Journal of Korean
Data & Information Science Society
2003, Vol. 14, No.2 pp. 355~366

Unified Non-iterative Algorithm for Principal Component Regression, Partial Least Squares and Ordinary Least Squares

Jong-Duk Kim[1)]

Abstract

A unified procedure for principal component regression (PCR), partial least squares (PLS) and ordinary least squares (OLS) is proposed. The process gives solutions for PCR, PLS and OLS in a unified and non-iterative way. This enables us to see the interrelationships among the three regression coefficient vectors, and it is seen that the so-called E-matrix in the solution expression plays the key role in differentiating the methods. In addition to setting out the procedure, the paper also supplies a robust numerical algorithm for its implementation, which is used to show how the procedure performs on a real world data set.

Keywords: Krylov matrix; Partial least squares; Principal component regression; Unified algorithm

1. Introduction

Regression analysis is one of the most common ways for finding linear relationships between measurables and variables. We consider the general linear regression model

$$\boldsymbol{y} = \beta_0 \mathbf{1} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$ is a $n\times 1$ vector of the dependent variable, β_0 and $\boldsymbol{\beta}' = (\beta_1, \cdots, \beta_p)$ are the unknown parameters of the model, $\mathbf{1}$ is a $n\times 1$ vector of ones, $\boldsymbol{X}$ is the $n\times p$

1) Professor, Department of Statistics, Pusan University of Foreign Studies, Busan, 608-738, Korea
E-mail: jdkim@pufs.ac.kr

matrix of the independent variables, and $\boldsymbol{\varepsilon}$ is a $n\times 1$ error vector normally distributed with mean zero and covariance matrix $\sigma^2 \boldsymbol{I}_p$. We shall also assume that the rank of $\boldsymbol{X}$ is r.

A serious problem arises with ordinary least squares (OLS) when the independent variables that comprises $\boldsymbol{X}$ are not independent but collinear. In such cases the model parameters are more sensitive to noise, causing a loss of full rank. Principal component regression (PCR) (Massy 1965) and partial least squares (PLS) (Wold et. al. 1984) circumvent the collinearity problem by using orthogonal latent variables.

One of the problems of PLS is that it is always given as an iterative algorithm in the literature and hence it is difficult to understand its structure and properties. Some efforts have been made to figure out its relationship with PCR and OLS (see, for example, Stone and Brooks 1990, Frank and Friedman 1993, Lang et al. 1998, Kim 2003). In this paper, we present an algorithm that provides solutions for PCR, PLS and OLS in a unified and non-iterative way and examine the structure of the solutions.

2. Unified Approach to PCR, PLS and OLS

A unified approach to PCR, PLS and OLS, called Cyclic Subspace Regression (CSR), was given by Lang et al. (1998). A numerically robust version of this algorithm was given by Brenchley et al. (1998).

A brief outline of the CSR procedure can be written as follows.

1. Autoscale $\boldsymbol{X}$ and $\boldsymbol{y}$.
2. Singular value decomposition of $\boldsymbol{X}$; i.e., $\boldsymbol{X}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}'$
3. Let k and l be fixed integers satisfying $1\le k\le l\le r$.
4. $\boldsymbol{P}_l=\boldsymbol{U}\boldsymbol{U}'=(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_l)(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_l)'$
5. $\boldsymbol{X}_l^1=\boldsymbol{X}$ and $\hat{\boldsymbol{y}}_l^1=\boldsymbol{P}_l\boldsymbol{y}$
6. For $i=1,\cdots,k$:

   $\boldsymbol{w}_l^i=\boldsymbol{X}_l^{i\prime}\hat{\boldsymbol{y}}_l^i/\|\boldsymbol{X}_l^{i\prime}\hat{\boldsymbol{y}}_l^i\|$

   $\boldsymbol{t}_l^i=\boldsymbol{X}_l^{i\prime}\boldsymbol{w}_l^i/\|\boldsymbol{X}_l^{i\prime}\boldsymbol{w}_l^i\|$

   $\boldsymbol{X}_l^{i+1}=\boldsymbol{X}_l^i-\boldsymbol{t}_l^i\boldsymbol{t}_l^{i\prime}\boldsymbol{X}_l^i$

