On Ridge Operators

Yoshio Takane [a,1], **Haruo Yanai** [b]

[a] *Department of Psychology, McGill University, 1205 Dr. Penfiled Avenue, Montreal, Quebec, H3A 1B1 Canada*

[b] *St. Luke's College of Nursing, 10-1 Akashi-cho, Chuo-ku, Tokyo 104-0044 Japan*

ABSTRACT

Let X be an n by p matrix, and define $R_X(\lambda) = X(X'X + \lambda P_{X'})^- X'$, which is called a ridge operator, where λ is a nonnegative constant (called the ridge parameter), and $P_{X'} = X'(XX')^- X$. Various properties of $R_X(\lambda)$ were discussed, including additive decompositions of this matrix similar to those of $P_X \equiv R_X(0) = X(X'X)^- X'$, the orthogonal projector onto the range space of X. These properties and decompositions are useful, especially in ridge estimation of reduced rank regression and multiple-set canonical correlation analyses.

AMS classification: 15A03; 15A09

Keywords: Metric matrix; Projector; Additive decompositions; Reduced rank regression; Multiple-set canonical correlation analysis

---

1 Introduction

Let X denote an n by p matrix. We define a ridge operator by

$$R_X(\lambda) = X(X'X + \lambda P_{X'})^- X', \tag{1}$$

where $^-$ indicates a g-inverse, λ (≥ 0) is called the ridge parameter, and $P_{X'} = X'(XX')^- X$ is the orthogonal projector onto $\text{Sp}(X')$, the row space of X. This is a linear operator for a fixed value of λ as will be assumed throughout this paper. Matrix $P_{X'}$ reduces to I_p, the identity matrix of order p, when X is columnwise nonsingular. A matrix of the above form most notably arises in ridge regression ([6], [15]; see also Groß [3] for an up-to-date account of the topic), where the vector of regression coefficients b in the linear model $y = Xb + e$ is estimated by minimizing the ridge least squares (RLS) criterion,

$$\phi_\lambda(b) = ||y - Xb||^2 + \lambda ||b||^2_{P_{X'}}, \tag{2}$$

[1] Corresponding author. Tel.: +1 514 398 6125; fax: +1 514 398 4896.

Email addresses: takane@psych.mcgill.ca (Y. Takane), hyanai@slcn.ac.jp (H. Yanai).

where $||y - Xb||^2 = (y - Xb)'(y - Xb)$, and $||b||^2_{P_{X'}} = b'P_{X'}b = b'b$. Throughout this paper we assume that the vector of regression coefficients is in the row space of predictor variables, namely $\mathrm{Sp}(b) \subset \mathrm{Sp}(X')$. This can be made without loss of generality: Suppose $b = b_0 + b_1$, where $b_0 \in \mathrm{Sp}(X')$ and $b_1 \in \mathrm{Ker}(X)$, where $\mathrm{Ker}(X)$ indicates the null space of X. Then, $Xb = Xb_0 + Xb_1 = Xb_0$. Thus, we may set $b = b_0 \in \mathrm{Sp}(X')$ without affecting the prediction vector Xb. An RLS estimate of b is given by $\hat{b}(\lambda) = X^-_\lambda y$, where

$$X^-_\lambda = (X'X + \lambda P_{X'})^- X' \tag{3}$$

is sometimes called a Tikhonov regularized inverse [5]. (To ensure $\hat{b}(\lambda) \in \mathrm{Sp}(X')$, we may premultiply $\hat{b}(\lambda)$ defined above by $P_{X'}$, which is equivalent to choosing $(X'X + \lambda P_{X'})^-_m$ (a minimum norm g-inverse of $X'X + \lambda P_{X'}$) for $(X'X + \lambda P_{X'})^-$ in (3). However, since the prediction vector $X\hat{b}(\lambda)$ is invariant no matter which g-inverse of $X'X + \lambda P_{X'}$ is used, we do not bother to require $\hat{b}(\lambda) \in \mathrm{Sp}(X')$ explicitly.) Matrix $R_X(\lambda)$ defined in (1) is an operator that turns y into $X\hat{b}(\lambda)$, that is, $X\hat{b}(\lambda) = R_X(\lambda)y$, where $R_X(\lambda) = XX^-_\lambda$.

Gulliksson and Wedin [5] called $R_X(\lambda)$ a Tikhonov filter matrix, and discussed some of its properties. In fact, they treated a special case of (1), in which X was assumed columnwise nonsingular. Note, however, that $(X'X + \lambda I_p)^{-1} \in \{(X'X + \lambda P_{X'})^-\}$, and that $R_X(\lambda)$ is invariant over the choice of g-inverse $(X'X + \lambda P_{X'})^-$ (see Theorem 1(i) below), so that $X(X'X + \lambda I_p)^{-1}X' = X(X'X + \lambda P_{X'})^- X' = R_X(\lambda)$. In this paper we present many other interesting properties of the ridge operator, including its additive decompositions analogous to the well known decompositions (e.g., [9]) of the orthogonal projector $P_X \equiv R_X(0) = X(X'X)^-X'$. We first discuss the simplest case in which X is a single (non-partitioned) matrix. We then discuss the situation in which X is partitioned into K disjoint row block matrices (section 3). We then focus on the special case of $K = 2$ and derive a number of decomposition formula for $R_X(\lambda)$ (section 4). In the final section we provide examples of application.

2 Non-partitioned matrix X

We begin by defining a matrix which plays a key role in this paper. Let X and λ be as introduced earlier. Define

$$M(\lambda) = J_n + \lambda(XX')^+, \tag{4}$$

where J_n is any symmetric matrix such that $J_nX = X$ (e.g., $J_n = sI_n + (1-s)P_X$ for any s, where $P_X \equiv X(X'X)^-X'$), and $(XX')^+$ is the Moore-Penrose inverse of XX'. However, to ensure nonnegative-definiteness (*nnd*) of $M(\lambda)$, we require $s \geq 0$. Matrix $M(\lambda)$ is called a ridge metric matrix. It can easily be observed that $M(\lambda)$ is invariant over any orthogonal transformations of X of the form XT, where $T'T = TT' = I_p$. Note that $(XX)^+$ can be expressed as

$$(XX')^+ = P_X(XX')^-P_X = X(X'X)^{+2}X', \tag{5}$$

where $(X'X)^{+2} = ((X'X)^+)^2$. The first equality follows from Note 3.3.8 of Rao and Mitra [8], and the second equality from the commutativity of $X'X$ and $(X'X)^+$. There are many interesting properties of $M(\lambda)$, of which the most relevant one in this paper is the following:

$$X'X + \lambda P_{X'} = X'M(\lambda)X, \tag{6}$$

so that $R_X(\lambda)$ defined in (1) can also be expressed as

$$R_X(\lambda) = X(X'M(\lambda)X)^- X'. \tag{7}$$

We also let

$$S_X(\lambda) = I_n - R_X(\lambda), \tag{8}$$

and

$$N(\lambda) = J_p + \lambda(X'X)^+, \tag{9}$$

where J_p is any matrix such that $XJ_p = X$, and hence can be any matrix of the form $sI_p + (1 - s)P_{X'}$ for any s. Again, we require $s \geq 0$ to ensure the *nnd*-ness of $N(\lambda)$. Similarly to (5), $(X'X)^+$ can be expressed as

$$(X'X)^+ = P_{X'}(X'X)^- P_{X'} = X'(XX')^{+2}X. \tag{10}$$

The following equalities hold.

