PCR/PLSR optimization based on noise covariance estimation and Kalman ¯ltering theory

Rolf Ergon[1] and Kim H. Esbensen[2]

[1]Telemark University College
P.O.Box 203, N-3901 Porsgrunn, Norway
e-mail: rolf.ergon.@hit.no
telephone: ++ 47 35 57 51 60
telefax: ++ 47 35 57 52 50

[2]Ålborg University Esbjerg, Denmark

Summary

The theoretical connection between principal component regression (PCR) and partial least squares regression (PLSR) on one hand, and Kalman ¯ltering (KF) on the other, is known from earlier work. In the present paper we investigate the possibilities to use latent variables modeling and KF theory as means for optimization of ordinary PLSR and PCR predictors, based on the prerequisite of prior X-noise covariance estimates facilitated e.g. by more X- than y-observations. The result is a new PLSR optimization method, while the PCR optimization turns out to be identical with an earlier known method. A simulation example and two real-world data examples supporting the theoretical development are presented. The treatment is limited to cases with only one response variable, although an extension to multiresponse cases is also possible.

KEY WORDS: PLSR/PCR, optimization, covariance estimation, Kalman ¯ltering

1 Introduction

The theoretical connection between static (timewise non-dynamical) multivariate calibration - principal component regression (PCR) and partial least squares regression (PLSR) - and Kalman ¯ltering (KF) was ¯rst recognized by Berntsen[1], and further developed by Ergon[2]. In the present paper we investigate the possibilities to use KF theory as a basis for optimization of ordinary PLSR and PCR predictors. In this endeavour we use latent variables modeling and the Helland predictor form[3], which is summarized in Section 2. In Section 3 we present the theory for KF-based optimal regularization, and in Section 4 this theory is used to ¯nd optimized PLSR and PCR predictor algorithms. This is based on the prior estimation of the X-noise covariance matrix, facilitated by more X- than y-observations. The PCR method actually turns out to be identical with a method proposed by Isaksson and N½s[4] and further discussed by Thomas[5], while the PLSR method to the best of our knowledge is new. A simulation example is given in Section 5, and two real-world data examples follow in Section 6. Conclusions are drawn in Section 7, while some theoretical details are given in an Appendix. We will limit the treatment to cases with only one response variable, although an extension to multiresponse cases is also possible.

2 Latent variables modeling and regularization

Ordinary least squares regression

Assuming experimental data from N observations, $X = \begin{pmatrix} x_1 & x_2 & \cdots & x_N \end{pmatrix}^T$ and $y = \begin{pmatrix} y_1 & y_2 & \cdots & y_N \end{pmatrix}^T$, and independent observation errors, we ¯nd the ordinary least squares (OLS) regression solution[6]

$$\hat{b}_{OLS} = \left(X^T X\right)^{-1} X^T y: \tag{1}$$

With a large number p of x variables relative to the number N of observations, this solution will be ill-conditioned , i.e. very noise and collinearity sensitive (for $p > N$ the OLS solution is in fact unde¯ned). As is well-known in all practical applications, satisfactory OLS results require that p is well below N. A general and detailed analysis of this problem for limited values of N is beyond the scope of the present paper[7]. In the ill-conditioned case there is a need for regularization, and this can be based on latent variables modeling as shown below.

Latent variables modeling

When a large number of x variables are signi¯cantly to highly collinear, the regressor information may be compressed into a much smaller number of latent variables[8;9] $\xi = \begin{pmatrix} \xi_1 & \xi_2 & \cdots & \xi_A \end{pmatrix}^T$. An analysis of such data compression may be based on the underlying model

$$\begin{aligned} \xi_{k+1} &= e_k \\ y_k &= Q\xi_k + v_{1;k} \\ x_k &= W\xi_k + v_{2;k}; \end{aligned} \tag{2}$$

where e_k, $v_{1;k}$ and $v_{2;k}$ are white noise sequences with covariances $R_e = Ee_k e_k^T$, $r_{11} = Ev_{1;k}^2$ and $R_{22} = Ev_{2;k}v_{2;k}^T$ (we assume centered data). This is a special case of a general dynamic model with $\xi_{k+1} = A\xi_k + Bu_k + Ge_k$, where u_k is a manipulated input, i.e. we use here $A = 0$, $Bu_k = 0$ and $G = I$. We will also assume that W is orthonormal, i.e. that $W^T W = I$.

