PLS score-loading correspondence and a bi-orthogonal factorization

Rolf Ergon
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

Published in Journal of Chemometrics 2002, 16: 368-373

Abstract

It is established industrial practice to use the correspondence between partial least square (PLS) scores and loadings or loading weights as a means for process monitoring and control. Deviations from the normal operating point in a score plot is then related to the influences from major process variables as shown in a loading or loading weight plot. These relations are often presented in a bi-plot, i.e. appropriately scaled scores and loadings or loading weights are displayed in the same plot. As shown in the present article, however, the orthogonal PLS algorithm of Wold gives no direct theoretical and graphical correspondence, i.e. the bi-plot will show an angle deviation that causes an interpretational problem. The alternative non-orthogonal PLS algorithm of Martens gives direct correspondence, but the correlated latent variables may then cause another interpretational problem. As a solution to these problems the article presents a PLS factorization where both scores and loadings are orthogonal (BPLS), and we show how the Wold and Martens factorizations can easily be transformed to this solution. The result is independent latent variables as well as direct score and loading correspondence. It is also shown that the transformations involved do not affect the predictor found by PLS regression. The score-loading correspondence properties for the different PLS factorizations are discussed using principal component analysis (PCA) as a reference case. An example using industrial paper plant data is included.

KEYWORDS: PLS, factorization, score-loading correspondence

1 Introduction and problem statement

It is established industrial practice to use the correspondence between partial least squares (PLS) scores and loadings or loading weights as a means for process monitoring and control.[1] Deviations from the normal operating point in a score plot is then related to the influences from major process variables as shown in a loading or loading weight plot. This is normally done in a bi-plot, i.e. appropriately scaled scores and loadings or loading weights are displayed in the same plot. Such correspondence is of interest also in a number of other application areas[2,3]

The existing PLS algorithms[4] are, however, not ideally suited for this purpose:

- The orthogonal algorithm of Wold uses independent latent variables, which in many cases reflects the underlying sources of variation. However, as shown in Section 3 the theoretical and graphical correspondence between scores and the variable representations in the loading plots is obscured by the fact that the loadings are non-orthogonal. It is also shown that the alternative use of loading weight plots is no solution to this problem. In both cases there will be an angle deviation and thus a certain lack of interpretability.

- The non-orthogonal algorithm of Martens uses correlated latent variables, which may be in conflict with a natural and simple interpretation. However, as shown in Section 3 this algorithm results in a direct graphical correspondence between scores and loading weights.

Each of the Wold and Martens factorizations thus have both good and less satisfying interpretational properties. A possible solution to this problem is to use a principal component analysis (PCA) factorization of the data matrix $\mathbf{X}$ instead of the PLS factorization, but this will in some cases give a less parsimonious model using more components, which in itself reduces the correspondence interpretability. A central problem of the present article is therefore to find how the two PLS factorizations can be transformed to a unified bi-orthogonal solution (BPLS), where the scores are orthogonal and the loadings orthonormal, just as in PCA. Simple transformations for this purpose based on a singular value decomposition (SVD) are presented in Section 2, and it is also shown that these transformations do not affect the final PLS regression predictor. As indicated in Section 2, the BPLS factorization might also have interesting properties other than the ones used in the correspondence context. However, a general investigation of these properties is beyond the scope of the present article.

Section 3 discusses correspondence properties of the different factorizations, using PCA as a reference case, Section 4 presents an industrial data example, and conclusions follow in Section 5.

2 A bi-orthogonal PLS factorization

Data matrix factorizations

A rather general factorization of a data matrix $\mathbf{X} \in \mathbb{R}^{N\times p}$ appearing in regression is

$$\mathbf{X} = \hat{\mathbf{U}}\hat{\mathbf{R}}\hat{\mathbf{V}}^T + \mathbf{E} = \hat{\mathbf{T}}\hat{\mathbf{V}}^T + \mathbf{E}, \tag{1}$$

where $\hat{\mathbf{U}} \in \mathbb{R}^{N\times A}$ and $\hat{\mathbf{V}} \in \mathbb{R}^{p\times A}$ are matrices with orthonormal columns and $\hat{\mathbf{R}} \in \mathbb{R}^{A\times A}$ is an invertible matrix.

In SVD/PCA the matrix $\hat{\mathbf{R}}$ is diagonal, resulting in

$$\mathbf{X} = \hat{\mathbf{U}}_{\mathrm{PCA}}\hat{\mathbf{R}}_{\mathrm{PCA}}\hat{\mathbf{V}}_{\mathrm{PCA}}^T + \mathbf{E}_{\mathrm{PCA}} = \hat{\mathbf{T}}_{\mathrm{PCA}}\hat{\mathbf{P}}_{\mathrm{PCA}}^T + \mathbf{E}_{\mathrm{PCA}}, \tag{2}$$

where the score matrix $\hat{\mathbf{T}}_{\mathrm{PCA}} = \hat{\mathbf{U}}_{\mathrm{PCA}}\hat{\mathbf{R}}_{\mathrm{PCA}} \in \mathbb{R}^{N\times A}$ has orthogonal columns, while the loading matrix $\hat{\mathbf{P}}_{\mathrm{PCA}} = \hat{\mathbf{V}}_{\mathrm{PCA}} \in \mathbb{R}^{p\times A}$ has orthonormal columns.

