Shrinkage structure in biased regression

Pierre Druilhet [a,*] Alain Mom [b]

[a] *CREST-ENSAI, Campus de Ker Lann, 35 170 BRUZ, France*
[b] *IRMAR et Université Rennes II, RENNES, France*

Abstract

Biased regression is an alternative to ordinary least squares (OLS) regression, especially when explanatory variables are highly correlated. In this paper, we examine the geometrical structure of the shrinkage factors of biased estimators. We show that, in most cases, shrinkage factors cannot belong to $[0,1]$ in all directions. We also compare the shrinkage factors of ridge regression (RR), principal component regression (PCR) and partial least squares regression (PLSR) in the orthogonal directions obtained by the signal-to-noise ratio (SNR) algorithm. In these directions, we find that PLSR and RR behave well whereas shrinkage factors of PCR have an erratic behaviour.

Key words: Biased regression, regression on components, shrinkage factors, signal-to-noise ratio, James-Stein estimator,
1991 MSC: 62J05, 62J07.

1 Introduction

Biased regressions are sometimes preferred to ordinary least squares (OLS) regression in order to improve the mean-squared error (MSE). When the explanatory variables are highly correlated, the most popular biased regression are Ridge regression (RR), principal component regression (PCR) and partial least square regression (PLSR). These three methods provide shrunk estimators in the sense that their Euclidean norms are lower than that of the OLS estimator. However, this overall feature does not give any indication of the

* Corresponding author
Email addresses: druilhet@ensai.fr (Pierre Druilhet), alain.mom@uhb.fr (Alain Mom).

shrinkage behaviour in specific directions. Frank and Friedman (1993) compared the shrinkage properties of these three estimators in the principal directions, i.e. in the directions given by the singular value decomposition of the design matrix. They pointed out that, in these directions, shrinkage factors for PLS estimators may be greater than 1 (in that case the term "shrinkage" is abusive).

In this paper, we show that, for any biased estimator, except for estimators proportional to the OLS estimator, there exist directions in which the shrinkage factors are greater than 1 and even equal to $+\infty$. Then, we study the shrinkage factors of PLSR, PCR and RR in the directions given by the SNR algorithm introduced by Druilhet and Mom (2006). Whereas principal directions can be seen as orthogonal directions that iteratively minimise the variance of the OLS estimator, the SNR algorithm constructs orthogonal directions that iteratively maximise the signal-to-noise ratio (SNR). Since the SNR appears when we seek optimal directional shrinkage factors, shrinkage behaviours in these directions are of interest. As expected, the peculiar behaviour of PLSR disappears, the shrinkage factors of PCR may become erratic and surprisingly those of RR behave well.

2 Shrinkage structure of biased estimators

We consider the centered linear model:

$$y = X\beta + \varepsilon. \tag{1}$$

where X is the (n,p) design matrix, β the p-vector of parameters and ε the n-vector of i.i.d. mean zero variance σ^2 errors. We write $S = X'X$ and $s = X'y$. For simplicity, from now on, we assume that S is of full rank. The non-full-rank case is equivalent except that we restrict the directions considered to those corresponding to estimable functions. We denote by $\widehat{\beta}^{\text{OLS}} = S^{-1}\,s$ the OLS estimator and by $\widehat{\beta}^*$ a competing estimator.

2.1 Geometrical structure of shrinkage factors

Let x be a p-vector. We define the shrinkage factor of $\widehat{\beta}^*$ in the direction x by:

$$\alpha_{x,\widehat{\beta}^*} = \frac{x'\widehat{\beta}^*}{x'\widehat{\beta}^{\text{OLS}}} \tag{2}$$

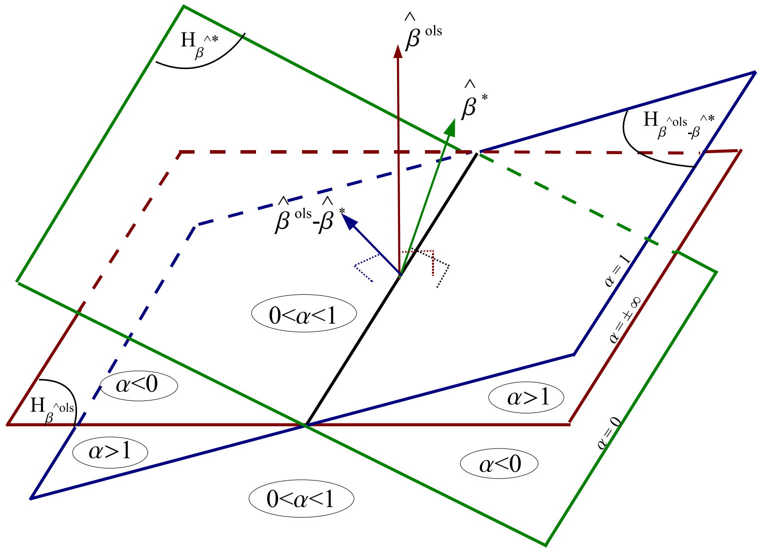

Fig. 1. Geometrical shrinkage structure of biased estimators

with the following convention:

$$\alpha_{x,\widehat{\beta}^*} = \frac{0}{0} \quad \text{if } x'\widehat{\beta}^{\text{OLS}} = 0 \text{ and } x'\widehat{\beta}^* = 0,$$

and

$$\alpha_{x,\widehat{\beta}^*} = \pm\infty \quad \text{if } x'\widehat{\beta}^{\text{OLS}} = 0 \text{ and } x'\widehat{\beta}^* \neq 0.$$

Note that in Druilhet and Mom (2006) the term "shrinkage factor" has another meaning: it corresponds to an optimal factor applied to the OLS estimator on a direction.

Since shrinkage factors are scale invariant, i.e. for non-zero scalar δ,

$$\alpha_{\delta x,\widehat{\beta}^*} = \alpha_{x,\widehat{\beta}^*}, \tag{3}$$

$\alpha_{x,\widehat{\beta}^*}$ depends only on the direction given by x, not on the exact value of x. Note that $\alpha_{x,\widehat{\beta}^*}$ is actually a shrinkage factor only if it belongs to $[0, 1]$. We want to determine when a competing estimator has its shrinkage factors in $[0, 1]$ in all directions and if it has not, what can we say about the shrinkage structure. We assume that $\widehat{\beta}^{\text{OLS}} \neq 0$, which arises with probability 1 in most cases, and that $\widehat{\beta}^* \neq 0$ (otherwise, all shrinkage factors are equal to 0 or $\frac{0}{0}$).