With N observations and $T = \begin{pmatrix} \xi_1 & \xi_2 & \cdots & \xi_N \end{pmatrix}^T$, $v_1 = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1N} \end{pmatrix}^T$ and $V_2 = \begin{pmatrix} v_{21} & v_{22} & \cdots & v_{2N} \end{pmatrix}^T$, the model (2) gives the static latent variables model

$$y = TQ^T + v_1 \tag{3}$$

$$X = TW^T + V_2: \tag{4}$$

Remark 1 If the regressor data are actually generated from an underlying state vector z_k as $x_k = C_2 z_k + v_{2;k}$, the output equation (4) will be replaced by $X = ZC_2^T + V_2$. In the special noise free case with $V_2 = 0$, we may then find T and W^T by a similarity transformation based on factorization of X by a number of alternative methods, e.g. PCA and PLSR. In practice, however, we will always observe some noise, and factorization of X then gives only estimates $\hat{T}$ and $\hat{W}^T$.

The Helland predictor

The PCR and PLSR regularizations are based on the latent variables model (3,4) above. The LS solution of (4) is

$$\hat{T} = XW; \tag{5}$$

and from (3) and (5) we thus find the LS predictor related to the latent variables

$$\hat{Q}^T = \left(\hat{T}^T\hat{T}\right)^{-1}\hat{T}^T y = \left(W^T X^T X W\right)^{-1} W^T X^T y; \tag{6}$$

which results in fitted experimental responses according to (3)

$$\hat{y} = \hat{T}\hat{Q}^T = XW\hat{Q}^T = XW\left(W^T X^T X W\right)^{-1} W^T X^T y; \tag{7}$$

and predictions of new responses

$$y_0 = \hat{z}_0^T\hat{Q}^T = x_0^T W\hat{Q}^T = x_0^T W\left(W^T X^T X W\right)^{-1} W^T X^T y: \tag{8}$$

The regularized latent variables predictor thus becomes

$$\hat{b}_{LV} = W\left(W^T X^T X W\right)^{-1} W^T X^T y: \tag{9}$$

This predictor was first presented by Helland[3], although there not explicitly based on an LV model.

The problem now is to find W, or more realistically a good estimate $\hat{W}$: In this endeavour we have in fact a number of possibilities. A simple choice is $W = I_p$, which brings us back to the OLS solution (1). Other choices give the standard statistical PCR and chemometrical PLSR solutions, or alternatively the new, optimized versions discussed in Section 4 below.

3 Optimal regularization

Assuming a known model (3,4), including noise covariance matrices R_e and R_{22}, the optimal loading weight matrix $W_{opt:}$ to use in the Helland predictor (9) may be found by use of general Kalman filtering theory[10]. We will, however, derive the optimal solution directly by introduction of the optimal state estimate related to the latent variables model (2),

$$\hat{z}_k = Kx_k; \tag{10}$$

where K is chosen such that the expectation

$$\begin{aligned} Z_k &= E(z_k - \hat{z}_k)(z_k - \hat{z}_k)^T = E\left[z_k - K(Wz_k + v_{2;k})\right]\left[z_k - K(Wz_k + v_{2;k})\right]^T \\ &= (I - KW)E\left(z_k z_k^T\right)(I - KW)^T + KE\left(v_{2;k}v_{2;k}^T\right)K^T \end{aligned} \tag{11}$$

is minimized. Using $Ez_k z_k^T = Ez_{k+1}z_{k+1}^T = Ee_k e_k^T = R_e$ and $Ev_{2;k}v_{2;k}^T = R_{22}$ we find[11]

$$\frac{\partial}{\partial K}\mathrm{trace}(Z_k) = -2(I - KW)R_e W^T + 2KR_{22}; \tag{12}$$

i.e. $\frac{@}{@K}\text{trace}(Z_k) = 0$ gives the optimal solution

$$K = R_e W^T \left(W R_e W^T + R_{22} \right)^{-1}. \tag{13}$$

This intermediate result, derived from general Kalman filtering theory, was first presented by Berntsen[1].

The resulting optimal response estimate is

$$\hat{y}_k = Q K x_k, \tag{14}$$

i.e.

$$y_k = Q K x_k + \varepsilon_k, \tag{15}$$

where it can be shown that ε_k is white noise[2]. Optimality here means that (14) gives the best linear unbiased estimate (BLUE), and the best possible estimator whatsoever assuming Gaussian noise distribution[10].

