Characterization of estimators uniformly shrinking on subspaces

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

[a] *IRMAR et Université Rennes II, Place du recteur Henri Le Moal, CS 24307/ 35043 RENNES, France*

[b] *Université Blaise Pascal, Clermont Universités, Campus des Cézeaux, B.P. 80026 63177 Aubière cedex France*

Abstract

Shrinkage factors play an important role in the behaviour of biased estimators. In this paper, we first show that the only way to have bounded shrinkage factors on a subspace is to shrink uniformly on this subspace. Then, we characterize regressions on components that shrink uniformly on the subspaces spanned by their associated weight vectors. We show that this problem is equivalent to solving a set of linear equations involving two different projectors. We define a class of matrices whose eigen decompositions give the solution.

Key words: Biased regression, eigenvectors, projector, regression on components, shrinkage factors,
1991 MSC: 62J05, 62J07.

1 Introduction

It is well established that some classes of biased estimators are often preferable to ordinary least squares estimators (OLS) in linear regression when the explanatory variables are highly correlated. The main classes of alternative estimators are Ridge Regression (Hoerl and Kennard, 1970), Liu estimators (Liu, 1993, Akdeniz, Styan and Werner, 2006) and regression on components such as Principal Component Regression or Partial Least-Squares Regression

* Corresponding author
Email addresses: alain.mom@univ-rennes2.fr (Alain Mom), pierre.druilhet@math.univ-bpclermont.fr (Pierre Druilhet).

Preprint submitted to Linear Algebra and its Applications *April 20, 2012*

(see Goutis, 1996, or Druilhet and Mom, 2006, for shrinkage properties of PLSR).

Druilhet and Mom (2008) have shown that in most cases, these estimators do not have bounded shrinkage factors in all the directions of the parameter space, which may lead to peculiar properties of these estimators (see Frank and Friedman, 1997, Butler and Denham, 2000, or Lingjærde and Christophersen, 2000). In the case of regression on components, a natural way to overcome this problem is to seek competing estimators that have bounded shrinkage factors in the directions generated by the weight vectors associated with the components.

The aim of this paper is to identify which regressions on components lead to estimators that shrink uniformly on the subspaces spanned by their associated weight vectors. In section 2, we show that, from a statistical point of view, this problem is equivalent to having bounded shrinkage factors on these subspaces. In Section 3, we show that, from an algebraic point of view, this problem is equivalent to finding orthonormal p-tuples $(w_1, ..., w_p)$ of p-vectors that satisfy some properties involving two projectors. Then, we give a solution and we characterize uniformly shrinking regression on components. In Section 4, we give another presentation of the results based on ridge estimators.

2 The statistical motivation

2.1 Uniformly shrinking estimators

We consider the standard linear model:

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

where X is the $n \times p$ 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 the sake of simplicity, from now on, we assume that S is of full rank. We denote by $\widehat{\beta}^{\text{OLS}} = S^{-1}\, s$, the OLS estimator.

Let w be a p-vector and $\widehat{\beta}$ a competing estimator. The shrinkage factor a_w associated with w is defined by:

$$a_w = \frac{w'\widehat{\beta}}{w'\widehat{\beta}^{\text{OLS}}}. \tag{2}$$

For w such that $w'\widehat{\beta}^{\text{OLS}} = 0$ and $w'\widehat{\beta} = 0$ or $w'\widehat{\beta}^{\text{OLS}} = 0$ and $w'\widehat{\beta} \neq 0$, a_w is not defined. In the first case, resp. the second case, we denote $a_w = \frac{0}{0}$, resp.

$a_w = \pm\infty$. Since the shrinkage factor associated with w is scale invariant, it depends on w only through the corresponding direction. Note that $w'\widehat{\beta}$ actually shrinks in the usual sense only when $0 \leq a_w \leq 1$.

