Journal of the Korean
Data & Information Science Society
2006, Vol. 17, No. 4, pp. 1169~1180

Expressions for Shrinkage Factors of PLS Estimator

Jong-Duk Kim[1)]

Abstract

Partial least squares regression (PLS) is a biased, non-least squares regression method and is an alternative to the ordinary least squares regression (OLS) when predictors are highly collinear or predictors outnumber observations. One way to understand the properties of biased regression methods is to know how the estimators shrink the OLS estimator. In this paper, we introduce an expression for the shrinkage factor of PLS and develop a new shrinkage expression, and then prove the equivalence of the two representations. We use two near-infrared (NIR) data sets to show general behavior of the shrinkage and in particular for what eigendirections PLS expands the OLS coefficients.

Keywords : 고유방향, 부분최소제곱회귀, 축소인자

1. 서론

다음의 선형회귀모형을 고려한다.

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1.1}$$

여기서 $\boldsymbol{y}$는 $n\times 1$ 반응벡터, $\boldsymbol{X}$는 $n\times p$ 설계행렬, $\boldsymbol{\beta}$는 $p\times 1$ 모수벡터, $\boldsymbol{\epsilon}$은 $n\times 1$ 오차(noise)벡터로 평균이 $\mathbf{0}$, 분산이 $\sigma^2\boldsymbol{I}$라고 가정한다. $\boldsymbol{X}$의 계수는 r로서 $\boldsymbol{X}$가 완전 열계수가 아닐 수 있는 일반적인 경우를 가정한다. $\boldsymbol{X}$와 $\boldsymbol{y}$는 중심화 또는 표준화된 것으로 간주한다.

비정칙값 분해(SVD, singular value decomposition)는 회귀추정량을 이해하는데 핵심 도구이며 이것을 요약하면 다음과 같다. 행렬 $\boldsymbol{X}$의 SDV는 $\boldsymbol{X}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}'$인데 여기서 $\boldsymbol{U}$는 $\boldsymbol{X}\boldsymbol{X}'$의 고유값으로 구성된 $n\times r$ 행렬, $\boldsymbol{D}$는 대각원소가 $\boldsymbol{X}$의 비정칙값이 크기순으로 나열된 $r\times r$ 대각행렬, $\boldsymbol{V}$는 $\boldsymbol{X}'\boldsymbol{X}$의 고유값으로 구성된 $p\times r$ 행렬이다. 즉 $\boldsymbol{U}=(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_r)$, $\boldsymbol{V}=(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_r)$, $\boldsymbol{D}=\mathrm{diag}(\delta_1,\cdots,\delta_r)$로서 $\boldsymbol{U}'\boldsymbol{U}=\boldsymbol{I}_r$, $\boldsymbol{V}'\boldsymbol{V}=\boldsymbol{I}_r$이다. 행렬 $\boldsymbol{U}$

1) 부산시 남구 우암동 산 55-1, 부산외국어대학교 응용통계학과 교수
E-mail : jdkim@pufs.ac.kr

는 $\boldsymbol{X}$의 주왼쪽비정칙벡터(principal left singular vectors)의 행렬, 행렬 $\boldsymbol{V}$는 $\boldsymbol{X}$의 주오른쪽비정칙벡터(principal right singular vectors)의 행렬이라 부른다. 그리고 $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}' = \sum_{i=1}^{r} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i'$이며 $\lambda_i = \delta_i^2$이다.

OLS 추정량은 선형이다. 회귀모형 (1.1)에서 $\boldsymbol{X}$가 완전열계수이면 OLS 추정량은 유일하나 완전열계수가 되지 못하면 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$은 과대모수화되어 무수히 많은 $\hat{\boldsymbol{\beta}}$가 존재한다. 그렇지만 여전히 $\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|$의

$$\hat{\boldsymbol{\beta}}_{ols} = (\boldsymbol{X}'\boldsymbol{X})^{+}\boldsymbol{X}'\boldsymbol{y}$$
$$\hat{\boldsymbol{y}}_{ols} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{+}\boldsymbol{X}'\boldsymbol{y}$$

을 사용할 수 있다. 여기서 첨자 +는 Moore-Penrose 일반화역행렬을 의미한다. $\hat{\boldsymbol{\beta}}_{ols}$은 유일하지 않지만 $\hat{\boldsymbol{y}}_{ols}$은 유일하다. OLS 추정량은 다음과 같이 고유벡터의 선형결합으로 나타낼 수 있다.

$$\hat{\boldsymbol{\beta}}_{ols} = \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}'\boldsymbol{y} = \sum_{j=1}^{r} \hat{\alpha}_j \boldsymbol{v}_j$$

여기서 $\hat{\alpha}_j = \boldsymbol{u}_j'\boldsymbol{y}/\delta_j$이다. MSE를 줄이기 위해 OLS 추정량에서 분산을 크게 만드는 고유방향을 축소시키는 방법을 생각해볼 수 있다. 이렇게 하면 당연히 편향이 발생한다. 따라서 우리는 편향의 증가량이 분산의 감소량에 비해 작기를 기대하면서 OLS 추정량의 축소를 시도한다.