In PLS1 (a single response variable) $\hat{\mathbf{R}}$ is right bi-diagonal[5]. The Wold factorization is

$$\mathbf{X} = \hat{\mathbf{U}}_{\mathrm{Wold}}\hat{\mathbf{\Lambda}}_{\mathrm{Wold}}^{\frac{1}{2}}\hat{\mathbf{P}}_{\mathrm{Wold}}^T\hat{\mathbf{W}}\hat{\mathbf{W}}^T + \mathbf{E}_{\mathrm{PLS}} = \hat{\mathbf{T}}_{\mathrm{Wold}}\hat{\mathbf{P}}_{\mathrm{Wold}}^T\hat{\mathbf{W}}\hat{\mathbf{W}}^T + \mathbf{E}_{\mathrm{PLS}}, \tag{3}$$

where $\hat{\mathbf{\Lambda}}_{\mathrm{Wold}} = \hat{\mathbf{T}}_{\mathrm{Wold}}^T\hat{\mathbf{T}}_{\mathrm{Wold}} \in \mathbb{R}^{A\times A}$ is diagonal, the score matrix $\hat{\mathbf{T}}_{\mathrm{Wold}} \in \mathbb{R}^{N\times A}$ has orthogonal columns, the loading matrix $\hat{\mathbf{P}}_{\mathrm{Wold}} \in \mathbb{R}^{p\times A}$ is non-orthogonal and the loading weight matrix $\hat{\mathbf{W}} \in \mathbb{R}^{p\times A}$ has orthonormal columns. Note that $\hat{\mathbf{V}}_{\mathrm{Wold}} = \hat{\mathbf{W}}\hat{\mathbf{W}}^T\hat{\mathbf{P}}_{\mathrm{Wold}}$ also is non-orthogonal. The Martens factorization is

$$\mathbf{X} = \hat{\mathbf{U}}_{\mathrm{Wold}}\hat{\mathbf{\Lambda}}_{\mathrm{Wold}}^{\frac{1}{2}}\hat{\mathbf{P}}_{\mathrm{Wold}}^T\hat{\mathbf{W}}\hat{\mathbf{W}}^T + \mathbf{E}_{\mathrm{PLS}} = \hat{\mathbf{T}}_{\mathrm{Martens}}\hat{\mathbf{W}}^T + \mathbf{E}_{\mathrm{PLS}}, \tag{4}$$

where $\mathbf{T}_{\mathrm{Martens}} = \hat{\mathbf{U}}_{\mathrm{Wold}}\hat{\mathbf{\Lambda}}_{\mathrm{Wold}}^{\frac{1}{2}}\hat{\mathbf{P}}_{\mathrm{Wold}}^T\hat{\mathbf{W}} \in \mathbb{R}^{N\times A}$ is non-orthogonal.

Unifying transformations

As pointed out in the introduction there is a need for a PLS factorization with an orthogonal score matrix and an orthonormal loading matrix, just as in PCA. Such a bi-orthogonal PLS factorization

(BPLS) may be found by use of SVD. After decomposition of $\hat{\mathbf{T}}_{\text{Martens}}$, the Martens factorization (4) can be transformed according to

$$\begin{aligned}\mathbf{X} &= \hat{\mathbf{T}}_{\text{Martens}}\hat{\mathbf{W}}^T + \mathbf{E}_{\text{PLS}} = \mathbf{U}_{\text{SVD}}\mathbf{S}_{\text{SVD}}\mathbf{V}_{\text{SVD}}^T\hat{\mathbf{W}}^T + \mathbf{E}_{\text{PLS}} \\ &= \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix}\begin{bmatrix} \mathbf{S}_1 \\ 0 \end{bmatrix}\mathbf{V}_{\text{SVD}}^T\hat{\mathbf{W}}^T + \mathbf{E}_{\text{PLS}} = (\mathbf{U}_1\mathbf{S}_1)\left(\hat{\mathbf{W}}\mathbf{V}_{\text{SVD}}\right)^T + \mathbf{E}_{\text{PLS}} = \hat{\mathbf{T}}_{\text{B}}\hat{\mathbf{V}}_{\text{B}}^T + \mathbf{E}_{\text{PLS}},\end{aligned} \quad (5)$$

resulting in $\hat{\mathbf{T}}_{\text{B}} = \mathbf{U}_1\mathbf{S}_1$ and $\hat{\mathbf{V}}_{\text{B}} = \hat{\mathbf{W}}\mathbf{V}_{\text{SVD}}$. The Wold factorization (3) can first be transformed to a Martens factorization according to

$$\mathbf{X} = \hat{\mathbf{T}}_{\text{Wold}}\hat{\mathbf{P}}_{\text{Wold}}^T\hat{\mathbf{W}}\hat{\mathbf{W}}^T + \mathbf{E}_{\text{PLS}} = \hat{\mathbf{T}}_{\text{Martens}}\hat{\mathbf{W}}^T + \mathbf{E}_{\text{PLS}}, \quad (6)$$

which may then be transformed to a bi–orthogonal factorization according to (5). Alternatively we may obtain $\hat{\mathbf{T}}_{\text{B}}$ and $\hat{\mathbf{V}}_{\text{B}}$ directly by an SVD of $\hat{\mathbf{X}} = \hat{\mathbf{T}}_{\text{Wold}}\hat{\mathbf{P}}_{\text{Wold}}^T\hat{\mathbf{W}}\hat{\mathbf{W}}^T = \hat{\mathbf{T}}_{\text{Martens}}\hat{\mathbf{W}}^T$ taken to the specified number of components.

