# Re-interpretation of NIPALS results solves PLSR inconsistency problem

Rolf Ergon

rolf.ergon@hit.no

Telemark University College

Porsgrunn, Norway

Published in Journal of Chemometrics 2009; Vol. 23/1: 72-75

**Abstract**

The well known NIPALS algorithm is commonly used for computation of components in partial least squares regression (PLSR) with orthogonalized score vectors. Based on generalized inverse formalism, Pell et al. [1] have recently claimed that the NIPALS results are inconsistent with respect to model spaces for residual-based outlier detection and prediction purposes. This theoretically important result is supported by the present paper, where it is also shown that a simple re-interpretation of the results from the NIPALS algorithm solves the problem. This is valid for cases with both one and several response variables (PLS1 and PLS2). It is also shown, however, that the price to pay for this solution is that the latent variables in the model will no longer be completely independent of the residual noise. Since the original PLSR model with non-orthogonal score vectors does not have the same inconsistency, a method for orthogonalization of this model is also included.

**Keywords:** PLSR; NIPALS; model inconsistency; orthogonalization

## 1 Introduction

As described by Martens [2], the two original algorithms for partial least squares regression (PLSR) were developed in parallel, in close cooperation between the two originators. Wold et al. [3] developed a solution with orthogonal score vectors and non-orthogonal loading vectors, and for that purpose they used the well known nonlinear iterative partial least squares (NIPALS) algorithm. Martens (see , e.g., [4]), on the other hand, developed an algorithm resulting in non-orthogonal score vectors and orthogonal loading vectors. In practice, the first of these algorithms is usually preferred, based on more or less realistic assumptions of independent latent variables. A nice property of the Wold solution is also that the latent variables are completely independent of the residual noise, which is not the case for the Martens solution. In some applications, however, as for example in correspondence analysis [5], orthogonal loading vectors are to be preferred. As shown below, it is an easy task to go from one resulting model to the other.

Helland [6] proved formally that the two algorithms are equivalent in the sense that the score vectors span the same space. However, based on Moore-Penrose generalized inverse formalism, Pell et al. [1] have recently claimed that the NIPALS results are inconsistent with respect to model spaces for residual-based outlier detection and prediction purposes. This is supported by the present paper. The problem is that the traditional NIPALS interpretation results in one model space for computation of the regression vector, and another model space for representation of the reconstructed data and thus for outlier detection based on residuals. When using the alternative algorithm of Martens, giving non-orthogonal score vectors, we do not have this problem.

As a conclusion of their findings, Pell et al. [1] seem to indicate that commercial programs based on the NIPALS algorithm ought to be replaced by programs based on the bidiagonalization (Bidiag2) algorithm of Golub and Kahan [7], which gives the same model as the Martens algorithm. This conclusion is far too dramatic. As indicated in several papers by Ergon [5,8,9,10], a simple re-interpretation of the NIPALS results will solve the problem. One of these papers is in fact referred to by Pell et al.. Another possibility is to orthogonalize the score vectors from the Bidiag2 or Martens algorithms, or some other equivalent non-orthogonalized PLSR solution.

The present paper shows that the residual inconsistency is present also when predictions are computed without explicit use of regression coefficients. It also presents the theoretical basis for the re-interpreted NIPALS (RE-NIPALS) solution to the problem, both in single response and multiresponse cases (PLS1 and PLS2). A simple procedure for orthogonalization of models with non-orthogonal score vectors is also included. Finally, orthogonality properties of the different models are investigated, and it is shown that when the residual inconsistency problem is solved, the complete independence between latent variables and residual noise is at the same time lost.

## 2 A re-interpretation of the NIPALS algorithm

**The traditional NIPALS algorithm**

As a starting point for further discussions a short summary and simplification of the well known NIPALS PLS1 algorithm for centered $\mathbf{X}$ and $\mathbf{y}$ data follows (see, e.g., [11] for a more detailed formulation, although for simplicity we here make use of the fact that deflation of $\mathbf{y}$ is not necessary):

