Chemometrics and Intelligent Laboratory Systems 149 (2015) 33–39

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

A sequential algorithm for multiblock orthogonal projections to latent structures

Bradley Worley, Robert Powers *

Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0304, United States

ARTICLE INFO

Article history:
Received 31 August 2015
Received in revised form 30 October 2015
Accepted 31 October 2015
Available online 7 November 2015

Keywords:
Multiblock data
CPCA-W
MB-PLS
OnPLS
MB-OPLS

ABSTRACT

Methods of multiblock bilinear factorizations have increased in popularity in chemistry and biology as recent increases in the availability of information-rich spectroscopic platforms have made collecting multiple spectroscopic observations per sample a practicable possibility. Of the existing multiblock methods, consensus PCA (CPCA-W) and multiblock PLS (MB-PLS) have been shown to bear desirable qualities for multivariate modeling, most notably their computability from single-block PCA and PLS factorizations. While MB-PLS is a powerful extension to the nonlinear iterative partial least squares (NIPALS) framework, it still spreads predictive information across multiple components when response-uncorrelated variation exists in the data. The OnPLS extension to O2PLS provides a means of simultaneously extracting predictive and uncorrelated variation from a set of matrices, but is more suited to unsupervised data discovery than regression. We describe the union of NIPALS MB-PLS with an orthogonal signal correction (OSC) filter, called MB-OPLS, and illustrate its equivalence to single-block OPLS for regression and discriminant analysis.

1. Introduction

The method of nonlinear iterative partial least squares (NIPALS) has firmly entrenched itself in the field of chemometrics. Implementations of principal component analysis (PCA) and projections to latent structures (PLS) that utilize NIPALS-type algorithms benefit from its numerical stability, as well as its flexibility and simplicity [1–3]. Only a few subroutines from level 2 of the basic linear algebra subprograms (BLAS) specification are required to construct a complete NIPALS-type algorithm [4,5], making it an attractive means of constructing PCA and PLS models of high-dimensional spectroscopic datasets.

One particularly recent addition to the NIPALS family of algorithms, called orthogonal projections to latent structures (OPLS), integrates an orthogonal signal correction (OSC) filter into NIPALS PLS [6,7]. By extracting variation from its computed PLS components that is uncorrelated (orthogonal) to the responses, OPLS produces a more interpretable regression model compared to PLS. In fact, when trained on the same data and responses, an OPLS model and a PLS model with the same total number of components will show no difference in predictive ability [8]. Despite its relative novelty to the field, the enhanced interpretability of OPLS over PLS has made it a popular method in exploratory studies of spectroscopic datasets of complex chemical mixtures (e.g., metabolomics [9], food and soil science [10], and chemical process control [11]).

Extensions of NIPALS PCA and PLS to incorporate blocking information that partitions the set of measured variables into multiple 'blocks' of data have recently gained attention in the field as more experimental designs involve the collection of data from multiple analytical platforms per sample. In such experiments, referred to as 'class II' multiblock schemes by Smilde et al. [12], correlated consensus directions are sought from the blocks that maximally capture block variation and (optionally) maximally predict a set of responses. Of the available extensions of NIPALS to multiblock modeling, a class of methods exists that bears attractive computational qualities, namely computability from single-block bilinear factorizations. When both super weights and block loadings are normalized in consensus PCA (i.e., CPCA-W), the obtained super scores are equivalent to those obtained from PCA of the concatenated matrix of blocks [13]. Likewise, scores obtained from PLS of the concatenated matrix are equivalent to super scores from multiblock PLS (MB-PLS) when super scores are used in the deflation step [13,14]. As a result, these multiblock bilinear factorizations inherit many of the useful properties of their single-block equivalents.