Note that after the unifying transformations above the loading weight matrix $\hat{\mathbf{W}}$ is replaced by the loading (weight) matrix $\hat{\mathbf{V}}_{\text{B}}$, i.e. there is no longer a need to distinguish between loadings and loading weights.

Permutations

As a result of the SVD decomposition used in (5) the ordering of components according to explaining power may get lost. In

$$\hat{\mathbf{X}} = \hat{\mathbf{t}}_{\text{B},1}\hat{\mathbf{v}}_{\text{B},1}^T + \hat{\mathbf{t}}_{\text{B},2}\hat{\mathbf{v}}_{\text{B},2}^T + \cdots + \hat{\mathbf{t}}_{\text{B},A}\hat{\mathbf{v}}_{\text{B},A}^T \quad (7)$$

the third component may for example explain more of the response variable y than the second component etc. This does not, however, affect the total explaining power of all A components, where A is determined through validation using an ordinary PLS procedure[6]. The ordering according to explaining power may be restored by augmenting (5) with a square and orthonormal permutation matrix, i.e.

$$\mathbf{X} = \hat{\mathbf{T}}_{\text{B}}\mathbf{Q}\mathbf{Q}^{-1}\hat{\mathbf{V}}_{\text{B}}^T + \mathbf{E}_{\text{PLS}} = \hat{\mathbf{T}}_{\text{B}}\mathbf{Q}\left(\hat{\mathbf{V}}_{\text{B}}\mathbf{Q}\right)^T + \mathbf{E}_{\text{PLS}} = \tilde{\mathbf{T}}_{\text{B}}\tilde{\mathbf{V}}_{\text{B}}^T + \mathbf{E}_{\text{PLS}}. \quad (8)$$

For the common case of a very low number A of total components the permutation to use is easily found by a systematic search (see example in Section 4). Other cases are of little interest in a correspondence context.

Final predictor

It can be shown[7] that the PLS predictor based on observations collected in an $\mathbf{X}$ matrix and a $\mathbf{y}$ vector (assuming a scalar response) can be written as

$$\hat{\mathbf{b}} = \hat{\mathbf{W}}\left(\hat{\mathbf{W}}^T\mathbf{X}^T\mathbf{X}\hat{\mathbf{W}}\right)^{-1}\hat{\mathbf{W}}^T\mathbf{X}^T\mathbf{y}, \quad (9)$$

where $\hat{\mathbf{W}}$ is found by either the Wold or the Martens algorithm. In the transformations above $\hat{\mathbf{W}}$ is replaced by $\hat{\mathbf{V}}_{\text{B}} = \hat{\mathbf{W}}\mathbf{V}_{\text{SVD}}$. Since $\mathbf{V}_{\text{SVD}} \in \mathbb{R}^{A\times A}$ is invertible we thus find

$$\hat{\mathbf{b}} = \hat{\mathbf{V}}_{\text{B}}\mathbf{V}_{\text{SVD}}^{-1}\left(\mathbf{V}_{\text{SVD}}^{-T}\hat{\mathbf{V}}_{\text{B}}^T\mathbf{X}^T\mathbf{X}\hat{\mathbf{V}}_B\mathbf{V}_{\text{SVD}}^{-1}\right)^{-1}\mathbf{V}_{\text{SVD}}^{-T}\hat{\mathbf{V}}_{\text{B}}^T\mathbf{X}^T\mathbf{y} = \hat{\mathbf{V}}_{\text{B}}\left(\hat{\mathbf{V}}_{\text{B}}^T\mathbf{X}^T\mathbf{X}\hat{\mathbf{V}}_{\text{B}}\right)\hat{\mathbf{V}}_{\text{B}}^T\mathbf{X}^T\mathbf{y}. \quad (10)$$

The predictor is thus unaltered after replacement of $\hat{\mathbf{W}}$ by $\hat{\mathbf{V}}_{\text{B}}$, and for the same reason it is also unaltered by the permutation matrix $\mathbf{Q}$ in (8).

Discussion on BPLS properties

In the same way as in PCA, the BPLS factorization results in a score matrix with orthogonal columns and a loading matrix with orthonormal columns. This makes a comparison with PCA natural.

The PCA factorization (2) may be found from solutions of the eigenvalue problem

$$\mathbf{X}^T\mathbf{X}\hat{\mathbf{p}}_i = \hat{\mathbf{p}}_i\hat{\lambda}_i, \tag{11}$$

associated with the spectral decomposition[8]

$$\mathbf{X}^T\mathbf{X} = \hat{\mathbf{p}}_1\hat{\lambda}_1\hat{\mathbf{p}}_1^T + \hat{\mathbf{p}}_2\hat{\lambda}_2\hat{\mathbf{p}}_2^T + \ldots + \hat{\mathbf{p}}_p\hat{\lambda}_p\hat{\mathbf{p}}_p^T = \hat{\mathbf{P}}\hat{\mathbf{\Lambda}}\hat{\mathbf{P}}^T, \tag{12}$$

where $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \ldots \leq \hat{\lambda}_p$, $\hat{\mathbf{P}}^T\hat{\mathbf{P}} = \mathbf{I}$ and $\hat{\mathbf{P}}\hat{\mathbf{P}}^T = \mathbf{I}$, and where $\hat{\mathbf{\Lambda}}$ is diagonal. Using A components this results in