   $\hat{\boldsymbol{y}}_l^{i+1}=\hat{\boldsymbol{y}}_l^i-\boldsymbol{t}_l^i\boldsymbol{t}_l^{i\prime}\hat{\boldsymbol{y}}_l^i$
7. Create matrices $\boldsymbol{W}_l^k=(\boldsymbol{w}_l^1,\cdots,\boldsymbol{w}_l^k)$, $\boldsymbol{T}_l^k=(\boldsymbol{t}_l^1,\cdots,\boldsymbol{t}_l^k)$ and

$$X_l^k = T_l^k T_l^{k\prime} X W_l^k W_l^{k\prime}.$$

8. $b_l^k = (X_l^k)^+ y$

Solutions obtained in step 8 correspond to PCR when $l = k$, PLS when $l = r$, and OLS when $l = k = r$. According to Kalivas (1999) the regression vectors for CSR can be expressed as

$$b = \sum_{i=1}^{l} \left(\frac{u_i' \hat{y}_l^k}{\delta_i} \right) v_i,$$

where $\hat{y}_l^k$ are the values calculated from step 6, based on l eigenvectors and k components. The CSR algorithm provides an extra loop in the Wold's PLS algorithm, modifying X and y, so the number of eigenvectors can be varied.
Although the above algorithm is a single method that incorporates all of the features of PCR, PLS and OLS, it is not easy to understand how the solutions for PCR, PLS and OLS are obtained.

We will now show another unified approach to PCR, PLS and OLS. Like CSR, this is a unified procedure which reveals the relationships and differences among the three solutions, though in a different way. Before explaining this we provide the following definition and proposition.

Definition 1. (Krylov matrix) Given matrices $X \in \mathbb{R}^{n \times p}$ and $y \in \mathbb{R}^n$, the Krylov matrix $K_p \in \mathbb{R}^{p \times p}$ for the pair $(X'X, X'y)$ is defined by

$$K_p = [X'y,\ X'XX'y,\ (X'X)^2X'y,\ \cdots,\ (X'X)^{p-1}X'y]$$

(see Golub and Van Loan 1996).

Proposition 1. The column space of the weighting matrix $W_k \in \mathbb{R}^{p \times k}$ obtained from Wold's PLS algorithm and the column space of the reduced Krylov matrix $K_k \in \mathbb{R}^{p \times k}$ coincide. That is, the columns of W_k and the columns of K_k span the same subspace.

Proof: The proof can be found in Helland (1988). ■

An alternative expression of the above reduced Krylov matrix K_k is

$$K_k = (VDU'y,\ VD^3U'y,\ \cdots,\ VD^{2k-1}U'y), \quad (1)$$

which can be easily verified by using $X = UDV'$, the singular value decomposition (SVD) of X. (Matrices U, D and V will be defined soon.)

The expression in proposition 2 below gives a unified and non-iterative solution of the regression vectors for OLS, PCR and PLS.

Proposition 2. Let the SVD of $X \in \mathbb{R}^{n \times p}$ be written as $X = UDV'$, where $U \in \mathbb{R}^{n \times r}$ represents the matrix of eigenvectors for XX', $V \in \mathbb{R}^{p \times r}$ symbolizes the matrix of basis eigenvectors for $X'X$, $D \in \mathbb{R}^{r \times r}$ denotes the diagonal matrix of singular values (δ_i), and r is the rank of X. Then the regression vectors for OLS, PCR and PLS are given by

$$b = V\phi, \quad \phi = D^{-1}EU'y,$$

where matrix $E \in \mathbb{R}^{r \times r}$ is

$E = I_r$ for OLS;

$E = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$ for PCR;

$E = Q_k(Q_k'Q_k)^{-1}Q_k'$ with $Q_k = (D^2U'y, D^4U'y, \cdots, D^{2k}U'y)$ for PLS.

Here I_a denotes an identity matrix of size $a \times a$ and k is the number of selected components for PCR and PLS.