A second class of multiblock methods exists in which every block is predicted in a regression model by every other block. In the first of such methods, known as nPLS, the MAXDIFF criterion [15] is optimized one component at a time (i.e., sequentially) to yield a set of predictive weight vectors for each block [16]. The recently described OnPLS algorithm also falls within this class [16]. OnPLS extends O2PLS to three or more matrices and may be considered a prefixing of nPLS with an OSC filtering step. OnPLS deflates non-globally predictive variation that

* Corresponding author at: University of Nebraska-Lincoln, Department of Chemistry, 722 Hamilton Hall, Lincoln, NE 68588-0304, United States. Tel.: +1 402 472 3039; fax: +1 402 472 9402.
E-mail address: rpowers3@unl.edu (R. Powers).

http://dx.doi.org/10.1016/j.chemolab.2015.10.018
0169-7439/© 2015 Elsevier B.V. All rights reserved.

may or may not be orthogonal to all blocks from each matrix, and then computes an nPLS model from the filtered result [16]. While fully symmetric OnPLS is a powerful and general addition to the existing set of multiblock modeling frameworks, it is arguably an over-complication when the regression of a single response matrix on multiple data blocks (i.e., MB-PLS) is sought. For such situations, a novel algorithm termed MB-OPLS for multiblock orthogonal projections to latent structures is introduced that embeds an OSC filter within NIPALS MB-PLS, thus solving an inherently different problem from OnPLS. It will be shown that MB-OPLS, in analogy to CPCA-W and MB-PLS, is computable from a single-block OPLS model of the matrix of concatenated data blocks. Thus, MB-OPLS forms a bridge between this special class of consensus component methods and the highly general symmetric regression framework of OnPLS.

2. Theory

MB-OPLS belongs to a set of multiblock methods that exhibit an equivalence to their single-block equivalents. A short discussion on these methods follows, in which the optimization criterion of each method is shown to belong to the MAXBET family of objective functions. This is contrasted to nPLS and OnPLS, which have been shown to optimize a MAXDIFF objective. The principal difference between MAXBET and MAXDIFF is one of explained variation: while MAXBET captures between-matrix covariances *and* within-matrix variances, MAXDIFF only captures the former [15,17]. Finally, the equivalence of MB-OPLS and OPLS is demonstrated, which highlights its differences from OnPLS.

In all following discussions, it will be understood that there exist *n* data matrices $\boldsymbol{X}_1$ to $\boldsymbol{X}_\text{n}$, each having *N* rows (observations) and $\boldsymbol{K}_\text{i}$ columns (variables). The matrix $\boldsymbol{X} = [\boldsymbol{X}_1| \dots |\boldsymbol{X}_n]$ of all concatenated blocks will be used in cases of single-block modeling. Finally, a response matrix $\boldsymbol{Y}$ having *N* rows and *M* columns will be assumed to exist for the purposes of regression (i.e., PLS-R, MB-PLS-R, etc.) or discriminant analysis (i.e., PLS-DA, MB-PLS-DA, etc.).

2.1. nPLS and OnPLS

In their initial description of the OnPLS modeling framework [16], Löfstedt and Trygg introduced nPLS as a generalization of PLS regression to cases where $n > 2$, and a model is sought in which each matrix $\boldsymbol{X}_1$ is predicted by all other matrices $\boldsymbol{X}_{j \neq i}$. The nPLS solution involves identifying a set of weight vectors $\boldsymbol{w}_i$ that simultaneously maximize covariances between each pair of resulting scores $\boldsymbol{t}_i = \boldsymbol{X}_i\boldsymbol{w}_i$ via the following objective function:

$$\sum_{\substack{i,\,j=1\\ i\neq j}}^{n} \boldsymbol{t}_i{}^T\boldsymbol{t}_j = \sum_{\substack{i,\,j=1\\ i\neq j}}^{n} \boldsymbol{w}_i{}^T\boldsymbol{X}_i{}^T\boldsymbol{X}_j\boldsymbol{w}_j \tag{1}$$

subject to the constraints $\|\boldsymbol{w}_i\| = 1$. This objective was subsequently recognized to be a member of the MAXDIFF family of functions, whose solution is obtainable using a general algorithm from Hanafi and Kiers [17]. After the identification of a set of weight vectors, the scores

$$\boldsymbol{t}_i = \boldsymbol{X}_i\boldsymbol{w}_i$$

and loadings

$$\boldsymbol{p}_i = \frac{\boldsymbol{X}_i{}^T\boldsymbol{t}_i}{\boldsymbol{t}_i{}^T\boldsymbol{t}_i}$$

may be computed for each matrix, which is then deflated prior to the computation of subsequent component weights:

$$\boldsymbol{X}_i \leftarrow \boldsymbol{X}_i - \boldsymbol{t}_i\boldsymbol{p}_i{}^T = \left(\boldsymbol{I} - \frac{\boldsymbol{t}_i\boldsymbol{t}_i{}^T}{\boldsymbol{t}_i{}^T\boldsymbol{t}_i}\right)\boldsymbol{X}_i. \tag{2}$$

