Linear Algebra and its Applications 434 (2011) 2539–2555

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

A new family of constrained principal component analysis (CPCA)

Yoshio Takane [a,*], Michael A. Hunter [b]

[a] Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, Quebec, Canada H3A 1B1
[b] Department of Psychology, University of Victoria, P.O. Box 3050 STN CSC, Victoria, British Columbia, Canada V8W 3P5

ARTICLE INFO

Article history:
Received 11 October 2010
Accepted 3 January 2011
Available online 1 February 2011

Submitted by R.A. Brualdi

AMS classification:
15A03
15A09

Keywords:
Orthogonal projectors
Singular value decomposition (SVD)
Generalized SVD
Ridge operators
Ridge metric matrix

ABSTRACT

Several decompositions of the orthogonal projector $P_X = X(X'X)^{-}X'$ are proposed with a prospect of their use in constrained principal component analysis (CPCA). In CPCA, the main data matrix X is first decomposed into several additive components by the row side and/or column side predictor variables G and H. The decomposed components are then subjected to singular value decomposition (SVD) to explore structures within the components. Unlike the previous proposal, the current proposal ensures that the decomposed parts are columnwise orthogonal and stay inside the column space of X. Mathematical properties of the decompositions and their data analytic implications are investigated. Extensions to regularized PCA are also envisaged, considering analogous decompositions of ridge operators.

1. Introduction

Multivariate data matrices analyzed by principal component analysis (PCA) are often accompanied by auxiliary information about the rows and columns of the data matrices. For example, the rows of a data matrix may represent subjects for whom some demographic information (e.g., gender, age, level of education, etc.) may be available. The columns, on the other hand, may represent stimuli defined by several attributes, and the values on the attributes characterizing the stimuli may be known. Constrained principal component analysis (CPCA; [1,2]) incorporates such information in PCA of the main data matrix.

* Corresponding author. Tel.: +1 514 398 6125; fax: +1 514 398 4896.
E-mail addresses: takane@psych.mcgill.ca (Y. Takane), mhunter@uvic.ca (M.A. Hunter).

doi:10.1016/j.laa.2011.01.002

In CPCA, a main data matrix is first decomposed into several components according to the external information (External Analysis). Columnwise and/or rowwise regression analyses are used for this purpose with the external information on the rows and/or columns of the data matrix as predictor variables. The decomposed components are then subjected to PCA to investigate structures within the components (Internal Analysis).

Let X denote an n by p data matrix, and let G and H denote, respectively, n by q and p by s matrices of row and column predictor variables. Consider the following three regression models:

$$X = GC_1 + E_1, \tag{1}$$

$$X = B_1H' + E_2, \tag{2}$$

and

$$X = GMH' + B_2H' + GC_2 + E_3, \tag{3}$$

where B's, C's, and M are matrices of regression coefficients, and E's are matrices of disturbance terms. Model (1) is for the case in which only G is available, Model (2) in which only H is available, and Model (3) in which both G and H are available. In (3) we require

$$G'B_2 = O, \tag{4}$$

and

$$C_2H = O \tag{5}$$

for model identification purposes. The ordinary least squares (OLS) estimation of the regression coefficients in these three models leads to the following decompositions of X:

$$X = P_GX + Q_GX, \tag{6}$$

$$X = XP_H + XQ_H, \tag{7}$$

and

$$X = P_GXP_H + Q_GXP_H + P_GXQ_H + Q_GXQ_H, \tag{8}$$

where $P_G = G(G'G)^-G'$ and $P_H = H(H'H)^-H'$ are, respectively, the orthogonal projectors onto $\mathrm{Sp}(G)$ (the range space of X) and $\mathrm{Sp}(H)$, and $Q_G = I_n - P_G$ and $Q_H = I_p - P_H$ are their orthogonal complements.

Once the data matrix X is decomposed according to the external information (External Analysis), the decomposed parts are subjected to PCA (Internal Analysis). Computationally, this amounts to singular value decomposition (SVD) of the terms in the decompositions. (A computational short cut for the SVD will be described in Proposition 6.) Internal Analysis allows us to investigate structures specific to a particular part in the decompositions. For example, $\mathrm{SVD}(P_GX)$ (the SVD of P_GX) specifically analyzes the portion of X that can be accounted for by G, $\mathrm{SVD}(Q_GX)$ the portion of X left unaccounted for by G, $\mathrm{SVD}(P_GXP_H)$ the portion of X that can be explained by both G and H, etc. In this way, CPCA can highlight certain aspects of the data matrix more clearly compared to the conventional PCA. CPCA has been widely used (e.g., [3–5]), and software for CPCA is available from at least two sources [6,7].

While there is nothing wrong with the basic idea of the conventional CPCA presented above, the decomposition given in (6) takes the data matrix X out of $\mathrm{Sp}(X)$ by projecting it onto $\mathrm{Sp}(G)$. The decomposition in (7) is not columnwise orthogonal. These facts are somewhat at odd with the spirit of PCA, which analyzes the variation in X into columnwise orthogonal components. In this paper, we propose columnwise orthogonal decompositions of X that also stay inside the range space of X. The proposed decompositions provide alternative ways of decomposing the data matrix in Eternal Analysis of CPCA.

This paper is organized as follows: In Section 2, we introduce basic results on duality useful in subsequent sections. In Section 3, we present our main results, basic decompositions of the data matrix that satisfy the above requirements (columnwise orthogonal decompositions within $\text{Sp}(X)$) along with their data analytic implications. In Section 4, we give a numerical example to illustrate these decompositions. In Section 5, we decompose the data matrix into finer components by combining some of the decompositions given in Section 3. In Section 6, we develop decompositions of the ridge operator analogous to those of the orthogonal projector. These decompositions can be used to extend the ordinary (nonregularized) CPCA to regularized CPCA. In the Appendix, we discuss relationships between our new proposal and the Wedderburn–Guttman decomposition [8–10].

2. Preliminaries

In this section, we introduce a notion of dual basis, which plays an important role in subsequent sections. Let X denote a matrix as introduced in the previous section. Let

$$S = X'X, \text{ and } T = XX', \tag{9}$$

and let

$$S^+ = (X'X)^+, \text{ and } T^+ = (XX')^+ \tag{10}$$

denote their Moore–Penrose inverses. Define the matrix X^* of dual basis of X by

$$X^* = (X')^+ = XS^+ = T^+X = X(X'X)^+. \tag{11}$$

Then,

$$X = (X^{*\prime})^+ = X^*S = TX^* = X^*(X^{*\prime}X^*)^+, \tag{12}$$

indicating that X in turn constitutes a matrix of dual basis of X^*. We may rewrite S, T, S^+, and T^+ as follows using X^*:

$$S = (X^{*\prime}X^*)^+, \quad T = (X^*X^{*\prime})^+, \tag{13}$$

and

$$S^+ = X^{*\prime}X^*, \quad T^+ = X^*X^{*\prime}. \tag{14}$$

We also define the orthogonal projector onto $\text{Sp}(X) = \text{Sp}(X^*)$ (the range spaces of X and X^*, which are identical) by

$$P_X = XX^{*\prime} = X^*X' = P_{X^*} = TT^+ = T^+T, \tag{15}$$

and the orthogonal projector onto $\text{Sp}(X') = \text{Sp}(X^{*\prime})$ by

$$P_{X'} = X'X^* = X^{*\prime}X = P_{X^{*\prime}} = SS^+ = S^+S. \tag{16}$$

Observe that $P_{X'}$ reduces to the identity matrix of order p if X is columnwise nonsingular (i.e., $\text{rank}(X) = p$). Note also that

$$S^+ = X'T^{+2}X, \text{ and } T^+ = XS^{+2}X', \tag{17}$$

where $T^{+2} = (T^+)^2$, and $S^{+2} = (S^+)^2$. In some of the definitions above, e.g., in the definition of P_X, the Moore–Penrose inverse of S is not necessary (any g-inverse S^- of S suffices), but unless otherwise stated, we consistently use the Moore–Penrose inverse that applies to all situations, since delineating which g-inverses may be used in which situations is not the main purpose of this paper.