For $u \in \mathbb{R}^p$, we denote $H_u = \{x \in \mathbb{R}^p \;/\; x'u = 0\} = u^\perp$, $H_u^+ = \{x \in \mathbb{R}^p \;/\; x'u > 0\}$ and $H_u^- = \{x \in \mathbb{R}^p \;/\; x'u < 0\}$. The three hyperplanes $H_{\widehat{\beta}^{\text{OLS}}}$, $H_{\widehat{\beta}^{\text{OLS}}-\widehat{\beta}^*}$ and $H_{\widehat{\beta}^*}$ correspond to the directions in which shrinkage factors are respectively either $\pm\infty$ or $\frac{0}{0}$, either 1 or $\frac{0}{0}$ and either 0 or $\frac{0}{0}$. They are displayed in Fig. 1. If $\widehat{\beta}^*$ is proportional to $\widehat{\beta}^{\text{OLS}}$ but different, the three hyperplanes coincide. If not,

$$H_{\widehat{\beta}^{\text{OLS}}} \cap H_{\widehat{\beta}^*} = H_{\widehat{\beta}^*} \cap H_{\widehat{\beta}^{\text{OLS}}-\widehat{\beta}^*} = H_{\widehat{\beta}^{\text{OLS}}} \cap H_{\widehat{\beta}^{\text{OLS}}-\widehat{\beta}^*} = \{x \ / \ \alpha_{x,\widehat{\beta}^*} = \frac{0}{0}\} \qquad (4)$$

is a linear subspace of dimension $p-2$ and the region where the shrinkage factors are greater than 1 is

$$(H^-_{\widehat{\beta}^{\text{OLS}}-\widehat{\beta}^*} \cap H^+_{\widehat{\beta}^{\text{OLS}}}) \cup (H^+_{\widehat{\beta}^{\text{OLS}}-\widehat{\beta}^*} \cap H^-_{\widehat{\beta}^{\text{OLS}}}) = \{x \ / \ 1 < \alpha_{x,\widehat{\beta}^*} < +\infty\} \neq \emptyset. \qquad (5)$$

These geometrical considerations show that, except in the case of an estimator proportional to $\widehat{\beta}^{\text{OLS}}$ such as, for instance, the James-Stein estimator (see Stein, 1956), it is hopeless to seek a competing estimator $\widehat{\beta}^*$ whose shrinkage factors belong to $[0,1]$ in all the directions of $\mathbb{R}^p$, as stated in the following proposition:

Proposition 1 *The shrinkage factors of a competing estimator* $\widehat{\beta}^*$ *are in* $[0,1]$ *in all directions if and only if* $\widehat{\beta}^*$ *is proportional to* $\widehat{\beta}^{\text{OLS}}$*, i.e.:*

$$\Updownarrow \begin{array}{l} a)\ \forall x \in \mathbb{R}^p, \quad 0 \leq \alpha_{x,\widehat{\beta}^*} \leq 1, \\ b)\ \exists a \in [0,1] \text{ such that } \widehat{\beta}^* = a\,\widehat{\beta}^{\text{OLS}}. \end{array}$$

where the scalar a may be random.

PROOF. $(b) \Rightarrow (a)$ is obvious. Now, suppose that (b) is false. Either $\widehat{\beta}^* = \alpha\widehat{\beta}^{\text{OLS}}$ with $\alpha \notin [0,1]$ and the result follows, or $\widehat{\beta}^*$ is not proportional to $\widehat{\beta}^{\text{OLS}}$. In that case, $H_{\widehat{\beta}^{\text{OLS}}}$, $H_{\widehat{\beta}^{\text{OLS}}-\widehat{\beta}^*}$ and $H_{\widehat{\beta}^*}$ are distinct and the region where the shrinkage factor is greater than 1 is not empty as seen in Eq. (5)

When $\widehat{\beta}^*$ is not proportional to $\widehat{\beta}^{\text{OLS}}$, it is possible to find shrinkage factors arbitrarily large when the direction is getting closer to $H_{\widehat{\beta}^{\text{OLS}}} \setminus H_{\widehat{\beta}^*}$.

2.2 Shrinkage factors for regression on components

Among biased regressions, regressions on components, such as PCR, PLSR and more generally continuum regression (Stone et Brooks, 1990), are widely used. We give here some general features of shrinkage factors for regression on components, useful for the following. Let $w_1, ..., w_q$ be q linearly independent p-vectors ($q \leq p$). The estimator $\widehat{\beta}_q$ of β obtained by regression on the q components $t_1 = X\,w_1, ..., t_q = X\,w_q$ is defined by

$$\widehat{\beta}_q = W_q\,(T'_{(q)}\,T(q))^{-1}T'_{(q)}\,Y = W_{(q)}\left(W_{(q)}{}'\,S\,W_{(q)}\right)^{-1}W_{(q)}{}'S\,\widehat{\beta}^{\text{OLS}} = \mathrm{P}^S_{W_{(q)}}\,\widehat{\beta}^{\text{OLS}}, \tag{6}$$

where $W_{(q)} = (w_1, ..., w_q)$, $T_{(q)} = (t_1, ..., t_q) = X\,W_{(q)}$ and $\mathrm{P}^S_{W_{(q)}}$ is the projection matrix onto range$(W_{(q)})$ w.r.t. the quadratic form S. In that context, $w_1, ..., w_q$ are often called weight vectors. Another interpretation of $\widehat{\beta}_q$ can be given in terms of constrained least squares estimator: denote by $w_{q+1}, ..., w_p$, $p-q$ linearly independent p-vectors orthogonal to $w_1, ..., w_q$, then

$$\widehat{\beta}_q = \underset{w_i'\beta=0,\, i=q+1,...,p}{\text{ArgMin}} ||Y - X\beta||^2 \tag{7}$$

Now, $w_{q+1}, ..., w_p$ can be seen as directions in $\mathbb{R}^p$ where the least squares estimator is constrained to be null and therefore, we have:

$$\alpha_{x,\widehat{\beta}_q} = 0 \text{ or } \frac{0}{0}, \quad \forall x \in \text{range}(W_{(q)})^{\perp}. \tag{8}$$

Since $(\mathrm{P}^S_{W_{(q)}})' = \mathrm{P}^{S^{-1}}_{SW_{(q)}}$, Eq. (6) gives

$$\alpha_{x,\widehat{\beta}_q} = 1 \text{ or } \frac{0}{0} \quad \forall x \in \text{span}(S\,w_1, ..., S\,w_q) \tag{9}$$

However, in the general case, the shrinkage factors of $\widehat{\beta}_q$ in the directions $w_1, ...w_q$ are not necessarily in $[0, 1]$. For example, consider $\widehat{\beta}^{\text{OLS}} = (0, 1)'$, $w_1 = (1, 0)'$ and $S = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. The estimator obtained by regression on the component $t_1 = X\,w_1$ is $\widehat{\beta}_1 = (1/2, 0)'$. The shrinkage factor $\alpha_{w_1,\widehat{\beta}_1}$ for $\widehat{\beta}_1$ in the direction w_1 is equal to $\pm\infty$.