This deflation scheme follows the precedent set by two-block PLS regression. Because their described approach used a distinct deflation scheme from single-component (sequential) MAXDIFF, it was given the name “nPLS” by the authors to distinguish it from MAXDIFF [16,18].

OnPLS extends nPLS by decomposing each matrix into a globally predictive part and a non-globally predictive (orthogonal) part using an orthogonal projection. By removing orthogonal variation from each block prior to constructing an nPLS model, OnPLS optimizes the following MAXDIFF-type objective function:

$$\sum_{\substack{i,\,j=1\\ i\neq j}}^{n} \boldsymbol{t}_i{}^T\boldsymbol{t}_j = \sum_{\substack{i,\,j=1\\ i\neq j}}^{n} \boldsymbol{w}_i{}^T\boldsymbol{X}_i{}^T\boldsymbol{X}_j\boldsymbol{w}_j \tag{3}$$

where $\boldsymbol{Z}_i$ represents the orthogonal projector identified by OnPLS for matrix *i*:

$$\boldsymbol{Z}_i = \boldsymbol{I} - \boldsymbol{T}_{\boldsymbol{o},i}\left(\boldsymbol{T}_{\boldsymbol{o},i}{}^T\boldsymbol{T}_{\boldsymbol{o},i}\right)^{-1}\boldsymbol{T}_{\boldsymbol{o},i}{}^T$$

where $\boldsymbol{T}_{\boldsymbol{o},i} = [\boldsymbol{t}_{\boldsymbol{o},i,1}|\dots|\boldsymbol{t}_{\boldsymbol{o},i,A_o}]$, the concatenation of all orthogonal score vectors for the block, and $\boldsymbol{t}_{\boldsymbol{o},i,a} = \boldsymbol{X}_i\boldsymbol{w}_{\boldsymbol{o},i,a}$. In OnPLS, each orthogonal weight $\boldsymbol{w}_{\boldsymbol{o},i,a}$ is chosen such that its score $\boldsymbol{t}_{\boldsymbol{o},i,a}$ contains maximal covariance with the variation in $\boldsymbol{X}_{j\neq i}$ that is not jointly predictive of $\boldsymbol{X}_1$. The OnPLS framework provides a powerful set of methods for unsupervised data mining and path modeling [16,19–21].

2.2. CPCA-W and MB-PLS

The consensus PCA method, introduced by Wold et al. as CPCA and modified by Westerhuis et al. as CPCA-W, identifies a set of weights $\boldsymbol{p}_i$ that maximally capture the within-block variances and between-block covariances of a set of *n* matrices [13]. It was further proven by Westerhuis, Kourti and MacGregor that the results of CPCA-W computed on matrices $\boldsymbol{X}_1$ to $\boldsymbol{X}_n$ are identical to those from PCA of the concatenated matrix $\boldsymbol{X} = [\boldsymbol{X}_1| \dots |\boldsymbol{X}_n]$. It immediately follows from this equivalence that the CPCA-W algorithm optimizes the following objective function:

$$\boldsymbol{t}^T\boldsymbol{t} = \boldsymbol{p}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{p} = \sum_{i,j=1}^{n} \boldsymbol{t}_i{}^T\boldsymbol{t}_j = \sum_{i,j=1}^{n} \boldsymbol{p}_i{}^T\boldsymbol{X}_i{}^T\boldsymbol{X}_j\boldsymbol{p}_j \tag{4}$$

subject to the constraint $\|p\| = 1$, where $\boldsymbol{p}^T = [\boldsymbol{p}_1{}^T|\dots|\boldsymbol{p}_n{}^T]$. Maximizing the above function yields a set of super scores $\boldsymbol{t}$ that relate the *N* observations in $\boldsymbol{X}$ to each other based on the extracted consensus directions in $\boldsymbol{p}$, as well as block scores $\boldsymbol{t}_i$ and loadings $\boldsymbol{p}_i$ that describe each block. This objective function is of the MAXBET variety, in contrast to the MAXDIFF objective of nPLS and OnPLS. As a result, the CPCA-W NIPALS algorithm may be considered a special case of the general algorithm from Hanafi and Kiers [17].