Lemma 1. Let $M(\lambda)$ and $N(\lambda)$ be as defined in (4) and (9). Then,

(i) $X'M(\lambda) = N(\lambda)X'$.

(ii) $M(\lambda)X = XN(\lambda)$.

(iii) $\mathrm{Sp}(M(\lambda)X) = \mathrm{Sp}(XN(\lambda)) = \mathrm{Sp}(X)$.

Proof. A proof for (i) is straightforward by noting (5) and (10). That is, $X'M(\lambda) = X' + \lambda X'P_X(XX')^- P_X = X' + \lambda X'(XX')^- X(X'X)^- X' = X' + \lambda P_{X'}(X'X)^- P_{X'}X' = (J_p + \lambda(X'X)^+)X' = N(\lambda)X'$. (ii) follows from (i) because of the symmetry of $M(\lambda)$ and $N(\lambda)$. The first equality in (iii) follows immediately from (ii). The second equality follows from $XN(\lambda) = X(I_p + \lambda(X'X)^+)$ and $\mathrm{rank}(I_p + \lambda(X'X)^+) = p$. □

In fact, somewhat more "general" results than Lemma 1 can be established, namely $X'M(\lambda)^s = N(\lambda)^s X$ and $M(\lambda)^s X = XN(\lambda)^s$ for any s. However, this generality is not relevant in the present paper. Lemma 1 implies the following corollary.

Corollary 1. Let $M(\lambda)$ and $N(\lambda)$ be as in Lemma 1. Then,

$$X'M(\lambda)X = N(\lambda)X'X = X'XN(\lambda) = N(\lambda)^s X'XN(\lambda)^{1-s} \tag{11}$$

for any s.

Proof. The first two equalities follow directly from Lemma 1. The second equality indicates

$X'X$ and $N(\lambda)$ commute, which implies the third equality. $\square$

We now give the first theorem.

Theorem 1. Let $R_X(\lambda)$, $S_X(\lambda)$, and $M(\lambda)$ be as defined above. Then, the following properties hold.

(i) $R_X(\lambda)$ is symmetric and invariant over the choice of g-inverse $(X'M(\lambda)X)^-$.

(ii) $R_X(\lambda)M(\lambda)R_X(\lambda) = R_X(\lambda)$. ($R_X(\lambda)$ is "idempotent" with respect to the metric matrix $M(\lambda)$.)

(iii) $R_X(\lambda) - R_X(\lambda)^2 = \lambda X(X'M(\lambda)X)^- P_{X'}(X'M(\lambda)X)^- X' \geq 0$, where "$\geq 0$" means the matrix on the left hand side is *nnd*. The strict equality means $R_X(\lambda)$ is a projector, which happens if $\lambda = 0$.

(iv) $R_X - R_X(\lambda)^2 = R_X(\lambda)S_X(\lambda) = S_X(\lambda)R_X(\lambda) = S_X(\lambda) - S_X(\lambda)^2 (\geq 0)$.

(v) $M(\lambda)^a R_X(\lambda) M(\lambda)^{1-a} = X(X'X)^- X' = P_X = R_X(0)$ for any $0 \leq a \leq 1$. In particular, $M(\lambda)^a R_X(\lambda) M(\lambda)^{1-a} = R_X(\lambda)M(\lambda)$ when $a = 0$, and $M(\lambda)^a R_X(\lambda) M(\lambda)^{1-a} = M(\lambda)R_X(\lambda)$ when $a = 1$.

(vi) $P_X R_X(\lambda) = R_X(\lambda)$, which implies $Q_X R_X(\lambda) = 0$, $Q_X S_X(\lambda) = Q_X$, and $P_X S_X(\lambda) = P_X - R_X(\lambda)$, where $Q_X = I - P_X$.

(vii) Let $\tilde{X}$ be any matrix such that $\mathrm{Sp}(\tilde{X}) \subset \mathrm{Sp}(X)$, and let $R_{\tilde{X}}(\lambda) = \tilde{X}(\tilde{X}'M(\lambda)\tilde{X})^- \tilde{X}'$. Then, $R_{\tilde{X}}(\lambda)M(\lambda)R_{\tilde{X}}(\lambda) = R_{\tilde{X}}(\lambda)$, and $R_{\tilde{X}}(\lambda)M(\lambda)R_X(\lambda) = R_X(\lambda)M(\lambda)R_{\tilde{X}}(\lambda) = R_{\tilde{X}}(\lambda)$.

(viii) $R_X(\lambda)^+ = M(\lambda)$, and $M(\lambda)^+ = R_X(\lambda)$.

(ix) $R_X(\lambda)$ is invariant over the orthogonal transformation of X of the form XT, where $T'T = TT' = I$.

Proof. (i) The invariance follows from $\mathrm{Sp}(X') \subset \mathrm{Sp}(X'M(\lambda)X)$, and Lemma 2.2.4(iii) (and Supplement 14) of Rao and Mitra [8]. The invariance implies symmetry. (ii) can be directly verified: $R_X(\lambda)M(\lambda)R_X(\lambda) = X(X'M(\lambda)X')^- X'M(\lambda)X(X'M(\lambda)X)^- X' = X(X'M(\lambda)X')^- X' = R_X(\lambda)$, since $X(X'M(\lambda)X')^- X'M(\lambda)X = X$ because $\mathrm{Sp}(X') \subset \mathrm{Sp}(X'M(\lambda)X)$. (ii) implies $R_X(\lambda)M(\lambda)$ is a projector. (iii) can also be directly verified: $R_X(\lambda)^2 + \lambda X(X'M(\lambda)X)^- P_{X'}(X'M(\lambda)X)^- X' = R_X(\lambda)^2 + \lambda R_X(\lambda)(XX')^+ R_X(\lambda) = R_X(\lambda)(J_n + \lambda(XX')^+)R_X(\lambda) = R_X(\lambda)M(\lambda)R_X(\lambda) = R_X(\lambda)$. (iii) indicates that $R_X(\lambda)$ is a contraction matrix with its eigenvalues all between 0 and 1 inclusive. $R_X(\lambda)$ is also semi-simple ($\mathrm{rank}(R_X(\lambda)) = \mathrm{rank}(R_X(\lambda)^2)$), so that it is diagonalizable by a similarity transformation. See also the second paragraph of Application 1 in the application section. (iv) is trivial, but it indicates that $S_X(\lambda)$ is also a contraction matrix. (v) (ii) indicates that $R_X(\lambda)M(\lambda)$ is the projector onto $\mathrm{Sp}(X)$ along $\mathrm{Ker}(X'M(\lambda))$ [15], since $\mathrm{rank}(M(\lambda)X) = \mathrm{rank}(X)$ by Lemma 1(iii) which also indicates that $\mathrm{Ker}(X'M(\lambda)) = \mathrm{Ker}(X')$, so that $R_X(\lambda)M(\lambda)$ is in fact the orthogonal projector onto $\mathrm{Sp}(X)$. (vi) By direct verification. (vii) is trivial. Note, however, that $R_{\tilde{X}}(\lambda)M(\lambda)$ is a projector, but in general $R_{\tilde{X}}(\lambda)M(\lambda) \neq P_{\tilde{X}}$.