1. Let $\mathbf{X}_0 = \mathbf{X}$. For $a = 1, 2, \cdots, A$ perform steps 2 to 6 below.
2. Compute $\mathbf{w}_a = \mathbf{X}_{a-1}^T\mathbf{y} / \left\|\mathbf{X}_{a-1}^T\mathbf{y}\right\|$ (with length 1).
3. Compute $\mathbf{t}_a = \mathbf{X}_{a-1}\mathbf{w}_a$.
4. Compute $q_a = \mathbf{y}^T\mathbf{t}_a(\mathbf{t}_a^T\mathbf{t}_a)^{-1}$
5. Compute $\mathbf{p}_a = \mathbf{X}_{a-1}^T\mathbf{t}_a(\mathbf{t}_a^T\mathbf{t}_a)^{-1}$
6. Compute the residual $\mathbf{X}_a = \mathbf{X}_{a-1} - \mathbf{t}_a\mathbf{p}_a^T$.
7. The resulting latent variables model is

$$\mathbf{y} = \mathbf{T}_{\mathrm{W}}\mathbf{q}_{\mathrm{W}} + \mathbf{f} \quad (1)$$

$$\mathbf{X} = \mathbf{T}_{\mathrm{W}}\mathbf{P}^T + \mathbf{E}_{\mathrm{W}}, \quad (2)$$

where $\mathbf{T}_W = \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_2 & \cdots & \mathbf{t}_A \end{bmatrix}$ is orthogonal, $\mathbf{P} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_A \end{bmatrix}$, $\mathbf{q}_W = \begin{bmatrix} q_1 & q_2 & \cdots & q_A \end{bmatrix}^T$ and $\mathbf{E}_W = \mathbf{X}_A$, and where the subscript W indicates the orthogonalized PLSR model of Wold et al. [3]. The matrix $\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_A \end{bmatrix}$ resulting from the algorithm is orthonormal.

**Remark 1** *In what follows we assume $A < N$, where $N$ is the number of rows in $\mathbf{X}$. Otherwise important properties of the results from the algorithm will break down, for example will $\mathbf{W}$ no longer be orthonormal, and $\mathbf{P}^T\mathbf{W}$ not bidiagonal (see Appendix A). Since $\mathbf{X}_p = \mathbf{0}$, where $p$ is the number of variables in $\mathbf{X}$, we must also assume $A \leq p$.*

### The traditional regression coefficients

The traditional formula for computation of the regression coefficients based on the NIPALS results is [11]

$$\hat{\mathbf{b}} = \mathbf{W}(\mathbf{P}^{\mathbf{T}}\mathbf{W})^{-1}\mathbf{q}_W, \tag{3}$$

and the major claim by Pell et al. [1] is that this is not consistent with the factorization (2). This result is supported by results developed below, where a simple solution to the problem is also presented.

### Non-orthogonalized PLSR

The orthogonalized PLSR model (1,2) may be compared with the non-orthogonalized PLSR model of Martens [4], given by

$$\mathbf{y} = \mathbf{T}_M\mathbf{q}_M + \mathbf{f} \tag{4}$$

$$\mathbf{X} = \mathbf{T}_M\mathbf{W}^T + \mathbf{E}_M, \tag{5}$$

where $\mathbf{T}_M = \mathbf{X}\mathbf{W}$ and $\mathbf{q}_M = \left(\mathbf{T}_M^T\mathbf{T}_M\right)^{-1}\mathbf{T}_M^T\mathbf{y}$. Note here that the final model spaces $\hat{\mathbf{X}}_W = \mathbf{T}_W\mathbf{P}^T$ and $\hat{\mathbf{X}}_M = \mathbf{T}_M\mathbf{W}^T$, and thus the residuals $\mathbf{E}_W$ and $\mathbf{E}_M$ are generally different. Also note that the model (4,5) can be established once $\mathbf{W}$ is found by use of the NIPALS algorithm, i.e. no special algorithm is needed. The regression coefficient vector follows as a simple least squares (LS) solution from the model (4,5),

$$\hat{\mathbf{b}} = \mathbf{W}\left(\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W}\right)^{-1}\mathbf{W}^T\mathbf{X}^T\mathbf{y}. \tag{6}$$

As shown in Appendix B this will give the same coefficients as the traditional formula (3).