The multiblock PLS (MB-PLS) method, when deflation is performed using super scores [14], shares an equivalence with single-block PLS as proven by Westerhuis et al. [13]. Therefore, the MB-PLS objective takes on a similar form as in CPCA-W, with the addition of a weighting matrix:

$$\begin{aligned}\boldsymbol{t}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{t} = \boldsymbol{w}^T\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{X}\boldsymbol{w} &= \sum_{i,j=1}^{n} \boldsymbol{t}_i{}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{t}_j \\ &= \sum_{i,j=1}^{n} \boldsymbol{w}_i{}^T\boldsymbol{X}_i{}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{X}_j\boldsymbol{w}_j\end{aligned} \tag{5}$$

where once again $\|w\|$ is constrained to unity. In analogy to Höskuldsson's interpretation of PLS as a regression on orthogonal components, where $\boldsymbol{Y}\boldsymbol{Y}^T$ is used to weight the covariance matrix, the above function corresponds to a MAXBET objective with an inner weighting of $\boldsymbol{Y}\boldsymbol{Y}^T$ [2]. Alternatively, Eq. (5) could be interpreted as a MAXBET computed on the *n* cross-covariance matrices $\boldsymbol{Y}^T\boldsymbol{X}_1$ to $\boldsymbol{Y}^T\boldsymbol{X}_n$.

2.3. MB-OPLS

Extension of prior multiblock NIPALS algorithms to incorporate an OSC filter rests on the observation that, in both the case of CPCA-W and MB-PLS, deflation of each computed component is accomplished using super scores. For any super score deflation method, a loading vector is computed for each block:

$$\boldsymbol{p}_i = \frac{\boldsymbol{X}_i{}^T\boldsymbol{t}}{\boldsymbol{t}^T\boldsymbol{t}}$$

and the super scores $\boldsymbol{t}$ and block loadings are then used to deflate their respective block:

$$\boldsymbol{X}_i \leftarrow \boldsymbol{X}_i - \boldsymbol{t}\boldsymbol{p}_i{}^T = \left(\boldsymbol{I} - \frac{\boldsymbol{t}\boldsymbol{t}^T}{\boldsymbol{t}^T\boldsymbol{t}}\right)\boldsymbol{X}_i \tag{6}$$

Eq. (6) differs from Eq. (2) used in nPLS and OnPLS, which uses block-specific scores and loadings during deflation. This method of super score deflation ensures that the super scores are an orthogonal basis, while allowing scores and loadings to become slightly correlated at the block level, and is a necessary condition for the equivalences between CPCA-W and MB-PLS and their single-block counterparts [13]. We shall employ this condition in MB-OPLS by deflating each matrix by a set of orthogonal super scores $\boldsymbol{T_o}$, which shall be shown to be equal to the orthogonal scores obtained from single-block OPLS. By constructing an MB-PLS model on the set of matrices after deflation by $\boldsymbol{T_o}$, we effectively arrive at another MAXBET objective:

$$\begin{aligned}\boldsymbol{t}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{t} = \boldsymbol{w}^T\boldsymbol{X}^T\boldsymbol{Z}\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{Z}\boldsymbol{X}\boldsymbol{w} &= \textstyle\sum_{i,j=1}^{n}\boldsymbol{t}_i{}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{t}_j \\ &= \textstyle\sum_{i,j=1}^{n}\boldsymbol{w}_i{}^T\boldsymbol{X}_i{}^T\boldsymbol{Z}\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{Z}\boldsymbol{X}_j\boldsymbol{w}_j\end{aligned} \tag{7}$$

where $\boldsymbol{w}$ is constrained to unit norm and $\boldsymbol{Z}$ is the orthogonal projector for the super scores $\boldsymbol{T_o}$:

$$\boldsymbol{Z} = \boldsymbol{I} - \boldsymbol{T_o}\left(\boldsymbol{T_o}{}^T\boldsymbol{T_o}\right)^{-1}\boldsymbol{T_o}{}^T$$