Definition 1 *An estimator $\widehat{\beta}$ is said to be* uniformly shrinking *on a subspace E if there exists a real $a \neq \pm\infty$ such that:*

$$w'\widehat{\beta} = a\, w'\widehat{\beta}^{\text{OLS}}, \quad \forall w \in E. \tag{3}$$

It should be remarked that $\hat{\beta}$ lies in a flat determined by the vector $a\widehat{\beta}^{\text{OLS}}$ and the vector space orthogonal to W. Note that when $\widehat{\beta}$ is proportional to $\widehat{\beta}^{\text{OLS}}$, this property holds for any subspace E (see e.g. Tian and Puntanen, 2009, for some examples of equality between the two estimators).

The following proposition shows that uniform shrinkage on E is the only way to obtain bounded shrinkage factors on each direction in E, i.e. the shrinkage factor associated with any direction w in E which is not equal to $\frac{0}{0}$ cannot be greater than some number M nor equal to $\pm\infty$.

Proposition 2 *Let $\widehat{\beta}$ be an estimator of β and E a subspace. Then, $\widehat{\beta}$ is uniformly shrinking on E if and only if the shrinkage factors in the directions belonging to E are bounded.*

Proof : If $\widehat{\beta}$ is uniformly shrinking on E, then, obviously, the shrinkage factors are bounded on E. Conversely, assume that $\widehat{\beta}$ is not uniformly shrinking on E. If there exists a direction in E whose shrinkage factor is equal to $\pm\infty$, then the shrinkage factors are not bounded. So, we assume that there exist w_1 and w_2 in E such that $a_{w_1} \neq a_{w_2}$, both belonging to $\mathbb{R}$. Let $\gamma = -\frac{w_1'\widehat{\beta}^{\text{OLS}}}{w_2'\widehat{\beta}^{\text{OLS}}} + \delta$ and $w = w_1 + \gamma w_2$, which belongs to E. We have

$$a_w = \frac{(a_{w_1} - a_{w_2})\, w_1'\widehat{\beta}^{\text{OLS}} + \delta\, w_2'\widehat{\beta}}{\delta\, w_2'\widehat{\beta}^{\text{OLS}}}.$$

When δ tends towards 0, a_w tends towards $+\infty$ or $-\infty$ and therefore cannot be bounded. □

2.2 Totally uniformly shrinking sequences of directions

In regression on components, we have p components $t_1 = X\, w_1, ..., t_p = X\, w_p$, where $(w_1, ..., w_p)$ are the corresponding linearly independent weight vectors. For $1 \leq q \leq p$, we consider $(i_1, ..., i_q)$ a subset of $(1, ..., p)$. We denote $W =$

$(w_{i_1}, ..., w_{i_q})$. The estimator $\widehat{\beta}_W$ obtained by regression on the components $t_{i_1}, ..., t_{i_q}$ is:

$$\widehat{\beta}_W = P_W^S \, \widehat{\beta}^{\text{OLS}}, \tag{4}$$

where $P_W^S = W(W'SW)^{-1}W'S$ is the projector onto $\text{span}(W)$ w.r.t. S. We name such estimators POD (Projection Onto Directions) estimators. If $q = 0$, i.e. if the set of directions is empty, we write $\widehat{\beta}_{\{0\}} = 0$.

Since $\widehat{\beta}_W$ is concentrated in the subspace spanned by the weight vectors, we want this estimator to have bounded shrinkage factors on this subspace, or equivalently, to shrink uniformly on this subspace as seen in Section 2.1.

Proposition 3 *The estimator $\widehat{\beta}_W$ defined by (4)is uniformly shrinking on* $\text{span}(W)$ *iff there exists a real a_W such that:*

$$\widehat{\beta}_W = a_W \, P_W \, \widehat{\beta}^{\text{OLS}}, \tag{5}$$

where $P_W = W(W'W)^{-1}W'$ is the orthogonal projector onto $\text{span}(W)$.