**A re-interpretation of the NIPALS results (PLS1)**

The re-interpretation is based on the fact that steps 5 to 6 in the NIPALS algorithm above serves the purpose of preparing for computation of the next orthogonal score vector. When the last score vector has been found, we may give up this orthogonality requirement, and stop after step 4 in the algorithm. The residual may thus be computed as $\mathbf{E} = \mathbf{X}_{A-1} - \mathbf{t}_A\mathbf{w}_A^T$, and the decomposition of $\mathbf{X}$ will then become

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \cdots + \mathbf{t}_{A-1}\mathbf{p}_{A-1}^T + \mathbf{t}_A\mathbf{w}_A^T + \mathbf{E}, \tag{7}$$

where as shown in Appendix A $\mathbf{E} = \mathbf{E}_{\mathrm{M}}$ in Eq. (5). In Appendix A it is also shown that the matrix formulation of this RE-NIPALS factorization results in the model (where Eq. (1) is repeated for clarity)

$$\mathbf{y} = \mathbf{T}_{\mathrm{W}}\mathbf{q}_{\mathrm{W}} + \mathbf{f} \tag{8}$$

$$\mathbf{X} = \mathbf{T}_{\mathrm{W}}\mathbf{P}^T\mathbf{W}\mathbf{W}^T + \mathbf{E}_{\mathrm{M}}. \tag{9}$$

From this follows the LS solution $\mathbf{T}_{\mathrm{W}} = \mathbf{X}\mathbf{W}\left(\mathbf{P}^T\mathbf{W}\right)^{-1}$ and thus

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}} = \mathbf{T}_{\mathrm{W}}\mathbf{q}_{\mathrm{W}} = \mathbf{X}\mathbf{W}\left(\mathbf{P}^T\mathbf{W}\right)^{-1}\mathbf{q}_{\mathrm{W}}. \tag{10}$$

This is consistent with the traditional regression coefficients $\hat{\mathbf{b}}$ in Eq. (3). It is also trivial to show that $\hat{\mathbf{b}}$ in Eq. (3) follows from Eq. (9) according to the Moore-Penrose generalized inverse formalism given in [1]. The simple solution to the inconsistency problem is thus to replace the factorization (2) in step 7 of the algorithm above with the factorization (9). Note, however, that we by this interpretation of the NIPALS algorithm give up the orthogonality requirement regarding the last score vector $\mathbf{t}_A$ and the residual $\mathbf{E}_{\mathrm{M}}$ (see Section 4 for a discussion of orthogonality properties).

**Special case with maximum number of components**

It is of special interest to study the case with $A = p$ components, where $p < N$ is the number of $x$-variables. Since $\mathbf{W}$ in this case is invertible, we find from Eq. (6) that $\hat{\mathbf{b}} = \mathbf{W}\left(\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W}\right)^{-1}\mathbf{W}^T\mathbf{X}^T\mathbf{y} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$, i.e. the ordinary LS solution, as must be expected. This also means that all data in $\mathbf{X}$ is utilized, such that $\mathbf{E}_{\mathrm{M}} = \mathbf{0}$. From Eq. (5) then follows $\mathbf{X} = \mathbf{X}\mathbf{W}\mathbf{W}^T$, and thus $\mathbf{W}\mathbf{W}^T = \mathbf{I}$. This means according to Eq. (9) that $\mathbf{X} = \mathbf{T}_{\mathrm{W}}\mathbf{P}^T\mathbf{W}\mathbf{W}^T + \mathbf{E}_{\mathrm{M}} = \mathbf{T}_{\mathrm{W}}\mathbf{P}^T$. According to Eq. (7) we must then have $\mathbf{p}_p = \mathbf{w}_p$, and from the final residual $\mathbf{E}_{\mathrm{M}} = \mathbf{X}_p = \mathbf{X}_{p-1} - \mathbf{t}_p\mathbf{w}_p^T = \mathbf{0}$ we thus find $\mathbf{X}_{p-1} = \mathbf{t}_p\mathbf{w}_p^T$. The NIPALS algorithm will accordingly give $\mathbf{p}_p = \mathbf{X}_{p-1}^T\mathbf{t}_p(\mathbf{t}_p^T\mathbf{t}_p)^{-1} = \mathbf{w}_p\mathbf{t}_p^T\mathbf{t}_p(\mathbf{t}_p^T\mathbf{t}_p)^{-1} = \mathbf{w}_p$. For $A = p$ the traditional NIPALS and the proposed RE-NIPALS interpretations will thus give the same result.