2.3.1. The MB-OPLS model

MB-OPLS constructs an OPLS model for each matrix $\boldsymbol{X}_i$, where the predictive and orthogonal loadings for each matrix are interrelated by a set of predictive and orthogonal super scores, respectively:

$$\boldsymbol{X}_i = \underbrace{TP_i^T}_{\boldsymbol{X_{p,i}}} + \underbrace{\boldsymbol{T_o}\boldsymbol{P_{o,i}}{}^T}_{\boldsymbol{X_{o,i}}} + \boldsymbol{E}_i \tag{8}$$

where each $\boldsymbol{E}_i$ is a data residual matrix that holds all variation in $\boldsymbol{X}_i$ not explained by the model. Concatenation of all block-level matrices together in Eq. (8) results in a top-level consensus model, which is in fact equivalent to an OPLS model trained on the partitioned data matrix $\boldsymbol{X}$:

$$\boldsymbol{X} = [\boldsymbol{X}_1|\ldots|\boldsymbol{X}_n] = \underbrace{\boldsymbol{T}\left[\boldsymbol{P}_1^T|\ldots|\boldsymbol{P}_n^T\right]}_{\boldsymbol{X_p}} + \underbrace{\boldsymbol{T_o}\left[\boldsymbol{P_{o,1}}{}^T|\ldots|\boldsymbol{P_{o,n}}{}^T\right]}_{\boldsymbol{X^o}} + \underbrace{[\boldsymbol{E}_1|\ldots|\boldsymbol{E}_n]}_{\boldsymbol{E}}. \tag{9}$$

Like PLS and MB-PLS, an MB-OPLS model contains a second equation that relates the predictive super scores and responses:

$$\boldsymbol{Y} = \boldsymbol{T}\boldsymbol{C}^T + \boldsymbol{F} \tag{10}$$

where $\boldsymbol{C}$ is the response loadings matrix that relates the super scores to the responses, and $\boldsymbol{F}$ is the response residual matrix that holds $\boldsymbol{Y}$-variation not captured by the model.