$\boldsymbol{\beta}$의 축소추정량의 일반식으로 다음을 고려한다(Frank와 Friedman, 1993).

$$\widehat{\boldsymbol{\beta}_{shr}} = \sum_{j=1}^{r} f_{j(.)} \widehat{\alpha_j} \boldsymbol{v_j}$$

$f_{j(.)}$는 축소인자(shrinkage factor)라 부르기로 한다. 축소인자가 $\boldsymbol{y}$에 의존하지 않으면 $\hat{\boldsymbol{\beta}}_{shr}$는 $\boldsymbol{y}$에 선형이며, $f_{j(.)} \neq 1$인 인자는 i번째 성분의 편향을 증가시킨다. j번째 성분의 분산은 $f_{j(.)} < 1$이면 감소하고 $f_{j(.)} > 1$이면 증가한다. 따라서 분산의 감소량에 비해 편향의 증가량이 작은 축소 방법이 바람직할 것이다. 이 축소추정량의 행렬형은 $\hat{\boldsymbol{\beta}}_{shr} = \boldsymbol{V}\boldsymbol{F}_{(.)}^{(m)}\hat{\boldsymbol{\alpha}}$으로 쓸 수 있는데, 여기서 $\boldsymbol{V}$는 $\boldsymbol{X}$의 주오른쪽비정칙벡터의 행렬, $\boldsymbol{F}_{(.)}^{(\boldsymbol{m})} = \mathrm{diag}(f_{1(.)}, \cdots, f_{r(.)})$, $\hat{\boldsymbol{\alpha}} = (\widehat{\alpha_1}, \cdots, \widehat{\alpha_r})'$이다.

2. 부분최소제곱회귀

부분최소제곱회귀(partial least squares regression, PLS)는 H. Wold(1975)에 의해 개발되었으며 반복 알고리듬인 NIPALS 알고리듬으로 주어졌다. 알고리듬에서 PLS의 인자벡터 $\boldsymbol{t}_k$는 각각 직교성을 유지하면서 순차적으로 한 개씩 계산되며 최대 r개 까지 구해질 수 있다. 인자수 m개의 PLS 추정량은

$$\hat{\boldsymbol{\beta}}_{pls}^{(m)} = \boldsymbol{W}_m(\boldsymbol{W}_m'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{W}_m)^{-1}\boldsymbol{W}_m'\boldsymbol{X}'\boldsymbol{y}$$

또는

$$\hat{\boldsymbol{\beta}}_{pls}^{(m)} = \boldsymbol{W}_m(\boldsymbol{P}_m'\boldsymbol{W}_m)^{-1}\boldsymbol{q}_m$$

으로 쓸 수 있는데, 여기서 $\boldsymbol{W}_m = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m)$은 가중값행렬, $\boldsymbol{P}_m = (\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m)$은 X의 적재행렬, $\boldsymbol{q}_m = (q_1, \cdots, q_m)'$은 y의 적재벡터이다. 이와 같이 PLS의 해는 반복 알고리즘으로 주어지고 인자를 구할 때 $\boldsymbol{X}$ 뿐만 아니라 $\boldsymbol{y}$의 정보도 이용하기 때문에 복잡한 비선형이 되어 그 통계적 특성을 이해하기 어렵다. 이들 특성을 규명하기 위한 많은 노력이 이루어졌다.

PLS의 비반복적 해도 구해졌다(Helland, 1988). $m \geq 1$에서 다음의 Krylov 공간을 고려한다.

$$\boldsymbol{K}_m = \mathrm{span}(\boldsymbol{X}'\boldsymbol{y}, (\boldsymbol{X}'\boldsymbol{X})\boldsymbol{X}'\boldsymbol{y}, \cdots, (\boldsymbol{X}'\boldsymbol{X})^{m-1}\boldsymbol{X}'\boldsymbol{y})$$

그러면 위의 PLS 알고리즘에서의 가중값벡터 집합 $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m\}$에 의해 생성되는 공간과 Krylov 공간은 동일하다. m개 잠재인자의 PLS 추정량은 다음의 최적화의 해이다.

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|$$
$$\text{such that } \boldsymbol{\beta} \in \boldsymbol{K}_m$$

$p \times m$ 행렬 $\boldsymbol{R}_m$을 열들이 부공간 $\boldsymbol{K}_m$을 생성하는 한 행렬이라 하면 m개 인자의 PLS 추정량은

$$\hat{\boldsymbol{\beta}}_{pls}^{(m)} = \boldsymbol{R}_m(\boldsymbol{R}_m'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{R}_m)^{-1}\boldsymbol{R}_m'\boldsymbol{X}'\boldsymbol{y}$$

이다.

m개 인자의 PLS의 비반복적 해는 다음의 형태도 가능하다(Kim, 2003; Kim과 Moon, 2005).

$$\hat{\boldsymbol{\beta}}_{pls}^{(m)} = \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{G}_{(m)}\boldsymbol{U}'\boldsymbol{y}$$

여기서 $\boldsymbol{G}_{(m)} = \boldsymbol{Q}_m(\boldsymbol{Q}_m'\boldsymbol{Q}_m)^{-1}\boldsymbol{Q}_m'$, $\boldsymbol{Q}_m = (\boldsymbol{D}^2\boldsymbol{U}'\boldsymbol{y}, \boldsymbol{D}^4\boldsymbol{U}'\boldsymbol{y}, \cdots, \boldsymbol{D}^{2m}\boldsymbol{U}'\boldsymbol{y})$이고, $\boldsymbol{V}$, $\boldsymbol{D}$, $\boldsymbol{U}$는 비정칙값 분해 $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}'$에서 나온 것이다.