We note parenthetically that X^* is a matrix analog of Green's function for X in functional analysis, and S^+ and T^+ that turn X into X^* (see (11)) are matrix analogs of reproducing kernel for X in the Hilbert space. Similarly, X is a matrix analog of Green's function for X^*, and S and T that turn X^* into X (see (12)) are matrix analogs of reproducing kernel for X^*.

The above relationship between X and X^* has an interesting implication in regression analysis. Let Y denote a matrix of criterion variables, and let X denote a matrix of predictor variables, which are usually not orthogonal to each other. (If X is columnwise orthogonal, $X^* = X$.) Let $\hat{Y}$ denote the matrix of predictions derived through the OLS estimation of regression coefficients. Then,

$$\hat{Y} = X(X^{*\prime}Y) \tag{18}$$

$$= X^*(X'Y). \tag{19}$$

Eq. (18) indicates that

$$A = X^{*\prime}Y \tag{20}$$

is the matrix of regression coefficients applied to X to obtain $\hat{Y}$. At the same time, (20) indicates that A represents the matrix proportional to the covariance between X^* and Y. (In the rest of this paper, we may simply refer the matrix of the form (20) as the covariance between X^* and Y.) Eq. (19), on the other hand, indicates that

$$B = X'Y \tag{21}$$

is the matrix of regression coefficients applied to X^* to obtain $\hat{Y}$, while (21) indicates that B is also the matrix of covariances between X and Y. In general, the matrix of regression coefficients is called a weight matrix, while the matrix of covariances a structure matrix. In this terminology, A is the weight matrix for the predictor variables X, and B the structure matrix. This relation is reversed for the predictor variables X^*, that is, B represents the weight matrix, and A the structure matrix.

The duality between X and X^* also has a similar implication in PCA. Let $X = UDV' = X\tilde{U}DV'$ represent SVD(X), where $U = X\tilde{U}$, and $X^{*'}U = \tilde{U}(= VD^{-1})$. That is, $\tilde{U}$ represents the matrix of weights applied to X to derive U, the matrix proportional to component scores, and at the same time it represents the matrix of covariances between X^* and U. (In the rest parts of this paper, we may simply refer U as the matrix of component scores.) The transpose of $\tilde{U}$, i.e., $\tilde{U}'$, is the matrix of weights applied to U to derive X^*. The SVD of X^*, on the other hand, is expressed as $X^* = UD^{-1}V' = X^*\tilde{U}^*D^{-1}V'$, where $U = X^*\tilde{U}^*$, and $X'U = \tilde{U}^* = S^+\tilde{U} = VD$. This means that $\tilde{U}^*$ represents the matrix of weights applied to X^* to derive U, and also the matrix of covariances between X and U. The transpose of $\tilde{U}^*$, i.e., $\tilde{U}^{*'}$ is also the matrix of weights applied to U to derive X.

Something analogous happens in common factor analysis with oblique factors. A matrix of factor scores X is called primary, while the corresponding X^* (most often appropriately scaled) reference. However, which one is called primary or reference is essentially arbitrary. If one is called primary, the other becomes reference. See [11].

3. Basic decompositions

In this section, we present basic decompositions of orthogonal projectors P_X and $P_{X'}$ onto Sp(X) and Sp(X'), respectively, and derive analogous decompositions of the data matrix. Let X, X^*, G, and H be as introduced earlier. Then, the following proposition holds:

Proposition 1. *Let K and L be such that $H'K = O$ and $Sp(H) \oplus Sp(K) = Sp(X')$, and $G'XL = O$ and $Sp(X'G) \oplus Sp(L) = Sp(X')$. Then, the following decompositions of P_X hold:*

$$\text{(A)} \quad P_X = P_{XH} + P_{X^*K}, \tag{22}$$

$$\text{(B)} \quad P_X = P_{X^*H} + P_{XK}, \tag{23}$$

$$\text{(C)} \quad P_X = P_{TG} + P_{X^*L}, \tag{24}$$

and

$$\text{(D)} \quad P_X = P_{P_X G} + P_{XL}. \tag{25}$$

The two terms on the right-hand side of the above decompositions are mutually orthogonal.

Proof. It can readily be verified that XH and X^*K in (A) are mutually orthogonal ($H'X'X^*K = H'P_{X'}K = HK = O$), and that $\text{Sp}(XH) \oplus \text{Sp}(X^*K) = \text{Sp}(X)$, since $\text{Sp}(H) \oplus \text{Sp}(K) = \text{Sp}(X')$. The other decompositions in the proposition can be similarly proven, since they can be systematically derived from (A). Decomposition (B) is obtained by interchanging X and X^* (or by interchanging H and K) in (A). Decomposition (C) is obtained by replacing H by $X'G$ in (A). Decomposition (D) is obtained by interchanging X and X^* (or by interchanging H and K) and replacing H by $X'G$ in (A). □

Decomposition (A) is the most basic one [12,13], and is motivated as follows. Suppose that we have a regression model,

$$Y = XA + E \tag{26}$$

similar to (1), and that we estimate the regression coefficients A under the hypothesis that H_0: $K'A = O$. This hypothesis can equivalently be expressed as H_0: $A = HA_0$ for some A_0. Estimation under H_0 splits P_X into the sum of P_{XH}, the portion of P_X that can be explained by H_0, and P_{X^*K}, the portion of P_X that cannot be explained by H_0. (Motivations for the other decompositions will be given later.)

Since $P_X X = X$, we obtain the corresponding decompositions of the data matrix X by premultiplying X by (22)–(25).

Proposition 2. *Let K and L be as defined in Proposition 1. Then,*

$$(\text{A}') \quad X = P_{XH}X + P_{X^*K}X, \tag{27}$$

$$(\text{B}') \quad X = P_{X^*H}X + P_{XK}X, \tag{28}$$

$$(\text{C}') \quad X = P_{TG}X + P_{X^*L}X, \tag{29}$$

and

$$(\text{D}') \quad X = P_{P_X G}X + P_{XL}X. \tag{30}$$

The two terms on the right-hand sides of the above decompositions are columnwise orthogonal, and they are all within Sp(X).

Proof. The orthogonality of the two terms in each of the above decompositions is assured by the orthogonality of the projectors premultiplied to X. That all of them reside within the column space of X may be seen by observing that they can all be written in the form of XW for some W, and that $\text{Sp}(XW) \subset \text{Sp}(X)$. (Concrete forms of W will be given in Proposition 3.) □

Decomposition (A′) is obtained by projecting X onto $\text{Sp}(XH)$ and taking its residual $\text{Sp}(X^*K) = \text{Sp}(X - P_{XH}X)$. This decomposition was implicitly used by Guttman [8,9] for extracting factors in the group centroid method of factor analysis. The relationship between Guttman's original formulation and the current one will be elaborated in the Appendix. This decomposition (and the equivalent decomposition (a) to be given in Proposition 3) is similar to (2). It is not identical ($P_{XH}X \neq XP_H$), however, despite the fact that $P_{XH}X = P_{XP_H}X$ and $\text{Sp}(P_{XH}X) = \text{Sp}(XP_H)$.