Making use of (15) with the assumptions that K is known and Q unknown, we may from experimental data find the LS estimate corresponding to (6)

$$\hat{Q}^T = \left(K X^T X K^T \right)^{-1} K X^T y, \tag{16}$$

and thus the fitted primary outputs corresponding to (7)

$$\hat{y} = X K^T \hat{Q}^T = X K^T \left(K X^T X K^T \right)^{-1} K X^T y. \tag{17}$$

The optimal choice of W in the Helland predictor (9) is thus in theory $W_{opt.} = K^T$, resulting in

$$\hat{b}_{KF+LS} = K^T \left(K X^T X K^T \right)^{-1} K X^T y. \tag{18}$$

4 Optimized PCR and PLSR

As was alluded to above measuring X data may often be much easier and less expensive than to obtain the corresponding y responses. In many practical situations we may thus obtain a large number of X observations, while the practically available number of y observations may be much lower. We emphasize here that this will always be an easily obtainable option regarding multivariate calibration[8;12]. These extra X observations may then be used for estimation of the X-noise covariance R_{22}, which together with an estimate of the latent variables covariance R_e may be used to improve the predictor according to the underlying Kalman filtering theory presented above. Note that if the extra X observations are not available at the time when the predictor is first determined, they will gradually appear during the normal use of the predictor, as long as the covariance structure of the data is unaltered.

In the development below we will use the notation X_{long} for all modeling observations and X_{short} for the observations corresponding to the available $y = y_{short}$ data only.

Optimized PLSR predictor

An optimized PLSR predictor based on the Kalman filtering theory presented above has the form given by (18), i.e.

$$\hat{b}_{PLSRopt} = \hat{K}_{PLS}^T \left(\hat{K}_{PLS} X_{short}^T X_{short} \hat{K}_{PLS}^T \right)^{-1} \hat{K}_{PLS} X_{short}^T y, \tag{19}$$

where $\hat{K}_{PLS}$ is found from (13) as

$$\hat{K}_{PLS} = \hat{R}_e \hat{W}_{PLS}^T \left(\hat{W}_{PLS} \hat{R}_e \hat{W}_{PLS}^T + \hat{R}_{22} \right)^{-1}. \tag{20}$$

Here the covariance estimates may be found as

$$\hat{R}_e = \frac{1}{N_{short} - 1}\hat{T}_{PLS}^T\hat{T}_{PLS} = \frac{1}{N_{short} - 1}\hat{W}_{PLS}^T X_{short}^T X_{short}\hat{W}_{PLS} \tag{21}$$

and

$$\hat{R}_{22} = \frac{1}{N_{long} - 1}\left(X_{long} - \hat{T}_{long}\hat{P}_{long}^T\right)^T\left(X_{long} - \hat{T}_{long}\hat{P}_{long}^T\right); \tag{22}$$

where $\hat{T}_{long}$ is given by the principal component analysis (PCA) factorization[6;8]

$$X_{long} = \hat{T}_{long}\hat{P}_{long}^T + E_{long}: \tag{23}$$

The number of PCA components A_{PCA} to use must here be determined through validation of the PCR predictor (24) presented below, preferentially using a test set[12]. Subsequently the number A of PLSR components to use must also be determined through proper validation, and it may well be that $A < A_{PCR}$ should be used. Also note that the inversion in (22) is ill-conditioned, such that a pseudo-inverse has to be used. The reason for this is that $\text{rank}(\hat{W}_{PLS}^T\hat{R}_e\hat{W}_{PLS}) = A$, and that also $\hat{R}_{22}$ according to (20) has a large condition number.

Optimized PCR predictor

Optimization of a PCR predictor may be performed along the lines of PLSR optimization discussed above. This will, however, result in the straightforward solution

$$\hat{b}_{PCRopt} = \hat{P}_{long}\left(\hat{P}_{long}^T X_{short}^T X_{short}\hat{P}_{long}\right)^{-1}\hat{P}_{long}^T X_{short}^T y; \tag{24}$$

where $\hat{P}_{long}$ is given by the factorization (23) (see Appendix for details). This is the same way of using extra X-observations for stabilization of the loading matrix as proposed by Isaksson and N½s[4] and further discussed by Thomas[5].