In the case of PCR, we denote by $w_1^{\text{PCR}}, ..., w_p^{\text{PCR}}$ the principal directions, i.e. the eigenvectors corresponding to decreasingly ordered eigenvalues $\lambda_1, ..., \lambda_q$ of S. We denote by $\widehat{\beta}_q^{\text{PCR}}$ the PCR estimator obtained by regression on the q principal components $t_1^{\text{PCR}} = X\,w_1^{\text{PCR}}, ..., t_q^{\text{PCR}} = X\,w_q^{\text{PCR}}$. Since the directions given by w_i^{PCR} and $S\,w_i^{\text{PCR}}$ are the same, Eq. (9) gives

$$\alpha_{w_i,\widehat{\beta}_q^{\text{PCR}}} = 1 \text{ or } \frac{0}{0} \quad \text{for } i \leq q,$$

and Eq. (8) gives

$$\alpha_{w_i,\widehat{\beta}_q^{\text{PCR}}} = 0 \text{ or } \frac{0}{0} \quad \text{for } i > q.$$

3 Shrinkage factors in the SNR directions

In the principal directions, shrinkage factors of PCR and RR belong to $[0,1]$. In the case of PLSR, Frank and Friedman (1993), Butler and Denham (2000), Lingjærde and Christophersen (2000) and Krämer (2005) showed that shrinkage factors in these directions can be outside of $[0,1]$. However, as explained in Section 2, biased regressions such as RR, PCR or PLSR cannot shrink in all directions and PLS estimators are not adapted to the principal directions whose construction only depends on the variance of the OLS estimator. The idea here is to compare the behaviour of PLSR, PCR, and RR in another orthogonal system of directions based on directional SNRs. The SNRs depend both on the variance and the actual value of the OLS estimator. They are related to optimal directional shrinking and are involved in the construction of PLSR estimators.

3.1 The SNR algorithm and related directions

The SNR arises in the following problem: consider a direction $x \in \mathbb{R}^p$. We want to improve the best linear unbiased estimator $x'\widehat{\beta}^{\text{OLS}}$ of $x'\beta$ by shrinking. We consider the class of estimators $a\,x'\widehat{\beta}^{\text{OLS}}$ for $a \in \mathbb{R}$. The scalar a_x^* that minimizes the quadratic risk $\mathbb{E}\left(a\,x'\widehat{\beta}^{\text{OLS}} - x'\beta\right)^2$ is equal to $\frac{\rho_x^2}{1+\rho_x^2}$ where $\rho_x = \frac{|x'\,\beta|}{\sigma\,\sqrt{x'\,S^{-1}x}}$. If σ is known, a_x^* can be estimated by $\widehat{a}_x^* = \frac{\widehat{\rho}_x^2}{1+\widehat{\rho}_x^2}$, where $\widehat{\rho}_x = \frac{|x'\,\widehat{\beta}^{\text{OLS}}|}{\sigma\,\sqrt{x'\,S^{-1}x}}$ is the SNR in the direction given by x. Note that $\widehat{a}_x^*$ is an increasing function of $\widehat{\rho}_x$, thus maximizing $\widehat{a}_x^*$ w.r.t. x is equivalent to maximizing $\widehat{\rho}_x$.

The SNR algorithm (Druilhet and Mom, 2006) seeks orthogonal directions that successively maximise $\widehat{\rho}_x$. At step one, the first direction w_1^{SNR} is:

$$w_1^{\text{SNR}} = \underset{w\in\mathbb{R}^p}{\text{ArgMax}}\,\widehat{\rho}_w \propto s. \tag{10}$$

Iteratively, at step i, w_i^{SNR} is the direction that maximises $\widehat{\rho}_w$ under the orthogonality constraint $w \perp \text{span}(w_1^{\text{SNR}}, ..., w_{i-1}^{\text{SNR}})$. Denote by q^* the greatest q such that $\widehat{\rho}_{w_q^{\text{SNR}}} \neq 0$. For $q \leq q^*$, w_q^{SNR} belongs to the Krylov subspace

$$K_q = \text{span}(s,\, S\,s,\, S^2\,s, ..., S^{q-1}\,s). \tag{11}$$

Note that q^* is also the smallest q satisfying $K_{q+1} = K_q$. For $q \geq q^*$, $\widehat{\rho}_{w_q} = 0$ and $(w_{q^*+1}^{\text{SNR}}, ..., w_p^{\text{SNR}})$ is any system of orthogonal directions that are orthogonal to K_{q^*}.

The PLS estimators $\widehat{\beta}_q^{\text{PLS}}$, $q = 1, ..., p$, can be obtained from the Krylov subspaces $K_1, ..., K_{q^*}$, and therefore from the SNR directions, by:

$$\widehat{\beta}_q^{\text{PLS}} = \underset{\beta \in K_q}{\text{ArgMin}} \, ||Y - X\beta||^2 = \underset{w_i^{\text{SNR}\prime}\beta=0,\; i=q+1,\dots,p}{\text{ArgMin}} \, ||Y - X\beta||^2, \tag{12}$$

(see Helland, 1988). For $q \geq q^*$, we have:

$$\widehat{\beta}_q^{\text{PLS}} = \widehat{\beta}^{\text{OLS}}. \tag{13}$$

Equivalently, $\widehat{\beta}_q^{\text{PLS}}$ can be obtained by regression on the components $t_1 = X\,w_1^{\text{SNR}}, ..., t_q = X\,w_q^{\text{SNR}}$. Denote $W_{(q)}^{\text{SNR}} = (w_1^{\text{SNR}}, ..., w_q^{\text{SNR}})$, then Eq. (6) gives:

$$\widehat{\beta}_q^{\text{PLS}} = \text{P}^S_{W_{(q)}^{\text{SNR}}} \, \widehat{\beta}^{\text{OLS}}. \tag{14}$$

SNR directions and PLS estimators can also be obtained simultaneously by the ΔSNR algorithm (Druilhet and Mom, 2006): at step one, we put $\widehat{\beta}_0^{\text{PLS}} = 0$ and we seek the direction w_1 that maximises $\Delta\text{SNR}_0(x) = \frac{|x'(\widehat{\beta}^{\text{OLS}} - \widehat{\beta}_0^{\text{OLS}})|}{\sigma\sqrt{x'S^{-1}x}}$. We find $w_1 = w_1^{\text{SNR}}$ and we define $\widehat{\beta}_1^{\text{PLS}}$ to be the least squares estimator constrained to belong to $\text{span}\, w_1^{\text{SNR}}$. Iteratively, at step q, w_q^{SNR} maximises $\Delta\text{SNR}_{q-1}(x) = \frac{|x'(\widehat{\beta}^{\text{OLS}} - \widehat{\beta}_{q-1}^{\text{PLS}})|}{\sigma\sqrt{x'S^{-1}x}}$ and $\widehat{\beta}_q^{\text{PLS}}$ is the least square estimator constrained to belong to $\text{span}(w_1^{\text{SNR}}, ..., w_q^{\text{SNR}}) = K_q$. This algorithm leads to interesting formulae. In particular, the vectors w_q may be chosen (up to a multiplicative constant) as

$$w_q^{\text{SNR}} = s - S\widehat{\beta}_{q-1}^{\text{PLS}} \qquad \text{for } q = 1, ..., q^*. \tag{15}$$

which gives:

$$\widehat{\rho}_{w_{q+1}^{\text{SNR}}} = \frac{1}{\sigma}\sqrt{s'\widehat{\beta}^{\text{OLS}} - s'\widehat{\beta}_q^{\text{PLS}}}. \tag{16}$$

We also have:

$$\Delta\text{SNR}_q(w_{q+1}^{\text{SNR}}) = \widehat{\rho}_{w_{q+1}^{\text{SNR}}}. \tag{17}$$

The following results establish a relationship between shrinkage factors of PLS estimators in a direction x and the SNR in the same direction.