3. PLS 추정량의 축소인자

Frank와 Friedman(1993)은 $n \times p$ 행렬 $\boldsymbol{X}$가 완전열계수인 경우에 PLS 해의 축소인자의 식을 증명 없이 보였다. 여기서는 $\boldsymbol{X}$가 불완전열계수일 수 있는 $\mathrm{rank}(\boldsymbol{X}) = r$인 경우의 축소인자의 식으로 바꾸고 증명을 첨부한다.

정리 1. m개 인자의 PLS 추정량의 축소인자는 다음과 같다.

$$f_{j(pls)}^{(m)} = \sum_{k=1}^{m} \delta_j^{2k} b_k, \quad j = 1, 2, \cdots, r \tag{3.1}$$

그리고 벡터 $\boldsymbol{b} = (b_1, \cdots, b_m)'$은 $\boldsymbol{b} = \boldsymbol{S}^{-1}\boldsymbol{s}$인데 여기서 $\boldsymbol{S}$는 (k,l)번째 원소가 $S_{kl} = \sum_{j=1}^{r} \delta_j^{2(k+l+1)} \hat{\alpha}_j^2$인 $m \times m$ 행렬이고 $\boldsymbol{s}$는 k번째 원소가 $s_k = \sum_{j=1}^{r} \delta_j^{2(k+1)} \hat{\alpha}_j^2$인 $m \times 1$ 벡터이다.

증명. m개 인자의 PLS 추정량

$$\hat{\boldsymbol{\beta}}_{pls}^{(m)} = \boldsymbol{W}_m(\boldsymbol{W}_m'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{W}_m)^{-1}\boldsymbol{W}_m'\boldsymbol{X}'\boldsymbol{y}$$

에서 가중값 행렬 $\boldsymbol{W}_m$은 Krylov 행렬 $\boldsymbol{K}_m$의 QR-분해인 $\boldsymbol{K}_m = \boldsymbol{W}_m\boldsymbol{R}$으로 얻어지며 따라서 $\boldsymbol{W}_m = \boldsymbol{K}_m\boldsymbol{R}^{-1}$이다(Di Ruscio, 2000). 그리고 행렬 $\boldsymbol{X}$가 불완전열계수일 때 $\boldsymbol{V}\boldsymbol{V}' \neq \boldsymbol{I}$이지만 $\boldsymbol{V}\boldsymbol{V}'\boldsymbol{W}_m = \boldsymbol{W}_m$은 성립하는데, 이것은 $\boldsymbol{K}_m = \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{Q}_m$을 이용하면

$$\boldsymbol{V}\boldsymbol{V}'\boldsymbol{W}_m = \boldsymbol{V}\boldsymbol{V}'\boldsymbol{K}_m\boldsymbol{R}^{-1} = \boldsymbol{V}\boldsymbol{V}'\boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{Q}_m\boldsymbol{R}^{-1} = \boldsymbol{K}_m\boldsymbol{R}^{-1} = \boldsymbol{W}_m$$

으로 바로 확인된다($\boldsymbol{Q}_m = (\boldsymbol{D}^2\boldsymbol{U}'\boldsymbol{y}, \boldsymbol{D}^4\boldsymbol{U}'\boldsymbol{y}, \cdots, \boldsymbol{D}^{2m}\boldsymbol{U}'\boldsymbol{y})$). 이상의 두 결과에 의해 PLS의 해벡터는

$$\hat{\boldsymbol{\beta}}_{pls}^{(m)} = \boldsymbol{V}\boldsymbol{V}'\boldsymbol{K}_m(\boldsymbol{K}_m'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{K}_m)^{-1}\boldsymbol{K}_m'\boldsymbol{X}'\boldsymbol{y}$$

으로 쓸 수 있다. 이제 $\boldsymbol{S} \equiv \boldsymbol{K}_m'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{K}_m$, $\boldsymbol{s} \equiv \boldsymbol{K}_m'\boldsymbol{X}'\boldsymbol{y}$로 두고 또한 $\boldsymbol{b} = \boldsymbol{S}^{-1}\boldsymbol{s}$로 두면

$$\hat{\boldsymbol{\beta}}_{pls}^{(m)} = \boldsymbol{V}\boldsymbol{V}'\boldsymbol{K}_m\boldsymbol{b}$$

이다. 여기서 $\boldsymbol{V}'\boldsymbol{K}_m\boldsymbol{b} = (\boldsymbol{D}\boldsymbol{U}'\boldsymbol{y}, \boldsymbol{D}^3\boldsymbol{U}'\boldsymbol{y}, \cdots, \boldsymbol{D}^{2m-1}\boldsymbol{U}'\boldsymbol{y})\boldsymbol{b}$이며 따라서 이것의 j번째 원소는 $(\delta_j, \delta_j^3, \cdots, \delta_j^{2m-1})\boldsymbol{u}_j'\boldsymbol{y}\boldsymbol{b}$이며 이것은 또한 $(\delta_j^2, \delta_j^4, \cdots, \delta_j^{2m})\boldsymbol{b}\boldsymbol{u}_j'\boldsymbol{y}/\delta_j$이다. 따라서