2.3.2. The MB-OPLS algorithm

The proposed MB-OPLS algorithm described herein admits a matrix of responses $\boldsymbol{Y}$, but also supports vector-$\boldsymbol{y}$ cases. Direct and normed assignment will be indicated by "←" and "∝", respectively. All assignments to block-specific structures (e.g., $\boldsymbol{w}_i$) that are to be performed over all values of i from 1 to n are suffixed with "$\forall\ i \in \{1, \ldots, n\}$".

```
1. For each $m \in \{1, \ldots, M\}$ do
   a. $\boldsymbol{v}_{i,m} \leftarrow \boldsymbol{X}_i{}^T\boldsymbol{y}_m \cdot (\boldsymbol{y}_m{}^T\boldsymbol{y}_m)^{-1}\ \forall i \in \{1, \ldots, n\}$
   b. $\boldsymbol{V}_i \leftarrow [\boldsymbol{V}_i|\boldsymbol{v}_{i,m}]\ \forall\ i \in \{1, \ldots, n\}$
2. Initialize $\boldsymbol{u}$ to a column of $\boldsymbol{Y}$
3. $\boldsymbol{w}_i \propto \boldsymbol{X}_i{}^T\boldsymbol{u}\ \forall i \in \{1, \ldots, n\}$
4. $\boldsymbol{t}_i \leftarrow \boldsymbol{X}_i\boldsymbol{w}_i\ \forall\ i \in \{1, \ldots, n\}$
5. $\boldsymbol{R} \leftarrow [\boldsymbol{t}_1| \ldots |\boldsymbol{t}_n]$
6. $\boldsymbol{w}_T \propto \boldsymbol{R}^T\boldsymbol{u}$
7. $\boldsymbol{t} \leftarrow \boldsymbol{R}\boldsymbol{w}_T$
8. $\boldsymbol{c} \leftarrow (\boldsymbol{Y}^T\boldsymbol{t}) \cdot (\boldsymbol{t}^T\boldsymbol{t})^{-1}$
9. $\boldsymbol{u} \leftarrow (\boldsymbol{Y}\boldsymbol{c}) \cdot (\boldsymbol{c}^T\boldsymbol{c})^{-1}$
10. If $\|\boldsymbol{u} - \boldsymbol{u}_{\text{old}}\|/\|\boldsymbol{u}_{\text{old}}\| > \varepsilon$, return to step (3).
    Otherwise, continue to step (11).
11. $\boldsymbol{p}_i \leftarrow (\boldsymbol{X}_i{}^T\boldsymbol{t}) \cdot (\boldsymbol{t}^T\boldsymbol{t})^{-1}\ \forall i \in \{1, \ldots, n\}$
12. To compute an orthogonal component, continue to step (13).
    Otherwise, proceed to step (21).
13. $\boldsymbol{w}_{\boldsymbol{o},i} \leftarrow \boldsymbol{p}_i\ \forall\ i \in \{1, \ldots, n\}$
14. For each $m \in \{1, \ldots, M\}$ do
   a. $\varphi \leftarrow (\sum_{i=1}^{n}\boldsymbol{v}_{i,m}{}^T\boldsymbol{w}_{\boldsymbol{o},i}) \cdot (\sum_{i=1}^{n}\boldsymbol{v}_{i,m}{}^T\boldsymbol{v}_{i,m})^{-1}$
   b. $\boldsymbol{w}_{\boldsymbol{o},i} \leftarrow \boldsymbol{w}_{\boldsymbol{o},i} - \varphi\boldsymbol{v}_{i,m}\ \forall\ i \in \{1, \ldots, n\}$
15. $\boldsymbol{w}_{\boldsymbol{o},i} \leftarrow \boldsymbol{w}_{\boldsymbol{o},i} \cdot (\sum_{i=1}^{n}\boldsymbol{w}_{\boldsymbol{o},i}{}^T\boldsymbol{w}_{\boldsymbol{o},i})^{-1/2}\ \forall i \in \{1, \ldots, n\}$
16. $\boldsymbol{t}_{\boldsymbol{o},i} \leftarrow \boldsymbol{X}_i\boldsymbol{w}_{\boldsymbol{o},i}\ \forall\ i \in \{1, \ldots, n\}$
17. $\boldsymbol{t_o} \leftarrow \sum_{i=1}^{n}\boldsymbol{t}_{\boldsymbol{o},i}$
18. $\boldsymbol{p}_{\boldsymbol{o},i} \leftarrow (\boldsymbol{X}_i{}^T\boldsymbol{t_o}) \cdot (\boldsymbol{t_o}{}^T\boldsymbol{t_o})^{-1}\ \forall i \in \{1, \ldots, n\}$
19. $\boldsymbol{X}_i \leftarrow \boldsymbol{X}_i - \boldsymbol{t_o}\boldsymbol{p}_{\boldsymbol{o},i}{}^T\ \forall i \in \{1, \ldots, n\}$
20. Return to step (2).
21. $\boldsymbol{X}_i \leftarrow \boldsymbol{X}_i - \boldsymbol{t}\boldsymbol{p}_i{}^T\ \forall i \in \{1, \ldots, n\}$
22. To compute another component, return to step (2).
    Otherwise, end.
```

In the above algorithm, the value of ε is set to a very small number, e.g., 10^{-9}. For each predictive component in the model, a set of orthogonal components is extracted. After the computation of a new orthogonal component, the current predictive component is updated to reflect the removal of orthogonal variation from the matrices $\boldsymbol{X}_i$. The MB-OPLS algorithm closely follows the matrix-$\boldsymbol{Y}$ OPLS algorithm presented by Trygg and Wold [6], but replaces the standard PLS computation (steps 4–10 in OPLS) with an MB-PLS computation (steps 2–11 above). However, as described below, the mechanism by which orthogonal variation is removed (steps 13–19 above) is identical to that of OPLS.