(viii) Another interpretation of (ii) is that $M(\lambda)$ is a g-inverse of $R_X(\lambda)$. That $M(\lambda)$ satisfies the other three Penrose conditions can be easily verified. See (v), and note that $M(\lambda)P_X = M(\lambda)$. The second equality follows from the first. (ix) This easily follows from the invariance of $M(\lambda)$ over any orthogonal transformations of the form XT. $\square$

The ridge operator defined in (1) can easily be generalized [6] to:

$$R_X^{(L)}(\lambda) = X(X'X + \lambda L)^{-}X', \tag{12}$$

where L is an *nnd* matrix such that $\text{Sp}(L) = \text{Sp}(X') = \text{Sp}(X'X)$. The above theorem can also be extended to the generalized ridge operator.

Theorem 2. Let X be as in Theorem 1, and let L be a p by p *nnd* matrix such that $\text{Sp}(L) = \text{Sp}(X')$. Define

$$M^{(L)}(\lambda) = J_n + \lambda(XL^{-}X')^{+}. \tag{13}$$

Then, $\text{Sp}(M^{(L)}(\lambda)X) = \text{Sp}(X)$, $X'M^{(L)}(\lambda)X = X'X + \lambda L$, $R_X^{(L)}(\lambda) = X(X'M^{(L)}(\lambda)X)^{-}X'$, and properties analogous to those that hold for $R_X(\lambda)$ stated in Theorem 1 also hold for $R_X^{(L)}(\lambda)$.

Proof. Note first that $(XL^{-}X')^{+}$ can be expressed as $(XL^{-}X')^{+} = X(X'X)^{+}L(X'X)^{+}X'$, and the rest of the theorem can be proved similarly to Theorem 1. If L does not satisfy $\text{Sp}(L) = \text{Sp}(X')$ initially, we may simply redefine it by $P_{X'}LP_{X'}$. $R_X(\lambda)$ and $M(\lambda)$ are considered as special cases of $P^{(L)}(\lambda)$ and $M^{(L)}(\lambda)$, respectively, where $L = P_{X'}$. Note that $X(P_{X'})^{-}X' = X(P_{X'})^{+}X' = XP_{X'}X' = XX'$. The property analogous to Theorem 1(ix) may require some elaboration. Let $T = (Z')_R^{-}UZ'$, where Z is a p by r matrix such that $L = ZZ'$ (where $r = \text{rank}(L)$), $(Z')_R^{-}$ is a right inverse of Z', and U is any orthogonal matrix of order r. Then, $R_X^{(L)}(\lambda)$ is invariant over the transformation of the form XT. Note that $A = (Z')_R^{-}U'Z' \in \{T_r^{-}\}$ (where T_r^{-} is a reflexive g-inverse of T), $XTA = X(Z')_R^{-}Z' = X$, and $AL^{-}A' = L^{-}$, so that $XL^{-}X' = XTAL^{-}A'T'X' = XTL^{-}T'X'$. $\square$

The above theorem can further be extended to the situation in which we have a non-identity weight matrix V on the column side of X. Let X and L be as defined in Theorem 2, and let V be an n by n *nnd* matrix such that $\text{rank}(VX) = \text{rank}(X)$ [15]. We define $M^{(L,V)}(\lambda) = J_n^* + \lambda(XL^{-}X'V)^{+}_{V,V}$, where J_n^* is any matrix such that $X'VJ_n^* = X'V$, and $(XL^{-}X'V)^{+}_{V,V}$ is the weighted Moore-Penrose inverse of $XL^{-}X'V$ with respect to the metric matrices V and V (i.e., $(XL^{-}X'V)^{+}_{V,V}$ is a reflexive g-inverse of $XL^{-}X'V$ such that both $(XL^{-}X'V)^{+}_{V,V}XL^{-}X'V$ and $XL^{-}X'V(XL^{-}X'V)^{+}_{V,V}$ are left-symmetric with respect to V). Then, $X'VM^{(L,V)}(\lambda)X = X'VX + \lambda L$, and we define $R_X^{(L,V)}(\lambda) = X(X'VM^{(L,V)}(\lambda)X)^{-}X'V$. Mitra [7] called the $(X'VX + \lambda L)^{-}X'V$ part of $R_X^{(L,V)}(\lambda)$ an optimal inverse. Obviously, an optimal inverse is not a g-inverse in the usual sense (e.g., [8]). Note that $(XL^{-}X'V)^{+}_{V,V}$ can be expressed as $(XL^{-}X'V)^{+}_{V,V} = X(X'VX)^{+}L(X'VX)^{+}X'V$, so that $M^{(L,V)}(\lambda)$ is not symmetric, but left-symmetric with respect to V.

3 When X is partitioned into K disjoint subsets

In this section we deal with the situation in which X is partitioned into K subsets. We assume that these submatrices are disjoint.

Theorem 3. Let $X = [X_1, \ldots, X_K]$ be an n by p row block matrix, where X_k (n by p_k), $k = 1, \ldots, K$, are disjoint, i.e., $\sum_{k=1}^{K} \mathrm{rank}(X_k) = \mathrm{rank}(X)$. Define $M(\lambda)$ as in (4). Then,

$$X_k' M(\lambda) X_j = \begin{cases} X_k' X_k + \lambda P_{X_k'} \equiv D_k(\lambda) & (k = j), \\ X_k' X_j & (k \neq j), \end{cases} \tag{14}$$

where $P_{X_k'} = X_k'(X_k X_k')^- X_k$ is the orthogonal projector onto $\mathrm{Sp}(X_k')$. Note that $P_{X_k'}$ reduces to I_{p_k} if X_k has full column rank.

Proof. From Theorem 1.2 of Anderson and Styan [2], $A_k A^- A_k = A_k$ and $A_k A^- A_j = 0$ for $k \neq j$ if and only if $\mathrm{rank}(A) = \sum_{k=1}^{K} \mathrm{rank}(A_k)$, where $A = \sum_{k=1}^{K} A_k$. By setting $A_k = X_k X_k'$, we obtain $A = XX' = \sum_{k=1}^{K} X_k X_k' = \sum_{k=1}^{K} A_k$, so that $\mathrm{rank}(A) = \sum_{k=1}^{K} \mathrm{rank}(A_k)$ is equivalent to $\mathrm{rank}(X) = \sum_{k=1}^{K} \mathrm{rank}(X_k)$. It also follows that ([16], Theorem 2)

$$X_k X_k' (XX')^- X_j X_j' = \begin{cases} X_k X_k' & (k = j), \\ 0 & (k \neq j). \end{cases} \tag{15}$$

By pre- and post-multiplying (15) by $(X_k' X_k)^+ X_k'$ and $X_j (X_j' X_j)^+$, respectively, we obtain

$$X_k' (XX')^- X_j = \begin{cases} P_{X_k'} & (k = j), \\ 0 & (k \neq j), \end{cases} \tag{16}$$

from which the theorem follows immediately. Note that $X_k'(XX')^- X_j = X_k' P_X (XX')^- P_X X_j$ $= X_k'(XX')^+ X_j$, where the second equality follows from (5). □

The following corollary can readily be derived from the above theorem. Matrix $R_{X_k}(\lambda)$ (to be defined below) has similar properties as $R_X(\lambda)$ discussed in the previous section.