Proposition 2

$$\Delta\text{SNR}_q(x) = \widehat{\rho}_x \; |1 - \alpha_{x,\widehat{\beta}_q^{\text{PLS}}}| \tag{18}$$

and

$$|1 - \alpha_{x,\widehat{\beta}_q^{\text{PLS}}}| \leq \frac{\widehat{\rho}_{w_{q+1}^{\text{SNR}}}}{\widehat{\rho}_x} \quad \forall q = 1, .., q^* - 1. \tag{19}$$

PROOF. Eq. (18) is straightforward. Since w_{q+1}^{SNR} maximises $\Delta\text{SNR}_q(x)$, we have $\Delta\text{SNR}(x) \leq \Delta\text{SNR}(w_{q+1}^{\text{SNR}})$. The result follows from Eq. (17) and (18).

We now examine the shrinkage behaviour of PLSR estimators onto the SNR directions. We saw in Section 2.2 that the fact that $\widehat{\beta}_q^{\text{PLS}}$ is obtained by regression on the components $t_1^{\text{PLS}} = X\,w_1^{\text{PLS}}, ..., t_q^{\text{PLS}} = X\,w_q^{\text{PLS}}$ does not necessarily imply that the shrinkage factors in the directions $w_1^{\text{PLS}}, ..., w_q^{\text{PLS}}$ belong to $[0,1]$. However, we shall show that this property holds for PLSR.

Consider the shrinkage factors given in the table below:

$$\begin{array}{c|cccccc}
 & \widehat{\beta}_1^{\text{PLS}} & \widehat{\beta}_2^{\text{PLS}} & \cdots & \cdots & \widehat{\beta}_{q^*-1}^{\text{PLS}} & \widehat{\beta}_{q^*}^{\text{PLS}} \\
\hline
w_1^{\text{SNR}} & \alpha_{w_1^{\text{SNR}},\widehat{\beta}_1^{\text{PLS}}} & \alpha_{w_2^{\text{SNR}},\widehat{\beta}_1^{\text{PLS}}} & \cdots & \cdots & \alpha_{w_1^{\text{SNR}},\widehat{\beta}_{q^*-1}^{\text{PLS}}} & 1 \\
w_2^{\text{SNR}} & 0 & \alpha_{w_2^{\text{SNR}},\widehat{\beta}_2^{\text{PLS}}} & \cdots & \cdots & \alpha_{w_2^{\text{SNR}},\widehat{\beta}_{q^*-1}^{\text{PLS}}} & 1 \\
\vdots & \vdots & 0 & & & \vdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & & \vdots & \vdots \\
w_{q^*}^{\text{SNR}} & 0 & 0 & \cdots & 0 & \alpha_{w_{q^*}^{\text{SNR}},\widehat{\beta}_{q^*-1}^{\text{PLS}}} & 1
\end{array} \tag{20}$$

We are interested in the "horizontal" and "vertical" behaviour of the shrinkage factors. Note that, by Eq. (8), the lower triangular matrix is 0.

Lemma 3 *For $q = 1, ..., q^*$ and $u = 0, ..., q^* - q$,*

$$\widehat{\beta}_{q+u}^{\text{PLS}}{}' S\, \widehat{\beta}_q^{\text{PLS}} = \widehat{\beta}_q^{\text{PLS}}{}' S\, \widehat{\beta}_q^{\text{PLS}} \tag{21}$$

Moreover, $q \longmapsto \widehat{\beta}_q^{\text{PLS}}{}' S\, \widehat{\beta}_q^{\text{PLS}}$ is decreasing for $q \in \{1, ..., q^*\}$.

PROOF.

Since $\text{range}(W_{(q)}) \subset \text{range}(W_{(q+u)})$, $P^S_{W^{\text{SNR}}_{(q)}} = P^S_{W^{\text{SNR}}_{(q)}} P^S_{W^{\text{SNR}}_{(q+u)}}$ and

$$\widehat{\beta}_q^{\text{PLS}} = \mathrm{P}^S_{W^{\text{SNR}}_{(q)}} \widehat{\beta}^{\text{OLS}} = \mathrm{P}^S_{W^{\text{SNR}}_{(q)}} \widehat{\beta}_{q+u}^{\text{PLS}}.$$

As $\widehat{\beta}_q^{\text{PLS}} = \mathrm{P}^S_{W^{\text{SNR}}_{(q)}} \widehat{\beta}_q^{\text{PLS}}$, $\widehat{\beta}_{q+u}^{\text{PLS}}{}'\, S\, \widehat{\beta}_q^{\text{PLS}} = (\mathrm{P}^S_{W_{(q)}} \widehat{\beta}_{q+u}^{\text{PLS}})'\, S\, \widehat{\beta}_q^{\text{PLS}}$ and Eq. (21) follows.

Since $\widehat{\beta}_q^{\text{PLS}}{}' S\, \widehat{\beta}_q^{\text{PLS}}$ is the norm of $\widehat{\beta}_q^{\text{PLS}}$ w.r.t. S, $\widehat{\beta}_{q+1}^{\text{PLS}}{}' S \widehat{\beta}_{q+1}^{\text{PLS}} \geq \widehat{\beta}_q^{\text{PLS}}{}' S\, \widehat{\beta}_q^{\text{PLS}}$. Equality holds iff $\widehat{\beta}_{q+1}^{\text{PLS}} = \widehat{\beta}_q^{\text{PLS}}$. In that case, $\Delta\text{SNR}_{q+1} = \Delta\text{SNR}_q$. Since w_{q+2} maximises ΔSNR_{q+1}, it also maximises ΔSNR_q, and then $w_{q+2} \in K_{q+1}$, i.e. $K_{q+2} = K_{q+1}$ and $q + 1 \geq q^*$. By Eq. (13), $\widehat{\beta}_{q+1}^{\text{PLS}} = \widehat{\beta}^{\text{OLS}}$. Therefore $\widehat{\beta}_q^{\text{PLS}} = \widehat{\beta}^{\text{OLS}}$ and $q \geq q^*$.

Lemma 4 *For $q = 1, ..., q^*$ and $u = 0, ..., q^* - q$,*

$$w_q^{\text{SNR}\prime}\, \widehat{\beta}_{q+u}^{\text{PLS}} = \widehat{\beta}_{q+u}^{\text{PLS}\prime}\, S\, \widehat{\beta}_{q+u}^{\text{PLS}} - \widehat{\beta}_{q-1}^{\text{PLS}\prime}\, S\, \widehat{\beta}_{q-1}^{\text{PLS}} \tag{22}$$

PROOF. By Eq. (15), $w_q^{\text{SNR}\prime}\, \widehat{\beta}_{q+u}^{\text{PLS}} = s'\, \widehat{\beta}_{q+u}^{\text{PLS}} - \widehat{\beta}_{q-1}^{\text{PLS}\prime}\, S\, \widehat{\beta}_{q+u}^{\text{PLS}} = \widehat{\beta}^{\text{OLS}} S \widehat{\beta}_{q+u}^{\text{PLS}} - \widehat{\beta}_{q-1}^{\text{PLS}\prime}\, S\, \widehat{\beta}_{q+u}^{\text{PLS}}$. By (21) applied at both terms, the last expression is equal to $\widehat{\beta}_{q+u}^{\text{PLS}\prime}\, S\, \widehat{\beta}_{q+u}^{\text{PLS}} - \widehat{\beta}_{q-1}^{\text{PLS}\prime}\, S\, \widehat{\beta}_{q-1}^{\text{PLS}}$.