$$\begin{aligned}\mathbf{X}^T\mathbf{X} &= \hat{\mathbf{p}}_1\hat{\lambda}_1\hat{\mathbf{p}}_1^T + \hat{\mathbf{p}}_2\hat{\lambda}_2\hat{\mathbf{p}}_2^T + \ldots + \hat{\mathbf{p}}_A\hat{\lambda}_A\hat{\mathbf{p}}_A^T + \mathbf{E}_{\mathrm{PCA}}^T\mathbf{E}_{\mathrm{PCA}} \\ &= \hat{\mathbf{P}}_{\mathrm{PCA}}\hat{\mathbf{\Lambda}}_{\mathrm{PCA}}\hat{\mathbf{P}}_{\mathrm{PCA}}^T + \mathbf{E}_{\mathrm{PCA}}^T\mathbf{E}_{\mathrm{PCA}} = \hat{\mathbf{P}}_{\mathrm{PCA}}\hat{\mathbf{T}}_{\mathrm{PCA}}^T\hat{\mathbf{T}}_{\mathrm{PCA}}\hat{\mathbf{P}}_{\mathrm{PCA}}^T + \mathbf{E}_{\mathrm{PCA}}^T\mathbf{E}_{\mathrm{PCA}},\end{aligned} \tag{13}$$

which is also found from (2).

The BPLS factorization (5), on the other hand, uses a loading matrix $\hat{\mathbf{V}}_B$ that is a linear combination

$$\hat{\mathbf{V}}_{\mathrm{B}} = \hat{\mathbf{P}}\mathbf{L}_{\mathrm{B}} = \left[\begin{array}{cccc} \hat{\mathbf{P}}\mathbf{l}_1 & \hat{\mathbf{P}}\mathbf{l}_2 & \cdots & \hat{\mathbf{P}}\mathbf{l}_A \end{array}\right] \tag{14}$$

such that

$$\hat{\mathbf{V}}_{\mathrm{B}}^T\hat{\mathbf{V}}_{\mathrm{B}} = \mathbf{L}_{\mathrm{B}}^T\hat{\mathbf{P}}^T\hat{\mathbf{P}}\mathbf{L}_{\mathrm{B}} = \mathbf{L}_{\mathrm{B}}^T\mathbf{L}_{\mathrm{B}} = \mathbf{I}, \tag{15}$$

and a score matrix

$$\hat{\mathbf{T}}_{\mathrm{B}} = \mathbf{X}\hat{\mathbf{V}}_{\mathrm{B}} = \mathbf{X}\hat{\mathbf{P}}\mathbf{L}_{\mathrm{B}} \tag{16}$$

such that

$$\hat{\mathbf{T}}_{\mathrm{B}}^T\hat{\mathbf{T}}_{\mathrm{B}} = \mathbf{L}_{\mathrm{B}}^T\hat{\mathbf{P}}^T\mathbf{X}^T\mathbf{X}\hat{\mathbf{P}}\mathbf{L}_{\mathrm{B}} \tag{17}$$

is diagonal, just as $\hat{\mathbf{T}}_{\mathrm{PCA}}^T\hat{\mathbf{T}}_{\mathrm{PCA}} = \hat{\mathbf{\Lambda}}_{\mathrm{PCA}}$. However, this does not imply that $\hat{\mathbf{V}}_{\mathrm{B}}$ can be found as a solution of an eigenvalue problem, except for $A = p$, in which case $\mathbf{L}_{\mathrm{B}} = \mathbf{I}$ and thus $\hat{\mathbf{V}}_{\mathrm{B}} = \hat{\mathbf{P}}$.

Note that $\hat{\mathbf{W}}$ in the ordinary PLS factorizations also is a linear combination of $\hat{\mathbf{P}}$ with $\mathbf{L}_{\mathrm{Martens}}^T\mathbf{L}_{\mathrm{Martens}} = \mathbf{I}$, but that $\hat{\mathbf{T}}_{\mathrm{Martens}}^T\hat{\mathbf{T}}_{\mathrm{Martens}}$ is non-diagonal[9]. Also $\hat{\mathbf{V}}_{\mathrm{Wold}} = \hat{\mathbf{W}}\hat{\mathbf{W}}^T\hat{\mathbf{P}}_{\mathrm{Wold}}$ is a linear combination of $\hat{\mathbf{P}}$, but then with $\mathbf{L}_{\mathrm{Wold}}^T\mathbf{L}_{\mathrm{Wold}} \neq \mathbf{I}$. Although in itself interesting, further relations between the BPLS and other factorizations are beyond the scope of the present correspondence context.

3 Score and loading correspondence

General discussion

As indicated in the introduction, correspondence between PLS scores and loadings is related to correspondence in several other multivariate display techniques used in PCA, correspondence factor analysis, spectral map analysis, factor analysis in the strict statistical sense etc.[2]. The common step in these methods is the factorization of the data matrix $\mathbf{X}$, but the methods differ with respect to the processing of the data prior to the factorization, and to the factorization method used.