Proof : Since $\widehat{\beta}_W$ and $P_W \, \widehat{\beta}^{\text{OLS}}$ belong to $\text{span}(W)$, (5) is equivalent to: $\forall w \in \text{span}(W)$, $w'\widehat{\beta}_W = a_W \, w' P_W \, \widehat{\beta}^{\text{OLS}}$, which is equivalent to $\forall w \in \text{span}(W)$, $w'\widehat{\beta}_W = a_W \, w' \widehat{\beta}^{\text{OLS}}$. □

Since we want these estimators to be uniformly shrinking for any subsets of components, we are led to the following concept:

Definition 4 *A p-tuple $(w_1, \cdots, w_p)$ of orthonormal directions in $\mathbb{R}^p$ is said to be* totally uniformly shrinking *if for any subset W the corresponding POD estimator is uniformly shrinking on* $\text{span}(W)$.

The next section is devoted to the characterization of totally uniformly shrinking p-tuples.

3 Characterization of totally uniformly shrinking p-tuples of directions

We deal with the following problem derived from the previous section: for any given $p \times p$ symmetric positive-definite matrix S and any given non-zero x in $\mathbb{R}^p$, we seek all the p-tuples $(w_1, \cdots, w_p)$ of orthonormal vectors in $\mathbb{R}^p$ such that, for any subset W, there exists a scalar a_W, called shrinkage factor, such that

$$P_W^S \; x = a_W \; P_W \; x \, . \tag{6}$$

Such a p-tuple $(w_1, \cdots, w_p)$ is said to be totally uniformly shrinking as in Section 2 with (5) corresponding to (6) and $\widehat{\beta}^{\text{OLS}}$ being replaced by x. Note that from Prop. 3, condition (6) is equivalent to: for any vector w in span(W),

$$w' \, P_W^S \; x = a_W \; w' \, x.$$

We also write $s = S\,x$.

3.1 Totally uniformly shrinking p-tuples of directions and matrices M_α

Consider the class of $p \times p$ symmetric matrices M_α defined by:

$$M_\alpha = S^{-1} + \alpha \; x \; x' \tag{7}$$

for $\alpha \in \mathbb{R}$. When $\alpha = -\frac{1}{s'\,x}$, M_α is denoted by H:

$$H = S^{-1} - \frac{1}{s'\,x} \, x \; x' \, . \tag{8}$$

Lemma 5 *All the matrices M_α, except H, are of full rank. The matrix H has a simple null eigenvalue which corresponds to the eigenvector s.*

Proof :

We have $M_\alpha u = 0 \Leftrightarrow S^{-1} u = -\alpha(u'\,x)\,x \Leftrightarrow u = -\alpha(u'\,x)s$. Then, s is the corresponding eigenvector and $S^{-1}s + \alpha(s'\,x)\,x = 0 \Leftrightarrow \alpha = -\frac{1}{s'\,x}$. $\square$

Proposition 6 *For any $\alpha \in \mathbb{R}$, the set of p orthonormal eigenvectors $u_1, ..., u_p$ of M_α is totally uniformly shrinking. Moreover, the shrinkage factor associated with $W = (u_{i_1}, ..., u_{i_k})$ is*

$$a_W = 1 + \alpha \parallel (I_p - P_W^S) \; x \parallel_S^2 \, . \tag{9}$$

Or equivalently,

$$a_W = \frac{\parallel P_W^S \; x \parallel_S^2}{\parallel x \parallel_S^2} \; \text{ if } \alpha = -\frac{1}{s'\,x}, \tag{10}$$

$$a_W = \frac{1 + \alpha(s'\,x)}{1 + \alpha(s' P_W \, x)} \; \text{ if } \alpha \neq -\frac{1}{s'\,x}. \tag{11}$$

Proof : For $1 \leqslant i \leqslant p$, let u_i be the normalized eigenvector of M_α associated with $\lambda_i(M_\alpha)$.

If $\lambda_i(M_\alpha) = 0$, Lemma 5 implies that $u_i \propto s$ and $\alpha = -\frac{1}{s'\,x}$. Then, Proposition 7 gives the result.