5 Monte Carlo simulation example

The practical case behind the following Monte Carlo simulation example could for example be a spectroscopic measurement of a solution with three different chemical substances. A typical simulation result is shown in Fig. 1. Note the overlapping peaks.

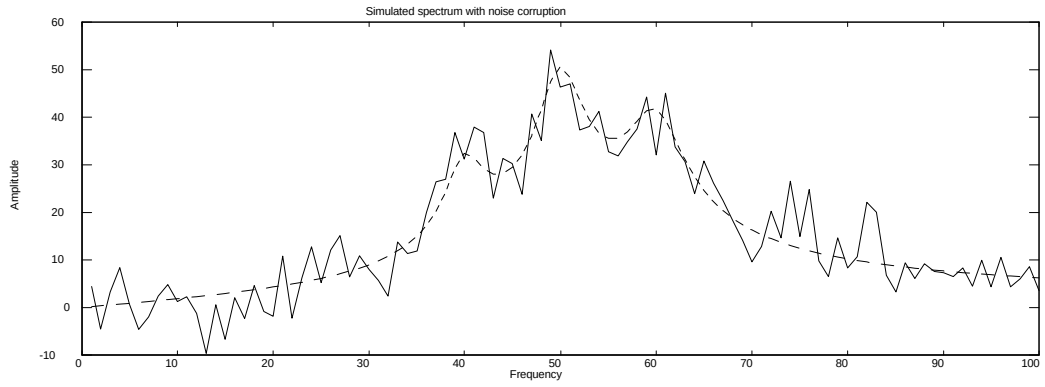

Figure 1. Simulation of noise free (dotted line) and noise corrupted spectrum (solid line) of three chemical species in solution. Note considerable overlap between peaks.

The Monte Carlo simulation is based on an assumed discrete frequency spectrum in the range $0 < f \cdot 100$ frequency units (f.u.),

$$x(f) = \sqrt{\frac{f_1 f}{(f_1^2 - f^2)^2 + (2\zeta f_1 f)^2}}(3 + z_1) + \sqrt{\frac{f_2 f}{(f_2^2 - f^2)^2 + (2\zeta f_2 f)^2}}(3 + z_2) + \sqrt{\frac{f_3 f}{(f_3^2 - f^2)^2 + (2\zeta f_3 f)^2}}(3 + z_3) + v_2(f) = 3C_2(f)\begin{bmatrix}1 & 1 & 1\end{bmatrix}^T + C_2(f)z + v_2(f); \quad (25)$$

with resonance frequencies $f_1 = 40$ f.u., $f_2 = 50$ f.u., $f_3 = 60$ f.u. and relative dampings $\zeta_1 = \zeta_2 = \zeta_3 = \zeta = 0{:}05$, and with $C_2(f)\ 2\ R^{1\times 3}$. It is also assumed that z_1, z_2 and z_3 are randomly generated zero mean numbers with normal distribution and variances $r_e = 1$, 1 and 0:5, while $v_2(f)$ are randomly generated zero mean numbers with normal distribution and equal variances $r_{22} = Ev_2^2 = 25$:

It is further assumed a response

$$y = z_2 = \begin{bmatrix}0 & 1 & 0\end{bmatrix} z + v_1; \quad (26)$$

with $r_{11} = Ev_1^2 = 0{:}01$: In the practical case this would mean that the primary response of interest would be the concentration of one of the three chemical substances, while the others would be treated as interferants.

The total model with centered data is then

$$\begin{aligned} z_{k+1} &= e_k \\ y_j &= \begin{bmatrix}0 & 1 & 0\end{bmatrix} z_j + v_{1;j} \\ x_k &= C_2 z_k + v_{2;k}; \end{aligned} \quad (27)$$

where $k = 1; 2; \ldots; N_{long}$ indicates a sequence of X observations corresponding to different concentrations of the three chemical substances, while the number of corresponding y_j observations is $N_{short} < N_{long}$, and where $C_2\ 2\ R^{100\times 3}$, $R_e = Ee_k e_k^T = \text{diag}\{1\ \ 1\ \ 0{:}5\}$, $r_{11} = Ew_k^2 = 0{:}01$ and $R_{22} = Ev_{2;k}v_{2;k}^T = \text{diag}\{25\ \ 25\ \ \ldots\ \ 25\}$.