Comparison of factorization methods

We will here use PCA as a reference. From the general factorization (1) and the relation $\hat{\mathbf{V}}_{\mathrm{Wold}} = \hat{\mathbf{W}}\hat{\mathbf{W}}^T\hat{\mathbf{P}}_{\mathrm{Wold}}$ used in (6) follow the least squares solutions

$$\hat{\mathbf{T}} = \mathbf{X}\hat{\mathbf{V}}\left(\hat{\mathbf{V}}^T\hat{\mathbf{V}}\right)^{-1} = \begin{cases} \mathbf{X}\hat{\mathbf{P}}_{\mathrm{PCA}} = \mathbf{X}\hat{\mathbf{V}}_{\mathrm{PCA}} & \text{PCA} \\ \mathbf{X}\hat{\mathbf{W}}\left(\hat{\mathbf{P}}_{\mathrm{Wold}}^T\hat{\mathbf{W}}\right)^{-1} \neq \mathbf{X}\hat{\mathbf{V}}_{\mathrm{Wold}} & \text{Wold PLS} \\ \mathbf{X}\hat{\mathbf{W}} = \mathbf{X}\hat{\mathbf{V}}_{\mathrm{Martens}} & \text{Martens PLS} \\ \mathbf{X}\hat{\mathbf{V}}_{\mathrm{B}} & \text{BPLS,} \end{cases} \tag{18}$$

where the orthonormality of $\hat{\mathbf{P}}_{\text{PCA}}$, $\hat{\mathbf{W}}$ and $\hat{\mathbf{V}}_{\text{B}}$ is used. Using the notation $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{\chi}_1 & \boldsymbol{\chi}_2 & \cdots & \boldsymbol{\chi}_N \end{bmatrix}^T$ and $\hat{\mathbf{T}} = \begin{bmatrix} \hat{\mathbf{t}}_1 & \hat{\mathbf{t}}_2 & \cdots & \hat{\mathbf{t}}_A \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\tau}}_1 & \hat{\boldsymbol{\tau}}_2 & \cdots & \hat{\boldsymbol{\tau}}_N \end{bmatrix}^T$ it follows that a given observation $\boldsymbol{\chi}_i^T$ results in scores

$$\hat{\boldsymbol{\tau}}_i^T = \begin{cases} \boldsymbol{\chi}_i^T \hat{\mathbf{P}}_{\text{PCA}} & \text{PCA} \\ \boldsymbol{\chi}_i^T \hat{\mathbf{W}} \left(\hat{\mathbf{P}}_{\text{Wold}}^T \hat{\mathbf{W}} \right)^{-1} & \text{Wold PLS} \\ \boldsymbol{\chi}_i^T \hat{\mathbf{W}} & \text{Martens PLS} \\ \boldsymbol{\chi}_i^T \hat{\mathbf{V}}_{\text{B}} & \text{BPLS,} \end{cases} \tag{19}$$

where $\hat{\mathbf{P}}_{\text{PCA}}$, $\hat{\mathbf{W}}$, and $\hat{\mathbf{V}}_{\text{B}}$ are orthonormal, while $\hat{\mathbf{W}} \left(\hat{\mathbf{P}}_{\text{Wold}}^T \hat{\mathbf{W}} \right)^{-1}$ is not.

Introducing the notation $\hat{\mathbf{P}}_{\text{PCA}} = \begin{bmatrix} \hat{\mathbf{p}}_1 & \hat{\mathbf{p}}_2 & \cdots & \hat{\mathbf{p}}_A \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\pi}}_1 & \hat{\boldsymbol{\pi}}_2 & \cdots & \hat{\boldsymbol{\pi}}_p \end{bmatrix}^T$, $\hat{\mathbf{W}} = \begin{bmatrix} \hat{\mathbf{w}}_1 & \hat{\mathbf{w}}_2 & \cdots & \hat{\mathbf{w}}_A \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\omega}}_1 & \hat{\boldsymbol{\omega}}_2 & \cdots & \hat{\boldsymbol{\omega}}_p \end{bmatrix}^T$, $\hat{\mathbf{V}}_B = \begin{bmatrix} \hat{\mathbf{v}}_{\text{B},1} & \hat{\mathbf{v}}_{\text{B},2} & \cdots & \hat{\mathbf{v}}_{\text{B},A} \end{bmatrix}$ $= \begin{bmatrix} \hat{\boldsymbol{\vartheta}}_{\text{B},1} & \hat{\boldsymbol{\vartheta}}_{\text{B},2} & \cdots & \hat{\boldsymbol{\vartheta}}_{\text{B},p} \end{bmatrix}^T$ and $\hat{\mathbf{V}}_{\text{Wold}} = \begin{bmatrix} \hat{\mathbf{v}}_{\text{Wold},1} & \hat{\mathbf{v}}_{\text{Wold},2} & \cdots & \hat{\mathbf{v}}_{\text{Wold},A} \end{bmatrix}$ $= \begin{bmatrix} \hat{\boldsymbol{\vartheta}}_{\text{Wold},1} & \hat{\boldsymbol{\vartheta}}_{\text{Wold},2} & \cdots & \hat{\boldsymbol{\vartheta}}_{\text{Wold},p} \end{bmatrix}^T$, and assuming centered data, a specific observation $\boldsymbol{\chi}_i^T = \begin{bmatrix} 0 & \cdots & 0 & \Delta x_{ij} & 0 & \cdots & 0 \end{bmatrix}$ results in

$$\hat{\boldsymbol{\tau}}_i = \begin{cases} \Delta x_{ij} \hat{\boldsymbol{\pi}}_j & \text{PCA} \\ \Delta x_{ij} \hat{\boldsymbol{\omega}}_j & \text{Martens PLS} \\ \Delta x_{ij} \hat{\boldsymbol{\vartheta}}_{\text{B},j} & \text{BPLS,} \end{cases} \tag{20}$$

while

$$\hat{\boldsymbol{\tau}}_i \neq \Delta x_{ij} \hat{\boldsymbol{\vartheta}}_{\text{Wold},j} \qquad \text{Wold PLS.} \tag{21}$$

Assuming orthogonal coordinate systems, the vector $\hat{\boldsymbol{\tau}}_i$ in the score plots thus has the same direction as the vector $\hat{\boldsymbol{\pi}}_j$, $\hat{\boldsymbol{\omega}}_j$ or $\hat{\boldsymbol{\vartheta}}_j$ in the corresponding loading or loading weight plots for PCA, Martens PLS and BPLS. For $\Delta x_{ij} = 1$ the vectors will coincide (see example in Section 4).