Corollary 2. Let X_k, $D_k(\lambda)$, and $M(\lambda)$ be as defined in Theorem 3. Further, let

$$R_{X_k}(\lambda) = X_k D_k(\lambda)^- X_k' \tag{17}$$

for $k = 1, \ldots, K$. Then, $R_{X_k}(\lambda)$ is symmetric and invariant over the choice of $D_k(\lambda)^-$, and

$$R_{X_k}(\lambda) M(\lambda) R_{X_j}(\lambda) = \begin{cases} R_{X_k}(\lambda) & (k = j), \\ R_{X_k}(\lambda) R_{X_j}(\lambda) & (k \neq j). \end{cases} \tag{18}$$

Proof. That $R_{X_k}(\lambda)$ is symmetric and invariant over the choice of $D_k(\lambda)^-$ follows from $\mathrm{Sp}(X_k') \subset \mathrm{Sp}(D_k(\lambda))$ and Lemma 2.2.4(iii) of Rao and Mitra [8], as in the case of $R_X(\lambda)$ in Theorem 1. Furthermore, we have $R_{X_k}(\lambda) M(\lambda) R_{X_j}(\lambda) = X_k D_k(\lambda)^- X_k' M(\lambda) X_j D_j(\lambda)^- X_j'$.

For $k = j$, $X_k' M(\lambda) X_k = D_k(\lambda)$, so that $R_{X_k}(\lambda) M(\lambda) R_{X_k}(\lambda) = X_k D_k(\lambda)^- D_k(\lambda) D_k(\lambda)^- X_k'$ $= X_k D_k(\lambda)^- X_k' = R_{X_k}(\lambda)$, since $X_k D_k(\lambda)^- D_k(\lambda) = X_k$. (This is no surprise if we set $\tilde{X} = X_k$ in Theorem 1(vii).) For $k \neq j$, $X_k' M(\lambda) X_j = X_k' X_j$, so that $R_{X_k}(\lambda) M(\lambda) R_{X_j}(\lambda) = R_{X_k}(\lambda) R_{X_j}(\lambda)$. □

Theorem 3 and Corollary 2 can be generalized into the following theorem in a manner in which Theorem 1 was generalized into Theorem 2.

Theorem 4. Let X be as introduced above, and let L_k be a p_k by p_k *nnd* matrix such that $\mathrm{Sp}(L_k) = \mathrm{Sp}(X_k')$. Let L be the block diagonal matrix with L_k as the k^{th} diagonal block. Define $M^{(L}(\lambda)$ as in (13). Then,

$$X_k' M^{(L)}(\lambda) X_j = \begin{cases} X_k' X_k + \lambda L_k \equiv D_k^{(L_k)}(\lambda) & (k = j), \\ X_k' X_j & (k \neq j). \end{cases} \tag{19}$$

Further, let

$$R_{X_k}^{(L_k)}(\lambda) = X_k D_k^{(L_k)}(\lambda)^- X_k' \tag{20}$$

for $k = 1, \ldots, K$. Then,

$$R_{X_k}^{(L_k)}(\lambda) M^{(L)}(\lambda) R_{X_j}^{(L_j)}(\lambda) = \begin{cases} R_{X_k}^{(L_k)}(\lambda) & (k = j), \\ R_{X_k}^{(L_k)}(\lambda) R_{X_j}^{(L_j)}(\lambda) & (k \neq j). \end{cases} \tag{21}$$

Proof. We set $A = XL^-X' = \sum_{k=1}^K X_k L_k^- X_k' = \sum_{k=1}^K A_k$ in Anderson and Styan's [2] Theorem 1.2. Then,

$$X_k L_k^- X_k' (XL^-X')^- X_j L_j^- X_j' = \begin{cases} X_k L_k^- X_k' & (k = j), \\ 0 & (k \neq j), \end{cases} \tag{22}$$

if and only if

$$\mathrm{rank}(XL^-X') = \sum_{k=1}^K \mathrm{rank}(X_k L_k^- X_k'), \tag{23}$$

which, in the light of $\mathrm{Sp}(L_k) = \mathrm{Sp}(X_k')$, is equivalent to $\mathrm{rank}(X) = \sum_{k=1}^K \mathrm{rank}(X_k)$. By pre- and post-multiplying (22) by $L_k(X_k'X_k)^+ X_k'$ and $X_k(X_k'X_k)^+ L_k$, respectively, and also taking into account that $\mathrm{Sp}(L_k) = \mathrm{Sp}(X_k')$, we obtain

$$X_k'(XL^-X')^- X_j = X_k'(XL^-X')^+ X_j = \begin{cases} L_k & (k = j), \\ 0 & (k \neq j), \end{cases} \tag{24}$$

from which (19) and (21) follow immediately. □

4 Decompositions of $R_X(\lambda)$ when $K = 2$

In this section we focus on the case in which $K = 2$ and derive various decompositions of the ridge operator $R_X(\lambda)$ analogous to those of the orthogonal projector P_X ([11], Lemma 3).

Theorem 5. Let $X = [X_1, X_2]$, where X_1 and X_2 are assumed disjoint except in (ii) below. Let $R_{X_1}(\lambda)$ and $R_{X_2}(\lambda)$ be as defined in (17). Then, the following statements hold.

(i) $R_X(\lambda) = R_{X_1}(\lambda) + R_{X_2}(\lambda)$ if and only if $X_1'M(\lambda)X_2 = X_1'X_2 = 0$.

(ii) $R_X(\lambda) = R_{X_1}(\lambda) + R_{X_2}(\lambda) - R_{X_1}(\lambda)M(\lambda)R_{X_2}(\lambda)$ if and only if $R_{X_1}(\lambda)M(\lambda)R_{X_2}(\lambda) = R_{X_2}(\lambda)M(\lambda)R_{X_1}(\lambda)$.

(iii) $R_X(\lambda) = R_{X_1}(\lambda) + R_{S_{X_1}(\lambda)X_2}(\lambda) = R_{X_2}(\lambda) + R_{S_{X_2}(\lambda)X_1}(\lambda)$, where $R_{S_{X_1}(\lambda)X_2}(\lambda) = S_{X_1}(\lambda)X_2(X_2'S_{X_1}(\lambda)X_2 + \lambda R_{X_2'})^- X_2'S_{X_1}(\lambda)$ and $R_{S_{X_2}(\lambda)X_1}(\lambda) = S_{X_2}(\lambda)X_1(X_1'S_{X_2}(\lambda)X_1 + \lambda P_{X_1'})^- X_1'S_{X_2}(\lambda)$. Note that $X_1'S_{X_2}(\lambda)M(\lambda)S_{X_2}(\lambda)X_1 = X_1'S_{X_2}(\lambda)X_1 + \lambda P_{X_1'}$ and $X_2'S_{X_1}(\lambda)M(\lambda)S_{X_1}(\lambda)X_2 = X_2'S_{X_1}(\lambda)X_2 + \lambda P_{X_2'}$.