Proof: It can be shown that

$$b_{OLS} = VD^{-1}U'y$$
$$b_{PCR} = V_kD_k^{-1}U_k'y$$
$$b_{PLS} = VD^{-1}E_kU'y$$

where $V = (v_1, v_2, \cdots, v_r)$, $D = \mathrm{diag}(\delta_1, \delta_2, \cdots, \delta_r)$ and $U = (u_1, u_2, \cdots, u_r)$; $V_k = (v_1, v_2, \cdots, v_k)$, $D_k = \mathrm{diag}(\delta_1, \delta_2, \cdots, \delta_k)$ and $U_k = (u_1, u_2, \cdots, u_k)$; $E_k = Q_k(Q_k'Q_k)^{-1}Q_k'$ with $Q_k = (D^2U'y, D^4U'y, \cdots, D^{2k}U'y)$ (for details, see Kim 2003). Then the result follows immediately. ■

The main advantage of this expression is that the regression vector is expressed as a linear combination of the eigenvectors v_i and that we can see how the weights differ in OLS, PCR and PLS by comparing the vector ϕ. It also shows that the difference in ϕ for the three methods depends on the matrix E.

We note that $\boldsymbol{E} = \boldsymbol{Q}_k(\boldsymbol{Q}_k'\boldsymbol{Q}_k)^{-1}\boldsymbol{Q}_k'$ is a symmetric idempotent matrix and $\mathrm{tr}(\boldsymbol{E}) = k$. We also note that $\mathrm{tr}(\boldsymbol{E}) = k$ for PCR and $\mathrm{tr}(\boldsymbol{E}) = r$ for OLS, where tr denotes the trace of a given matrix. From now on we will call the $\boldsymbol{E}$ matrix $\boldsymbol{E}$-matrix.

The summarization of this procedure is as follows.

Step 1. Autoscale $\boldsymbol{X}$ and $\boldsymbol{y}$:

$$\boldsymbol{X} \leftarrow \boldsymbol{X}, \quad \boldsymbol{y} \leftarrow \boldsymbol{y}$$

Step 2. SVD of $\boldsymbol{X}$:

$$\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}' = \boldsymbol{X}$$

Step 3. Compute $\boldsymbol{E}$-matrix:

For OLS, set

$$\boldsymbol{E} = \boldsymbol{I}_r$$

For PCR with the first k components, set

$$\boldsymbol{E} = \begin{pmatrix} \boldsymbol{I}_k & 0 \\ 0 & 0 \end{pmatrix}$$

For PLS with the first k components, set

$$\boldsymbol{E} = \boldsymbol{Q}_k(\boldsymbol{Q}_k'\boldsymbol{Q}_k)^{-1}\boldsymbol{Q}_k' \quad \text{with} \quad \boldsymbol{Q}_k = (\boldsymbol{D}^2\boldsymbol{U}'\boldsymbol{y}, \boldsymbol{D}^4\boldsymbol{U}'\boldsymbol{y}, \cdots, \boldsymbol{D}^{2k}\boldsymbol{U}'\boldsymbol{y})$$

Step 4. Compute $\boldsymbol{\phi}$:

$$\boldsymbol{\phi} = \boldsymbol{D}^{-1}\boldsymbol{E}\boldsymbol{U}'\boldsymbol{y}$$

Step 5. Compute regression coefficient vector:

$$\boldsymbol{b} = \boldsymbol{V}\boldsymbol{\phi}$$

Thus once $\boldsymbol{\phi} = (\phi_1, \phi_2, \cdots, \phi_r)'$ is obtained for each method, the regression vector is written as

$$\begin{aligned} \boldsymbol{b} &= \boldsymbol{V}\boldsymbol{\phi} \\ &= \phi_1\boldsymbol{v}_1 + \phi_2\boldsymbol{v}_2 + \cdots + \phi_r\boldsymbol{v}_r \end{aligned}$$

which is a linear combination of eigenvectors $\boldsymbol{v}_i$ and each ϕ_i is a corresponding weight. Also the investigation on the values of $\boldsymbol{E}$ helps us to see the differences of $\boldsymbol{\phi}$ among the methods, since vector $\boldsymbol{\phi}$ depends on $\boldsymbol{E}$.

The algebraic expression of each regression vector can be written as follows.