From Lemma 5 and Lemma 4, $w_{q^*}^{\text{SNR}\prime} \widehat{\beta}^{\text{OLS}} \neq 0$ for $q = 1, ..., q^*$ and therefore, the shrinkage factors in the SNR directions are well defined.

Proposition 5 *The shrinkage factors displayed in Table 20 belong to* $[0, 1]$. *Moreover, for $q = 1, ..., q^*$, $\alpha_{w_q^{\text{SNR}}, \widehat{\beta}_u^{\text{PLS}}}$ is increasing w.r.t. u and decreasing w.r.t. q.*

PROOF. For $u = 0, ..., q-1$, $\alpha_{w_q^{\text{SNR}}, \widehat{\beta}_u^{\text{PLS}}} = 0$. We now consider the case $q \leq u \leq q^*$. By Lemma 4, we have $\alpha_{w_q^{\text{SNR}}, \widehat{\beta}_u^{\text{PLS}}} = \frac{w_q^{\text{SNR}\prime}\, \widehat{\beta}_u^{\text{PLS}}}{w_q^{\text{SNR}\prime}\, \widehat{\beta}^{\text{OLS}}} = \frac{\widehat{\beta}_u^{\text{PLS}\prime}\, S\, \widehat{\beta}_u^{\text{PLS}} - \widehat{\beta}_{q-1}^{\text{PLS}\prime}\, S\, \widehat{\beta}_{q-1}^{\text{PLS}}}{\widehat{\beta}_{q^*}^{\text{PLS}\prime}\, S\, \widehat{\beta}_{q^*}^{\text{PLS}} - \widehat{\beta}_{q-1}^{\text{PLS}\prime}\, S\, \widehat{\beta}_{q-1}^{\text{PLS}}}$. By Lemma 3, $\widehat{\beta}_u^{\text{PLS}\prime}\, S\, \widehat{\beta}_u^{\text{PLS}}$ is increasing w.r.t. u, thus $0 \leq \alpha_{w_q^{\text{SNR}}, \widehat{\beta}_u^{\text{PLS}}} \leq 1$ and $\alpha_{w_q^{\text{SNR}}, \widehat{\beta}_u^{\text{PLS}}}$ is increasing w.r.t. u. Now, u is fixed. Since $\alpha_{w_q^{\text{SNR}}, \widehat{\beta}_u^{\text{PLS}}} = 1 - \frac{\widehat{\beta}_{q^*}^{\text{PLS}\prime}\, S\, \widehat{\beta}_{q^*}^{\text{PLS}} - \widehat{\beta}_u^{\text{PLS}\prime}\, S\, \widehat{\beta}_u^{\text{PLS}}}{\widehat{\beta}_{q^*}^{\text{PLS}\prime}\, S\, \widehat{\beta}_{q^*}^{\text{PLS}} - \widehat{\beta}_{q-1}^{\text{PLS}\prime}\, S\, \widehat{\beta}_{q-1}^{\text{PLS}}}$, it is easy to see that $\alpha_{w_q^{\text{SNR}}, \widehat{\beta}_u^{\text{PLS}}}$ is also decreasing w.r.t. q.

Proposition 5 provides a simple proof of a result established independently by De Jong (1995) and Goutis (1996):

Corollary 6 *We have*

$$||\widehat{\beta}_1^{\text{PLS}}|| < ||\widehat{\beta}_2^{\text{PLS}}|| < ... < ||\widehat{\beta}_{q^*}^{\text{PLS}}|| = ||\widehat{\beta}^{\text{OLS}}||.$$

PROOF. From Eq. (12), $w_q^{\text{SNR}\,\prime} \widehat{\beta}_u^{\text{PLS}} = 0$ for $i > q^*$ and $q = 1, ..., q^*$. So we have $||\widehat{\beta}_u^{\text{PLS}}||^2 = \sum_{q=1}^{q^*} \frac{(w_q^{\text{SNR}\,\prime} \widehat{\beta}_u^{\text{PLS}})^2}{||w_q^{\text{SNR}}||^2}$. For $q = 1, ..., q^*$, Proposition 5 implies that $w_q^{\text{SNR}\,\prime} \widehat{\beta}_u$ is increasing w.r.t. u. By Lemmas 4 and 3, $w_q^{\text{SNR}\,\prime} \widehat{\beta}_u \geq 0$ and therefore $(w_q^{\text{SNR}\,\prime} \widehat{\beta}_u)^2$ is also increasing. The result follows.

3.3 Shrinkage factors for RR

Here, we investigate the shrinkage factors of the RR estimators $\widehat{\beta}^{\mathrm{RR}}_{\gamma} = (S + \gamma I)^{-1} s$ in the SNR directions. Surprisingly, their behaviours are comparable to those of PLS.

Proposition 7 *For $q = 1, ..., q^*$ and $\gamma \geq 0$, the shrinkage factors $\alpha_{w_q^{\mathrm{SNR}}, \widehat{\beta}_{\gamma}^{\mathrm{RR}}}$ belong to $[0, 1]$. They are increasing w.r.t. γ.*

The proof is technical and is given in the appendix. The key point is that the Krylov subspaces generated by (S, s) and by $(S + \gamma I, s)$ are the same. In the light of Section 3.4, we conjecture that $\alpha_{w_q^{\mathrm{SNR}}, \widehat{\beta}_{\gamma}^{\mathrm{RR}}}$ is decreasing w.r.t. q.

3.4 Examples

We compare the shrinkage factors of PLSR, PCR and RR in the SNR directions for several artificial and real data sets. They are represented in Figures 2-4. We use the canonical model to describe the data, i.e. we express S and $\widehat{\beta}^{\mathrm{OLS}}$ by using the spectral decomposition $S = \sum_i \lambda_i w_i^{\mathrm{PCR}} w_i^{\mathrm{PCR}\prime}$. We denote $\widehat{\beta}_i = w_i^{\mathrm{PCR}\prime} \widehat{\beta}^{\mathrm{OLS}}$. The two artificial data comes from Frank and Friedman (1993): $\{\widehat{\beta}_i = 1\}_j$ $\{\lambda_i = 1/j\}_j$ (neutral $\widehat{\beta}^{\mathrm{OLS}}$, moderate collinearity) and $\{\widehat{\beta}_i = 1/j\}_j$ $\{\lambda_i = 1/j^2\}_j$ (favourable $\widehat{\beta}^{\mathrm{OLS}}$, high collinearity), where $j = 1, ..., 10$. The real data come from calibration experiments to determine the chemical composition of liquid detergent by using mid-infra-red spectroscopy (Brown, 1990 and Butler and Denham, 2000). The data in their canonical form are $(\lambda_1, ..., \lambda_{12})$ =(8.0059, 6.0324, 1.4529, 0.8665, 0.0201, 0.0122, 0.0053, 0.0033, 0.0020, 0.0017, 0.0014) and $(\widehat{\beta}_1, ..., \widehat{\beta}_{12})$ = (2.7837, -1.3266, -7.2850, 4.21118, -0.3483, -2.7951, 0.0755, -0.5455, -1.95, 0.2941, 2.7857).

