda_3' + \Psi_3 \end{bmatrix} = \Lambda\Lambda' + \Psi.$$

Assume further that $\text{rank}(\Lambda_1\Lambda_2') = \text{rank}(\Lambda_1) = \text{rank}(\Lambda_2)$, and $\text{Sp}(\Lambda_3') \subset \text{Sp}(\Lambda_2')$. Then, Ψ is determined uniquely.

Proof. $\text{Sp}(\Lambda_3') \subset \text{Sp}(\Lambda_2')$ implies $\Lambda_3' = \Lambda_2' W$ for some W. $\text{rank}(\Lambda_1\Lambda_2') = \text{rank}(\Lambda_1) = \text{rank}(\Lambda_2)$ implies $\Lambda_2'(\Lambda_1\Lambda_2')^-\Lambda_1$ is the projector onto $\text{Sp}(\Lambda_2')$ along $\text{Ker}(\Lambda_1)$ (Property 9), and it is thus invariant over any choice of g-inverse of $\Lambda_1\Lambda_2'$. We then have

$$\Sigma_{32}\Sigma_{12}^-\Sigma_{13} = \Lambda_3\Lambda_2'(\Lambda_1\Lambda_2')^-\Lambda_1\Lambda_3' = \Lambda_3\Lambda_2'(\Lambda_1\Lambda_2')^-\Lambda_1\Lambda_2'W = \Lambda_3\Lambda_3' = \Sigma_{33} - \Psi_3, \qquad (76)$$

establishing $\Psi_3 = \Sigma_{33} - \Sigma_{32}\Sigma_{12}^-\Sigma_{13}$, which is invariant over any choice of g-inverse of Σ_{12}. □

This establishes the desired result. Observe that Lemma 6 covers the result of Anderson and Rubin (1956), where it was assumed that $\text{rank}(\Lambda_1\Lambda_2') = \text{rank}(\Lambda_1) = \text{rank}(\Lambda_2) = m$, which automatically implies $\text{Sp}(\Lambda_3') \subset \text{Sp}(\Lambda_2')$. Lemma 6 also covers Kano (1989), since

$$m = \text{rank}(P_{\Lambda_1'}P_{\Lambda_2'}) \leq \text{rank}(\Lambda_1\Lambda_2') \leq \text{rank}(\Lambda_j) = m$$

for $j = 1, 2$. Note that $\text{rank}(P_{\Lambda_1'}) = \text{rank}(P_{\Lambda_2'}) = m$, and that $\text{Sp}(\Lambda_3') \subset \text{Sp}(\Lambda_2')$ also holds.

Note 3. As a reviewer of this manuscript has pointed out (see also Takeuchi, Yanai and Mukherjee, 1982, section 7.2.2), there are in general an infinite number of possible Ψ's that satisfy the factor analysis model if Λ is allowed to change in such a way that $\text{Sp}(\Lambda_3') \subset \text{Sp}(\Lambda_2')$

no longer holds.

Acknowledgment. We are deeply indebted to an anonymous reviewer of this article who read the manuscript carefully and gave us some useful comments which lead to Note 3.

References

Anderson, T. W. (2003). *A introduction to multivariate statistical analysis*, (Third Ed.). New York: Wiley.

Anderson, T. W., and Gupta, D. (1963). Some inequalities on characteristic roots of matrices. *Biometrika*, **59**, 522-524.

Anderson, T. W., and Rubin, H. (1956). Statistical inferences in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 5 (pp. 111-150). Berkeley: University of California Press.

Baksalary, J. K. (1987). Algebraic characterizations and statistical implications of the commutativity of orthogonal projectors. In T. Pukkila, and S. Puntanen (Eds.), *Proceedings of the Second International Tampere Conference in Statistics* (pp. 113-142). University of Tampere, Finland.

Baksalary, J. K., and Styan, G. P. H. (1990). Around a formula for the rank of a matrix product with some statistical applications. In R. S. Rees (Ed.), *Graphs, matrices, and designs* (pp. 1-18). New York: Marcel Dekker.

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology*, **28**, 97-104.

Beckenbach, E. F., and Bellman, R. (1961). *Inequalities.* New York: Springer.