(iv) $R_X(\lambda) = R_{X_1/S_{X_2}(\lambda)}(\lambda) + R_{X_2/S_{X_1}(\lambda)}(\lambda)$, where $R_{X_1/S_{X_2}(\lambda)}(\lambda) = X_1(X_1'S_{X_2}(\lambda)X_1 + \lambda P_{X_1'})^- X_1'S_{X_2}(\lambda)$ and $R_{X_2/S_{X_1}(\lambda)}(\lambda) = X_2(X_2'S_{X_1}(\lambda)X_2 + \lambda P_{X_2'})^- X_2'S_{X_1}(\lambda)$.

(v) $R_X(\lambda) = R_{XH}(\lambda) + R_{XG}(\lambda)$, where G is such that $\mathrm{Sp}(G) = \mathrm{Ker}(H'X'M(\lambda)X)$.

Proof. We prove the theorem in the order of (i), (iii), (v), (iv) and (ii).

(i) It can easily be verified that $R_X(\lambda)M(\lambda)$, $R_{X_1}(\lambda)M(\lambda)$, and $R_{X_2}(\lambda)M(\lambda)$ are orthogonal projectors, and that $\mathrm{Sp}(R_X(\lambda)M(\lambda)) = \mathrm{Sp}(X) = \mathrm{Sp}([X_1, X_2]) = \mathrm{Sp}(R_{X_1}(\lambda)M(\lambda)) \oplus \mathrm{Sp}(R_{X_2}(\lambda)M(\lambda))$. Further, X_1 and X_2 are $M(\lambda)$-orthogonal. We thus have $R_X(\lambda)M(\lambda) = R_{X_1}(\lambda)M(\lambda) + R_{X_2}(\lambda)M(\lambda)$ (by Lemma 4(i) in Takane and Yanai [11], for example). By postmultiplying this equation by $M(\lambda)^+ = R_X(\lambda)$, and taking into account Theorem 1(v) and (vii), we obtain (i). Conversely, $R_X(\lambda) = R_{X_1}(\lambda) + R_{X_2}(\lambda)$, and that $R_X(\lambda)M(\lambda)$ and $R_{X_k}(\lambda)M(\lambda)$ ($k = 1, 2$) are projectors imply $R_{X_1}(\lambda)R_{X_2}(\lambda) = -R_{X_2}(\lambda)R_{X_1}(\lambda)$. By premultiplying both sides of this equation by $R_{X_1}(\lambda)M(\lambda)$ we obtain $R_{X_1}(\lambda)R_{X_2}(\lambda) = -R_{X_1}(\lambda)R_{X_2}(\lambda)R_{X_1}(\lambda)$, and by postmultiplying them by $M(\lambda)R_{X_1}(\lambda)$ we obtain $R_{X_1}(\lambda)R_{X_2}(\lambda)R_{X_1}(\lambda) = -R_{X_2}(\lambda)R_{X_1}(\lambda)$, which together imply $R_{X_1}(\lambda)R_{X_2}(\lambda) = R_{X_2}(\lambda)R_{X_1}(\lambda)$. Since $R_{X_1}(\lambda)R_{X_2}(\lambda) = -R_{X_2}(\lambda)R_{X_1}(\lambda)$ also holds, this implies $R_{X_1}(\lambda)R_{X_2}(\lambda) = R_{X_2}(\lambda)R_{X_1}(\lambda) = 0$, which in turn implies $X_1'X_2 = 0$. This decomposition is analogous to $P_X = P_{X_1} + P_{X_2}$ if and only if $X_1'X_2 = 0$.

(iii) It can easily be verified that X_1 and $S_{X_1}(\lambda)X_2$ are $M(\lambda)$-orthogonal, i.e., $X_1'M(\lambda)S_{X_1}(\lambda)X_2 = 0$, and so are X_2 and $S_{X_2}(\lambda)X_1$. (The former may be seen from $X_1'M(\lambda)S_{X_1}(\lambda)X_2 = X_1'M(\lambda)X_2 - X_1'M(\lambda)X_1(X_1'M(\lambda)X_1)^- X_1'X_2 = X_1'X_2 - X_1'X_2 = 0$. The latter can also be shown in essentially the same way.) We show $\mathrm{Sp}(X) = \mathrm{Sp}([X_1, S_{X_1}(\lambda)X_2]) = \mathrm{Sp}([S_{X_2}(\lambda)X_1, X_2])$. Then, this case reduces to (i). To show the first equality, we note that

$$[X_1, S_{X_1}(\lambda)X_2] = [X_1, X_2] \begin{bmatrix} I_{p_1} & -(X_1'M(\lambda)X_1)^- X_1'X_2 \\ 0 & I_{p_2} \end{bmatrix},$$

where the second matrix on the right hand side is nonsingular. The second equality can be similarly proven. This decomposition is analogous to $P_X = P_{X_1} + P_{Q_{X_2}X_1} = P_{X_2} + P_{Q_{X_1}X_2}$, where $P_{Q_{X_2}X_1}$ and $P_{Q_{X_1}X_2}$ are the orthogonal projectors onto $\mathrm{Sp}(Q_{X_2}X_1)$ and $\mathrm{Sp}(Q_{X_1}X_2)$, respectively. This decomposition is useful when we fit one of X_1 and X_2 first, and then fit the other to the residual.

(v) It can readily be seen that XH and XG are $M(\lambda)$-orthogonal, and that $\mathrm{Sp}(X) = \mathrm{Sp}(XH) \oplus \mathrm{Sp}(XG)$. By setting $X_1 = XH$ and $X_2 = XG$, this case reduces to (i). This decomposition is analogous to $P_X = P_{XH}+P_{XG}$, where G is such that $\mathrm{Sp}(G) = \mathrm{Ker}(H'X'X)$. In [11], G is parameterized as C, where $C = X'M(\lambda)XG$, and hence $\mathrm{Sp}(C) \subset \mathrm{Sp}(X'M(\lambda)X)$. The decomposition in (v) is then written as $R_X(\lambda) = R_{XH}(\lambda) + R_{X(X'M(\lambda)X)^-C}(\lambda)$. This form is often more convenient, since two forms of constraints on the vector b of regression coefficients, $b = Hb^*$ for some b^* and $C'b = 0$, are equivalent when $\mathrm{Ker}(C') = \mathrm{Sp}(H)$.

(iv) is obtained by a direct expansion of $R_X(\lambda) = X(X'M(\lambda)X)^-X'$. Note that $(X'M(\lambda)X)^-$ has the following expression, since $X'M(\lambda)X$ is *nnd*:

$$(X'M(\lambda)X)^- = \begin{bmatrix} T_1^- & -T_1^- X_1'X_2D_2(\lambda)^- \\ -T_2^- X_2'X_1D_1(\lambda)^- & T_2^- \end{bmatrix},$$

where $T_1 = D_1(\lambda) - X_1'R_{X_2}(\lambda)X_1$, and $T_2 = D_2(\lambda) - X_2'R_{X_1}(\lambda)X_2$. Note further that $-T_2^- X_2'X_1D_1(\lambda)^-$ and $-T_1^- X_1'X_2D_2(\lambda)^-$ can be made equal to the transpose of each other, and T_2^- can be further expanded into $T_2^- = D_2(\lambda)^- + D_2(\lambda)^- X_2'X_1T_1^- X_1'X_2D_2(\lambda)^-$. ($T_1^-$ can be similarly expanded.) This decomposition is analogous to $P_X = P_{X_1/Q_{X_2}} + P_{X_2/Q_{X_1}}$, where $P_{X_1/Q_{X_2}} = X_1(X_1'Q_{X_2}X_1)^-X_1'Q_{X_2}$ and $P_{X_2/Q_{X_1}} = X_2(X_2'Q_{X_1}X_2)^-X_2'Q_{X_1}$. The decomposition is useful when we fit both X_1 and X_2 simultaneously. Note that $X_1'S_{X_2}(\lambda)M(\lambda)X_2 = 0$ and $X_2'S_{X_1}(\lambda)M(\lambda)X_1 = 0$. Note also that $X_1'(S_{X_1}(\lambda)M(\lambda)S_{X_1}(\lambda)+ S_{X_2}(\lambda)M(\lambda)S_{X_2}(\lambda))X_2 = 0$, so that the two terms in this decomposition are orthogonal with respect to $S_{X_1}(\lambda)M(\lambda)S_{X_1}(\lambda) + S_{X_2}(\lambda)M(\lambda)S_{X_2}(\lambda)$.