**Alternative way of finding predictions**

When parameter values in $\mathbf{q}_{\mathrm{W}}$, $\mathbf{W}$ and $\mathbf{P}$ have been determined by use of the NIPALS algorithm and modeling data $\mathbf{X}$ and $\mathbf{y}$, new score vectors for new data $\mathbf{X}_{\mathrm{new}}$ with one or several objects can be found by again using the NIPALS algorithm, but now with these parameters fixed. The score vectors then found will according to Eq. (1) give

$$\hat{\mathbf{y}}_{\mathrm{new}} = \mathbf{T}_{\mathrm{new}}\mathbf{q}_{\mathrm{W}}, \tag{11}$$

i.e. without explicit use of the regression coefficients. As shown in Appendix C, however, one is then implicitly using regression coefficients according to Eq. (3), based on the factorization (9), i.e. $\mathbf{T}_{\text{new}} = \mathbf{X}_{\text{new}}\mathbf{W}\left(\mathbf{P}^T\mathbf{W}\right)^{-1}$. This result is mentioned without details in Pell et al. [1], and it is in agreement with their inconsistency claim.

**The multiresponse case (PLS2)**

In the NIPALS algorithm for multiresponse data (PLS2), $\mathbf{w}_a$, $\mathbf{t}_a$ and $\mathbf{p}_a$ are found in an inner iterative loop, together with an appropriate combination $\mathbf{u}_a$ of the $y$-variables (see, e.g., [11] for details). The regression coefficients will be given by Eq. (6), with the vector $\mathbf{y}$ replaced by a matrix $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_m \end{bmatrix}$. The inconsistency problem will be the same as for PLS1, and a simple solution is also here a NIPALS re-interpretation as described above.

## 3 Conversion from one model to the other

Since the two models (4,5) and (8,9) have the same residuals they are equivalent, and it should therefore be possible to convert one of the models into the other. A comparison of the factorizations shows that

$$\mathbf{T}_{\text{M}} = \mathbf{T}_{\text{W}}\mathbf{P}^T\mathbf{W}, \tag{12}$$

i.e. conversion from the modified orthogonalized form to the non-orthogonalized form is straightforward, since $\mathbf{T}_{\text{W}}$, $\mathbf{P}$ and $\mathbf{W}$ are then known. Conversion the other way around is somewhat more complicated, since $\mathbf{P}$ is then not known, but using for example Matlab it is easily done by the QR factorization [12]

$$\mathbf{T}_{\text{M}} = \mathbf{QR} = \mathbf{QZ}^{-1}\mathbf{ZR}. \tag{13}$$

Here $\mathbf{Z}$ is a diagonal scaling matrix giving $\mathbf{ZR} = \mathbf{P}^T\mathbf{W}$ as an upper triangular matrix with ones along the main diagonal, where we make use of the known bidiagonal structure of $\mathbf{P}^T\mathbf{W}$ [13] (see also Appendix A). From Eqs. (12) and (13) then also follows the orthogonalized score matrix

$\mathbf{T}_{\mathrm{W}} = \mathbf{T}_{\mathrm{M}} \left(\mathbf{P}^T\mathbf{W}\right)^{-1} = \mathbf{Q}\mathbf{Z}^{-1}$. As an alternative to the NIPALS algorithm, with re-interpretation as shown above, we may thus start by using the algorithm of Martens [4] for non-orthogonalized PLSR, and then perform an orthogonalization of the score matrix. We may also start by using the Bidiag2 algorithm [7], or, e.g., the alternative algorithm of Ergon and Esbensen [14]. When the model (8,9) is found, the original orthogonalized model (1,2) is obtained by removal of $\mathbf{W}\mathbf{W}^T$.

# 4 Orthogonality properties

One aim of orthogonalized PLSR is obviously to obtain orthogonal score vectors, corresponding to independent latent variables. The same orthogonal score vectors are obtained also with the consistent model (8,9), using $\mathbf{T}_{\mathrm{W}}$ as score matrix and $\mathbf{W}\mathbf{W}^T\mathbf{P}$ as loading matrix.

Another aim of the orthogonalized PLSR model (1,2) is presumably to obtain independence between latent variables and noise in the residuals, and this is in fact obtained since $\mathbf{T}_{\mathrm{W}}^T\mathbf{E}_{\mathrm{W}} = \mathbf{0}$ (which follows trivially from the NIPALS algorithm). Using $\mathbf{T}_{\mathrm{W}}$ as score matrix in the consistent model (8,9), on the other hand, we obtain $\mathbf{T}_{\mathrm{W}}^T\mathbf{E}_{\mathrm{M}} = \mathbf{T}_{\mathrm{W}}^T \left(\mathbf{E}_{\mathrm{W}} + \mathbf{t}_A \left(\mathbf{p}_A^T - \mathbf{w}_A^T\right)\right)$, where we make use of Eq. (7). Since the score vectors are orthogonal, all rows except the last one, $\mathbf{t}_A^T\mathbf{E}_{\mathrm{M}}$, will here be zero. The residual noise is thus independent of all latent variables except the last one.