$$\boldsymbol{\phi}_{OLS} = \boldsymbol{D}^{-1}\boldsymbol{U}'\boldsymbol{y} = \begin{pmatrix} \frac{1}{\delta_1}\boldsymbol{u}_1'\boldsymbol{y} \\ \vdots \\ \frac{1}{\delta_r}\boldsymbol{u}_r'\boldsymbol{y} \end{pmatrix}$$

$$\boldsymbol{\phi}_{PCR} = \boldsymbol{D}_k^{-1}\boldsymbol{U}_k'\boldsymbol{y} = \begin{pmatrix} \frac{1}{\delta_1}\boldsymbol{u}_1'\boldsymbol{y} \\ \vdots \\ \frac{1}{\delta_k}\boldsymbol{u}_k'\boldsymbol{y} \end{pmatrix}$$

$$\boldsymbol{\phi}_{PLS} = \boldsymbol{D}^{-1}\boldsymbol{E}_k\boldsymbol{U}'\boldsymbol{y} = \begin{pmatrix} \frac{1}{\delta_1}\sum_{j=1}^{r} e_{1j}\boldsymbol{u}_j'\boldsymbol{y} \\ \vdots \\ \frac{1}{\delta_r}\sum_{j=1}^{r} e_{rj}\boldsymbol{u}_j'\boldsymbol{y} \end{pmatrix}$$

where δ_i is the ith diagonal element of $\boldsymbol{D}$, e_{ij} is the (i, j)th element of $\boldsymbol{E}$, and $\boldsymbol{u}_i$ is the ith vector of $\boldsymbol{U}$.

Thus we have

$$\begin{aligned} \boldsymbol{b}_{OLS} &= \boldsymbol{V}\boldsymbol{\phi}_{OLS} = \sum_{i=1}^{r}\left(\frac{\boldsymbol{u}_i'\boldsymbol{y}}{\delta_i}\right)\boldsymbol{v}_i \\ \boldsymbol{b}_{PCR} &= \boldsymbol{V}\boldsymbol{\phi}_{PCR} = \sum_{i=1}^{k}\left(\frac{\boldsymbol{u}_i'\boldsymbol{y}}{\delta_i}\right)\boldsymbol{v}_i \\ \boldsymbol{b}_{PLS} &= \boldsymbol{V}\boldsymbol{\phi}_{PLS} = \sum_{i=1}^{r}\left(\frac{1}{\delta_i}\sum_{j=1}^{r} e_{ij}\boldsymbol{u}_j'\boldsymbol{y}\right)\boldsymbol{v}_i \end{aligned} \tag{2}$$

Therefore the regression vector for the three methods can be described as follows:

1. An OLS coefficient vector requires summation of all r weighted $\boldsymbol{v}_i$ eigenvectors, where each weight is $\phi_i = \frac{\boldsymbol{u}_i'\boldsymbol{y}}{\delta_i}$.
2. A PCR coefficient vector is based on summation of the first k weighted $\boldsymbol{v}_i$ eigenvectors, where each weight is the same as that of OLS.
3. A PLS coefficient vector requires summation of all r weighted $\boldsymbol{v}_i$ eigenvectors, where each weight is $\phi_i = \frac{1}{\delta_i}\sum_{j=1}^{r} e_{ij}\boldsymbol{u}_j'\boldsymbol{y}$.

An alternative form of Equation 2 is

$$\boldsymbol{b}_{OLS} = \sum_{i=1}^{r}\gamma_i\frac{\boldsymbol{v}_i}{\delta_i} \quad \text{where} \quad \gamma_i = \boldsymbol{u}_j'\boldsymbol{y}$$

$$\boldsymbol{b}_{PCR} = \sum_{i=1}^{k} \gamma_i \frac{\boldsymbol{v}_i}{\delta_i} \quad \text{where} \quad \gamma_i = \boldsymbol{u}_j' \boldsymbol{y}$$

$$\boldsymbol{b}_{PLS} = \sum_{i=1}^{r} \gamma_i \frac{\boldsymbol{v}_i}{\delta_i} \quad \text{where} \quad \gamma_i = \sum_{j=1}^{r} e_{ij} \boldsymbol{u}_j' \boldsymbol{y}$$

Here each γ_i represents the above ϕ_i weight with the singular value information removed.

Thus by investigating ϕ_i or γ_i weights, further information can be provided on when PCR and PLS produce the same or different results and why. Another way to see the interrelationships among the three regression vectors are as follows.

1. If $k = r$, then $\boldsymbol{b}_{PCR}$ becomes $\boldsymbol{b}_{OLS}$.
2. If $e_{ij} = \begin{cases} 1, & \forall\ i = j \\ 0, & \forall\ i \neq j \end{cases}$, then $\boldsymbol{b}_{PLS}$ becomes $\boldsymbol{b}_{OLS}$.
3. If $e_{ij} = \begin{cases} 1, & \text{for } i = j,\ i, j = 1, \cdots, k\ (k < r) \\ 0, & \text{otherwise} \end{cases}$, then $\boldsymbol{b}_{PLS}$ becomes $\boldsymbol{b}_{PCR}$.