$r\times 1$ 벡터 $\boldsymbol{V'K_m b}$는 j번째 원소가 $(\delta_j^2, \delta_j^4, \cdots, \delta_j^{2m})\boldsymbol{b}$인 $r\times r$ 대각행렬과 원소가 $\hat{\alpha_j}=\boldsymbol{u_j'y}/\delta_j$인 $r\times 1$인 벡터의 곱으로 나타낼 수 있다. 이 대각행렬과 벡터를 각각 $\boldsymbol{F^{(m)}}$와 $\hat{\boldsymbol{\alpha}}$으로 표시하면 PLS 추정량은

$$\hat{\boldsymbol{\beta}}_{pls}^{(m)}=\boldsymbol{VF}^{(m)}\hat{\boldsymbol{\alpha}}$$

으로 표현되며 따라서 $\boldsymbol{b}=(b_1, b_2, \cdots, b_m)'$로 두면 축소인자의 식은

$$f_j^{(m)}=\sum_{k=1}^{m}\delta_j^{2k}b_k,\quad j=1,2,\cdots,r$$

이다. $\boldsymbol{b}=\boldsymbol{S}^{-1}\boldsymbol{s}$에서 행렬 $\boldsymbol{S}$의 (k,l)번째 원소는 다음과 같이 구해진다.

$$\boldsymbol{S}=\boldsymbol{K_m}'\boldsymbol{X'XK_m}=(\boldsymbol{K_m}'\boldsymbol{VD})(\boldsymbol{DV'K_m})$$

에서 $\boldsymbol{K_m}'\boldsymbol{VD}$의 k번째 행은 $(\delta_1^{2k+1}\hat{\alpha}_1, \cdots, \delta_r^{2k+1}\hat{\alpha}_r)$이고 따라서 $\boldsymbol{S}$의 (k,l)번째 원소는

$$(\delta_1^{2k+1}\hat{\alpha}_1)(\delta_1^{2l+1}\hat{\alpha}_1)+\cdots+(\delta_r^{2k+1}\hat{\alpha}_r)(\delta_r^{2l+1}\hat{\alpha}_r)=\sum_{j=1}^{r}\delta_j^{2(k+l+1)}\hat{\alpha}_j^2$$

이다. 행렬 $\boldsymbol{s}$의 k번째 원소는 다음과 같이 구해진다.

$$\boldsymbol{s}=\boldsymbol{K_m}'\boldsymbol{X'y}=\boldsymbol{K_m}'\boldsymbol{VDU'y}=\boldsymbol{K_m}'\boldsymbol{VD}^2\hat{\boldsymbol{\alpha}}$$

이고, $\boldsymbol{K_m}'\boldsymbol{VD}^2$은 앞서 나온 $\boldsymbol{K_m}'\boldsymbol{VD}$와 대각행렬 $\boldsymbol{D}$의 곱이므로 이것의 k번째 행은 $(\delta_1^{2(k+1)}\hat{\alpha}_1, \cdots, \delta_r^{2(k+1)}\hat{\alpha}_r)$이고 따라서 $\boldsymbol{K_m}'\boldsymbol{VD}^2\hat{\boldsymbol{\alpha}}$의 k번째 값은

$$\delta_1^{2(k+1)}\hat{\alpha}_1^2+\cdots+\delta_r^{2(k+1)}\hat{\alpha}_r^2=\sum_{j=1}^{r}\delta_j^{2(k+1)}\hat{\alpha}_j^2$$

이다. ■

Frank와 Friedman(1993)은 축소인자 $f_{j(pls)}>1$인 것도 있음을 언급하였다. 즉 어떤 고유방향으로는 팽창하는 것도 있다는 것이다. PLS의 축소인자는 다음과 같이 쓸 수도 있다.

정리 2. m개 인자의 PLS의 축소인자는 다음과 같다.

$$f_{j(pls)}^{(m)}=\frac{\sum_{l=1}^{r}g_{jl}\boldsymbol{u}_l'\boldsymbol{y}}{\boldsymbol{u}_j'\boldsymbol{y}},\quad j=1,2,\cdots,r \tag{3.2}$$

단 g_{jl}은 $\boldsymbol{G}^{(m)}$의 (j,l)번째 원소, $\boldsymbol{u}_j$는 $\boldsymbol{U}$의 i번째 열, $\boldsymbol{G}_{(m)}=\boldsymbol{Q}_m(\boldsymbol{Q}_m{}'\boldsymbol{Q}_m)^{-1}\boldsymbol{Q}_m{}'$, $\boldsymbol{Q}_m=(\boldsymbol{D}^2\boldsymbol{U}'\boldsymbol{y},\ \boldsymbol{D}^4\boldsymbol{U}'\boldsymbol{y},\ \cdots,\ \boldsymbol{D}^{2m}\boldsymbol{U}'\boldsymbol{y})$이다.