(ii) *Sufficiency.* When X_1 and X_2 are not disjoint, $\mathrm{Sp}(X)$ is partitioned into three disjoint subspaces, the subspace unique to X_1, the subspace unique to X_2, and the subspace commonly shared by both X_1 and X_2. Let $R_{11}(\lambda)$, $R_{22}(\lambda)$, and $R_{12}(\lambda)$ denote the ridge operators on these three subspaces. Then, $R_{X_1}(\lambda) = R_{11}(\lambda) + R_{12}(\lambda)$, where the two terms on the right hand side are $M(\lambda)$-orthogonal. Similarly, $R_{X_2}(\lambda) = R_{22}(\lambda)+R_{12}(\lambda)$, where the two terms on the right hand side are $M(\lambda)$-orthogonal. By expanding $R_{X_1}(\lambda)M(\lambda)R_{X_2}(\lambda)$ and $R_{X_2}(\lambda)M(\lambda)R_{X_1}(\lambda)$, which are by assumption equal, we obtain $R_{11}(\lambda)M(\lambda)R_{22}(\lambda) = R_{11}(\lambda)R_{22}(\lambda) = R_{22}(\lambda)R_{11}(\lambda) = R_{22}(\lambda)M(\lambda)R_{11}(\lambda)$, which is further equal to 0, since $R_{11}(\lambda)$ and $R_{22}(\lambda)$ are disjoint. This implies $R_{12}(\lambda) = R_{X_1}(\lambda)M(\lambda)R_{X_2}(\lambda) = R_{X_2}(\lambda)M(\lambda)R_{X_1}(\lambda)$. Thus, $R_X(\lambda) = R_{11}(\lambda) + R_{22}(\lambda) + R_{12}(\lambda) = (R_{11}(\lambda) + R_{12}(\lambda)) + (R_{22}(\lambda) + R_{12}(\lambda)) - R_{12}(\lambda) = R_{X_1}(\lambda) + R_{X_2}(\lambda) - R_{X_1}(\lambda)M(\lambda)R_{X_2}(\lambda)$.

Necessity. $R_X(\lambda)$, $R_{X_1}(\lambda)$, and $R_{X_2}(\lambda)$ are all symmetric, and so is $R_{X_1}(\lambda)M(\lambda)R_{X_2}(\lambda)$, which immediately implies $R_{X_1}(\lambda)$ and $R_{X_2}(\lambda)$ are $M(\lambda)$-commutative. Note that $M(\lambda)$-commutativity is equivalent to $R_{X_1}(\lambda)M(\lambda)R_{X_2}(\lambda)M(\lambda)$ being idempotent. It is also equivalent to
$R_{X_2}(\lambda)M(\lambda)R_{X_1}(\lambda)M(\lambda)$ being idempotent. This decomposition is analogous to $P_X = P_{X_1} + P_{X_2} - P_{X_1}P_{X_2}$ if and only if $P_{X_1}P_{X_2} = P_{X_2}P_{X_1}$ [9], and is useful in two-way ANOVA without interactions. □

The following two decompositions are obtained by combining two decompositions ((iii) and (v)) in Theorem 5.

Corollary 3. Let X be as in Theorem 5, and let H be a p_1 by t ($\leq p_1$) matrix.

(1) Further, let A, B, and W be such that (a) $\mathrm{Sp}(A) = \mathrm{Ker}(H'X_1'R_{X_2}(\lambda)X_1)$, (b) $\mathrm{Sp}(B) = \mathrm{Ker}(H'X_1'S_{X_2}(\lambda)M(\lambda)X_1)$, and (c) $\mathrm{Sp}(W) = \mathrm{Ker}(X_1'X_2)$. Then,

$$\begin{aligned} R_X(\lambda) = R_{R_{X_2}(\lambda)X_1H}(\lambda) + \\ R_{R_{X_2}(\lambda)X_1A}(\lambda) + R_{S_{X_2}(\lambda)X_1H}(\lambda) + R_{S_{X_2}(\lambda)X_1B}(\lambda) + R_{X_2W}(\lambda). \end{aligned} \tag{25}$$

The five terms in the decomposition are all $M(\lambda)$-orthogonal, and $M(\lambda)$-projectors. ($R_X(\lambda)$ is said to be an $M(\lambda)$-projector when $R_X(\lambda)M(\lambda)$ and $M(\lambda)R_X(\lambda)$ are projectors.)

(2) Let G, H, U, and V be such that (a) $\mathrm{Sp}(G) = \mathrm{Ker}(H'X_1'M(\lambda)X_1)$, (b) $\mathrm{Sp}(U) = \mathrm{Ker}(X_2'X_1H)$, and (c) $\mathrm{Sp}(V) = \mathrm{Ker}(X_2'X_1G)$. Then,

$$R_X(\lambda) = R_{R_{X_1H}(\lambda)X_2}(\lambda) + R_{X_1HU}(\lambda) + R_{R_{X_1G}(\lambda)X_2}(\lambda) + R_{X_1GV}(\lambda) + R_{S_{X_1}(\lambda)X_2}(\lambda). \tag{26}$$

The five terms in the decomposition are all $M(\lambda)$-orthogonal, and $M(\lambda)$-projectors.

Proof. (1) $M(\lambda)$-orthogonalities among the terms in the decomposition can be shown by tracing the derivation of these terms. $R_X(\lambda)$ is first split into $R_{X_2}(\lambda)$ and $R_{S_{X_2}(\lambda)X_1}(\lambda)$ by (iii) of Theorem 5. (References to Greek numbers such as (iii) are to those in Theorem 5). Then, $R_{X_2}(\lambda)$ is split into $R_{R_{X_2}(\lambda)X_1}(\lambda)$ and $R_{X_2}(\lambda) - R_{R_{X_2}(\lambda)X_1}(\lambda) = R_{X_2W}(\lambda)$ using (1c) and (v). (References to Arabic numerals like (1a) are to those in Corollary 3.) Finally, $R_{R_{X_2}(\lambda)X_1}(\lambda)$ is split into $R_{R_{X_2}(\lambda)X_1H}(\lambda)$ and $R_{R_{X_2}(\lambda)X_1}(\lambda) - R_{R_{X_2}(\lambda)X_1H}(\lambda) = R_{R_{X_2}(\lambda)X_1A}(\lambda)$ using (1a) and (v), and $R_{S_{X_2}(\lambda)X_1}(\lambda)$ is split into $R_{S_{X_2}(\lambda)X_1H}(\lambda)$ and $R_{S_{X_2}(\lambda)X_1}(\lambda) - R_{S_{X_2}(\lambda)X_1H}(\lambda) = R_{S_{X_2}(\lambda)X_1B}(\lambda)$ using (1b) and (v). All of these decompositions are $M(\lambda)$-orthogonal, so that the resultant terms are all mutually $M(\lambda)$-orthogonal. That they all become projectors when postmultiplied by $M(\lambda)$ can be directly verified.