For the Wold PLS solution, on the other hand, the vector $\hat{\boldsymbol{\tau}}_i$ and the corresponding vector in any of the possible loading or loading weight plots ($\hat{\mathbf{V}}_{\text{Wold}}$, $\hat{\mathbf{W}}$ or $\hat{\mathbf{P}}_{\text{Wold}}$) will not have the same directions. The reason for this is that the $\hat{\mathbf{V}}_{\text{Wold}}$ matrix used in the factorization is not orthogonal, and plotting projections of $\hat{\mathbf{W}}$ or $\hat{\mathbf{P}}_{\text{Wold}}$ instead of $\hat{\mathbf{V}}_{\text{Wold}}$ does not remedy the situation (see example in Section 4).

Relation to predictive power

The correspondence discussion and results above are limited to the different factorization methods, and are thus not related to the predictive power of the different regression methods. This means that the good interpretational properties of PCA and BPLS to a certain extent may be undermined by prediction errors.

4 Industrial data example

The example uses multivariate regression data from a paper production plant[10,11]. The problem considered here is to monitor a given paper quality y_i (the second column in the first data set) from six known process variables $\boldsymbol{\chi}_i^T = \begin{bmatrix} \chi_{i1} & \chi_{i2} & \chi_{i3} & \chi_{i4} & \chi_{i5} & \chi_{i6} \end{bmatrix}$ (columns 14 to 19 in the first data set), and for the purpose of finding PLS factorizations all $N = 29$ samples of $\boldsymbol{\chi}_i^T$ and y_i are used. The first three process variables χ_{i1}, χ_{i2} and χ_{i3} were varied systematically through an experiment, taking the values 1, 0 and -1. The next three variables were constructed as $\chi_{i4} = \chi_{i1}^2$, $\chi_{i5} = \chi_{i2}^2$ and $\chi_{i6} = \chi_{i3}^2$. The three constructed variables $\chi_{i1}\chi_{i2}$, $\chi_{i1}\chi_{i2}$ and $\chi_{i2}\chi_{i3}$ are also included in the data set, but for the paper quality chosen they have little predictive power, and for clarity of presentation they are not used in the present example.

Prediction

Although prediction as such is not the main topic in the present context, some results are included as a background for the correspondence results presented below. As a first step samples 1 to 20 were used to find PLS regression (PLSR and BPLSR) and principal component regression (PCR) predictors using different numbers of components, while the samples 21 to 29 were used for validation. Centered and standardized data were used, and the validation results are given in Table 1. The BPLSR results were obtained by use of three components and a permutation matrix $\mathbf{Q}$ such that after the permutation (8) the ordering was 2, 3, 1 (the best possible ordering found by trial and error). The fact that the two first BPLSR components explain more than the two first PLSR components may be due to the very limited number of samples.

Table 1: RMSEP results for different predictors.

No. of components	$\text{RMSEP}_{\text{PLSR}}$	$\text{RMSEP}_{\text{BPLSR}}$	$\text{RMSEP}_{\text{PCR}}$
0	0.9428	0.9428	0.9428
1	0.6023	0.7525	0.9206
2	0.4822	0.4111	0.9168
3	0.4106	0.4106	0.7337
4	0.4220	-	0.7561

Correspondence

In a second step all $N = 29$ samples were used to find PLS and BPLS factorizations and the corresponding loading and loading weight matrices using $A = 3$ components. In accordance with (8) the BPLS score and loading matrices after the component permutation are denoted $\tilde{\mathbf{T}}_{\text{B}}$ and $\tilde{\mathbf{V}}_{\text{B}}$. New $\mathbf{X}$ data were subsequently introduced as

$$\mathbf{X}^{\text{test}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{22}$$

and the new scores together with the predictor loadings and loading weights for the two first components were plotted (Fig. 1). To ease the interpretation of the results $\hat{\mathbf{T}}^{\text{test}}_{\text{Wold}}$ and $\hat{\mathbf{W}}$ etc. are plotted in the same plots (bi-plots). For the Wold algorithm there is generally a distinction between $\hat{\mathbf{V}}_{\text{Wold}} = \hat{\mathbf{W}}\hat{\mathbf{W}}^T\hat{\mathbf{P}}_{\text{Wold}}$ and $\hat{\mathbf{P}}_{\text{Wold}}$, although $\hat{\mathbf{V}}_{\text{Wold}} = \hat{\mathbf{P}}_{\text{Wold}}$ for the first two components (all except the last). The results are in agreement with the theoretical discussion in Section 3 above, i.e. only the Martens PLS and the BPLS factorizations show total correspondence between scores and loadings/loading weights.