For each data set, we have displayed the shrinkage factors w.r.t. the directions w_i^{SNR}, $i = 1, ..., 10$ or 11 for $\widehat{\beta}_q^{\mathrm{PLS}}$, $\widehat{\beta}_q^{\mathrm{PCR}}$ and $\widehat{\beta}_{\gamma}^{\mathrm{RR}}$, for q varying from 1 to 8 and γ chosen such that $||\widehat{\beta}_q^{\mathrm{PLS}}|| = ||\widehat{\beta}_{\gamma}^{\mathrm{RR}}||$.

In all cases, we see that the shrinkage factors of PLS estimators in the SNR directions have a good behaviour as expected. Ridge estimators have a smooth behaviour and the decrease of the shrinkage factors is slower than for PLS. We also observe that the shrinkage factors for PCR are more erratic and can take very large positive or negative values.

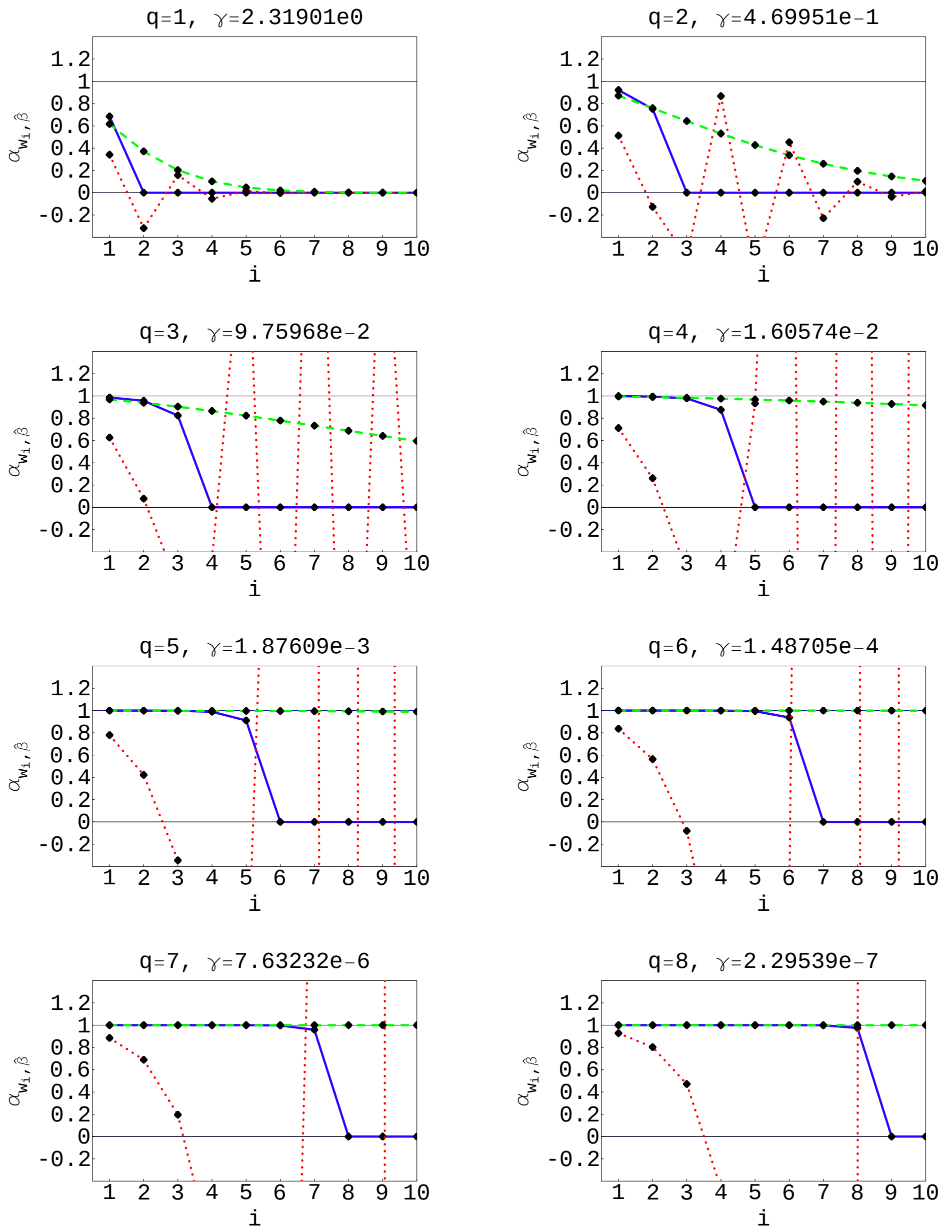

Fig. 2. Shrinkage factors in SNR directions for PLSR (solid), RR (dashed) and PCR (dotted) for neutral $\widehat{\beta}^{\text{OLS}}$ and moderate collinearity.

4 Discussion

In this paper, we have demonstrated that biased estimators such as RR, PCR or PLSR cannot shrink in all directions. Therefore, they favour certain directions to the detriment of others, and a peculiar behaviour on a specific direction does not necessarily lead to a bad overall behaviour. PCR estimators are constructed to shrink in the principal directions and have peculiar