(2) Again, $M(\lambda)$-orthogonalities of the five terms in the decompostion can be readily shown by tracing the decomposition. $R_X(\lambda)$ is first split into $R_{X_1}(\lambda)$ and $R_{S_{X_1}(\lambda)X_2}(\lambda)$ by the second half of (iii). Then, $R_{X_1}(\lambda)$ is split into $R_{X_1H}(\lambda)$ and $R_{X_1G}(\lambda)$ by (2a) and (v). Then, X_2 is projected onto both $R_{X_1H}(\lambda)$ and $R_{X_1G}(\lambda)$. The former splits $R_{X_1H}(\lambda)$ into $R_{R_{X_1H}(\lambda)X_2}(\lambda)$ and $R_{X_1H}(\lambda) - R_{R_{X_1H}(\lambda)X_2}(\lambda) = R_{X_1HU}(\lambda)$ by (2b) and (v), and the latter $R_{X_1G}(\lambda)$ into $R_{R_{X_1G}(\lambda)X_2}(\lambda)$ and $R_{X_1G}(\lambda) - R_{R_{X_1G}(\lambda)X_2}(\lambda) = R_{X_1GV}(\lambda)$ by (2c) and (v).

Again, all these decompositions are $M(\lambda)$-orthogonal, and consequently the resultant terms are all mutually $M(\lambda)$-orthogonal. $\square$

The decompositions in the above corollary are analogous to those derived by Takane, Yanai, and Hwang [12] for P and were used to construct a variety of constrained/partial canonical correlation analyses (CANO). With the above decompositions it is now possible to develop a variety of ridge constrained/partial CANO methods.

5 Applications

Application 1. Let Y $(n \times p)$ and X $(n \times q)$ denote matrices of criterion variables and predictor variables, respectively, in a multivariate regression model,

$$Y = XB + E, \tag{27}$$

where B is the matrix of regression coefficients and E is the matrix of disturbance terms. In reduced rank regression analysis [1], B is subject to the rank restriction of the form, $\text{rank}(B) \leq \min(p, q)$. In the reduced rank ridge regression analysis, an estimate of B is obtained by minimizing the RLS criterion,

$$\phi_\lambda(B) = ||Y - XB||^2 + \lambda ||B||^2_{P_{X'}}, \tag{28}$$

where $||Y - XB||^2 = \text{tr}(Y - XB)'(Y - XB)$ (the Frobenius norm), and $||B||^2_{P_{X'}} = \text{tr}(B' P_{X'} B) = \text{tr}(B'B)$. Let $\hat{B}(\lambda) = (X'M(\lambda)X)^- X'Y$ be a ridge LS estimate of B without rank restriction. Then, using Theorem 1(iii), we can resplit $\phi_\lambda(B)$ into:

$$\phi_\lambda(B) = ||Y||^2_{S_X(\lambda)} + ||\hat{B}(\lambda) - B||^2_{X'M(\lambda)X}. \tag{29}$$

Since the first term on the right hand side is unrelated to B, $\phi_\lambda(B)$ can be minimized by minimizing the second term. This can be achieved by the generalized singular value decomposition (GSVD; e.g., [10]) of B with the metrics $X'M(\lambda)X$ and I_p, which is denoted as $\text{GSVD}(\hat{B}(\lambda))_{X'M(\lambda)X, I_p}$.

Let

$$\tilde{Y} = \begin{bmatrix} Y \\ 0 \end{bmatrix}, \quad \text{and} \quad \tilde{X} = \begin{bmatrix} X \\ \lambda^{1/2} P_{X'} \end{bmatrix}. \tag{30}$$

Then, the RLS criterion (28) can be rewritten entirely in the form of a LS criterion:

$$\phi_\lambda(B) = ||\tilde{Y} - \tilde{X}B||^2. \tag{31}$$

Minimizing this criterion with respect to B leads to the orthogonal projector onto $\text{Sp}(\tilde{X})$, partitioned as follows:

$$P_{\tilde{X}} = \begin{bmatrix} R_X(\lambda) & A \\ A' & C \end{bmatrix}, \tag{32}$$

where $A = \lambda^{1/2}X(X'M(\lambda)X)^+$, and $C = \lambda(X'M(\lambda)X)^+$. From the idempotency of a projector, it follows that

$$R_X(\lambda) - R_X(\lambda)^2 = AA' \geq 0, \tag{33}$$

$$C - C^2 = A'A \geq 0, \tag{34}$$

and

$$R_X(\lambda)A + AC = A. \tag{35}$$

(33) is identical to Theorem 1(iii), and (34) shows C is also a contraction matrix.

Application 2. Anderson [1] proposed an extended reduced rank regression model with X divided into two subsets, X_1 and X_2, (and B into B_1 and B_2), where the rank restriction is imposed only on B_1. The model may be written as

$$Y = X_1B_1 + X_2B_2 + E, \tag{36}$$

where X_1 may be considered as predictor variables with reduced rank coefficients B_1 while X_2 with full rank coefficients B_2 may be viewed as extraneous variables (or covariates) whose effects are to be eliminated. This model may be called partial reduced rank regression model. In accordance with the decomposition in Theorem 5(iii), (36) may be rewritten as

$$Y = S_{X_2}(\lambda)X_1B_1 + X_2B_2^* + E, \tag{37}$$

where $B_2^* = B_2 + (X_2'M(\lambda)X_2)^- X_2'X_1B_1$. Observe that the first two terms on the right hand side of (37) are $M(\lambda)$-orthogonal. Let $\hat{B}_1(\lambda) = (X_1'S_{X_2}(\lambda)M(\lambda)S_{X_2}(\lambda)X_1)^- X_1'S_{X_2}(\lambda)Y$ is an RLS estimate of B_1 without rank restriction, and $\hat{B}_2^*(\lambda) = (X_2'M(\lambda)X_2)^- X_2'Y$ is an RLS estimate of B_2^*. Then,

$$||\hat{B}_1(\lambda) - B_1||^2_{X_1'S_{X_2}(\lambda)M(\lambda)S_{X_2}(\lambda)X_1} + ||\hat{B}_2^*(\lambda) - B_2^*||^2_{X_2'M(\lambda)X_2} = ||\hat{B}(\lambda) - B||^2_{X'M(\lambda)X}. \tag{38}$$

The second term on the left hand side of (38) can always be made equal to zero by setting $B_2 = (X_2'M(\lambda)X_2)^- X_2'(Y - X_1B_1)$, so that an RLS estimate of B_1 with rank restriction can be obtained by minimizing the first term, which is achieved by $\text{GSVD}(\hat{B}_1(\lambda))_{X_1'S_{X_2}(\lambda)M(\lambda)S_{X_2}(\lambda)X_1, I_p}$.