Bentler, P. M., and Yuan, K. H. (1997). Optimal conditionally equivalent factor score estimation, In M. Berkane (Ed.), *Latent variable modeling and application to causality* (pp. 259-281). New York: Springer.

Guttman, L. (1940). Multiple rectilinear prediction and the resolution into components. *Psychometrika*, **5**, 75-99.

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, **18**, 277-296.

Harman, H. H. (1967). *Modern factor analysis* (revised 1976). Chicago: University of Chicago Press.

Hoerl, A. F., and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthgonal problems. *Technometrics*, **12**, 55-67.

Horst P. (1965). *Factor analysis of data matrices.* New York: Holt, Rinehart and Winston.

Ihara, M, and Kano, Y. (1986). A new estimator of the uniqueness in factor analysis. *Psychometrika*, **51**, 563-566.

Kahtri, C. G. (1976). A note on multiple and canonical correlation matrix of a singular covariance matrix. *Psychometrika*, **41**, 465-470.

Kaiser, H. F. (1976). Image and anti-image covariance matrices from a correlation matrix that may be singular. *Psychometrika*, **41**, 295-300.

Kano, Y. (1989). A new estimation procedure using g-inverse matrix in factor analysis. *Mathematica Japonica*, **34**, 43-52.

Kano Y., and Harada A. (2000). Stepwise variable selection in factor analysis. *Psychometrika*, **65**, 7-22.

Kristof, W. (1970). A theorem on the trace of certain matrix product and some application. *Journal of Mathematical Psychology*, **7**, 515-530.

McDonald, R P., and Burr, E. J. (1967). A comparison of four methods constructing factor scores. *Psychometrika*, **32**, 381-401.

Rao, C. R. (1955). Estimation and tests of significance in factor analysis. *Psychometrika*, **20**, 93-111.

Rao, C. R. (1979). Separation theorem for singular values of matrices and their applications in multivariate analysis. *Journal of Multivariate Analysis*, **9**, 362-377.

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

Roff, M. (1936). Some properties of the communality in multiple factor theory. *Psychometrika*, **1**, 1-6.

Takane, Y. (2004). Matrices with special reference to applications in psychometrics. *Linear Algebra and Its Applications*, **388C**, 341-361.

Takane, Y., and Yanai, H. (2005). On ridge operators. Submitted for publication.

Takeuchi, K., Yanai, H., and Mukherjee, B. N. (1982). *The foundation of multivariate analysis.* New Delhi: Wiley Eastern (New York: Halsted Press).

ten Berge, J. M. F. (1993). *Least squares optimization in multivariate analysis.* Leiden: DSWO Press.

ten Berge, J. M. F. (1999). Some new results on correlation-preserving factor score prediction methods. *Linear Algebra and Its Applications*, **289**, 311-318.

Thomson, G. H. (1946). *The factorial analysis of human ability.* (2nd ed.) New York: Houghton Mifflin.

Thurstone, L. L. (1935). *Vector of mind.* Chicago: The University of Chicago.

Yanai, H. (1980). A proposition of generalized method for forward selection of variables, *Behaviormetrika*, **7**, 95-107.

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

Yanai, H., and Ichikawa, M. (1990). New lower and upper bounds for communality in factor analysis, *Psychometrika*, **55**, 405-409.

Yanai, H., and Mukherjee, B. N. (1987). A generalized method of image analysis from an intercorrelation matrix which may be singular, *Pychometrika*, **52**, 555-564.

Yanai, H., and Puntanen, S. (1993). Partial canonical correlation associated with symmetric reflexive generalized inverses of the dispersion matrix. In K. Matsushita, M. L. Puri, and T. Hayakawa (Eds.), *Statistical sciences and data analysis* (pp. 253-264). Utrecht, The Netherlandws: VSP.

Table 1: Communalities, SMC's, and upper and lower bounds for communalities in the numerical example (Yanai & Ichikawa, 1990).

Variable	Communality	NLB	UB	SMC
1	.770	.735	···	.654
2	.680	.645	···	.573
3	.650	.568	···	.518
4	.650	.510	.735	.498
5	.560	···	.645	.453
6	.450	···	.568	.384

Table 2: Results of the stepwise selection method in principal factor analysis.