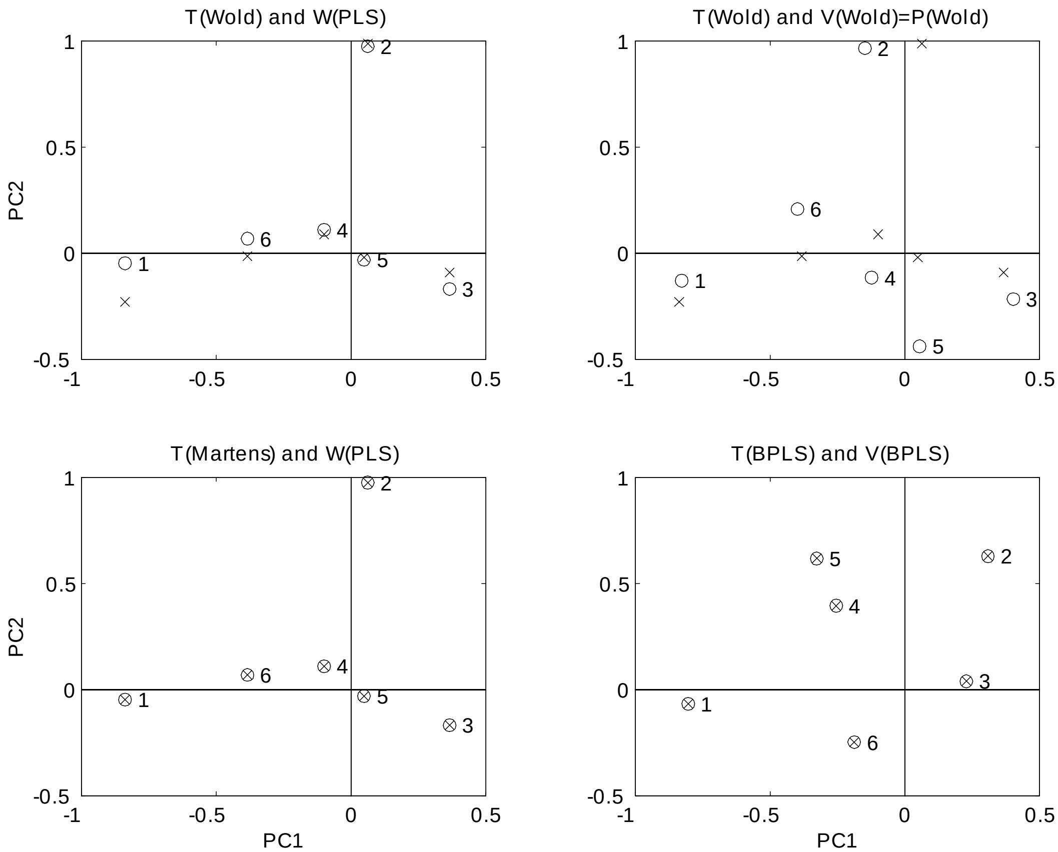

Figure 1. Loadings/loading weights $\hat{\mathbf{V}}_{\text{Wold}}$, $\hat{\mathbf{W}}$ and $\tilde{\mathbf{V}}_{\text{B}}$ (o) for the modeling data, and scores $\hat{\mathbf{T}}_{\text{Wold}}^{\text{test}}$, $\hat{\mathbf{T}}_{\text{Martens}}^{\text{test}}$ and $\tilde{\mathbf{T}}_{\text{B}}^{\text{test}}$ (x) for the $\mathbf{X}^{\text{test}}$ data (22) with the Wold PLS, Martens PLS and BPLS factorizations. Note the total correspondence for the Martens PLS and BPLS factorizations only.

Since the $\mathbf{X}$-variables are correlated, the test data (22) are not realistic in the present case. However, a realistic test observation is

$$\boldsymbol{\chi}_{\text{test}}^{T} = \left[\begin{array}{cccccc} 1 & 0 & 0 & 1 & 0 & 0 \end{array} \right]. \tag{23}$$

The result of this is shown in Fig. 2, where the deficiency of the $\hat{\boldsymbol{\tau}}_{\text{Wold}}^{\text{test}}$ and $\hat{\mathbf{W}}$ plot is clearly demonstrated. Use of $\hat{\boldsymbol{\tau}}_{\text{Wold}}^{\text{test}}$ and $\hat{\mathbf{V}}_{\text{Wold}} = \hat{\mathbf{P}}_{\text{Wold}}$ gives in fact a somewhat more correct picture of the influences of variables 1 and 4, although total correspondence is found only by use of $\hat{\boldsymbol{\tau}}_{\text{Martens}}^{\text{test}}$ and $\hat{\mathbf{W}}$ or $\tilde{\boldsymbol{\tau}}_{\text{B}}^{\text{test}}$ and $\tilde{\mathbf{V}}_{\text{B}}$.