증명. 이것의 증명은 2절에서 언급된 PLS의 해벡터 $\hat{\boldsymbol{\beta}}_{pls}^{(m)}=\boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{G}_{(m)}\boldsymbol{U}'\boldsymbol{y}$, $\boldsymbol{G}_{(m)}=\boldsymbol{Q}_m(\boldsymbol{Q}_m{}'\boldsymbol{Q}_m)^{-1}\boldsymbol{Q}_m{}'$, $\boldsymbol{Q}_m=(\boldsymbol{D}^2\boldsymbol{U}'\boldsymbol{y},\ \boldsymbol{D}^4\boldsymbol{U}'\boldsymbol{y},\ \cdots,\ \boldsymbol{D}^{2m}\boldsymbol{U}'\boldsymbol{y})$에서 시작한다. 이 해벡터 식에서 $\boldsymbol{G}_{(m)}$의 j번째 행벡터를 $\boldsymbol{g}_j{}'$이라 하면 $\boldsymbol{D}^{-1}\boldsymbol{G}_{(m)}$은 j번째 행벡터가 $\boldsymbol{g}_j{}'/\delta_j$인 $r\times r$인 행렬이고, $\boldsymbol{D}^{-1}\boldsymbol{G}_{(m)}\boldsymbol{U}'\boldsymbol{y}$는 j번째 원소가 $\boldsymbol{g}_j{}'\boldsymbol{U}'\boldsymbol{y}/\delta_j$인 $r\times 1$인 벡터이다. 이것은 j번째 대각원소가 $\boldsymbol{g}_j{}'\boldsymbol{U}'\boldsymbol{y}/\delta_j\boldsymbol{u}_j{}'\boldsymbol{y}$인 $r\times r$ 대각행렬과 j번째 원소가 $\boldsymbol{u}_j{}'\boldsymbol{y}$인 $r\times 1$인 벡터로 분해되며, 전자의 대각행렬은 다시 j번째 대각원소가 $\boldsymbol{g}_j{}'\boldsymbol{U}'\boldsymbol{y}/\boldsymbol{u}_j{}'\boldsymbol{y}$인 $r\times r$ 대각행렬과 j번째 원소가 $1/\delta_j$인 $r\times r$ 대각행렬로 분해된다. 여기서

$$\boldsymbol{F}^{(m)}=\mathrm{diag}(f_1^{(m)},\cdots,f_r^{(m)}),\ \text{단}\ f_j^{(m)}=\frac{\boldsymbol{g}_j{}'\boldsymbol{U}'\boldsymbol{y}}{\boldsymbol{u}_j{}'\boldsymbol{y}}$$

로 두면 바로

$$\hat{\boldsymbol{\beta}}_{pls}^{(m)}=\boldsymbol{V}\boldsymbol{F}^{(m)}\boldsymbol{D}^{-1}\boldsymbol{U}'\boldsymbol{y}$$

을 얻는다. g_{jl}을 $\boldsymbol{G}_{(m)}$의 (j,l)번째 원소라 하면 $\boldsymbol{g}_j{}'\boldsymbol{U}'\boldsymbol{y}=\sum_{l=1}^{r}g_{jl}\boldsymbol{u}_l{}'\boldsymbol{y}$이고 따라서 증명된다. ■

식 (3.2)의 축소인자와 식 (3.1)의 축소인자의 동일성은 다음과 같이 보일 수 있다. j번째 축소인자는

$$f_j^{(m)}=\frac{\sum_{l=1}^{r}g_{jl}\boldsymbol{u}_l{}'\boldsymbol{y}}{\boldsymbol{u}_j{}'\boldsymbol{y}}=\frac{\sum_{l=1}^{r}g_{jl}\delta_l\hat{\alpha}_l}{\delta_j\hat{\alpha}_j}$$

으로 쓸 수 있다. 그리고 $\boldsymbol{q}_j{}'$을 $\boldsymbol{Q}_m$의 j번째 행이라 하면

$$g_{jl}=\boldsymbol{q}_j{}'(\boldsymbol{Q}_m{}'\boldsymbol{Q}_m)^{-1}\boldsymbol{q}_l$$

이다. 따라서

$$\begin{aligned}f_j^{(m)}&=\frac{1}{\delta_j\hat{\alpha}_j}\sum_{l=1}^{r}\boldsymbol{q}_j{}'(\boldsymbol{Q}_m{}'\boldsymbol{Q}_m)^{-1}\boldsymbol{q}_l\delta_l\hat{\alpha}_l\\&=\frac{1}{\delta_j\hat{\alpha}_j}\boldsymbol{q}_j{}'(\boldsymbol{Q}_m{}'\boldsymbol{Q}_m)^{-1}\sum_{l=1}^{r}\boldsymbol{q}_l\boldsymbol{u}_l{}'\boldsymbol{y}\end{aligned}$$

이다. 여기서 $\boldsymbol{Q}_m = (\boldsymbol{D}^2\boldsymbol{U}'\boldsymbol{y}, \boldsymbol{D}^4\boldsymbol{U}'\boldsymbol{y}, \cdots, \boldsymbol{D}^{2m}\boldsymbol{U}'\boldsymbol{y}) = \boldsymbol{D}\boldsymbol{V}'\boldsymbol{K}_m$ 이므로 $\boldsymbol{Q}_m{}'\boldsymbol{Q}_m$ 은 정리 1에서의 $\boldsymbol{S}$와 동일하고, $\sum_{l=1}^{r}\boldsymbol{q}_l\boldsymbol{u}_l'\boldsymbol{y} = \boldsymbol{Q}_m{}'\boldsymbol{U}'\boldsymbol{y} = \boldsymbol{K}_m{}'\boldsymbol{V}\boldsymbol{D}\boldsymbol{U}'\boldsymbol{y} = \boldsymbol{K}_m{}'\boldsymbol{X}'\boldsymbol{y}$는 정리 1에서의 $\boldsymbol{s}$와 동일하다. 그리고 남은 앞부분은

$$\frac{1}{\delta_j\hat{\alpha}_j}\boldsymbol{q}_j' = \frac{1}{\delta_j\hat{\alpha}_j}(\delta_j^3\hat{\alpha}_j, \delta_j^5\hat{\alpha}_j, \cdots, \delta_j^{2m+1}\hat{\alpha}_j) = (\delta_j^2, \delta_j^4, \cdots, \delta_j^{2m})$$

으로 정리 1의 $\boldsymbol{V}'\boldsymbol{K}_m$와 동일하다. 따라서 이 축소인자와 정리 1의 축소인자는 동일함이 확인된다.