Decomposition (B′) is obtained by applying Decomposition (B) to X. Decomposition (B) in turn is obtained by interchanging the roles of X and X^* in (A), but why do we want to do that? There are at least two possible reasons: the first one is related to the duality between X and X^*. If we write

$P_{XH}X = XJ$, and $P_{X^*H}X = X^*J^*$, J represents the weights applied to X to derive $P_{XH}X$ in (A′), and similarly J^* represents the weights applied to X^* to derive $P_{X^*H}X$ in (B′). At the same time, the former represents the covariances between X^* and $P_{XH}X$, since $X^{*\prime}P_{X^*H}X = X^{*\prime}XJ = J$, while the latter the covariances between X and $P_{X^*H}X$, since $X'P_{X^*H}X = X'X^*J^* = J^*$. This means that H used with X regulates the weights applied to X to derive $P_{XH}X$, while H used with X^* regulates the covariances between X and $P_{X^*H}X$. For example, suppose H is a vector of ones (i.e., $H = \mathbf{1}_p$). This H applied to X enforces the weights applied to X to be a constant, while that applied to X^* enforces the covariances between X and $P_{X^*H}X$ to be a constant. This point will be illustrated by a numerical example in the next section.

Second, consider (2) as a special case of the growth curve model (GCM: [14]), where G is missing (or set equal to I_n). The maximum likelihood estimate of B_1 under the distributional assumption of $\mathrm{vec}(E) \sim \mathcal{N}(\mathbf{0}, \Sigma \otimes I_n)$, where Σ may be singular, is given by

$$\hat{B}_1 = XS^+H(H'S^+H)^-, \tag{31}$$

so that

$$\hat{B}_1H' = X^*H(H'X^{*\prime}X^*H)^-H'X^{*\prime}X = P_{X^*H}X, \tag{32}$$

which is identical to the first term in Decomposition (B′). Note that $H'X^{*\prime}X = H'P_{X'} = H'$ in the above derivation.

In Decomposition (C′), we take $X'G$ as the weight matrix to be applied to X, rather than projecting X directly onto $\mathrm{Sp}(G)$ as in (1). This keeps the subspace of projection inside $\mathrm{Sp}(X)$. Rao [15, Section 11] tried to find a subspace in $\mathrm{Sp}(X)$ orthogonal to a given G, and obtained what amounts to the following decomposition of X:

$$X = XP_{X'G} + XQ_{X'G}. \tag{33}$$

This decomposition is similar to (C′), but not identical (i.e., $P_{XQ_{X'G}}X \neq XQ_{X'G}$, despite the fact that $P_{Q_{X'G}}X = P_{XL}$ and $\mathrm{Sp}(P_{XQ_{X'G}}X) = \mathrm{Sp}(XQ_{X'G})$). Furthermore, this decomposition is not columnwise orthogonal in the usual identity metric. The above decomposition may be regarded as being derived by setting $H = X'G$ in (2) instead of (27).

Note 1. Decomposition (33) may be rewritten as $X = P_{TG/T^+}X + P_{XL/T^+}X$, where the two terms on the right-hand side are term by term equal to those on the right-hand side of (33), and where $P_{XH/T^+} + P_{XK/T^+} = P_X$. The two terms on the left-hand are columnwise orthogonal only with respect to the nonidentity metric T^+. Similarly, (2) may be rewritten as $X = P_{XH/T^+}X + P_{XK/T^+}X$. Again, the two terms on the right-hand side are columnwise orthogonal only with respect to T^+. (See Proposition 3 for more details of this kind of projectors.)

In Decomposition (D′), we project G onto $\mathrm{Sp}(X)$ (rather than the other way round as in (1)) to obtain the subspace $\mathrm{Sp}(P_XG)$ in $\mathrm{Sp}(X)$, and then project X onto this subspace. This is equivalent to weighting X by $X^{*\prime}G$, the matrix of regression coefficients obtained by regressing G onto X, which is also proportional to the matrix of covariances between X^* and G (or to the group means if G is a group indicator matrix). When G represents a matrix of dummy variables indicating group memberships, this is similar to defining a discriminant subspace inside $\mathrm{Sp}(X)$. The CPCA of $P_{P_XG}X$ and canonical discriminant analysis (CDA) of X with G are not identical, however. Whereas the former obtains the SVD of $P_{P_XG}X$ under the identity row metric, CDA typically obtains the SVD under a nonidentity metric, namely, $\mathrm{GSVD}(P_{P_XG}X)_{I,S^+}$ (generalized SVD with the column metric S^+), where S^+ is as defined in (10). (The matrix S is proportional to the total covariance matrix in CDA.)

In the above proposition, projectors are always applied to X from the left. These effects may be transferred to the right-hand side of X by defining appropriate projectors that operate in the row space of X. (These are the W matrices mentioned in the proof of Proposition 2.) More specifically, we have the following decompositions of $P_{X'}$, the orthogonal projector onto $\text{Sp}(X')$, which, when applied to X from the right yield decompositions of X equivalent to those given in Proposition 2.

Proposition 3. *Let K and L be as defined in Proposition 1. Then, the following decompositions hold:*

$$(\text{a}) \quad P_{X'} = P'_{SH/S^+} + P'_{K/S^+}, \tag{34}$$

$$(\text{b}) \quad P_{X'} = P'_{H/S^+} + P'_{SK/S^+}, \tag{35}$$

$$(\text{c}) \quad P_{X'} = P'_{SX'G/S^+} + P'_{L/S^+}, \tag{36}$$

and

$$(\text{d}) \quad P_{X'} = P'_{X'G/S^+} + P'_{SL/S^+}, \tag{37}$$

where the matrices of the form $P_{A/M} = A(A'MA)^- A'M$, where M is a symmetric nonnegative definite matrix such that $\text{rank}(A) = \text{rank}(MA)$, are projectors onto $Sp(A)$ along $Ker(A'M)$, as A takes different matrix arguments. The two terms on the right-hand sides of the above decompositions are orthogonal with respect to the metric matrix S.

Proof. The above decompositions can be readily verified by a generalization of Khatri's [16] lemma. See Khatri [17, Theorem 1]. Also, see Yanai and Takane [13, Lemma 2.4(ii)]. The columnwise orthogonality of the two terms on the right-hand sides with respect to S can be directly verified. In (a), for example, $P_{SH/S^+}SP'_{K/S^+} = O$, since $H'SS^+SS^+K = H'P_{X'}K = HK = O$. The other decompositions in the proposition are similar. □

Note that in the above proposition we actually use projectors of the form $P'_{A/M}$, the transpose of $P_{A/M}$. We can always turn them into projectors without transpositions, e.g., $P'_{A/M} = P_{MA/M^+}$. However, as will be seen in Proposition 4, the transposed projectors are always applied to X from the right, implying that they operate on the rows of X. If the rows of X are brought into column vectors by transposing X, then the projector to be applied to them must be of the form $P_{A/M}$ (i.e., we obtain $P_{A/M}X'$).