Since the solutions for the three regression methods depend on the e_{ij} values, we can see the differences among OLS, PCR and PLS better by investigating the e_{ij} values.

3. Computational Aspect

The algorithm in the previous section is theoretically valid but numerically inappropriate, and hence should not be implemented as described. The columns of $\boldsymbol{Q}_k$ consists of power series and hence computations of $\boldsymbol{Q}_k' \boldsymbol{Q}_k$ and its inverse have a serious ill-conditioning problem. One way to tackle this problem is to orthogonalize the columns of $\boldsymbol{Q}_k$ by the Gram-Schmidt procedure. Although this reduces the computational error, the ill-conditioning problem still remains.
This can be solved as follows. Consider

$$\boldsymbol{Q}_k = (\boldsymbol{D}^2 \boldsymbol{U}' \boldsymbol{y},\ \boldsymbol{D}^4 \boldsymbol{U}' \boldsymbol{y},\ \cdots,\ \boldsymbol{D}^{2k} \boldsymbol{U}' \boldsymbol{y})$$

and

$$\boldsymbol{K}_k = (\boldsymbol{V}\boldsymbol{D}\boldsymbol{U}' \boldsymbol{y},\ \boldsymbol{V}\boldsymbol{D}^3 \boldsymbol{U}' \boldsymbol{y},\ \cdots,\ \boldsymbol{V}\boldsymbol{D}^{2k-1} \boldsymbol{U}' \boldsymbol{y})$$

of Definition 1. It can be easily verified that $\boldsymbol{K}_k = \boldsymbol{V}\boldsymbol{D}^{-1} \boldsymbol{Q}_k$, and hence we have $\boldsymbol{Q}_k = \boldsymbol{D}\boldsymbol{V}' \boldsymbol{K}_k$. Using Wold's algorithm we can obtain matrix $\boldsymbol{W}_k$ that is

numerically stable and spans the same subspace as K_k does by proposition 1. We now set $R_k = DV'W_k$. Then we can obtain numerically robust result by using R_k instead of Q_k.

To further increase the numerical accuracy, we perform a SVD of R_k, i.e., $R_k = U_R D_R V_R'$ and substitute this into $E = R_k(R_k'R_k)^{-1}R_k'$ to obtain $E = U_R U_R'$ since V_R is nonsingular. The latter E-matrix gives numerically more accurate result than the former E-matrix.

Therefore for computational purposes, Step 3 of the algorithm in Section 2 should be replaced by the following procedure.

Step 3. Compute E-matrix:

For OLS, set

$$E = I_r$$

For PCR with the first k components, set

$$E = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$$

For PLS with the first k components, set

For $i = 1, \cdots, k$,

$$w_i = X_{i-1}'y_{i-1} / \| X_{i-1}'y_{i-1} \|$$

$$t_i = X_{i-1}w_i / \| X_{i-1}w_i \|$$

$$X_i = (I_n - t_i t_i')X_{i-1}$$

$$y_i = (I_n - t_i t_i')y_{i-1}$$

end for

$R_k = DV'W_k$ with $W_k = (w_1, \cdots, w_k)$

$U_R D_R V_R' = R_k$

$E = U_R U_R'$

4. Example

To illustrate how our algorithm performs on real world data, we use data from Fearn (1983). They are the result of an experiment performed to calibrate a near infrared (NIR) reflectance instrument for the measurement of protein content in ground wheat samples. The protein content measurements were made by the

standard Kjeldahl method, and the six values $x_1, x_2, \cdots, x_6$ are measurements of the reflectance of NIR radiation by the wheat samples at six different wavelengths in the range 1680-2310 nm. The number of observations is 24. The aim of the calibration is to find a good regression model which predicts protein content.

In these data, $p=6$ and $r=6$ and it can be shown that we have an optimal model when the number of components $k=4$ by the PRESS criterion using leave-one-out cross-validation for both PCR and PLS.

The $\boldsymbol{E}$-matrix of $\boldsymbol{\phi} = \boldsymbol{D}^{-1}\boldsymbol{E}\boldsymbol{U}'\boldsymbol{y}$ for $k=4$ is as follows.