Frank와 Friedman(1993)은 m개 인자의 PLS는 m번째 고유값과 근처에서 계수가 팽창한다고 결론을 내렸다. 그러나 Butler과 Denham(2000)은 다른 예를 통해 항상 그렇게 되는 것은 아니라고 하였다.

축소인자식 (3.2)에서 j번째 항은 $f_{j(pls)}^{(m)} = g_{jj} + \sum_{l\neq j}^{r} g_{jl}\boldsymbol{u_l}'\boldsymbol{y}/\boldsymbol{u_j}'\boldsymbol{y}$로 쓸 수 있다. 그리고 $\boldsymbol{G}_{(m)}$은 대칭 멱등행렬이므로 (i) $0 \le g_{jj} \le 1$ (ii) $g_{jj} = \sum_{l=1}^{r} g_{jl}^2$이 성립된다. j번째 항 $f_{j(pls)}^{(m)}$의 값이 1보다 크면 그 고유방향으로 팽창된다.

4. 수치 예

두 개의 실제 데이터 셋으로 PLS의 축소인자의 값을 구해보도록 한다. 이 값의 계산은 본 논문에서 유도된 식 (3.2)를 이용하였으며, 모든 계산은 SAS/IML로 수행되었다.

데이터 셋 1 (Fearn, 1983): x 값은 NIR 반사의 여섯 파장(설명변수)에서 나온 것이고 y 값은 밀에 포함된 단백질 비율(%)이다. 총 표본 수는 24이다. 모든 변수는 사전 처리로서 단위길이표준화(상관변환)를 하였다. 이 변환된 $\boldsymbol{X}'\boldsymbol{X}$의 고유값은

(5.8679606, 0.1007471, 0.0186959, 0.0122150, 0.0002301, 0.0001513)

으로 매우 높은 공선성의 데이터임을 알 수 있다. <표 1>은 PLS의 각 인자 수에서 고유방향에 따른 축소인자의 값이다.

<표 1> Fearn 데이터에 적용된 PLS의 축소인자 $f_{j(pls)}^{(m)}$의 값 (볼드체로 한 것은 1보다 큰 값)

고유방향 j	인자수 m					
	1	2	3	4	5	6
1	**1.0218501**	0.9999097	1.0000000	1.0000000	1.0000000	1.0000000
2	0.0175442	**1.4820444**	0.9997565	**1.0000011**	1.0000000	1.0000000
3	0.0032557	0.2788941	**1.0049619**	0.9998805	**1.0000001**	1.0000000
4	0.0021271	0.1824156	0.6995192	**1.0112845**	0.9999923	1.0000000
5	0.0000401	0.0034425	0.0146740	0.0338496	**1.2184397**	1.0000000
6	0.0000264	0.0022647	0.0096600	0.0223391	0.8107309	1.0000000

데이터 셋 2 (Brereton, 2003): 각 샘플은 27개의 파장(220nm에서 350nm 사이의 파장을 5nm 간격)과 화합물에서의 Pyrene의 농도(10 PAHs)를 측정한 것으로 구성되어 있다. 총 샘플(EAS 스펙트럼)수는 25이다. 여기서도 모든 변수는 사전 처리로서 상관변환을 하였다. 이 변환된 데이터의 $\boldsymbol{X}'\boldsymbol{X}$의 양의 고유값은

(20.953006, 2.5443417, 1.3252614, 1.0724578, 0.6091839, 0.2301074, 0.0833184, 0.0651115, 0.0374254, 0.0324147, 0.0201557, 0.0134545, 0.0061830, 0.0022937, 0.0018606, 0.0011580, 0.0009120, 0.0005756, 0.0002793, 0.0002632, 0.0001596, 0.0000396, 0.0000229, 0.0000144)

이다. 따라서 크기 25×27의 $\boldsymbol{X}$의 계수(rank)는 24로 불완전계수이다. 모든 경우의 축소인자의 값은 <표 2>에 주어져 있다.