If $\lambda_i(M_\alpha) \neq 0$ for $1 \leqslant i \leqslant p$, then $u_i + \alpha(u_i'\,x)s = \lambda_i(M_\alpha)Su_i$.
$\forall j \neq i$, $u_j'Su_i = \alpha\frac{(u_i'\,x)(u_j's)}{\lambda_i(M_\alpha)}$ since $u_j'u_i = 0$. For $j = i$, $u_i'Su_i = \alpha\frac{(u_i'\,x)(u_i's)}{\lambda_i(M_\alpha)} + \frac{1}{\lambda_i(M_\alpha)}$. Let $d_i = \alpha\frac{(u_i'\,x)}{\lambda_i(M_\alpha)}$ and $W = (u_{i_1}, \cdots, u_{i_k})$, $1 \leqslant k \leqslant p$. Denote by $d \in \mathbb{R}^k$ the vector whose l^{th} entry is d_{i_l} and D the $k \times k$ diagonal matrix whose l^{th} diagonal entry is $\frac{1}{\lambda_{i_l}(M_\alpha)}$ for $1 \leqslant l \leqslant k$, then $W'SW = D + (W's)d'$. Since $d'D^{-1}(W's) = \alpha(s'P_W\,x)$ and $W'SW$ is always of full rank, we have $1 + \alpha(s'P_W\,x) \neq 0$. Then,

$$(W'SW)^{-1}W'S\,x = \frac{1}{1 + \alpha(s'P_W\,x)}D^{-1}W's.$$

Furthermore, for $1 \leqslant i \leqslant p$, $S^{-1}u_i = (\lambda_i(M_\alpha)u_i) - \alpha(u_i'\,x)\,x$, thus $W'S^{-1}s = D^{-1}W's - \alpha(s'\,x)W'\,x \Rightarrow D^{-1}W's = (1 + \alpha(s'\,x))W'\,x$. Since $W'W = I_k$, (6) holds with:

$$a_W = \frac{1 + \alpha(s'\,x)}{1 + \alpha(s'P_W\,x)}.$$

For the last part of the proof, simply note that (10) is equivalent to (9) if $\alpha = -\frac{1}{s'\,x} = -\frac{1}{\|x\|_S^2}$.
If $\alpha \neq -\frac{1}{s'\,x}$, (6) and the symmetry of the projector $I_p - P_W^S$ *w.r.t* S imply that (11) is equivalent to (9).
□

3.2 Preliminary results

In order to prove Theorem 9 giving the main result of the paper, we need the following proposition and lemma.

Proposition 7 *Any set of p orthonormal eigenvectors of H is totally uniformly shrinking. Furthermore, for any subset W, a_W is given by (10).*

Proof : From (6), straightforward algebra gives the first part of Proposition 7. By Lemma 5, s is an eigenvector of H. Thus, if $s \in \text{span}(W)$, multiplying (6) by s' gives (10). If $s \notin \text{span}(W)$, then $P_W^S\,x = 0$ and (6) implies that $a_W = 0 = \frac{\|P_W^S\,x\|_S^2}{\|x\|_S^2}$. □

Lemma 8 *Let $\alpha \in \mathbb{R}$ and $(w_1, \cdots, w_p)$ be the eigenvectors of M_α. Denote by q the number of eigenvectors w_i such that $w_i'\,x \neq 0$. Without loss of generality, assume that these eigenvectors are the q first ones. For, $q + 1 \leqslant j \leqslant p$,*

$w_j' x = 0$. *Let* $(u_{q+1}, \cdots, u_p)$ *be any orthonormal basis of* $\mathrm{span}(w_1, \cdots, w_q)^\perp$, *then* $(w_1, \cdots, w_q, u_{q+1}, \cdots, u_p)$ *is totally uniformly shrinking.*

Proof : First, from the proof of Proposition 6, for any subset W_1 of $(w_1, \cdots, w_q)$, there exists $a_{W_1} \in \mathbb{R}$, such that $P^S_{W_1}\ x = a_{W_1}\, P_{W_1}\ x$.
For, $q+1 \leqslant j \leqslant p$, since w_j is an eigenvector of M_α and $w_j' x = 0$, w_j is an eigenvector of S and thus $\lambda_j(M_\alpha) \neq 0$. Then, for $1 \leqslant i \leqslant q$ and $q+1 \leqslant j \leqslant p$,