1. For $\boldsymbol{b}_{OLS}$, $\boldsymbol{E} = \boldsymbol{I}_6$. That is, $e_{11}=e_{22}=e_{33}=e_{44}=e_{55}=e_{66}=1$ and the rest of the elements equal 0.
2. For $\boldsymbol{b}_{PCR}$, $\boldsymbol{E} = \begin{pmatrix} \boldsymbol{I}_4 & 0 \\ 0 & 0 \end{pmatrix}$. That is, $e_{11}=e_{22}=e_{33}=e_{44}=1$ and the remaining diagonal elements (e_{55}, e_{66}) plus all off-diagonal elements equal 0.
3. For $\boldsymbol{b}_{PLS}$,

$$\boldsymbol{E} = \begin{pmatrix} 1 & 8.816E-18 & -7.33E-16 & 1.053E-14 & 3.286E-13 & 2.88E-13 \\ 8.816E-18 & 1 & -1.334E-8 & 1.914E-7 & 5.9774E-6 & 5.2346E-6 \\ -7.33E-16 & -1.334E-8 & 0.9999972 & 0.0000404 & 0.0012638 & 0.0011029 \\ 1.053E-14 & 1.914E-7 & 0.0000404 & 0.99942 & -0.018158 & -0.01581 \\ 3.286E-13 & 5.9774E-6 & 0.0012638 & -0.018158 & 0.0003315 & 0.0002886 \\ 2.88E-13 & 5.2346E-6 & 0.0011029 & -0.01581 & 0.0002886 & 0.0002513 \end{pmatrix}.$$

That is, elements $e_{11}, e_{22}, e_{33}, e_{44}$ are equal to 1 or slightly smaller than 1 and the remaining diagonal elements e_{55} and e_{66} become slightly bigger than 0 in such a way that the sum of the diagonal elements is 4 (remember that $\sum e_{ii} = k$ for PLS); the remaining elements are small values (positive or negative).

Thus for OLS, the diagonal elements are all 1 and all the remaining elements are 0. Since $k=4$, PCR has elements $e_{11}, e_{22}, e_{33}, e_{44}$ which equal 1 and all the remaining elements are 0, and PLS has elements $e_{11}, e_{22}, e_{33}, e_{44}$ which equal 1 or slightly smaller values while the remaining elements are 0 or small nonzero values.

It can be easily shown that the general form of the weight vector $\boldsymbol{\phi}$ for this example is given by

$$\boldsymbol{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_6 \end{pmatrix} = \begin{pmatrix} 1/\delta_1(e_{11}\boldsymbol{u}_1'\boldsymbol{y} + e_{12}\boldsymbol{u}_2'\boldsymbol{y} + e_{13}\boldsymbol{u}_3'\boldsymbol{y} + e_{14}\boldsymbol{u}_4'\boldsymbol{y} + e_{15}\boldsymbol{u}_5'\boldsymbol{y} + e_{16}\boldsymbol{u}_6'\boldsymbol{y}) \\ 1/\delta_2(e_{21}\boldsymbol{u}_1'\boldsymbol{y} + e_{22}\boldsymbol{u}_2'\boldsymbol{y} + e_{23}\boldsymbol{u}_3'\boldsymbol{y} + e_{24}\boldsymbol{u}_4'\boldsymbol{y} + e_{25}\boldsymbol{u}_5'\boldsymbol{y} + e_{26}\boldsymbol{u}_6'\boldsymbol{y}) \\ \vdots \\ 1/\delta_6(e_{61}\boldsymbol{u}_1'\boldsymbol{y} + e_{62}\boldsymbol{u}_2'\boldsymbol{y} + e_{63}\boldsymbol{u}_3'\boldsymbol{y} + e_{64}\boldsymbol{u}_4'\boldsymbol{y} + e_{65}\boldsymbol{u}_5'\boldsymbol{y} + e_{66}\boldsymbol{u}_6'\boldsymbol{y}) \end{pmatrix}$$

.