<표 2> Brereton의 데이터에 적용된 PLS의 축소인자 $f_{j(pls)}^{(m)}$의 값 (볼드체로 한 것은 1보다 큰 값)

고유방향 j	인자수 m							
	1	2	3	4	5	6	7	8
1	**1.9923154**	0.9966199	**1.0000510**	0.9999996	1.0000000	1.0000000	1.0000000	1.0000000
2	0.2419286	**1.0346828**	0.9942856	**1.0006016**	0.9999287	**1.0000094**	0.9999999	1.0000000
3	0.1260124	0.5704465	**1.2793579**	0.8342482	**1.0552426**	0.9820824	**1.0003673**	0.9999981
4	0.1019746	0.4669183	**1.1723000**	**1.2627035**	0.8496363	**1.0698040**	0.9982125	**1.0000115**
5	0.0579242	0.2707269	0.8131543	**1.6260574**	**1.2531818**	0.4276831	**1.0269797**	0.9996865
6	0.0218797	0.1039634	0.3543623	**1.0085407**	**1.1819695**	**1.1777862**	0.9730427	**1.0009072**
7	0.0079223	0.0378821	0.1350767	0.4308210	0.5842390	0.8254315	**1.2171041**	0.9676150
8	0.0061911	0.0296272	0.1062199	0.3433620	0.4737220	0.6951742	**1.0864045**	**1.2294962**
9	0.0035586	0.0170496	0.0616325	0.2032826	0.2877938	0.4465271	0.7585403	**1.2990434**
10	0.0030821	0.0147701	0.0534716	0.1770031	0.2517504	0.3944651	0.6798825	**1.2408421**
11	0.0019165	0.0091890	0.0333873	0.1114949	0.1603729	0.2573277	0.4591870	0.9682218
12	0.0012793	0.0061356	0.0223374	0.0749513	0.1084702	0.1762897	0.3204679	0.7268335
13	0.0005879	0.0028205	0.0102903	0.0347071	0.0505613	0.0833118	0.1544751	0.3775991
14	0.0002181	0.0010465	0.0038224	0.0129277	0.0188994	0.0313693	0.0587751	0.1492925
15	0.0001769	0.0008489	0.0031010	0.0104912	0.0153434	0.0254877	0.0478103	0.1219533
16	0.0001101	0.0005284	0.0019305	0.0065344	0.0095626	0.0159058	0.0298924	0.0767686
17	0.0000867	0.0004161	0.0015205	0.0051475	0.0075347	0.0125384	0.0235793	0.0606994
18	0.0000547	0.0002626	0.0009598	0.0032500	0.0047586	0.0079238	0.0149146	0.0385186
19	0.0000266	0.0001274	0.0004657	0.0015774	0.0023103	0.0038490	0.0072506	0.0187787
20	0.0000250	0.0001201	0.0004389	0.0014865	0.0021772	0.0036274	0.0068334	0.0177010
21	0.0000152	0.0000728	0.0002661	0.0009014	0.0013203	0.0022001	0.0041458	0.0107498
22	3.7695E-6	0.0000181	0.0000661	0.0002240	0.0003281	0.0005469	0.0010308	0.0026758
23	2.1751E-6	0.0000104	0.0000382	0.0001292	0.0001893	0.0003156	0.0005949	0.0015445
24	1.3682E-6	6.5655E-6	0.0000240	0.0000813	0.0001191	0.0001985	0.0003742	0.0009716

<표 2> (연속)

고유방향	인자수 m							
j	9	10	11	12	13	14	15	16
1	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
2	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
3	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
4	0.9999998	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
5	**1.0000077**	0.9999998	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
6	0.9999310	**1.0000052**	0.9999999	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
7	**1.0153703**	0.9960846	**1.0003526**	0.9999750	**1.0000040**	0.9999999	1.0000000	1.0000000
8	0.5718795	**1.1554998**	0.9801095	**1.0021593**	0.9995240	**1.0000160**	0.9999999	1.0000000
9	0.9440552	0.8567916	**1.5867934**	0.2623218	**1.3752734**	0.9768609	**1.0001560**	0.9999975
10	**1.0516061**	0.9184716	**1.0194313**	**1.0153412**	0.9897570	**1.0007440**	0.9999942	**1.0000001**
11	**1.1599374**	**1.1420655**	0.5666853	0.9726728	**1.0805571**	0.9896328	**1.0001324**	0.9999959
12	**1.0287419**	**1.1462505**	**1.1239025**	0.9904235	0.9703944	**1.0065972**	0.9998709	**1.0000063**
13	0.6294814	0.8086885	**1.5402409**	**1.6016306**	**1.4284747**	0.5262981	**1.0220557**	0.9972744
14	0.2698295	0.3735544	0.9346358	**1.1834574**	**1.2117964**	**1.1760879**	0.9577953	**1.0336192**
15	0.2223497	0.3103555	0.7981199	**1.0326740**	**1.0761210**	**1.1416687**	**1.0824059**	0.7858279
16	0.1419477	0.2007686	0.5392025	0.7222548	0.7745486	0.9395089	**1.1260308**	0.9495891
17	0.1127851	0.1602589	0.4368573	0.5922269	0.6415391	0.8131933	**1.0518961**	**1.0723384**
18	0.0720490	0.1030221	0.2865294	0.3948116	0.4336212	0.5824223	0.8317614	**1.1114234**
19	0.0353311	0.0507994	0.1437722	0.2009436	0.2233895	0.3151033	0.4885267	0.8163537
20	0.0333140	0.0479136	0.1357323	0.1898528	0.2111988	0.2986851	0.4650820	0.7862053
21	0.0202728	0.0292134	0.0832590	0.1170343	0.1307458	0.1880095	0.3009302	0.5471239
22	0.0050582	0.0073052	0.0209655	0.0296391	0.0332742	0.0487651	0.0805295	0.1587408
23	0.0029205	0.0042192	0.0121206	0.0171487	0.0192651	0.0283084	0.0469497	0.0935731
24	0.0018375	0.0026551	0.0076311	0.0108011	0.0121383	0.0178599	0.0296852	0.0594933