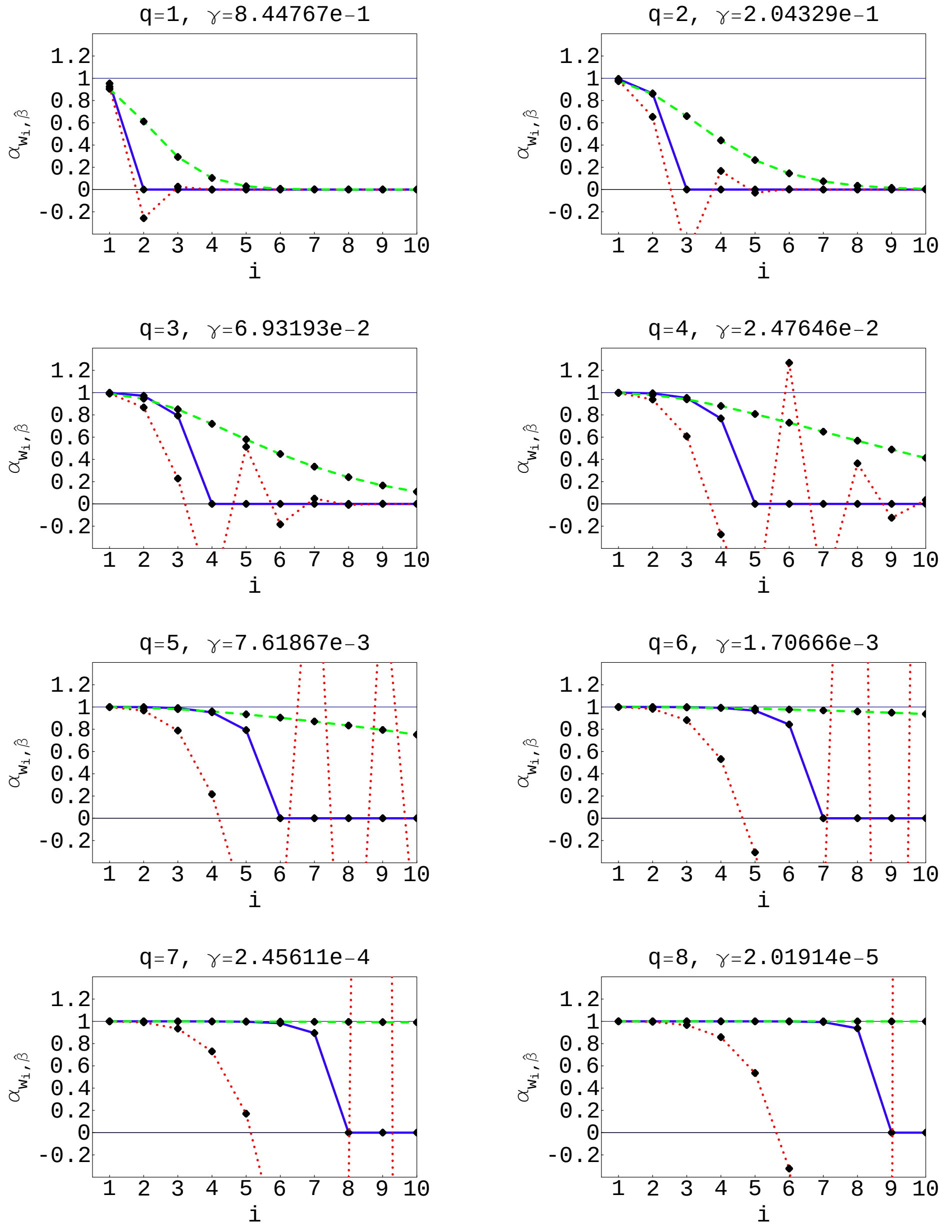

Fig. 3. Shrinkage factors in SNR directions for PLSR (solid), RR (dashed) and PCR (dotted) for favourable $\widehat{\beta}^{\text{OLS}}$ and high collinearity.

shrinkage behaviour in the SNR directions. Conversely, PLS estimators are based on the SNR directions and have peculiar shrinkage behaviour in the principal directions. The RR estimators are known to minimise a bayesian risk and surprisingly shrink in both the SNR and principal directions in a smooth way.

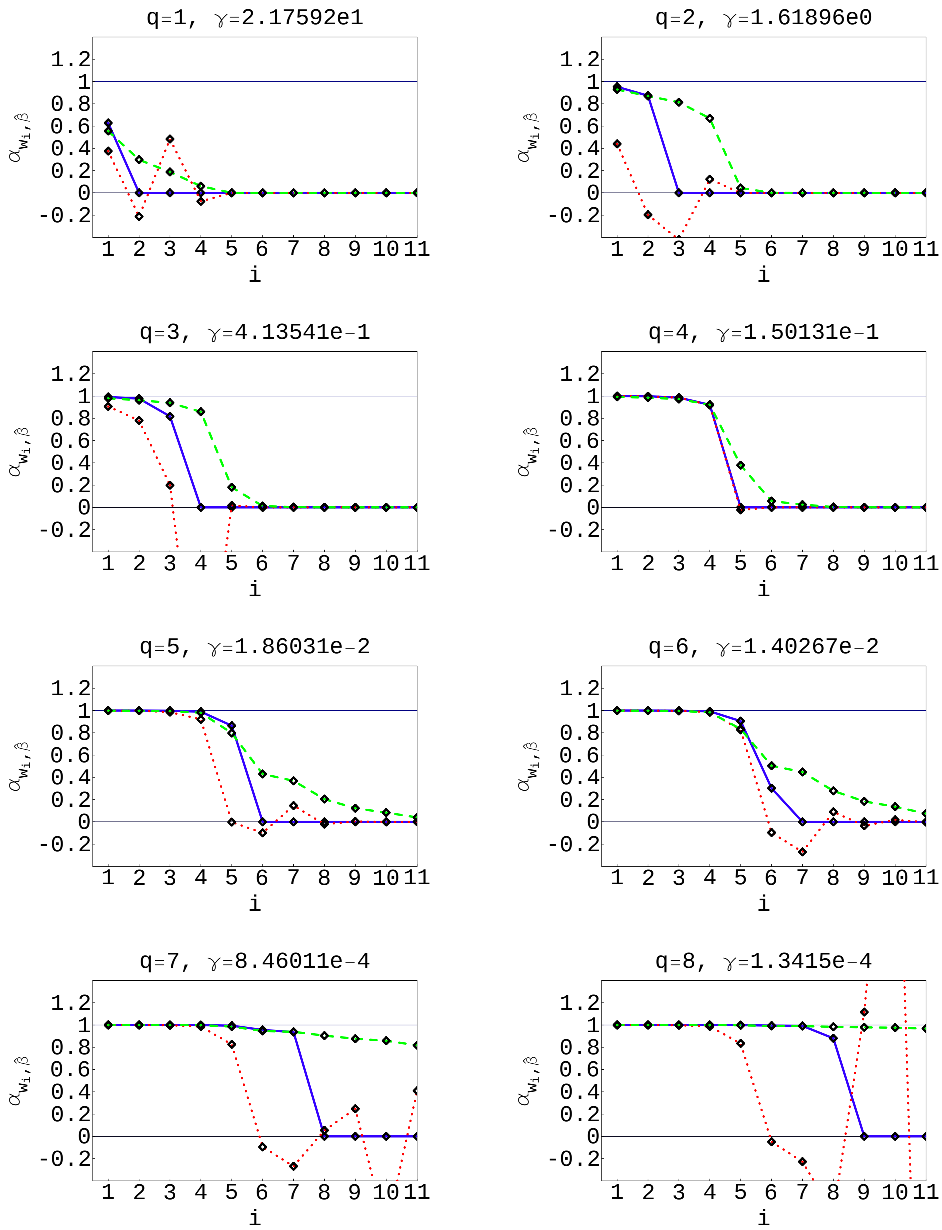

Fig. 4. Shrinkage factors in SNR directions for PLSR (solid), RR (dashed) and PCR (dotted) for detergent data.

5 Appendix

We give here the proof of Proposition 7.