As has been alluded to earlier, we obtain the corresponding data decompositions by postmultiplying X by the above decompositions.

Proposition 4. *Let K and L be as defined in Proposition 1. Then,*

$$(\text{a}') \quad X = XP'_{SH/S^+} + XP'_{K/S^+}, \tag{38}$$

$$(\text{b}') \quad X = XP'_{H/S^+} + XP'_{SK/S^+}, \tag{39}$$

$$(\text{c}') \quad X = XP'_{SX'G/S^+} + XP'_{L/S^+}, \tag{40}$$

and

$$(\text{d}') \quad X = XP'_{X'G/S^+} + XP'_{SL/S^+}. \tag{41}$$

The two terms on the right-hand sides of the above decompositions are columnwise orthogonal.

Proof. The decompositions themselves trivially follow from Proposition 3. The columnwise orthogonality of the two terms on the right-hand sides of the decompositions is assured by the orthogonality of the corresponding projectors in Proposition 3 with respect to $S = X'X$. □

The decompositions in Proposition 2 and the corresponding decompositions in Proposition 4 are identical. Specifically, we have:

Proposition 5.

$$P_{XH}X + P_{X^*K}X = XP'_{SH/S^+} + XP'_{K/S^+}, \tag{42}$$

$$P_{X^*H}X + P_{XK}X = XP'_{H/S^+} + XP'_{SK/S^+}, \tag{43}$$

$$P_{TG}X + P_{X^*L}X = XP'_{SX'G/S^+} + XP'_{L/S^+}, \tag{44}$$

and

$$P_{P_XG}X + P_{XL}X = XP'_{X'G/S^+} + XP'_{SL/S^+}. \tag{45}$$

The left-hand and right-hand sides of the above equations are term by term equal.

Proof. The term by term equality can be shown directly. In (42), for example, $P_{XH}X = XH(H'X'XH)^-H'X'X = XS^+SH(H'SS^+SH)^-H'S = XP'_{SH/S^+}$, and $P_{X^*K}X = X^*K(K'X^{*\prime}X^*K)^-K'X^{*\prime}X = XS^+(K'S^+K)^-K' = KP'_{K/S^+}$. The other decompositions are similar. As has already been pointed out, the fact that in all cases the effects of projectors applied to the left-hand side of X can be transferred to the right-hand side is an ultimate proof that all terms in the above decompositions stay inside $\mathrm{Sp}(X)$. □

Once the data matrix X is decomposed according to the external information (External Analysis), each term in the decompositions may be subject to SVD (Internal Analysis). The SVD of the terms of the form P_AX can be carried out economically by the following computational procedure.

Proposition 6. *Let F denote a matrix of any orthogonal basis vectors of A. Then, $P_A = FF'$. Let*

$$F'X = U^*DV' \tag{46}$$

denote the SVD of $F'X$. Then, the SVD of $P_AX = FF'X$ is obtained by

$$P_AX = UDV', \tag{47}$$

where

$$U = FU^*. \tag{48}$$

A proof the above proposition is rather rudimentary, and will not be given here. See Theorem 1 of Takane and Hunter (2001) for some detail.

Note 2. We may develop decompositions analogous to those in Proposition 1 for the projector $P_{X/W} = X(X'WX)^-X'W$, where W is a known symmetric nonnegative definite matrix such that $\mathrm{rank}(X) = \mathrm{rank}(WX)$. This projector is of the same form as those in Proposition 3, but it is applied to X from the left (as opposed to those in Proposition 3) and arises from the weighted LS estimation in regression analysis. We will not, however, elaborate on this any further in this paper.

4. A numerical example

In this section we illustrate the new family of CPCA up to hitherto presented with a numerical example. In the data set we use, thirty seven samples in three species of the *Lauraceae* family of wood, 1, *Ocotea bullata* (Obul); 2, *Ocotea kenyensis* (Oken); 3, *Ocotea Porosa* (Opor), are measured on six anatomical properties: (1) VesD (tangential vessel diameter in μm), (2) VesL (vessel element length in μm), (3) FibL (fibre length in μm), (4) RayH (Ray height in μm), (5) RayW (ray width in μm), and (6) NumVes (number of vessels per square mm). The first two species are indigenous to South Africa while the third is an imported wood used as a substitute for *O. bullata* in the manufacture of high quality furniture. The data are displayed in Table 1.

We employ a matrix of dummy variables indicating which tree species samples of wood belong to as G. We use another matrix of dummy variables indicating which groups of variables the six observed variables belong to as H. Specifically,

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

Table 1
Burden et al.'s [18] data.

Sample	Species	(1) VesD	(2) VesL	(3) FibL	(4) RayH	(5) RayW	(6) NumVes
1	1	78	346	961	223	24	31
2	1	129	406	1165	428	44	11
3	1	111	448	1096	379	40	13
4	1	82	361	1039	316	27	25
5	1	79	324	1048	369	29	26
6	1	103	371	1165	326	26	10
7	1	74	281	1175	324	26	11
8	1	104	387	1290	381	22	12
9	1	91	372	1234	375	26	11
10	1	85	418	1051	347	34	14
11	1	113	314	1253	466	23	10
12	1	94	437	1271	336	36	10
13	1	76	320	1130	347	29	13
14	1	119	359	1280	412	32	11
15	1	79	383	941	333	30	17
16	1	102	567	1221	395	40	11
17	1	114	569	1369	568	52	11
18	1	93	541	1267	347	34	14
19	1	141	621	1527	419	34	15
20	1	95	415	1225	416	38	10
21	2	156	401	1588	512	42	11
22	2	162	502	1591	369	42	8
23	2	147	402	1391	440	32	9
24	2	142	393	1468	443	35	6
25	2	125	322	1530	459	34	11
26	2	103	378	1655	441	34	11
27	2	126	414	1759	459	42	8
28	3	130	471	1072	409	39	15
29	3	139	133	993	342	33	14
30	3	115	352	1048	300	36	14
31	3	153	419	1077	392	48	20
32	3	112	309	1044	358	47	8
33	3	130	325	1166	428	36	12
34	3	130	368	1005	356	39	16
35	3	127	331	1027	473	38	20
36	3	135	370	1104	531	38	15
37	3	122	346	981	393	40	14

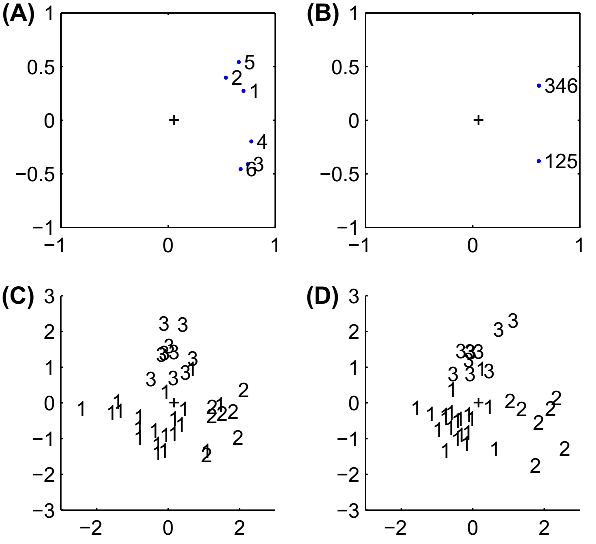

Fig. 1. (A) CPCA of $P_{XH}X$: component loadings; (B) CPCA of $P_{X^*H}X$: component loadings; (C) CPCA of $P_{XX'G}X$: component scores; (D) CPCA of $P_{P_GX}X$: component scores.

which indicates that variables 1, 2, and 5 belong to group 1, while variables 3, 4, and 6 belong to group 2. The sixth variable x_6 was reflected to make it positively correlated with most of the other variables. The data matrix X was then columnwise centered and normalized so that $\text{diag}(X'X) = I$.