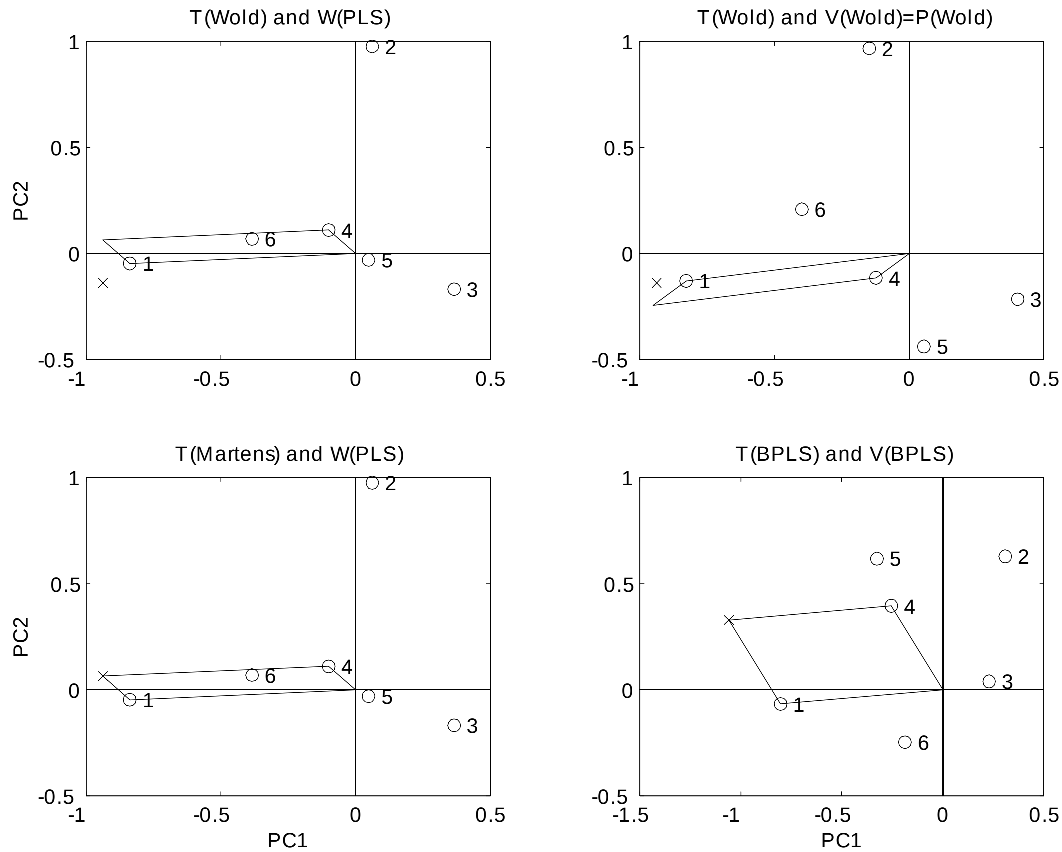

Figure 2. Loadings/loading weights $\hat{\mathbf{V}}_{\mathrm{Wold}}$, $\hat{\mathbf{W}}$ and $\tilde{\mathbf{V}}_{\mathrm{B}}$ (o) for the modeling data, and scores $\hat{\boldsymbol{\tau}}^{\mathrm{test}}_{\mathrm{Wold}}$, $\hat{\boldsymbol{\tau}}^{\mathrm{test}}_{\mathrm{Martens}}$ and $\tilde{\boldsymbol{\tau}}^{\mathrm{test}}_{\mathrm{B}}$ (x) for the $\boldsymbol{\chi}^T_{\mathrm{test}}$ data (23) with the Wold PLS, Martens PLS and BPLS factorizations. The parallelograms indicate the target score vector for $\boldsymbol{\chi}^T_{\mathrm{test}}$ assuming total score-loading/loading weight correspondence. Note that the Martens PLS and BPLS scores only are on target.

5 Conclusions

The existing PLS factorizations causes some interpretational problems with respect to score-loading correspondence (orthogonal PLS of Wold) or latent variables covariance (non-orthogonal PLS of Martens). As a solution a new PLS factorization (BPLS) has been developed, which just as the PCA factorization has both an orthogonal score matrix and an orthonormal loading matrix. The two well-known PLS algorithms of Wold and Martens can easily be transformed into a BPLS algorithm, without altering the final predictor for the chosen number of components. The score-loading/loading weight correspondence properties have been analyzed for the PCA, $\mathrm{PLS_{Wold}}$, $\mathrm{PLS_{Martens}}$ and BPLS factorizations, and it has been shown that all of these except the $\mathrm{PLS_{Wold}}$ factorization show total correspondence. The $\mathrm{PLS_{Martens}}$ solution, however, has the drawback of using correlated latent variables, while the new BPLS factorization uses independent latent variables. An example using industrial paper plant data illustrates the potential BPLS advantages in process monitoring applications.

References

[1] Skagerberg B, Sundin L. Multidimensional monitoring of complex industrial processes, *ABB Review* 1993;4/93:31-38

[2] Thielemans A, Lewi PJ, Massart DL. Similarities and Differences among Multivariate Display Techniques Illustrated by Belgian Cancer Mortality Distribution Data. *Chemometrics Intell. Lab. Syst.* 1988;**3**:277-300.

[3] Kvalheim OM, Karstang TV. Interpretation of Latent-Variable Regression Models. *Chemometrics and Intelligent Laboratory Systems* 1989;**7**:39-51.

[4] Martens H, Næs T. *Multivariate Calibration*, Wiley: New York, 1989;121-125.

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

[6] Esbensen KH. *Multivariate Data Analysis - in practice*, Camo ASA: Trondheim, Norway, 2000;155-168.

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

[8] Johnson AJ, Wichern DW. *Applied Multivariate Statistical Analysis*, Prentice-Hall: Englewood Cliffs, NJ, 1992;48.

[9] Kalivas JH. Interrelationships of multivariate regression methods using eigenvector basis sets. *J. Chemometrics* 1999; **13**:111-132.

[10] Aldrin M. Moderate projection pursuit regression for multivariate response data. *Computational Statistics and Data Analysis* 1996; **21**:501-531.

[11] StatLib-Datasets Archive Website. http://lib.stat.cmu.edu/datasets/papir [14 June 1999].