$$w_i' S w_j = \frac{1}{\lambda_j(M_\alpha)} w_i' w_j = 0 \tag{12}$$

and $(w_1, \cdots, w_q) \perp_S (w_{q+1}, \cdots, w_p)$. Thus, for any subset W_1 of $(w_1, \cdots, w_q)$, and any W_2 such that $\mathrm{span}(W_2) \subset \mathrm{span}(w_{q+1}, \cdots, w_p)$, $P^S_W = P^S_{W_1} + P^S_{W_2}$ where $W = (W_1, W_2)$. Since, for $q+1 \leqslant j \leqslant p$, $w_j' x = 0$, on the one hand, $P_{W_2}\ x = 0 \Rightarrow P_W\ x = P_{W_1}\ x$. On the other hand, for $q+1 \leqslant j \leqslant p$, $0 = w_j' x = w_j' S^{-1} s = \lambda_j(M_\alpha) w_j' s \Rightarrow w_j' s = 0$ since $\lambda_j(M_\alpha) \neq 0$. Thus $s \in \mathrm{span}(w_1, \cdots, w_q) \Rightarrow P^S_{W_2}\ x = 0 \Rightarrow P^S_W\ x = P^S_{W_1}\ x$.

Then,

$$P^S_W\ x = P^S_{W_1}\ x = a_{W_1}\, P_{W_1}\ x = a_{W_1}\, P_W\ x\,.$$

Thus, for any orthonormal basis of $\mathrm{span}(w_1, \cdots, w_q)^\perp$, say $(u_{q+1}, \cdots, u_p)$, $(w_1, \cdots, w_q, u_{q+1}, \cdots, u_p)$ is an orthonormal basis of $\mathbb{R}^p$ which is totally uniformly shrinking. □

3.3 Main result

The following result, which is the main one, is a reciprocal of Proposition 6. Roughly speaking, a totally uniformly shrinking p-tuple of directions is a set of eigenvectors of a matrix M_α for some α.

When $x = \widehat{\beta}^{\text{OLS}}$ and $S = X'X$, a totally uniformly shrinking p-tuple of Theorem 9, say $(w_1, \cdots, w_p)$, turns into a totally uniformly shrinking p-tuple of directions given in Definition 4. Thus, any subset W of $(w_1, \cdots, w_p)$ gives a POD estimator defined in (4) and the statistical problem dealt with in section 2 is solved.

Theorem 9 *Let* $(w_1, \cdots, w_p)$ *be a p-tuple of orthonormal directions.*
If, for $1 \leqslant i \leqslant p$, $w_i' x \neq 0$, *then* $(w_1, \cdots, w_p)$ *is totally uniformly shrinking if and only if there exists* $\alpha \in \mathbb{R}$ *such that, for* $1 \leqslant i \leqslant p$, w_i *are the eigenvectors of* M_α .
If there are $q \leqslant p-1$ *directions* w_i *among* $(w_1, \cdots, w_p)$ *such that* $w_i' x \neq 0$, *(assumed to be the q first ones without loss of generality) then* $(w_1, \cdots, w_p)$ *is*

totally uniformly shrinking if and only if $\exists \alpha \in \mathbb{R}$ such that $(w_1, \cdots, w_q)$ are eigenvectors of M_α and $(w_{q+1}, \cdots, w_p)$ are any orthonormal basis of $\mathrm{span}(w_1, \cdots, w_q)^\perp$.

Proof :

For the first part of Theorem 9, the necessary part follows from Proposition 6. For the sufficient part, we search p orthonormal directions $(w_1, \cdots, w_p)$ such that any subset W satisfies (6). Consider w_i one of these directions and denote by $W_{(-i)} = \mathrm{span}(w_i)^\perp$ its orthogonal subspace. Then, there exists $a_{(-i)} \in \mathbb{R}$ such that

$$P^S_{W_{(-i)}}\, x = a_{(-i)}\, P_{W_{(-i)}}\, x$$

$$\Leftrightarrow (I_p - P^S_{W_{(-i)}})\, x = (I_p - a_{(-i)}\, P_{W_{(-i)}})\, x = (1 - a_{(-i)})\, x + a_{(-i)} P_{w_i}\, x$$