The first analysis pertains to $\text{SVD}(P_{XH}X)$ (the SVD of the first term in Decomposition (A′)). Let $P_{XH}X = UDV' = X\tilde{U}DV'$, where $U = X\tilde{U}$, denote the SVD of $P_{XH}X$. Fig. 1(A) presents the plot of columns of DV', namely component loadings. In this case, two linear composites were formed, each with equal weights applied to all variables within groups, and principal components were derived from the two composite variables. The loadings indicate the correlations between the derived components and the variables in X. This can be seen by observing that $U'X = U'P_{XH}X = DV'$. Note that $\tilde{U}$ is the matrix of weights applied to X to derive U, but it also represents the matrix of covariances between X^* and U, since $X^{*\prime}U = \tilde{U}$.

The second analysis involves $\text{SVD}(P_{X^*H}X)$ (the SVD of the first term in Decomposition (B′)), which is denoted by $P_{X^*H}X = U^*D^*V^{*\prime} = X^*\tilde{U}^*D^*V^{*\prime}$, where $U^* = X^*\tilde{U}^*$, and $X'U^* = \tilde{U}^*$. Fig. 1(B) displays the plot of component loadings, i.e., columns of $D^*V^{*\prime}$. In this case, two composite variables were formed in such a way that they correlate equally with all variables within groups. The component loadings indicating correlations between the principal components U^* and the variables in X (i.e., $U^{*\prime}X = U^{*\prime}P_{X^*H}X = D^*V^{*\prime}$) coincide within groups, because U^* is so constructed. Note that $\tilde{U}^*$ represents the matrix of weights applied to X^* to derive U^*, while it also represents the covariance matrix between X and U^*.

The third analysis concerns $\text{SVD}(P_{XX'G}X)$ (the SVD of the first term in Decomposition (C′)). Since this decomposition involves G, Fig. 1(C) presents component scores rather than component loadings. In this decomposition, linear composites are formed by weighting variables in X in proportion to the covariances between X and G, and X is projected onto these composites. Three tree species are separated fairly well in the space of derived component scores. (To conform to the usual convention, U was scaled up by $n^{1/2}$ and plotted in Fig. 1(C and D).)

The fourth analysis pertains to Decomposition (D′), in which G is first regressed onto X, and the matrix of regression coefficients was used as weights to form linear composites of X, onto which X is projected. The derived component scores plotted in Fig. 1(D) again discriminate among the three tree species fairly well, perhaps even slightly better than in the previous analysis.

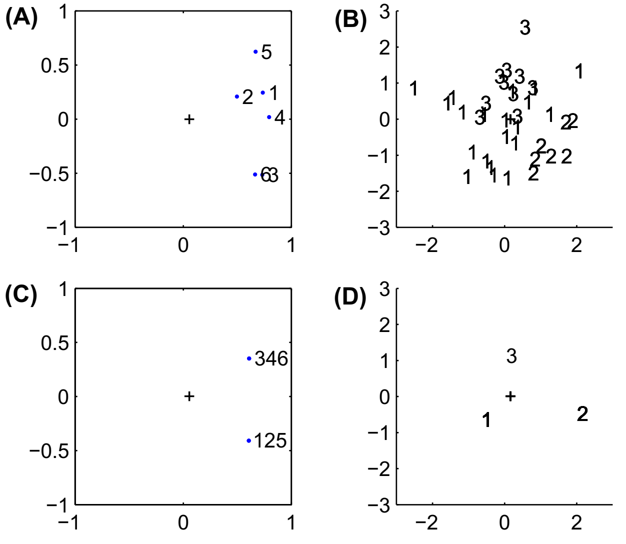

Fig. 2. (A) PCA of X: component loadings; (B) PCA of X: component scores; (C) CPCA of XP_H: component loadings; (D) CPCA of P_GX: component scores.

For comparisons, results of the conventional PCA of X are presented in the top panel of Fig. 2. Fig. 2(A) displays component loadings, which indicates that the groupings of variables are less clear than those obtained by CPCA, variable 4 located closer to group 1 (consisting of variables 1, 2, and 5) than to group 2 (consisting of variables 3, 4, and 6). Fig. 2(B), which displays component scores, also indicates that the three tree species are less well separated than in the corresponding CPCA.

Fig. 2(C) shows component loadings obtained by the conventional CPCA, namely SVD(XP_H), the SVD of the first term in Decomposition (2). This is strikingly similar to Fig. 1(B), where the loadings within the same variable groups take equal values. However, they are quite distinct conceptually. Whereas in Fig. 1(B) linear composites were formed in such a way that they correlate equally with variables within the same groups, thereby producing equal loadings within groups, in Fig. 2(C) the observed data X is totally replaced by the linear composites XP_H, and the loadings indicate the correlations between XP_H and the component scores U, which naturally take equal values within the same variable groups.

Fig. 2(D) displays component scores obtained by the CPCA of P_GX, the first term in Decomposition (1). The three integers in the figure indicate the map of the centroids of the three species groups in the space of observed variables to the space of component scores. These three points coincide with the centroids of the component scores of the wood samples in the three groups plotted in Fig. 1(D). This may be seen by observing that $P_GP_{P_XG}X = P_GX$.

5. Finer decompositions

By combining some of the two-term decompositions in Proposition 1, we can generate the following four-term decompositions of P_X. The symbol like (AC) in the following proposition indicates that it is a combination of Decompositions (A) and (C) in Proposition 1.

Proposition 7. *(i) Let B and C be such that $G'XHB = O$ and $Sp(H'X'G) \oplus Sp(B) = Sp(H'X')$, and $G'X^*KC = O$ and $Sp(K'X^{*\prime}G) \oplus Sp(C) = Sp(K'X^{*\prime})$. (Note that $Sp(B)$ is null if $rank(XH) = rank(G'XH)$, and $Sp(C)$ is null if $rank(X^*K) = rank(G'X^*K)$.) Then, the following two decompositions hold:*

$$(\mathrm{AC}) \quad P_X = P_{XHH'X'G} + P_{XH(H'SH)^-B} + P_{X^*KK'X^{*\prime}G} + P_{X^*K(K'S^+X^*K)^-C}, \tag{49}$$

and

$$(\mathrm{AD})\quad P_X = P_{P_{XH}G} + P_{XHB} + P_{P_{X^*K}G} + P_{X^*KC}. \tag{50}$$

(ii) Let B and C be such that $G'X^*HB = O$ *and* $Sp(H'X^{*'}G) \oplus Sp(B) = Sp(H'X^{*'})$, *and* $G'XKC = O$ *and* $Sp(K'X'G) \oplus Sp(C) = Sp(K'X')$. *(Note that* $Sp(B)$ *is null if* $rank(X^*H) = rank(G'X^*H)$, *and* $Sp(C)$ *is null if* $rank(XK) = rank(G'XK)$.*) Then, the following two decompositions hold:*

$$(\mathrm{BC})\quad P_X = P_{X^*HH'X^{*'}G} + P_{X^*H(H'S^+H)^-B} + P_{XKK'X'G} + P_{XK(K'SK)^-C}, \tag{51}$$

and

$$(\mathrm{BD})\quad P_X = P_{P_{X^*H}G} + P_{X^*HB} + P_{P_{XK}G} + P_{XKC}. \tag{52}$$

The four terms on the right-hand sides of the above four decompositions are mutually orthogonal.