Based on a modeling set with $N_{long} = 200$ and $N_{short} = 40$ (every 5th observation) the average root mean square error of prediction (RMSEP) results from 100 Monte Carlo runs using independent validation sets with $N_{val} = 200$ (for both X and y) are shown in Fig. 2.

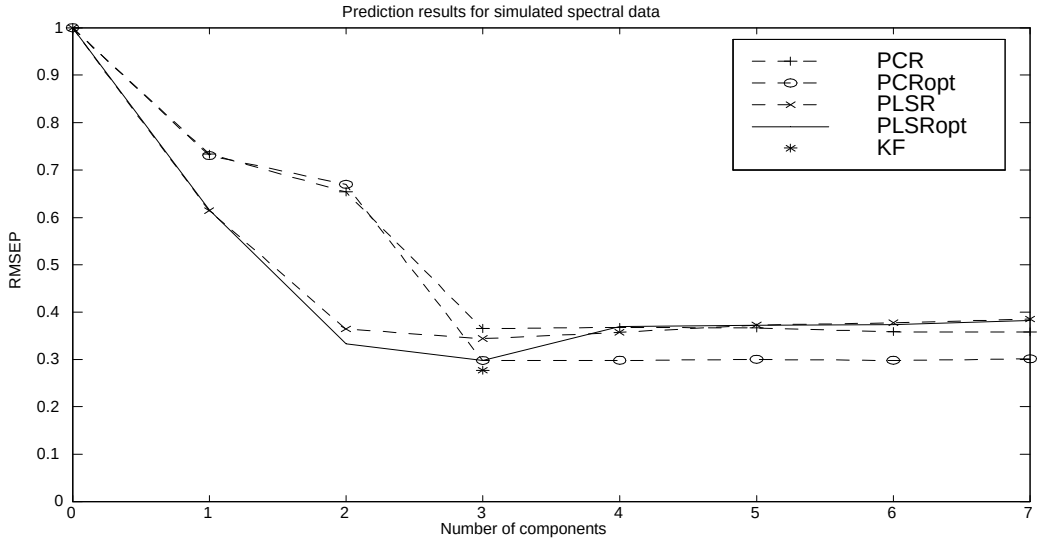

Figure 2. Simulation mean results from 100 Monte Carlo runs, based on test set validation. The mean KF solution (*) is shown for $A = 3$:

Here the covariance estimate $\hat{R}_{22}$ as given by (22) is based on a PCA of X_{long} using a ¯xed number $A_{PCA} = 3$ components, while $\hat{R}_e$ as given by (21) is based on the $A = 1, 2, \ldots, 7$ components used for the di®erent predictors. Note the following:

- The $A_{PCA} = 3$ ¯xation originates from the minimum RMSEP value for the optimized PCR predictor (24).
- $A = 3$ gives the best result, just as expected with the assumed sum of three independent spectra.
- The best result is fairly close to the theoretical KF result according to (17).
- The fact that $A = A_{PCA} = 3$ results in $\hat{\beta}_{PLSRopt} = \hat{\beta}_{PCRopt}$ can be explained using the Kalman ¯ltering theory (see Appendix).
- The improvements using estimated covariances is clearly at hand also for $A = 2$, where the PLSR predictor from the outset is much better than the PCR predictor. The PLSRopt estimator using $A = 2$ components gives a 14 % increase in RMSEP as compared with the $A = 3$ solution, but may still be a good choice in applications where a low number of components is preferred for interpretational purposes.

6 Real world data examples

Metal ion mixtures

The optimization methods developed in Section 4 are tested on a data set made available from the Wentzell Group[13]. The data set, labeled "inorfull", was "obtained through a carefully designed experiment involving three-component mixtures of metal ions (Co(II), Cr(III), Ni(II))". The X measurements were absorbances at $p = 176$ frequencies, while the concentration of Co was used as the response variable y, and the subsets used had $N_{long} = 104$, $N_{short} = 20$ (every 5th observation) and $N_{val} = 26$. The data were autoscaled and the results are shown in Fig. 3.