Since $W^{\perp_S}_{(-i)} = \mathrm{span}(S^{-1}w_i)$, $(I_p - P^S_{W_{(-i)}})\, x = P^S_{S^{-1}w_i}\, x$. Then,

$$P^S_{S^{-1}w_i}\, x = (1 - a_{(-i)})\, x + a_{(-i)} P_{w_i}\, x$$

Thus, for $1 \leqslant i \leqslant p$, $S^{-1}w_i \in \mathrm{span}(x, w_i)$. More precisely, if $w_i' x \neq 0$ for $1 \leqslant i \leqslant p$, w_i verifies:

$$M_{\alpha(w_i)} w_i = S^{-1}w_i + \alpha(w_i)(w_i' x)\, x = \lambda_i(M_{\alpha(w_i)}) w_i \tag{13}$$

with $\alpha(w_i) = -(1 - a_{(-i)})\frac{w_i' S^{-1} w_i}{(w_i' x)^2}$ and

$$\lambda_i(M_{\alpha(w_i)}) = a_{(-i)} \frac{w_i' S^{-1} w_i}{w_i' w_i}.$$

For any other direction of $(w_1, ..., w_p)$, say w_j, such that $w_j' x \neq 0$, we have:

$$M_{\alpha(w_j)} w_j = S^{-1}w_j + \alpha(w_j)(w_j' x)\, x = \lambda_j(M_{\alpha(w_j})w_j. \tag{14}$$

Since $w_i' w_j = 0$, multiplying (13) by w_j' and (14) by w_i' gives:

$$\alpha(w_i) = \alpha(w_j) = -\frac{w_i' S^{-1} w_j}{(w_i' x)(w_j' x)} = \alpha.$$

Then all the w_i, for $1 \leqslant i \leqslant p$, such that $w_i' x \neq 0$, are the eigenvectors of the same matrix M_α.
For the second part of Theorem 9, the necessary part follows from Lemma 8. For the sufficient part, let $(w_1, \cdots, w_p)$ be a totally uniformly shrinking p-tuple with for $1 \leqslant i \leqslant q$, $w_i' x \neq 0$ and for $q + 1 \leqslant i \leqslant p$, $w_i' x = 0$. The first part of the proof shows that $\exists \alpha \in \mathbb{R}$ such that, for $1 \leqslant i \leqslant q$, w_i are the

eigenvectors of M_α. For $1 \leqslant i \leqslant q$,we have:

$$w_i + \alpha(w_i' x)s = \lambda_i(M_\alpha)Sw_i. \tag{15}$$

Since $w_j' x = 0$ for $q+1 \leqslant j \leqslant p$, $x \in \text{span}(W_1)$ where $W_1 = (w_1, \cdots, w_q)$ and, by (6), $w_j' s = 0$ for $q+1 \leqslant j \leqslant p$. Thus, $s \in \text{span}(W_1)$.

For $1 \leqslant i \leqslant q$ and for $q+1 \leqslant j \leqslant p$, multiplying (15) by w_j' gives $w_j' S w_i = 0$ because $\lambda_i(M_\alpha) \neq 0$. Thus,

$$W_1 \bot_S W_2 = (w_{q+1}, \cdots, w_p). \tag{16}$$

Furthermore, the set W_2 is necessarily an orthonormal basis of $\text{span}(W_1)^\perp$. To complete the proof, we have to prove that it could be any one. Since (16) holds, the remainder of the proof is the same as in Lemma 8 after (12). □

Now, we seek for which values of α the associated shrinkage factor actually shrinks, i.e. belongs to $[0, 1]$.