Hence we have

$$\boldsymbol{\phi}_{OLS} = \begin{pmatrix} 1/\delta_1(\boldsymbol{u}_1'\boldsymbol{y}) \\ 1/\delta_2(\boldsymbol{u}_2'\boldsymbol{y}) \\ 1/\delta_3(\boldsymbol{u}_3'\boldsymbol{y}) \\ 1/\delta_4(\boldsymbol{u}_4'\boldsymbol{y}) \\ 1/\delta_5(\boldsymbol{u}_5'\boldsymbol{y}) \\ 1/\delta_6(\boldsymbol{u}_6'\boldsymbol{y}) \end{pmatrix} = \begin{pmatrix} 0.1914298 \\ -1.28772 \\ 5.5857538 \\ 1.0500725 \\ -2.277065 \\ -3.696978 \end{pmatrix}$$

$$\boldsymbol{\phi}_{PCR} = \begin{pmatrix} 1/\delta_1(\boldsymbol{u}_1'\boldsymbol{y}) \\ 1/\delta_2(\boldsymbol{u}_2'\boldsymbol{y}) \\ 1/\delta_3(\boldsymbol{u}_3'\boldsymbol{y}) \\ 1/\delta_4(\boldsymbol{u}_4'\boldsymbol{y}) \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.1914298 \\ -1.28772 \\ 5.5857538 \\ 1.0500725 \\ 0 \\ 0 \end{pmatrix}$$

$$\boldsymbol{\phi}_{PLS} = \begin{pmatrix} 1/\delta_1(e_{11}\boldsymbol{u}_1'\boldsymbol{y} + e_{12}\boldsymbol{u}_2'\boldsymbol{y} + e_{13}\boldsymbol{u}_3'\boldsymbol{y} + e_{14}\boldsymbol{u}_4'\boldsymbol{y} + e_{15}\boldsymbol{u}_5'\boldsymbol{y} + e_{16}\boldsymbol{u}_6'\boldsymbol{y}) \\ 1/\delta_2(e_{21}\boldsymbol{u}_1'\boldsymbol{y} + e_{22}\boldsymbol{u}_2'\boldsymbol{y} + e_{23}\boldsymbol{u}_3'\boldsymbol{y} + e_{24}\boldsymbol{u}_4'\boldsymbol{y} + e_{25}\boldsymbol{u}_5'\boldsymbol{y} + e_{26}\boldsymbol{u}_6'\boldsymbol{y}) \\ \cdots \\ \cdots \\ \cdots \\ 1/\delta_6(e_{61}\boldsymbol{u}_1'\boldsymbol{y} + e_{62}\boldsymbol{u}_2'\boldsymbol{y} + e_{63}\boldsymbol{u}_3'\boldsymbol{y} + e_{64}\boldsymbol{u}_4'\boldsymbol{y} + e_{65}\boldsymbol{u}_5'\boldsymbol{y} + e_{66}\boldsymbol{u}_6'\boldsymbol{y}) \end{pmatrix} = \begin{pmatrix} 0.1914298 \\ -1.287722 \\ 5.5850863 \\ 1.061922 \\ -0.077078 \\ -0.082587 \end{pmatrix}$$

(Each e_{ij} is defined as the (i,j)th element of $\boldsymbol{E}$ for PLS.)

Now it becomes clearer when $\boldsymbol{\phi}_{PCR}$ and $\boldsymbol{\phi}_{PLS}$ differ. The reason why the first element value of both vectors is identical is that $e_{11} = 1$ and all the remaining elements in that row are practically 0 for $\boldsymbol{\phi}_{PLS}$. Similar statement can be made for the second element value of both vectors. The third and fourth elements become slightly different in the two methods since e_{33} and e_{44} for $\boldsymbol{\phi}_{PLS}$ become slightly smaller than 1 and the remaining elements in each row are not all equal to 0. The difference in the fifth or sixth element in the two methods is due to the fact that e_{55} and e_{66} are all zeros for $\boldsymbol{\phi}_{PCR}$ and are neither 0 nor the remaining elements in each row equal 0 for $\boldsymbol{\phi}_{PLS}$.

Thus the smaller the values $e_{11}, e_{22}, e_{33}, e_{44}$ are compared to 1 and the further from 0 the remaining element values are, the greater the difference in ϕ values for PCR and PLS.