The first and second terms in (49) add up to the first term on the right-hand side of Decomposition (A), and the third and fourth terms to the second term. The first and third terms in (49) add up to the first term on the right-hand side of in Decomposition (C), and the second and the fourth terms to the second term. Similar relations hold for the other decompositions in Proposition 7. Note that Decomposition (AD) is a special case of the decomposition presented in Theorem 2 of Takane et al. [19], which was introduced in the context of constrained canonical correlation analysis.

As before, the above decompositions of P_X may be applied to X from the left to obtain the corresponding decompositions of X. Again, the projectors applied from the left-hand side of X are transferrable to the right-hand side by defining the right-hand side projectors appropriately. Each term in the decompositions may be subjected to SVD for Internal Analysis.

6. Ridge operators and their decompositions

If we use the ridge least squares estimation in External Analysis rather than OLS, we obtain a ridge operator defined by

$$R_X(\lambda) = X(X'X + \lambda P_{X'})^- X' = X(X'M_X(\lambda)X)^- X' \tag{53}$$

similar to the orthogonal projector we have been dealing with so far, where λ is the ridge parameter (usually assuming a small positive value), and

$$M_X(\lambda) = P_X + \lambda T^+ \tag{54}$$

is called the left-hand side ridge metric matrix [20,21]. Let X^* be redefined as

$$X^* = X(X'M_X(\lambda)X)^+. \tag{55}$$

Then, we have

$$R_X(\lambda) = XX^{*'} = X^*X', \tag{56}$$

and

$$X'X^* = X^{*'}X = R_{X'}(\lambda) = S(X'M(\lambda)X)^+ = (X'M^*X)^+S = X'(XN_X(\lambda)X')^-X, \tag{57}$$

where

$$N_X(\lambda) = P_{X'} + \lambda S^+ \tag{58}$$

is called the right-hand side ridge metric matrix.

We may derive decompositions of $R_X(\lambda)$ similar to those of the orthogonal projector in Proposition 1:

Proposition 8. *Let K and L be as defined in Proposition 1. Then, we have the following decompositions of the ridge operator $R_X(\lambda)$ defined in (53):*

$$(\ddot{A})\quad R_X(\lambda) = R_{XH}(\lambda) + R_{X^*K}(\lambda), \tag{59}$$

$$(\ddot{B})\quad R_X(\lambda) = R_{X^*H}(\lambda) + R_{XK}(\lambda), \tag{60}$$

$$(\ddot{C})\quad R_X(\lambda) = R_{TG}(\lambda) + R_{X^*L}(\lambda), \tag{61}$$

and

$$(\ddot{D})\quad R_X(\lambda) = R_{R_X(\lambda)G}(\lambda) + R_{XL}(\lambda). \tag{62}$$

The two terms on the right-hand sides of the above decompositions are orthogonal with respect to $M_X(\lambda)$.

Decomposition (Ä) has been shown to hold in [20,21]. Proofs for the other decompositions are similar.

By premultiplying X by $R_X(\lambda)$ and applying SVD to the resultant $R_X(\lambda)X$, we in effect obtain a method of regularized PCA, in which component loadings are systematically shrunk toward zero. Let $X = UDV'$ denote the SVD of X. Then, $R_X(\lambda)$ can be expressed as $R_X(\lambda) = UD(D + \lambda I)^{-1}U'$, so that

$$R_X(\lambda)X = UD^2(D + \lambda I)^{-1}V', \tag{63}$$

which gives the SVD of $R_X(\lambda)X$. Since $D^2(D+\lambda I)^{-1} \leqslant D$ for $\lambda \geqslant 0$, the elements of the loading matrix are shrunk toward zero relative to the loadings obtained from the original X.

By applying the decompositions of $R_X(\lambda)$ given in the above propositions to X, we obtain a variety of decompositions of $R_X(\lambda)X$. By applying SVD to the terms in the decompositions, we obtain a variety of regularized CPCA's. As before, the effects of $R_X(\lambda)$ and its decompositions applied to the left of X can be transferred to the right by appropriately defining the right-hand ridge operators.

Note 3. The ridge LS estimation is not invariant over the scale of predictor variables. It is desirable to choose a scale for H, for example, that produces XH comparable in scale to the original X. This may be achieved by redefining a new H by U_rV_r', where $H = U_rD_rV_r'$ is the SVD of the original H matrix, and $r = \text{rank}(H)$. We must also redefine the matrix of regression coefficients by $V_rD_rV_r'B$, where B is the original matrix of regression coefficients, to make up for the redefinition of H. Note that with the new definition of $H = U_rV_r'$, we have $H'P_{X'}H = H'H = V_rV_r' = P_{H'} = P_{(XH)'}$, since $\text{Sp}(H) \subset \text{Sp}(X')$. Essentially the same procedure can be used for K to obtain a new K with similar properties.

Again, combining some of the two-term decompositions given above, we may derive four-term decompositions of $R_X(\lambda)$ analogously to Proposition 7.

Proposition 9. *(i) Let B and C be such that $G'XHB = O$, and $Sp(H'X'G) \oplus Sp(B) = Sp(H'X')$, and $G'X^{*\prime}KC = O$, and $Sp(K'X^{*\prime}G) \oplus Sp(C) = Sp(K'X^{*\prime})$, respectively. Then, the following two decompositions hold:*

$$\begin{aligned}(\ddot{A}\ddot{C})\quad R_X(\lambda) &= R_{XHH'X'G}(\lambda) + R_{XH(H'X'M_X(\lambda)XH)^-B}(\lambda)\\ &\quad + R_{X^*KK'X^{*\prime}G}(\lambda) + R_{X^*K(K'X^{*\prime}M_X(\lambda)X^*K)^-C}(\lambda),\end{aligned} \tag{64}$$

and

$$(\ddot{A}\ddot{D})\quad R_X(\lambda) = R_{R_{XH}(\lambda)G}(\lambda) + R_{XHB}(\lambda) + R_{R_{X^*K}(\lambda)G}(\lambda) + R_{X^*KC}(\lambda). \tag{65}$$

(ii) Let B and C be such that $G'X^*HB = O$, and $Sp(H'X^{*\prime}G) \oplus Sp(B) = Sp(H'X^{*\prime})$, and $G'X'KC = O$, and $Sp(K'X'G) \oplus Sp(C) = Sp(K'X')$, respectively. Then, the following two decompositions hold:

$$\text{(}\ddot{B}\ddot{C}\text{)} \quad R_X(\lambda) = R_{X^*HH'X^{*\prime}G}(\lambda) + R_{X^*H(H'X^{*\prime}M_X(\lambda)X^*H)^-B}(\lambda) + R_{XKK'X'G}(\lambda) + R_{XK(K'X'M_X(\lambda)XK)^-C}(\lambda), \tag{66}$$

and

$$\text{(}\ddot{B}\ddot{D}\text{)} \quad R_X(\lambda) = R_{R_{X^*H}(\lambda)G}(\lambda) + R_{X^*HB}(\lambda) + R_{R_{XK}(\lambda)G}(\lambda) + R_{XKC}(\lambda). \tag{67}$$

The four terms on the righ-hand side of the above decompositions are orthogonal with respect to $M_X(\lambda)$.