<표 2> (연속)

고유방향 j	인자수 m							
	17	18	19	20	21	22	23	24
1	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
2	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
3	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
4	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
5	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
6	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
7	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
8	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
9	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
10	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
11	**1.0000001**	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
12	0.9999999	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
13	**1.0001276**	0.9999896	**1.0000006**	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
14	0.9940270	**1.0016872**	0.9997091	**1.0000071**	1.0000000	1.0000000	1.0000000	1.0000000
15	**1.0576994**	0.9774143	**1.0051769**	0.9998403	**1.0000007**	1.0000000	1.0000000	1.0000000
16	0.8954743	**1.1192965**	0.9410151	**1.0031443**	0.9999758	**1.0000026**	1.0000000	1.0000000
17	**1.0070854**	0.9591509	**1.0341812**	0.9975541	**1.0000250**	0.9999965	1.0000000	1.0000000
18	**1.2314917**	**1.0170466**	0.5046608	**1.0667690**	0.9987404	**1.0003233**	0.9999972	1.0000000
19	**1.1098764**	**1.2008956**	**1.1287361**	0.8792057	**1.0289581**	0.9721995	**1.0005145**	1.0000000
20	**1.0810069**	**1.1922335**	**1.1904082**	0.9715945	0.9444149	**1.0627281**	0.9987612	1.0000000
21	0.8083214	**1.0100389**	**1.4399625**	**1.6983135**	0.9947972	0.9564494	**1.0015162**	1.0000000
22	0.2542736	0.3663714	0.7302764	**1.2517745**	**1.3727341**	**1.2592333**	0.9230022	1.0000000
23	0.1515541	0.2226464	0.4623342	0.8285242	0.9918333	**1.0038498**	**1.0215410**	1.0000000
24	0.0968961	0.1437455	0.3045902	0.5578935	0.6970128	0.7417211	0.8691119	1.0000000

PLS의 $f_{j(pls)}$는 복잡한 비선형이어서 주어진 데이터에 따라 형태가 달라지며 그 성질을 알기가 어렵다. 그러나 <표 1>과 <표 2>의 축소인자 $f_{j(pls)}$ 값을 보면 PLS의 특이한 축소구조가 어느 정도 드러난다. m개 인자의 모형에서, m번째 고유방향과 그 전의 몇 고유방향에서 확대와 축소를 반복하는 패턴이 있고 m번째와 그 이후 한두 개의 고유방향은 연속적으로 확대하고 그 이후는 급속히 축소하는 경향을 보였다. m개 인자의 모형에서 m번째 고유방향으로는 뚜렷하게 확대하는 패턴을 보였으나, 일부 그렇지 않는 경우도 나타났으며 데이터 셋 2에서 인자수 $m = 9, 10, 11, 12, 16, 20, 21$에서는 각각의 9, 10, 11, 12, 16, 20, 21번째 고유방향으로 확대하지 않고 오히려 축소하였다. 전체적으로, 각 고유방향으로 팽창하는 경우 1 보다 약간 큰 경우가 대부분이나 팽창이 상대적으로 매우 큰 것도 일부 나타났다. <표 1>에서는 $m = 2$의 두 번째 고유방향의 $f_{j(pls)}$ 값은 1.482이고 <표 2>에서는 $m = 1$

의 첫 번째 고유방향과 $m=20$의 21번째 고유방향의 $f_{j(pls)}$ 값이 각각 1.992와 1.698로 상당한 팽창을 보였다. 인자수가 행렬 $\boldsymbol{X}$의 계수인 24일 때는 모든 $f_{j(pls)}$ 값이 1이 되어 OLS가 됨이 확인된다.

참고문헌

1. Brereton, R. (2003). *Chemometrics: Data Aanalysis for the Laboratory and Chemical Plant*, John Wiley & Sons.
2. Butler, N. and Denham, M. (2000). The Peculiar Shrinkage Properties of Partial Least Squares Regression, *Journal of Royal Statistical Society*, B, 62, Part 3, 585-593.
3. Di Ruscio, D. (2000). A Weighted View on the Partial Least Squares Algorithm, *Automatica*, 36, 831-850.
4. Fearn, T. (1983). A Misuse of Ridge Regression in the Calibration of a Near Infrared Reflectance Instrument, *Journal of Applied Statistics*, 32, 73-79.
5. Frank, I. E. & Friedman, J. H. (1993). A Statistical View of Some Chemometrics Regression Tools (with Discussion), *Technometrics*, 35, 109-148.
6. Helland, I. S. (1988). On the structure of partial least squares regression, *Communications in Statistics - Simulation and Computation*, 17, 581-607.
7. Kim, J. D. (2003). Unified Non-iterative algorithm for Principal Component Regression, Partial Least Squares and Ordinary Least Squares, *Journal of Korean Data & Information Science Society*, 14, 355-366.
8. Kim, J. D. and S. Moon (2005). Connecting Partial Least Squares and Generalized Ridge Regression, *Journal of the Korean Data Analysis Society*, 7, 61-71.
9. Wold, H. (1975). Soft Modelling by Latent Variables; the Nonlinear Iterative Partial Least Squares Approach, in *Perspectives in Probability and Statistics*. Papers in Honour of M. S. Bartlett (ed. J. Gani). Academic Press, New York.

[2006년 9월 접수, 2006년 11월 채택]