Remark : From (9), when $x = \widehat{\beta}^{\text{OLS}}$, the least-squares estimator constrained to belong to W actually shrinks, i.e. $0 \leq a_W \leq 1$, iff:

$$-\frac{1}{\| (I_p - P_W^S)\, x \|_S^2} \leqslant \alpha \leqslant 0. \tag{17}$$

Let maxnorm $= \max_{W_{p-1}}(\| P_{W_{p-1}}^S\; x \|_S^2)$ where W_{p-1} is one of the p subsets of $p-1$ directions of $(w_1, ..., w_p)$. Therefore, it follows from (17) that if $-\frac{1}{\text{maxnorm}} \leqslant \alpha \leqslant 0$, then the 2^p estimators defined by (4) actually shrink on their respective subspaces. It is the case when $-\frac{1}{\|x\|_S^2} = -\frac{1}{s' x} \leqslant \alpha \leqslant 0$. Note that $\alpha = -\frac{1}{s' x}$ corresponds to $M_\alpha = H$. If $\alpha = 0$, the eigenvectors of M_α are the directions of PCR and the corresponding shrinkage coefficients are all equal to 1 as shown by (9).

Remark : Since a regular symmetric matrix has the same eigenvectors as its inverse, all matrices M_α, except H, can be replaced in Theorem 9 by their inverses:

$$M_\alpha^{-1} = S - \frac{\alpha}{1 + \alpha s' x} ss'.$$

In these expressions, S does not need to be inverted and the matrices are easier to diagonalize.

4 Totally uniformly shrinking p-tuples of directions obtained from Ridge regression

In this section, we show that the directions w_i in Theorem 9 can also be obtained from a set of matrices based on $(S+kI_p)^{-1}s$ where $s = Sx$. The matrix $S+kI_p$ is better conditioned than S and is preferable from a numerical point of view. In the context of section 2 , when $x = \widehat{\beta}^{\text{OLS}}$ then $(S+kI_p)^{-1}s$ corresponds to a ridge estimator: see e.g. Groß and Markiewicz (2004) for properties of general ridge estimators.

Denote by $\mathcal{M}$ the set of $p \times p$ matrices

$$M(\alpha_1, \alpha_2) = \alpha_1 S^{-1} + \alpha_2\ x\ x'$$

where $\alpha_1, \alpha_2 \in \mathbb{R}$. When $\alpha_1 = 0$, x is an eigenvector of $M(0, \alpha_2)$ and the other eigenvectors span its orthogonal subspace. Therefore, if W is any subspace spanned by some eigenvectors of $M(0, \alpha_2)$, $P_W^S\ x = P_W\ x = x$ if $x \in W$. If $x \notin W$, $P_W^S\ x \neq 0$ and $P_W\ x = 0$ and there does not exist a real a_W such that (6) holds. Except in this case, there exists a real $\alpha = \frac{\alpha_2}{\alpha_1}$ such that the eigenvectors of $M(\alpha_1, \alpha_2)$ are also those of $S^{-1} + \alpha\ x\ x'$.

Now denote by $\mathcal{N}$ the set of $p \times p$ matrices $N(\gamma_1, \gamma_2, k)$ defined by:

$$N(\gamma_1, \gamma_2, k) = \gamma_1 (S+kI_p)^{-1} + \gamma_2 ((S+kI_p)^{-1}s)((S+kI_p)^{-1}s)' \tag{18}$$

where $\gamma_1,\ \gamma_2 \in \mathbb{R}$, $k \in \mathbb{R} - \{-\lambda_i(S), \forall i = 1, \cdots, p\}$ and $s = Sx$.
Note that in the context of section 2 when $x = \widehat{\beta}^{\text{OLS}}$,
$N(\gamma_1, \gamma_2, k) = \gamma_1 (S+kI_p)^{-1} + \gamma_2 \widehat{\beta}_k^{ridge} \widehat{\beta}_k^{ridge\prime}$, where $\widehat{\beta}_k^{ridge} = (S+kI_p)^{-1}s$.
Then:

Proposition 10 *To each matrix belonging to $\mathcal{M}$ corresponds a matrix of $\mathcal{N}$ with the same eigenvectors.*

Proof :