The first and second terms in Decompositions (ÄB̈) add up to the first term in Decomposition (Ä), and the third and the four terms to the second term in (Ä). The first and the third terms add up to the first term in (C̈), and the second and fourth terms to the second term in (C̈). Similar relationships hold for the other decompositions. As before, these decompositions can be applied to X to obtain various decompositions of $R_X(\lambda)X$, and terms in the decompositions may be subjected to SVD to obtain a variety of regularized CPCA's.

Note 4. The ridge operator defined in (53) may be generalized into a generalized ridge operator (GRO) by:

$$R_X^{(W,Q)}(\lambda) = X\left(X'WX + \lambda Q\right)^- XW = X\left(X'WM_X^{(W,Q)}(\lambda)X\right)^- X'W, \tag{68}$$

where W is an n by n nonnegative definite matrix of weights such that $\text{rank}(WX) = \text{rank}(X)$, and

$$M_X^{(W,Q)}(\lambda) = P_{X/W} + \lambda X(X'WX)^- Q(X'WX)^- X'W \tag{69}$$

is the generalized ridge metric matrix. By allowing $W \neq I_n$ and $Q \neq P_{X'}$, a greater variety of regularization are enabled. Let $X = UDV'$ denote the GSVD of X with the metric matrices W and Q^+. Then, $R_X^{(W,Q)}(\lambda)X = UD^2(D+\lambda I)^{-1}V'$. This shows that the GRO has the effect of shrinking the (generalized) singular values of X with respect to W and Q^+. The above equation appears identical in form to (63). Note, however, the key quantities involved (i.e., U's, V's, and D's) are different. In the former they were derived from the ordinary SVD of X, while in the present case they were derived from the GSVD of X with respect to W and Q^+.

7. Concluding remarks

In this paper, we presented several decompositions of the orthogonal projector P_X with the intention of their use in CPCA, in which the decompositions were first applied to the data matrix X to derive analogous decompositions of X (External Analysis), and the terms in the decompositions were then subjected to SVD to examine possible underlying structures within them (Internal Analysis). A numerical example was given to illustrate the basic procedure. The basic idea of decompositions has been extended to ridge operators to effectively obtain regularized CPCA.

The numerical example given in this paper is hardly sufficient to explain all important aspects of the proposed method. Many more realistic examples are necessary to demonstrate its usefulness in its full capacity. In particular, more specific guidelines as to "which decompositions are useful when" with concrete examples would be of enormous help for application oriented researchers.

Appendix A. Relationships with the Wedderburn–Guttman Decomposition

Let X denote an n by p matrix as before, and let M and N be matrices such that $M'XN$ is square and nonsingular. It has been shown by Guttman [8–10] that $\text{rank}(X - XN(M'XN)^{-1}M'X) = \text{rank}(X) - \text{rank}(XN(M'XN)^{-1}M'X) = \text{rank}(X) - \text{rank}(M'XN)$. This is called the Wedderburn–Guttman (WG) theorem. Takane and Yanai [22,23] generalized the theorem by identifying the necessary and sufficient condition under which the regular inverse of $M'XN$ could be replaced by a generalized inverse in the above formula. It turned out that $\text{rank}(X - XN(M'XN)^{-}M'X) = \text{rank}(X) - \text{rank}(M'XN)$ held unconditionally. However, for $\text{rank}(X - XN(M'XN)^{-}M'X) = \text{rank}(X) - \text{rank}(XN(M'XN)^{-}M'X)$ to hold (or equivalently for $\text{rank}(XN(M'XN)^{-}M'X) = \text{rank}(M'XN)$ to hold) requires a condition. This condition along with the rank formula is called the extended WG theorem. These theorems (the original and extended WG theorems) imply a decomposition of X of the form

$$X = XN(M'XN)^{-}M'X + (X - XN(M'XN)^{-}M'X), \tag{70}$$

which we call the Wedderburn–Guttman decomposition. In this Appendix, we discuss relationships between our proposal and the above decomposition.

We begin by rewriting the second term in the above decomposition as a single matrix (rather than a difference between two matrices).

Theorem. *Let X be as introduced earlier, and let M, N, $\tilde{M}$, and $\tilde{N}$ be such that*

$$\text{rank}(M'XN) + \text{rank}(\tilde{N}'X^{-}\tilde{M}) = \text{rank}(X), \tag{71}$$

where X^{-} is any g-inverse of X,

$$M'XX^{-}\tilde{M} = O, \tag{72}$$

and

$$\tilde{N}'X^{-}XN = O. \tag{73}$$

Then, the following decomposition holds:

$$X = XN(M'XN)^{-}M'X + \tilde{M}(\tilde{N}'X^{-}\tilde{M})^{-}\tilde{N}'. \tag{74}$$

Proof. Let $Z_1 = [N, X\tilde{M}]$, and $Z_2 = [M, (X^{-})'\tilde{N}]$. Then, we have

$$Z_2'XZ_1 = \begin{bmatrix} M'XN & O \\ O & \tilde{N}'X^{-}\tilde{M} \end{bmatrix}, \tag{75}$$

and

$$XZ_1(Z_2'XZ_1)^{-*}Z_2'X = XN(M'XN)^{-}M'X + \tilde{M}(\tilde{N}'X^{-}\tilde{M})^{-}\tilde{N}', \tag{76}$$

where

$$(Z_2'XZ_1)^{-*} = \begin{bmatrix} (M'XN)^{-} & O \\ O & (\tilde{N}'X^{-}\tilde{M})^{-} \end{bmatrix}. \tag{77}$$

Clearly, $(Z_2'XZ_1)^{-*} \in \{(Z_2'XZ_1)^{-}\}$. From (75) and (71) we have

$$\text{rank}(Z_2'XZ_1) = \text{rank}(M'XN) + \text{rank}(\tilde{N}'X^{-}\tilde{M}) = \text{rank}(X). \tag{78}$$

(74) follows from Theorem 2.1 of Mitra [24], which states that $XZ_1(Z_2'XZ_1)^-Z_2'X = X$ (i.e., $Z_1(Z_2'XZ_1)^-Z_2' \in \{X\}$) if and only if $\text{rank}(Z_2'XZ_1) = \text{rank}(X)$.

Decomposition (74) may be called a rectangular version of Khatri's [16] lemma. It was initially assumed that X was symmetric and positive-definite, and that M and N were identical. In Khatri [17], this condition was relaxed somewhat, but X was still assumed square. □

Corollary 1. *Let X, M, N, $\tilde{M}$, and $\tilde{N}$ be as defined in the above theorem. Let*

$$rank(M'XN) = rank(XN) = rank(M'XX^-), \tag{79}$$

and

$$rank(\tilde{N}'X^-\tilde{M}) = rank(\tilde{M}) = rank(\tilde{N}'X^-). \tag{80}$$

Then,

$$XX^- = XN(M'XN)^-M'XX^- + \tilde{M}(\tilde{N}'X^-\tilde{M})^-\tilde{N}'X^-, \tag{81}$$

where the first term on the right is the projector onto $Sp(XN)$ along $Ker(M'XX^-)$, and the second term the projector onto $Sp(\tilde{M})$ along $Ker(\tilde{N}'X^-)$.