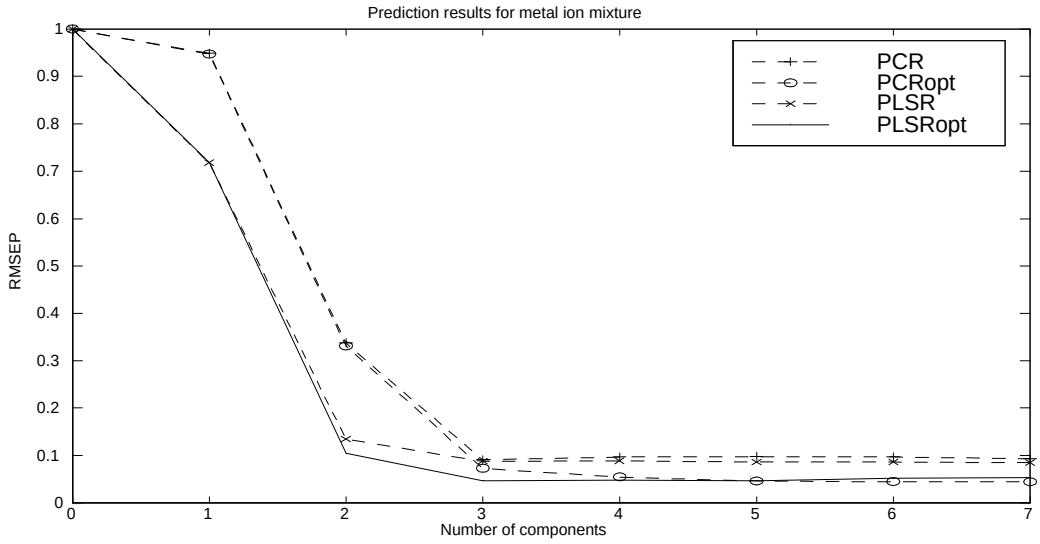

Figure 3. Results for predicted Co concentration from absorbance spectrum, based on test set validation.

The R_{22} estimate used in the optimized PLSR solution was here based on $A_{PCA} = 5$ components (the PCRopt minimum). The relative improvement from ordinary to optimized PLSR predictor at $A = 3$ components is 46 %, while the corresponding result at $A = 2$ is 22 %. Note that the overall result has a lot in common with the previous simulation result.

Acoustic °ow measurement

Acoustic chemometrics is based on signals from an acoustic sensor (accelerometer) placed for example on, or slightly downstream of, a standard ori¯ce plate. Observations of the power spectrum of the sensor signal is collected in the X matrix, and calibrated against physical y primary quantities like multi-component mixture concentrations, density etc., using for example a standard PLSR method[14].

In an experiment on a test rig at Telemark University College, the °ow rate of ordinary drinking water was measured by use of an ultrasonic °owmeter and used as the response variable y, while the acoustic power spectral densities at 512 frequencies originating from an accelerometer placed on a standard ori¯ce plate were used as x variables. The °ow rate was varied by means of a control valve with a control signal generated as ¯ltered white noise. Each power spectrum representing an X observation was formed as the mean value of 100 consecutive power spectra in the sampling interval, and the resulting y and X signals were recorded for N_{long} = 230 observations. This standard procedure in static multivariate calibration based on acoustic data is necessary in order to obtain a reasonable noise level. Only every 10th y sample was used in the estimator identi¯cation procedure, and the number of y observations in the modeling set were accordingly N_{short} = 23. The calibration was based on centered data.

The resulting predictors were validated against a separate data set with 170 X and y observations, with normalized RMSEP results as shown in Fig. 4.