Let $M(\alpha_1, \alpha_2) = \alpha_1 S^{-1} + \alpha_2\ x\ x'$ be any matrix of $\mathcal{M}$. First, if $\alpha_1 = 0$, then x is an eigenvector of $M(0, \alpha_2)$ and the other eigenvectors span its orthogonal subspace. These vectors can also be obtained as eigenvectors of the matrix $N(1, \gamma, k)$ belonging to $\mathcal{N}$ with $\gamma = \frac{1}{s'(S^{-1}-(S+kI_p)^{-1})s}$. Now, if $\alpha_1 \neq 0$, the eigenvectors of $M(\alpha_1, \alpha_2)$ are also those of $M_\alpha = S^{-1} + \alpha\ x\ x'$ with $\alpha = \frac{\alpha_2}{\alpha_1}$. Then $M_\alpha u_i = \lambda_i(M_\alpha) u_i$ is equivalent to $(1 + k\lambda_i(M_\alpha))(S+kI_p)^{-1}u_i + \alpha(u_i' x)(S+kI_p)^{-1}s = \lambda_i(M_\alpha) u_i$.
If there exists i such that $1 + k\lambda_i(M_\alpha) = 0$, then $u_i \propto (S+kI_p)^{-1}s$ which

is an eigenvector of $N(\gamma_1, \gamma_2, k)$ with $\gamma_1 = 0$ for any $\gamma_2 \neq 0$. Since the other eigenvectors of M_α are orthogonal to u_i, then the eigenvectors of M_α are also the eigenvectors of any matrix $N(0, \gamma_2, k)$ with $\gamma_2 \neq 0$.

If for all $1 \leq i \leq p,\ 1+k\lambda_i(M_\alpha) \neq 0$, we consider two cases for the eigenvectors of M_α :

Case 1: If u_i satisfies

$$s'(S + kI_p)^{-1}u_i \neq 0, \tag{19}$$

then u_i is an eigenvector of $N(1, \gamma, k)$ with:

$$\gamma = \frac{\alpha(u_i' x)}{(1 + k\lambda_i(M_\alpha))(s'(S + kI_p)^{-1}u_i)}.$$

As in the proof of Theorem 9, it can be shown that γ does not depend on the eigenvector of M_α satisfying (19).

Case 2: If u_i satisfies $s'(S + kI_p)^{-1}u_i = 0$ then for all $j \neq i$,

$$(1 + k\lambda_j(M_\alpha))(S + kI_p)^{-1}u_j + \alpha(u_j' x)(S + kI_p)^{-1}s = \lambda_j(M_\alpha)u_j. \tag{20}$$

Since $u_i'u_j = 0$, multiplying (20) by u_i' gives: $(1+k\lambda_j(M_\alpha))(u_i'(S+kI_p)^{-1}u_j) = 0$. So, for all $j \neq i$, $u_i'(S+kI_p)^{-1}u_j = 0$ and $(S+kI_p)^{-1}u_i \in \text{span}(u_i)$. Therefore u_i is an eigenvector of $(S + kI_p)^{-1}$. Since $s'(S + kI_p)^{-1}u_i = 0$, u_i is then an eigenvector of any matrix $N(1, \gamma_2, k)$.

Therefore, if there exist any eigenvectors of M_α in case 1, then all the eigenvectors of M_α are eigenvectors of $N(1, \gamma, k)$ where γ is defined in case 1. Otherwise, the eigenvectors of M_α are the eigenvectors of any matrix $N(1, \gamma_2, k)$.

□

References

[1] Akdeniz, F., Styan, G. P. H. and Werner, H. J. (2006). The general expressions for the moments of the stochastic shrinkage parameters of the Liu type estimator *Communication in Statistics - Theory and Methods 35*, 423–437

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

[3] Druilhet, P. and Mom, A. (2006). PLS regression: a directional signal-to-noise ratio approach. *Journal of Multivariate Analysis 97*, 1313–1329.

[4] Druilhet, P. and Mom, A. (2008). Shrinkage structure in biased regression. *Journal of Multivariate Analysis 99*, 232–244.

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

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

[7] Groß, J. and Markiewicz, A. (2004). Characterizations of admissible linear estimators. in the linear model *Linear Algebra and Its Applications 388*, 239-248.

[8] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics 12*, 55-67

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

[10] Tiao, Y. and Puntanen, S. (2009). On the equivalence of estimations under a general linear model and its transformed models. *Linear Algebra and Its Applications 430*, 2622-2641