Application 3. Suppose the constraint, $C'B = 0$, is imposed on B as well as the rank restriction in the reduced rank regression model. As noted in the proof of Theorem 5(v), this constraint can be reparameterized as $B = HB^*$ for some B^*, where H is such that $\text{Sp}(H) = \text{Ker}(C')$. An RLS estimate of B under this constraint without rank restriction is obtained by $\hat{B}_c(\lambda) = H'(HX'M(\lambda)XH)^- H'X'Y$. Then, the RLS criterion $\phi_\lambda(B)$ can be split into two parts:

$$\phi_\lambda(B) = ||Y||^2_{R_{XH}(\lambda)} + ||\hat{B}_c(\lambda) - B||^2_{X'M(\lambda)X}, \tag{39}$$

where the first term on the right can be further split into:

$$||Y||^2_{R_{XH}(\lambda)} = ||Y||^2_{S_X(\lambda)} + ||Y||^2_{R_{X(X'M(\lambda)X)^-C}(\lambda)}, \tag{40}$$

where $R_{X(X'M(\lambda)X)^-C}(\lambda) = R_X(\lambda) - R_{XH}(\lambda)$ according to Theorem 5(v). Since the first term on the right hand side of (39) has nothing to do with B, $\phi_\lambda(B)$ can be minimized by minimizing the second term, which is achieved by $\text{GSVD}(\hat{B}_c(\lambda))_{X'M(\lambda)X,I_p}$. As in Application 2, the rank restriction can be imposed on only parts of B.

Groß[4] considered a ridge type of estimation under a slightly more general form of the constraint $C'B = Z$. Let H be as introduced earlier. Then, B is reparameterized as $B = HB^* + B_0$, where $B_0 = C(C'C)^-Z$, and a constrained ridge estimate of B without rank restriction is obtained by $\hat{B}_c(\lambda) = F(\lambda)X'Y - F(\lambda)X'XB_0 + B_0$, where $F(\lambda) = H(H'X'M(\lambda)XH)^-H'$. A corresponding reduced rank estimate is obtained by $\text{GSVD}(\hat{B}_c(\lambda))_{X'M(\lambda)X,I_p}$. (Note, however, that this estimate does not satisfy the original constraint, $C'B = Z$. We may apply the GSVD to only the first two terms of $\hat{B}_c(\lambda)$ to obtain an estimate that satisfies this constraint.)

Application 4. Let X_k $(k = 1, \cdots, K)$ and X be as defined in Theorem 3. In multiple-set canonical correlation analysis (GCANO) data matrices X_k's are either columnwise standardized or centered. Let W denote a matrix of weights (applied to X to derive canonical variates) partitioned conformably to the partition of X. In ridge estimation of GCANO, we obtain the weight matrix that maximizes

$$\psi_\lambda(W) = \text{tr}(W'X'M(\lambda)XW) \tag{41}$$

subject to the ortho-normalization restriction that $W'D(\lambda)W = I_d$, where $D(\lambda)$ is a block diagonal matrix with $D_k(\lambda) = X_k'M(\lambda)X_k$ (as defined in (14)) as the k^{th} diagonal block, and d is the dimensionality of the solution. This leads to the generalized singular value decomposition of the form $\text{GSVD}(XD(\lambda)^-)_{M(\lambda),D(\lambda)}$. In this decomposition we obtain a matrix of left singular vectors F^* such that $F^{*\prime}M(\lambda)F^* = I_r$ (where r is the rank of X), a matrix of right generalized singular vectors W^* such that $W^{*\prime}D(\lambda)W^* = I_r$, and an r by r positive-definite diagonal matrix of generalized singular values Δ^* such that $XD^-(\lambda) = F^*\Delta^*W^{*\prime}$. Matrix W is obtained by retaining only the portions of W^* corresponding to the d largest generalized singular values (assuming $d \leq r$). Matrix of canonical scores F is likewise obtained by retaining the portions of F^* pertaining to the largest d singular values. Essentially equivalent results can also be obtained by the generalized eigen-decomposition of $X'M(\lambda)X$ with respect to $D(\lambda)$, or that of $XD(\lambda)^-X' = \sum_{k=1}^K R_{X_k}(\lambda)$ with respect to $M(\lambda)$, where $R_{X_k}(\lambda)$ is as defined in (17). Note that $XD(\lambda)^-X'$ is invariant over the choice of $D(\lambda)^-$ because $\text{Sp}(X') \subset \text{Sp}(D(\lambda))$ ([8], Lemma 2.2.4(iii)), and that $D(\lambda)^-_* \in \{D(\lambda)^-\}$, where $D(\lambda)^-_*$ is a block diagonal matrix with $D_k(\lambda)^-$ as the k-th diagonal block.

Just as GCANO reduces to the usual two-set canonical correlation analysis when $K = 2$, the ridge GCANO reduces to the two-set ridge canonical correlation analysis [14] when

$K = 2$.

Acknowledgement

The work reported in this paper has been supported by grant A6394 from the Natural Sciences and Engineering Research Council of Canada to the first author.

References

[1] T. W. Anderson, Estimating linear restrictions on regression coefficients for multivariate normal distribution, Annals of Mathematical Statistics 22 (1951) 327-351.

[2] T. W. Anderson, G. P. H. Styan, Cochran's theorem, rank additivity and tripotent matrices, in G. Kallianpur, P. R. Krishnaiah, J. K. Ghosh (Eds.), Statistics and probability: Essays in honor of C. R. Rao, North Holland, Amsterdam, 1982, pp. 1-23.

[3] J. Groß, Linear regression, Springer, Berlin, 2003a.

[4] J. Groß, Restricted ridge estimation, Statistics and Probability Letters 65 (2003b) 57-64.

[5] M. Gulliksson, P. Wedin, The use and properties of Tikhonov filter matrices, SIAM Journal on Matrix Analysis and Applications 22 (2000) 276-281.

[6] A. F. Hoerl, R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, Technometrics 12 (1970) 55-67.

[7] S. K. Mitra, Optimal inverse of a matrix, Sankhyā, Series A 37 (1975) 550-563.

[8] C. R. Rao, S. K. Mitra, Generalized inverse of matrices and its applications, Wiley, New York, 1971.

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

[10] Y. Takane, M. A. Hunter, Consrained principal component analysis: A comprehensive theory, Applicable Algebra in Engineering, Communication and Computing 12 (2001) 391-419.

[11] Y. Takane, H. Yanai, On oblique projectors, Linear Algebra and Its Applications 289 (1999) 297-310.

[12] Y. Takane, H. Yanai, H. Hwang, An improved method for generalized constrained canonical correlation analysis. Computational Statistics and Data Analysis 50 (2006) 221-241.

[13] A. N. Tikhonov, V. Y. Arsenin, Solutions of ill-posed problems, Winston, Washington, D. C., 1977.

[14] H. D. Vinod, Canonical ridge and econometrics of joint production, Journal of Econometrics 4 (1976) 147-166.

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

[16] H. Yanai, S. Mayekawa, Some extensions of inequalities concerning diagonal elements of orthogonal projectors and conditions for equalities, Japanese Journal of Applied Statistics 17 (1988) 131-138 (in Japanese).