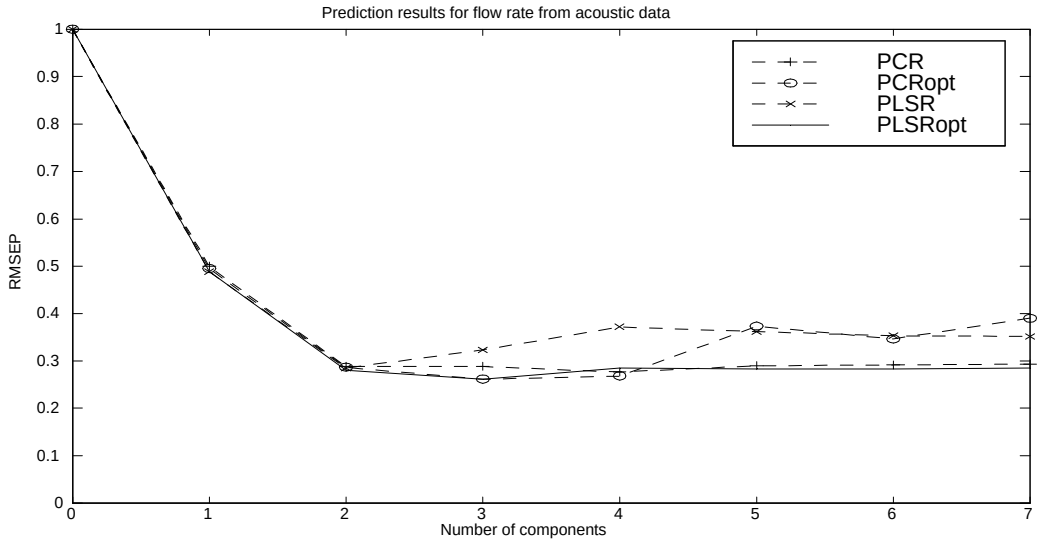

Figure 4. Results for °ow prediction from acoustic power spectrum, based on test set validation.

The R_{22} estimate used in the optimized PLSR solution was in this case based on A_{22} = 3 components (the PCRopt minimum). The relative RMSEP reduction from ordinary PLSR with A = 2 to the optimized PLSR and PCR solutions with A = A_{22} = 3 components is 8 %. The relative PLSR improvement at A = 2 is 1 %.

7 Conclusions

Methods for KF based PLSR and PCR optimizations utilizing estimated covariance matrices have been derived. The PCR method turns out to be identical with an earlier proposed method for stabilization of the loading matrix[4;5], while the PLSR method to the best of our knowledge is new. The theoretical results have been substantiated by both a Monte Carlo simulation and two applications on real-world calibration data sets. The results show clearly reduced RMSEP values at the optimal number of components, and in some cases also at a lower number of components. The basis for these improvements is only that more X- than y-observations are obtained, a trivial option for most multivariate calibration situations. However,

further tests on real-world data sets with extra X-observations are needed before a final judgement on the practical value of this methodology can be drawn.

Appendix

We give here some details concerning the optimized PCR and PLSR predictors.

Optimized PCR predictor

The optimized PCR predictor is given by (18), with K replaced by

$$\hat{K}_{PCR} = \hat{R}_e \hat{P}_{long}^T \left(\hat{P}_{long} \hat{R}_e \hat{P}_{long}^T + \hat{R}_{22} \right)^{-1}; \tag{28}$$

where PCA of X_{long} gives

$$\hat{R}_e = \frac{1}{N_{long} - 1} \hat{T}_{long}^T \hat{T}_{long} = \frac{1}{N_{long} - 1} \hat{P}_{long}^T X_{long}^T X_{long} \hat{P}_{long} \tag{29}$$

and from (22) and (5)

$$\hat{R}_{22} = \frac{1}{N_{long} - 1} \left(I - \hat{P}_{long} \hat{P}_{long}^T \right) X_{long}^T X_{long} \left(I - \hat{P}_{long} \hat{P}_{long}^T \right). \tag{30}$$

Since $\hat{P}_{long}^T \hat{P}_{long} = I$ it follows from (30) that $\hat{P}_{long}^T \hat{R}_{22} = 0$, and (28) therefore gives

$$\begin{aligned} \hat{K}_{PCR} &= \hat{P}_{long}^T \hat{P}_{long} \hat{R}_e \hat{P}_{long}^T \left(\hat{P}_{long} \hat{R}_e \hat{P}_{long}^T + \hat{R}_{22} \right)^{-1} \\ &= \hat{P}_{long}^T \left(\hat{P}_{long} \hat{R}_e \hat{P}_{long}^T + \hat{R}_{22} - \hat{R}_{22} \right) \left(\hat{P}_{long} \hat{R}_e \hat{P}_{long}^T + \hat{R}_{22} \right)^{-1} \\ &= \hat{P}_{long}^T - \hat{P}_{long}^T \hat{R}_{22} \left(P_{long} \hat{R}_e P_{long}^T + \hat{R}_{22} \right)^{-1} = \hat{P}_{long}^T. \end{aligned} \tag{31}$$

The optimized predictor is thus given by (24).