# 5 Discussion

Looking at both residual inconsistency and noise independence properties following from the NIPALS algorithm, it appears that no ideal solution is available, assuming continued use of the traditional regression coefficients formula $\hat{\mathbf{b}} = \mathbf{W}\left(\mathbf{P}^T\mathbf{W}\right)^{-1}\mathbf{q}_{\mathrm{W}}$:

- When using the traditional interpretation and the model (1,2), one obtains $\mathbf{T}_{\mathrm{W}}^T\mathbf{E}_{\mathrm{W}} = \mathbf{0}$, i.e. complete independence between score vectors and noise in the residuals. The price to pay for this is the residual inconsistency found by Pell et al. [1], which as shown by their example may cause considerable errors in residual based outlier detection. It should be mentioned, however, that common outlier detection methods based on the score vectors will not be affected by this.

- When using a re-interpretation of the NIPALS results as presented above, the residual inconsistency problem is solved. The price to pay is then that $\mathbf{t}_A^T\mathbf{E}_{\mathrm{M}} \neq \mathbf{0}$, i.e. the independence between latent variables and noise in the residuals will no longer be complete. The importance of this in different applications is left as an open question.

From this follows that one has to make a choice between residual inconsistency using $\mathbf{X} = \mathbf{T}_{\mathrm{W}}\mathbf{P}^T + \mathbf{E}_{\mathrm{W}}$, or latent variable and residual noise covariance using $\mathbf{X} = \mathbf{T}_{\mathrm{W}}\mathbf{P}^T\mathbf{W}\mathbf{W}^T + \mathbf{E}_{\mathrm{M}}$. The best choice to make might well be application dependent.

# 6 Conclusion

The well known NIPALS algorithm for computation of PLSR components with orthogonalized score vectors, is recently claimed to give mathematically inconsistent results with respect to model spaces for residual-based outlier detection and prediction purposes. Results in the present paper support this inconsistency result. It is also shown that the inconsistency problem is easily solved by a simple re-interpretation of the results from the NIPALS algorithm, and this is valid for cases with both one and several response variables (PLS1 and PLS2). The price to pay for this is that the last latent variable will no longer be completely independent of the residual noise. A corresponding modification of commercial programs using the NIPALS algorithm should be easy to perform.

Alternative algorithms giving models with non-orthogonal score vectors do not give the same inconsistency problem, and a procedure for orthogonalization of such models is therefore included. The results will then be the same as after the described re-interpretation of the NIPALS results, including the last latent variable and residual noise covariance.

# A Detailed RE-NIPALS results

Comparing the two factorizations (5) and (9), we must show that $\mathbf{T}_{\mathrm{W}}\mathbf{P}^T\mathbf{W} = \mathbf{T}_{\mathrm{M}} = \mathbf{X}\mathbf{W}$. Making use of the known bidiagonal structure of the product $\mathbf{P}^T\mathbf{W}$ [13], we find

$$\begin{aligned}
\mathbf{T}_{\mathrm{W}}\mathbf{P}^T\mathbf{W} &= \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_2 & \cdots & \mathbf{t}_a & \mathbf{t}_{a+1} & \cdots & \mathbf{t}_A \end{bmatrix} \\
&\times \begin{bmatrix}
1 & \mathbf{p}_1^T\mathbf{w}_2 & 0 & \cdots & 0 & \cdots & 0 \\
0 & 1 & \ddots & \ddots & & & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & & \vdots \\
0 & & \ddots & \ddots & \mathbf{p}_a^T\mathbf{w}_{a+1} & \ddots & \vdots \\
0 & & & \ddots & 1 & \ddots & 0 \\
\vdots & & & & \ddots & \ddots & \mathbf{p}_{A-1}^T\mathbf{w}_A \\
0 & 0 & \cdots & \cdots & \cdots & 0 & 1
\end{bmatrix} \\
&= \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_1\mathbf{p}_1^T\mathbf{w}_2 + \mathbf{t}_2 & \cdots & \mathbf{t}_a\mathbf{p}_a^T\mathbf{w}_{a+1} + \mathbf{t}_{a+1} & \cdots & \mathbf{t}_{A-1}\mathbf{p}_{A-1}^T\mathbf{w}_A + \mathbf{t}_A \end{bmatrix} \quad (14)
\end{aligned}$$