Proof. Postmultiplying (74) by X, we obtain (81). Conditions (79) and (80) ensure that the two terms on the right-hand side of (81) are the projectors with the prescribed onto and along spaces. □

Corollary 2. *Let X, M, N, $\tilde{M}$, and $\tilde{N}$ be as defined in the above theorem. Let*

$$rank(M'XN) = rank(X^-XN) = rank(\tilde{M}'X), \tag{82}$$

and

$$rank(\tilde{N}'X^-\tilde{M}) = rank(X^-\tilde{M}) = rank(\tilde{N}). \tag{83}$$

Then,

$$X^-X = X^-XN(M'XN)^-M'X + X^-\tilde{M}(\tilde{N}'X^-\tilde{M})^-\tilde{N}', \tag{84}$$

where the first term on the right is the projector onto $Sp(X^-XN)$ along $Ker(M'X)$, and the second term the projector onto $Sp(X^-\tilde{M})$ along $Ker(\tilde{N})$.

Proof. A proof is similar to that of Corollary 1, and so is omitted. □

Note that (79) and (82) are equivalent. On the other hand, (80) and (83) are not, unless $\text{Sp}(\tilde{M}) \subset \text{Sp}(X)$ and $\text{Sp}(\tilde{N}) \subset \text{Sp}(X^-)$.

Note 5. In (81) $M'XX^-$ may be redefined as new M', and in (84), X^-XN may be redefined as as new N. They can be done without loss of generality.

Corollary 3. *Let $X^- = X^{*\prime}$. Then, (74), (81), and (84) can be rewritten as:*

$$X = XN(M'XN)^-M'X + \tilde{M}(\tilde{N}'X^{*\prime}\tilde{M})^-\tilde{N}', \tag{85}$$

$$P_X = XN(M'XN)^-M'P_X + \tilde{M}(\tilde{N}'X^{*\prime}\tilde{M})^-\tilde{N}'X^{*\prime}, \tag{86}$$

and

$$P_{X'} = P_{X'}N(M'XN)^-M'X + X^{*\prime}\tilde{M}(\tilde{N}'X^{*\prime}\tilde{M})^-\tilde{N}', \tag{87}$$

respectively.

Note 6. $P_X M = M$ if $\mathrm{Sp}(M) \subset \mathrm{Sp}(X)$, and $P_{X'} N = N$ if $\mathrm{Sp}(N) \subset \mathrm{Sp}(X')$.

Decompositions (A′), (A), and (a) in Propositions 2, 1, and 3 follow from (85), (86), and (87), respectively, by setting $M = XH$, $N = H$, $\tilde{M} = X^*K$, and $\tilde{N} = K$. The other decompositions in the three propositions can be derived similarly.

Acknowledgements

The work reported in this paper has been supported by grants from the Social Sciences and Humanities Research Council of Canada and from the Natural Sciences and Engineering Research Council of Canada to the first author.

References

[1] Y. Takane, T. Shibayama, Principal component analysis with external information on both subjects and variables, Psychometrika 56 (1991) 97–120.
[2] Y. Takane, M.A. Hunter, Constrained principal component analysis: a comprehensive theory, Appl. Algebra Engrg. Comm. Comput. 12 (2001) 391–419.
[3] M.A. Hunter, Y. Takane, Constrained principal component analysis: various applications, J. Educ. Behav. Stat. 27 (2002) 41–81.
[4] P. Metzak, E. Feredoes, Y. Takane, L. Wang, S. Weinstein, T. Cairo, E.T.C. Ngan, T.S. Woodward, Constrained principal component analysis reveals functionally connected load-dependent networks involved in multiple stages of working memory, Hum. Brain Mapp., in press, doi:10.1002/hbm.21072. Available from: < http://onlinelibrary.wiley.com/doi/10-1002/hbm.21072/pdf>.
[5] T.S. Woodward, T.A. Cairo, C.C. Ruff, Y. Takane, M.A. Hunter, E.T.C. Ngan, Load dependent neural systems underlying encoding and maintenance in verbal working memory, Neuroscience 139 (2006) 317–325.
[6] M.A. Hunter, Y. Takane, CPCA: a program for principal component analysis with external information on subjects variables, Behav. Res. Methods Instrum. Comput. 30 (1998) 506–516.
[7] T.S. Woodward, Y. Takane, L. Wang, P.D. Metzak, J. C. Whitman, Constrained principal component analysis for fMRI data (fMRI CPCA), Organ. Hum. Brain Mapp., in press. Available from: < www.aievolution.com/hbm1001/index.cfm?do=abs.viewAbs&abs=3845>.
[8] L. Guttman, General theory and methods of matric factoring, Psychometrika 9 (1944) 1–16.
[9] L. Guttman, Multiple group methods for common-factor analysis: their basis computation and interpretation, Psychometrika 17 (1952) 209–222.
[10] L. Guttman, A necessary and sufficient formula for matric factoring, Psychometrika 22 (1957) 79–81.
[11] M.W. Browne, W. Kristof, On the oblique rotation of a factor matrix to a specified pattern, Psychometrika 34 (1969) 237–248.
[12] Y. Takane, H. Yanai, S. Mayekawa, Relationships among several methods of linearly constrained correspondence analysis, Psychometrika 56 (1991) 667–684.
[13] H. Yanai, Y. Takane, Canonical correlation analysis with linear constraints, Linear Algebra Appl. 176 (1992) 75–82.
[14] R.F. Potthoff, S.N. Roy, A generalized multivariate analysis of variance model useful especially for growth curve problems, Biometrika 51 (1964) 313–326.
[15] C.R. Rao, The use and interpretation of principal component analysis in applied research, Sankhya Ser. A 26 (1964) 329–358.
[16] C.G. Khatri, A note on a MANOVA model applied to problems in growth curves, Ann. Inst. Statist. Math. 18 (1966) 75–86.
[17] C.G. Khatri, Some properties of BLUE in a linear model and canonical correlations associated with linear transformations, J. Multivariate Anal. 34 (1990) 211–226.
[18] M. Burden, S. Gardner, N.J. Le Roux, J.P.J. Swart, Ou-Kaapse meubels en stinkhout-identifikasie: moontlikhede met kanoniece veranderlike-analise en bistippings, South African J. Cultur. Hist. 15 (2001) 50–73.
[19] Y. Takane, H. Yanai, H. Hwang, An improved method for generalized constrained canonical correlation analysis, Comput. Stat. Data Anal. 50 (2006) 221–241.
[20] Y. Takane, H. Yanai, On ridge operators, Linear Algebra Appl. 428 (2008) 1778–1790.
[21] Y. Takane, More on regularization and (generalized) ridge operators, in: K. Shigemasu, A. Okada, T. Imaizumi, T. Hoshino (Eds.), New Trends in Psychometrics, University Academic Press, Tokyo, 2008, pp. 443–452.
[22] Y. Takane, H. Yanai, On the Wedderburn–Guttman theorem, Linear Algebra Appl. 410 (2005) 267–278.
[23] Y. Takane, H. Yanai, Alternative characterizations of the extended Wedderburn–Guttman theorem, Linear Algebra Appl. 422 (2007) 701–711.
[24] S.K. Mitra, A new class of g-inverse of square matrices, Sankhya Ser. A 30 (1968) 322–330.