Optimized PLSR predictor

In the results in Section 5 and 6 it may be noticed that the optimized PLSR predictor at the number of components used for finding $\hat{R}_{22}$ is equal with the optimized PCR predictor (24). In order to prove that this is always the case we thus assume that the estimates $\hat{R}_e$ and $\hat{R}_{22}$ are based on the same number of components, i.e. $A = A_{PCR}$. From (20) follows

$$\hat{K}_{PLS} \left(\hat{W}_{PLS} \hat{R}_e \hat{W}_{PLS}^T + \hat{R}_{22} \right) = \hat{R}_e \hat{W}_{PLS}^T; \tag{32}$$

and thus

$$\hat{K}_{PLS} \hat{W}_{PLS} \hat{R}_e \hat{W}_{PLS}^T \hat{P}_{long} + \hat{K}_{PLS} \hat{R}_{22} \hat{P}_{long} = \hat{R}_e \hat{W}_{PLS}^T \hat{P}_{long}. \tag{33}$$

Since $\hat{R}_{22} \hat{P}_{long} = 0$ and since $\hat{R}_e \hat{W}_{PLS}^T \hat{P}_{long}$ is invertible it follows that $\hat{K}_{PLS} \hat{W}_{PLS} = I$. From (32) then follows

$$\hat{R}_e \hat{W}_{PLS}^T + \hat{K}_{PLS} \hat{R}_{22} = \hat{R}_e \hat{W}_{PLS}^T; \tag{34}$$

such that also $R_{22} \hat{K}_{PLS}^T = 0$:

Both $\hat{P}_{long}$; $\hat{R}_{22}$ and $\hat{K}^T$ are determined by use of X_{long}, and thus have bases which are subspaces of the row space of X_{long}. From (22) further follows that $\text{rank}(\hat{R}_{22}) = \min(N_{long}; p) - A$, and since

$\hat{R}_{22}\hat{K}^T = 0$ and $\hat{R}_{22}\hat{P}_{long} = 0$, while $\text{rank}(\hat{P}) = \text{rank}(\hat{K}^T) = A$, the columns of both $\hat{K}^T$ and $\hat{P}$ spans the nullspace of $\hat{R}_{22}$. From this ¯nally follows that $\hat{K} = \hat{M}\hat{P}_{long}^T$, where $\hat{M}$ is an invertible matrix that rotates the $\hat{P}_{long}$ column vectors without a®ecting the predictor. The optimized PLSR predictor at the number of components $A = A_{PCR}$ used for ¯nding $\hat{R}_{22}$ is thus equal with the optimized PCR predictor (24).

References

[1] Berntsen H. Utvidet Kalman¯lter og multivariabel kalibrering. Report STF48 A88019, SINTEF, Trondheim, Norway, 1988

[2] Ergon R. Dynamic System Multivariate Calibration for Optimal Primary Output Estimation. Ph.D. thesis at the Norwegian University of Science and Technology/Telemark University College, Trondheim/Porsgrunn, Norway, 1999

[3] Helland I.S. On the structure of partial least squares regression. Communications in statistics 1988;17:581-607.

[4] Isaksson T, N½s T. Selection of Samples for Calibration in Near-Infrared Spectroscopy. Part II: Selection Based on Spectral Measurements. Applied Spectroscopy 1990;44:1152-1158.

[5] Thomas E.V. Incorporating Auxiliary Predictor Variation in Principal Component Regression Models. J. Chemometrics 1995;9:471-481.

[6] Johnson R.A, Wichern D.W. Applied Multivariate Statistical Analysis, Prentice Hall: New Jersey, 1992.

[7] Belsley D.A. Conditioning Diagnostics: Collinearity and Weak Data in Regression, Wiley: New York, 1992

[8] Martens H, N½s T. Multivariate Calibration, Wiley: New York, 1989

[9] Burnham A.J, Viveros R, MacGregor J.F. Frameworks for Latent Variable Multivariate Regression. J. Chemometrics 1996;10:31-45.

[10] Grewal M.S, Andrews A.P. Kalman Filtering: Theory and Practice, Prentice Hall: New Jersey, 1993.

[11] Gelb A. Applied Optimal Estimation, MIT Press: Mass., 1974.

[12] Esbensen K.H. Multivariate Data Analysis - in practice, Camo ASA: Oslo, Norway, 2000.

[13] http://www.dal.ca/~pdwentze/download.htm

[14] Esbensen K.H, Hope B, Lied T.T, Halstensen M, Gravermoen T, Sundberg K. Acoustic chemometrics for °uid °ow quanti¯cations-II: A small constriction will go a long way. J. Chemometrics 1999;13:209-236.