In addition to the fact that step 3 in the NIPALS algorithm gives $\mathbf{t}_1 = \mathbf{X}\mathbf{w}_1$, we must thus also show that $\mathbf{t}_a\mathbf{p}_a^T\mathbf{w}_{a+1} + \mathbf{t}_{a+1} = \mathbf{X}\mathbf{w}_{a+1}$. Also this can be shown by use of the bidiagonal structure of $\mathbf{P}^T\mathbf{W}$, which in step 3 of the NIPALS algorithm generally gives

$$\mathbf{t}_{a+1} = \mathbf{X}_a\mathbf{w}_{a+1} = \left(\mathbf{X} - \mathbf{t}_1\mathbf{p}_1^T - \cdots - \mathbf{t}_a\mathbf{p}_a^T\right)\mathbf{w}_{a+1} = \mathbf{X}\mathbf{w}_{a+1} - \mathbf{t}_a\mathbf{p}_a^T\mathbf{w}_{a+1}, \quad (15)$$

and thus

$$\mathbf{t}_a\mathbf{p}_a^T\mathbf{w}_{a+1} + \mathbf{t}_{a+1} = \mathbf{X}\mathbf{w}_{a+1}. \quad (16)$$

# B Two formulas for the regression coefficients

We here show that the two formulas (3) and (6), $\hat{\mathbf{b}} = \mathbf{W}(\mathbf{P}^{\mathbf{T}}\mathbf{W})^{-1}\mathbf{q}_{\mathrm{W}}$ and $\hat{\mathbf{b}} = \mathbf{W}\left(\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W}\right)^{-1}\mathbf{W}^T\mathbf{X}^T\mathbf{y}$, give the same regression coefficients. Using the LS solution of Eq. (1), $\mathbf{q}_{\mathrm{W}} = \left(\mathbf{T}_{\mathrm{W}}^T\mathbf{T}_{\mathrm{W}}\right)^{-1}\mathbf{T}_{\mathrm{W}}^T\mathbf{y}$, and the relation $\mathbf{T}_{\mathrm{W}} = \mathbf{T}_{\mathrm{M}} \left(\mathbf{P}^T\mathbf{W}\right)^{-1}$ shown in Appendix A, together with $\mathbf{T}_{\mathrm{M}} = \mathbf{X}\mathbf{W}$, we find

$$\begin{aligned}
\hat{\mathbf{b}} &= \mathbf{W}\left(\mathbf{P}^T\mathbf{W}\right)^{-1}\mathbf{q}_{\mathrm{W}} = \mathbf{W}\left(\mathbf{P}^T\mathbf{W}\right)^{-1}\left(\mathbf{T}_{\mathrm{W}}^T\mathbf{T}_{\mathrm{W}}\right)^{-1}\mathbf{T}_{\mathrm{W}}^T\mathbf{y} \\
&= \mathbf{W}\left(\mathbf{P}^T\mathbf{W}\right)^{-1}\left(\left(\mathbf{W}^T\mathbf{P}\right)^{-1}\mathbf{T}_{\mathrm{M}}^T\mathbf{T}_{\mathrm{M}}\left(\mathbf{P}^T\mathbf{W}\right)^{-1}\right)^{-1}\left(\mathbf{W}^T\mathbf{P}\right)^{-1}\mathbf{T}_{\mathrm{M}}^T\mathbf{y} \\
&= \mathbf{W}\left(\mathbf{T}_{\mathrm{M}}^T\mathbf{T}_{\mathrm{M}}\right)^{-1}\mathbf{T}_{\mathrm{M}}^T\mathbf{y} = \mathbf{W}\left(\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W}\right)^{-1}\mathbf{W}^T\mathbf{X}^T\mathbf{y}. \qquad (17)
\end{aligned}$$

# C Prediction using implicit regression coefficients

Based on $\mathbf{q}_{\mathrm{W}}$, $\mathbf{W}$ and $\mathbf{P}$ from the NIPALS algorithm applied on modeling data $\mathbf{X}$ and $\mathbf{y}$, new score vectors and predictions for new data $\mathbf{X}_{\mathrm{new}}$ with one or several objects can be found by the following procedure:

1. Let $\mathbf{X}_0 = \mathbf{X}_{\mathrm{new}}$. For $a = 1, 2, \cdots, A$ perform steps 2 to 3 below.

2. Compute $\mathbf{t}_a^{\mathrm{new}} = \mathbf{X}_{a-1}\mathbf{w}_a$.

3. Compute the residual $\mathbf{X}_a = \mathbf{X}_{a-1} - \mathbf{t}_a^{\mathrm{new}}\mathbf{p}_a^T$.

This gives

$$\mathbf{t}_{a+1}^{\mathrm{new}} = \left(\mathbf{X}_{\mathrm{new}} - \mathbf{t}_1^{\mathrm{new}}\mathbf{p}_1^T - \cdots - \mathbf{t}_a^{\mathrm{new}}\mathbf{p}_a^T\right)\mathbf{w}_{a+1} = \mathbf{X}_{\mathrm{new}}\mathbf{w}_{a+1} - \mathbf{t}_a^{\mathrm{new}}\mathbf{p}_a^T\mathbf{w}_{a+1}, \qquad (18)$$

where we again make use of the bidiagonal structure of $\mathbf{P}^T\mathbf{W}$. This is the same relation as in Eq. (16) in Appendix A above, and from the main result in Appendix A thus follows that

$$\mathbf{T}_{\mathrm{W}}^{\mathrm{new}} = \mathbf{T}_{\mathrm{M}}^{\mathrm{new}}(\mathbf{P}^{\mathbf{T}}\mathbf{W})^{-1} = \mathbf{X}^{\mathrm{new}}\mathbf{W}(\mathbf{P}^{\mathbf{T}}\mathbf{W})^{-1}, \qquad (19)$$

which further gives the predictions

$$\hat{\mathbf{y}}^{\mathrm{new}} = \mathbf{T}_{\mathrm{W}}^{\mathrm{new}} \mathbf{q}_{\mathrm{W}} = \mathbf{X}^{\mathrm{new}} \mathbf{W} (\mathbf{P}^{\mathbf{T}} \mathbf{W})^{-1} \mathbf{q}_{\mathrm{W}}, \tag{20}$$

and thus as with explicit use of regression coefficients

$$\hat{\mathbf{b}} = \mathbf{W} (\mathbf{P}^{\mathbf{T}} \mathbf{W})^{-1} \mathbf{q}_{\mathrm{W}}. \tag{21}$$

## References


[1] Pell R.J., Ramos L.S., Manne R. The model space in partial least squares regression. *J. Chemometrics* 2007; **21**: 165-172.

[2] Martens H. Reliable and relevant modelling of real world data: a personal account of the development of PLS Regression. *Chemometrics Intell. Lab. Syst.* 2001; **58**: 85-95.

[3] Wold S., Martens H., Wold H. The multivariate calibration-problem in chemistry solved by the PLS method. *Lecture notes in mathematics*. Plenum Publ. Corp.: New York, 1983; **973**: 286-293

[4] Martens H., Næs T. Multivariate calibration by data compression. *Near-infrared technology in agricultural and food industries*. Am. Assoc. Cerial Chem.: St. Paul, Minnesota, 1987: 57-87

[5] Ergon R. PLS score-loading correspondence and a bi-orthogonal factorization. *J. Chemometrics* 2002; **16**: 368-373.

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

[7] Golub G.H., Kahan W. Calculating the singular values and pseudo-inverse of a matrix. *SIAM J. Scientific. Comput.* 1984; **2**: 205-224

[8] Ergon R. Dynamic system multivariate calibration. *Chemometrics Intell. Lab. Syst.* 1998; **44**: 135-146.

[9] Ergon R. Noise Handling Capabilities of Multivariate Calibration Methods. *Modeling, Identification and Control* 2002; **23**: 259-273

[10] Ergon R. Finding **Y**-relevant parts of **X** by use of PCR and PLSR model reduction methods. *J. Chemometrics* 2007; **21**: 537-546.

[11] Martens H., Næs T. *Multivariate Calibration.* Wiley: New York, 1989.

[12] Golub G.H, Van Loan C.F. *Matrix Computations.* The Johns Hopkins University Press: Baltimore, 1996.

[13] Manne R. Analysis of two partial-least-squares algorithms for multivariate calibration. *Chemometrics Intell. Lab. Syst.* 1987; **2**: 187-197.

[14] Ergon R., Esbensen K.H. A didactically motivated PLS algorithm. *Modeling, Identification and Control* 2001; **22**: 131-13