Now each regression vector is determined by $\boldsymbol{b} = \boldsymbol{V}\boldsymbol{\phi}$.

$$\boldsymbol{b}_{OLS} = \boldsymbol{V}\boldsymbol{\phi}_{OLS} = \sum_{i=1}^{6} \phi_{i(OLS)}\boldsymbol{v}_i = \begin{pmatrix} 0.6577155 \\ 0.0333157 \\ 5.0247563 \\ -5.181638 \\ 0.370844 \\ -0.426699 \end{pmatrix}$$

$$\boldsymbol{b}_{PCR} = \boldsymbol{V}\boldsymbol{\phi}_{PCR} = \sum_{i=1}^{4} \phi_{i(PCR)}\boldsymbol{v}_i = \begin{pmatrix} -0.755196 \\ 3.0947336 \\ 2.6913879 \\ -3.907986 \\ 0.4373464 \\ -1.06924 \end{pmatrix}$$

$$\boldsymbol{b}_{PLS} = \boldsymbol{V}\boldsymbol{\phi}_{PLS} = \sum_{i=1}^{6} \phi_{i(PLS)}\boldsymbol{v}_i = \begin{pmatrix} -0.730327 \\ 3.0147931 \\ 2.7652155 \\ -3.928654 \\ 0.4354395 \\ -1.06574 \end{pmatrix}$$

Each regression vector $\boldsymbol{b}$ is obtained as a linear combination of eigenvectors $\boldsymbol{v}_i$ using $\boldsymbol{\phi}$ as weights. Weights $\boldsymbol{\phi}$ are a function of singular values, left singular vectors, dependent variable vector, and e_{ij} values. In forming regression vector $\boldsymbol{b}$, OLS uses all 6 eigenvectors with appropriate weights, PCR uses the first 4 eigenvectors with appropriate weights, and PLS uses all 6 eigenvectors with appropriate weights.

5. Conclusions

In this paper, a unified procedure for OLS, PCR and PLS was presented. The process gives solutions for OLS, PCR and PLS and possibly other intermediate regression methods in a unified and basically non-iterative way. This unified solution consists of weighted linear combinations of the eigenvectors of $\boldsymbol{X}'\boldsymbol{X}$, provides concise matrix expressions, and makes it easy to see the differences between the solutions. The difference is determined by the so-called $\boldsymbol{E}$-matrix, and it has been illustrated on how the $\boldsymbol{E}$-matrix makes differences among the three solutions in the data set discussed. In addition, it brings an implication that

we may be able to obtain other and hopefully better biased estimators than PCR or PLS by choosing appropriate e_{ij} values. That could be another research project.

References

1. Brenchley, J. M., Lang, P. M., Nieves, R. G., and Kalivas, J. H. (1998). Stabilization of Cyclic Subspace Regression, *Chemometrics and Intelligent Laboratory Systems*, 41, 127-134.
2. Fearn, T. (1983). A Misuse of Ridge Regression in the Calibration of a Near Infrared Reflectance Instrument, *Journal of Applied Statistics*, 32, 73-79.
3. Frank, I. E. and Friedman, J. H. (1993). A Statistical View of Some Chemometrics Regression Tools, *Technometrics*, 35, 109-135.
4. Golub, G. H. and Van Loan, C. F. (1996). Matrix Computations, 3rd edition, The Johns Hopkins University Press.
5. Helland, I. S. (1988). On the Structure of Partial Least Squares Regression, *Communications in Statistics - Simulation and Computation*, 17, 581-607.
6. Kalivas, J. H. (1999). Cyclic Subspace Regression with Analysis of the Hat Matrix, *Chemometrics and Intelligent Laboratory Systems*, 45, 215-224.
7. Kim, J. D. (2003). Alternative Expressions of Regression Vector for Ordinary Least Squares, Principal Component Regression and Partial Least Squares Regression, *Journal of the Korean Data Analysis Society*, 5, 17-26.
8. Lang, P. M., Brenchley, J. M., Nieves, R. G. and Kalivas, J. H. (1998). Cyclic Subspace Regression, *Journal of Multivariate Analysis*, 65, 58-70.
9. Massy, W. F. (1965). Principal Components Regression in Exploratory Statistical Research, *Journal of the American Statistical Association*, 60, 234-246.
10. Stone, M. and Brooks, R. J. (1990). Continuum Regression: Cross-validated Sequentially Constructed Prediction embracing Ordinary Least Squares, Partial Least Squares and Principal Component Regression, *Journal of the Royal Statistical Society*, B 52, 237-269.
11. Wold, S., Wold, J., Dunn, W. J. and Ruhe, A. (1984). The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses, *SIAM Journal on Scientific and Statistical Computing*, 5, 735-743.

[received date : Feb. 2003, accepted date : Apr. 2003]