Step 1: fix $q \leq q^*$ and $\gamma > -\lambda_p$. We define

$$G_0(\gamma) = \frac{w_q^{\text{SNR}\prime}(S+\gamma I)^{-1}s}{w_q^{\text{SNR}\prime}S^{-1}s} = \alpha_{w_q^{\text{SNR}},\widehat{\beta}^{\text{RR}}_{\gamma}}.$$

Note that γ may be negative. We have $G_0(0) = 1$ and from Section 3.1, $w_q^{\text{SNR}\prime} S^{-1} s \neq 0$. So by continuity, $G_0(\gamma) > 0$ for γ around 0. The derivative is

$$G_0'(\gamma) = -\frac{w_q^{\text{SNR}\prime}(S+\gamma\, I)^{-2} s}{w_q^{\text{SNR}\prime} S^{-1} s}.$$

We can choose w_q^{SNR} as in Eq. (15). By Lemma 4, we have $w_q^{\text{SNR}\prime} S^{-1} s > 0$ and

$$w_q^{\text{SNR}\prime}\, S^{-2}\, s = s' S^{-2} s - \widehat{\beta}_{q-1}^{\text{PLS}\prime} S^{-1}\, s = ||\widehat{\beta}^{\text{OLS}}||^2 - \widehat{\beta}^{\text{OLS}\prime}\, \widehat{\beta}_{q-1}^{\text{PLS}}.$$

By the Cauchy-Schwarz inequality and Corollary 6, $\widehat{\beta}^{\text{OLS}\prime}\widehat{\beta}_{q-1}^{\text{PLS}} \leq ||\widehat{\beta}^{\text{OLS}\prime}||\, ||\widehat{\beta}_{q-1}^{\text{PLS}}|| < ||\widehat{\beta}^{\text{OLS}\prime}||^2$. Therefore, $G_0'(0) < 0$ and, by continuity, $G_0'(\gamma) < 0$ for γ around 0, i.e. there exists $0 < r_0 < \lambda_p$ such that, for $\gamma \in B_0 =]-r_0, r_0[$, $G_0(\gamma)$ is decreasing.

Step 2: for $\delta > 0$ and $\gamma > -\lambda_p$, we define

$$G_\delta(\gamma) = \frac{w_q^{\text{SNR}\prime}(S+\gamma\, I)^{-1} s}{w_q^{\text{SNR}\prime}(S+\delta\, I)^{-1} s}.$$

We want to prove that the behaviour of $G_\delta(\gamma)$ around δ is similar to that of $G_0(\gamma)$ around 0. The idea is to replace S by $S + \delta\, I$ in the SNR and ΔSNR algorithms. This leads to a sequence of directions $w_1^\delta, w_2^\delta, ..., w_p^\delta$ and a sequence of estimators $\widehat{\beta}_1^\delta, \widehat{\beta}_2^\delta, ..., \widehat{\beta}_p^\delta = \widehat{\beta}_\delta^{\text{RR}}$. The Krylov subspaces generated by S and s are equal to those generated by $S+\delta\, I$ and s. Then, $w_i^\delta \propto w_i^{\text{SNR}}$, for $i = 1, ..., q^*$. Moreover, $\widehat{\beta}_{q^*}^\delta = \widehat{\beta}_\delta^{\text{RR}}$. Since $G_\delta(\gamma)$ is scale invariant w.r.t. w_q,

$$G_\delta(\gamma) = \frac{w_q^{\delta\prime}(S+\gamma\, I)^{-1}\, s}{w_q^{\delta\prime}(S+\delta\, I)^{-1}\, s} = \frac{w_q^{\delta\prime}((S+\delta\, I) + (\gamma-\delta)\, I)^{-1} s}{w_q^{\delta\prime}(S+\delta\, I)^{-1} s}.$$

The direction w_i^δ plays the same role in this step as w_i^{SNR} in step 1. Therefore, we can apply the results obtained for G_0, i.e. there exists $0 < r_\delta < \lambda_p$ such that, for $\gamma \in B_\delta =]\delta - r_\delta, \delta + r_\delta[$, $G_\delta(\gamma)$ is decreasing and positive.

Step 3: fix $0 \leq \gamma_a < \gamma_b$. Since $[\gamma_a, \gamma_b]$ is a compact set, the open covering $\{B_\delta\, ;\, \gamma_a \leq \delta \leq \gamma_b\}$ admits a finite subcover $\{B_{\delta_i}\, ;\, i = 1, ..., n\}$ with $\delta_1 = \gamma_a < \delta_2 < ... < \delta_n = \gamma_b$. This subcover can be chosen minimal, i.e. for all $i \neq j$, $B_i \not\subset B_j$, which implies here that for $i = 2, ..., n$, $B_{i-1} \cap B_i \neq \emptyset$. Let $\gamma_i \in B_{i-1} \cap B_i$. We have $\gamma_1 < ... < \gamma_n$ and

$$\frac{\alpha_{w_q^{\text{SNR}}, \widehat{\beta}_{\gamma_a}^{\text{RR}}}}{\alpha_{w_q^{\text{SNR}}, \widehat{\beta}_{\gamma_b}^{\text{RR}}}} = \frac{G_{\delta_1}(\gamma_1)}{G_{\delta_1}(\gamma_2)}\, \frac{G_{\delta_2}(\gamma_2)}{G_{\delta_2}(\gamma_3)}\, \frac{G_{\delta_3}(\gamma_3)}{G_{\delta_3}(\gamma_4)} \cdots \frac{G_{\delta_n}(\gamma_{n-1})}{G_{\delta_n}(\gamma_n)}$$

From step 2, we have $0 < G_{\delta_i}(\gamma_{i+1})/G_{\delta_i}(\gamma_i) < 1$. Therefore, $\alpha_{w_q^{\text{SNR}}, \widehat{\beta}_{\gamma_a}^{\text{RR}}} \geq \alpha_{w_q^{\text{SNR}}, \widehat{\beta}_{\gamma_b}^{\text{RR}}}$ and $G_0(\gamma) = \alpha_{w_q^{\text{SNR}}, \widehat{\beta}_{\gamma}^{\text{RR}}}$ is decreasing w.r.t. γ on $[0, +\infty)$. Since $G_0(0) =$

1 and $\lim_{\gamma\to\infty} G(\gamma) = 0$, $\alpha_{w_q^{\text{SNR}},\widehat{\beta}_\gamma^{\text{RR}}}$ belongs to $[0, 1]$.

References

Brown, P. B. (1990). Partial least squares in perspective. *Anal. Proc. R. Soc. Chem. 27*, 303–306.

Butler, N. A. and M. C. Denham (2000). The peculiar shrinkage properties of partial least squares regression. *Journal of the Royal Statistical Society, Series B, Methodologica 62*, 585–593.

De Jong, S. (1995). PLS shrinks. *Journal of Chemometrics 9*, 323–326.

Druilhet, P. and A. Mom (2006). Pls regression: a directional signal-to-noise ratio approach. *to appear in Journal of Multivariate Analysis*.

Frank, I. E. and J. H. Friedman (1993). A statistical view of some chemometrics regression tools (Disc: p136-148). *Technometrics 35*, 109–135.

Goutis, C. (1996). Partial least squares algorithm yields shrinkage estimators. *The Annals of Statistics 24*, 816–824.

Helland, I. S. (1988). On the structure of partial least squares regression. *Communications in Statistics, Part B – Simulation and Computation [Split from: @J(CommStat)] 17*, 581–607.

Krämer, N. (2005). On the behaviour of partial least squares regression. *Bericht-Nr 2005-05, ISSN 1436-9915*.

Lingjærde, O. C. and N. Christophersen (2000). Shrinkage structure of partial least squares. *Scandinavian Journal of Statistics 27*(3), 459–473.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I*, Berkeley and Los Angeles, pp. 197–206. University of California Press.

Stone, M. and R. J. Brooks (1990). Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression (Corr: V54 p906-07). *Journal of the Royal Statistical Society, Series B, Methodological 52*, 237–269.