Step	Scale	Factor 1	Factor 2	Factor 3	Factor 4	Communality	s_j	$s_{(j)}$
1	D	**.684**	-.298	.254	-.326	.728	.728	.728
2	A	.183	**.821**	-.040	.098	.715	.698	1.426
3	Co	**.751**	.186	-.189	.199	.674	.451	1.877
4	C	.466	-.102	**.534**	-.018	.512	.265	2.412
5	R	.032	**.457**	.209	.438	.445	.190	2.333
6	G	-.082	**.677**	-.030	-.117	.479	.170	2.503
7	S	-.135	**.795**	.012	.030	.651	.099	2.602
8	N	**.808**	-.157	.045	.059	.684	.088	2.690
9	Ag	-.056	.404	**.407**	.026	.332	.057	2.747
10	O	**.837**	.073	-.020	.011	.707	.053	2.800
11	T	.157	.092	-.171	**.471**	.284	.034	2.834
12	I	.383	**-.535**	.185	.000	.461	.020	2.854

Table 3: List of twelve scales.

No.	Symbol	Scale
1)	D	Depression
2)	A	Ascendance
3)	Co	Lack of cooperativeness
4)	C	Cyclic tendency
5)	R	Rhathymia
6)	G	General activity
7)	S	Social extraversion
8)	N	Nervousness
9)	Ag	Lack of agreeableness
10)	O	Lack of objectivity
11)	T	Thinking extraversion
12)	I	Inferiority feelings

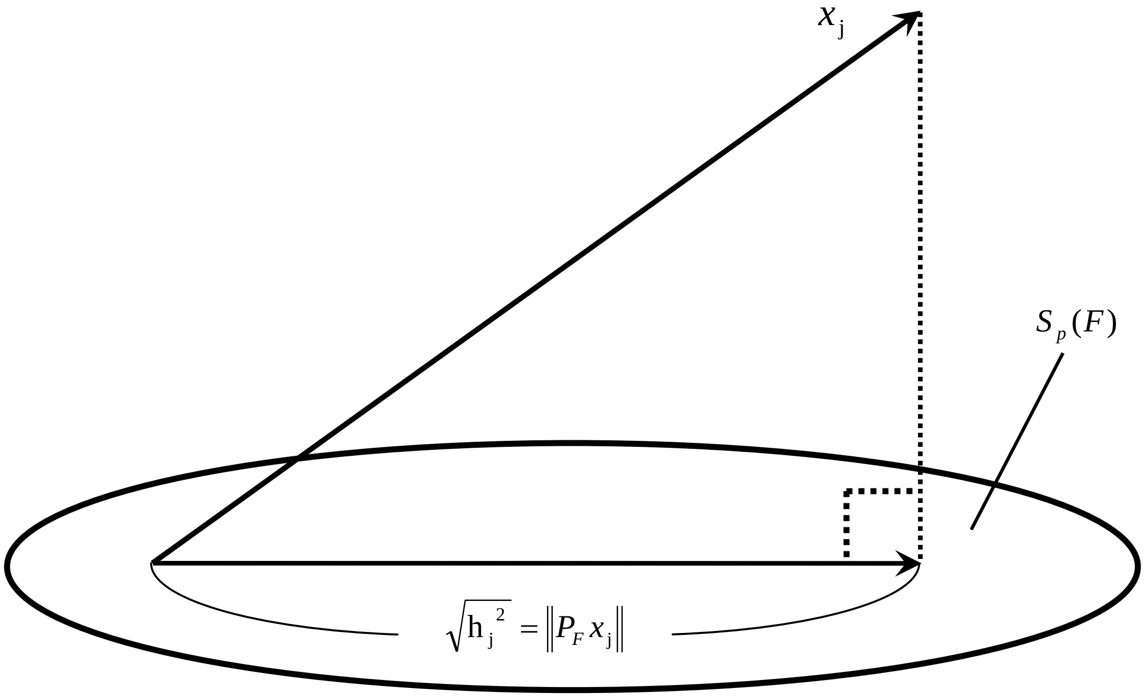

Fig. 1. Representation of communality h_j^2 of a vector x_j in terms of orthogonal projection assuming $\|x_j\| = 1$.