2.3.3. Equivalence to OPLS

In both the vector-$\boldsymbol{y}$ and matrix-$\boldsymbol{Y}$ OPLS algorithms proposed by Trygg and Wold [6], a basis $\boldsymbol{V}$ for the response-correlated variation in $\boldsymbol{X}$ is constructed by regressing the data onto each column of responses:

$$\boldsymbol{v}_m \leftarrow \frac{\boldsymbol{X}^T\boldsymbol{y}_m}{\boldsymbol{y}_m{}^T\boldsymbol{y}_m}\ \forall m \in \{1, \ldots, M\} \tag{11}$$

where $\boldsymbol{y}_m$ and $\boldsymbol{v}_m$ denote the m-th columns of $\boldsymbol{Y}$ and $\boldsymbol{V}$, respectively. When $\boldsymbol{X}$ is partitioned into multiple blocks, the computed basis also bears the same partitioning, i.e., $\boldsymbol{V}^T = [\boldsymbol{V}_1{}^T|\ldots|\boldsymbol{V}_n{}^T]$, where each of the n submatrices corresponds to the regression of its respective block $\boldsymbol{X}_i$ onto the responses:

$$\boldsymbol{v}_{i,m} \leftarrow \frac{\boldsymbol{X}_i{}^T\boldsymbol{y}_m}{\boldsymbol{y}_m{}^T\boldsymbol{y}_m}\ \forall m \in \{1, \ldots, M\} \tag{12}$$

where $\boldsymbol{v}_{i,m}$ is the m-th column of $\boldsymbol{V}_i$. Given a single-block PLS loading vector $\boldsymbol{p}$, the OPLS algorithm computes an orthogonal weight $\boldsymbol{w_o}$ by orthogonalizing p to the columns of $\boldsymbol{V}$:

$$\boldsymbol{w_o} \leftarrow \boldsymbol{w_o} - \left(\frac{\boldsymbol{v}_m{}^T \boldsymbol{w_o}}{\boldsymbol{v}_m{}^T \boldsymbol{v}_m} \right) \boldsymbol{v}_m \ \forall m \in \{1, \ldots, M\} \tag{13}$$

after $\boldsymbol{w_o}$ has been initialized to $\boldsymbol{p}$. From the proof of Westerhuis et al. [13], it is known that the single-block PLS loading $\boldsymbol{p}$ equals the concatenation of all block loadings from MB-PLS, i.e., that $\boldsymbol{p}^T = [\boldsymbol{p}_1{}^T|\ldots|\boldsymbol{p}_n{}^T]$. Expansion of all vector terms in the above equation into their partitioned forms results in the following new assignment rule:

$$\boldsymbol{w}_{\boldsymbol{o},i} \leftarrow \boldsymbol{w}_{\boldsymbol{o},i} - \left(\frac{\sum_{i=1}^{n} \boldsymbol{v}_{i,m}{}^T \boldsymbol{w}_{\boldsymbol{o},i}}{\sum_{i=1}^{n} \boldsymbol{v}_{i,m}{}^T \boldsymbol{v}_{i,m}} \right) \boldsymbol{v}_{i,m} \ \forall m \in \{1, \ldots, M\} \tag{14}$$

The scalar term in Eq. (14) should be recognized as φ in the MB-OPLS algorithm. By the same reasoning, step (15) in the algorithm is equivalent to scaling $\boldsymbol{w_o}$ to unit norm. In effect, by computing φ as the fraction of orthogonal variation to remove from its loadings, MB-OPLS yields the same orthogonal weights ($\boldsymbol{w_o}$) as OPLS of the concatenated matrix. Therefore, because $\boldsymbol{w_o}$ equals the column-wise concatenation of all weights $\boldsymbol{w}_{\boldsymbol{o},i}$, it is then apparent that the orthogonal super scores extracted by MB-OPLS are identical to those from OPLS of the concatenated matrix $\boldsymbol{X}$, as illustrated in the following equation:

$$\boldsymbol{t_o} = \boldsymbol{E}\boldsymbol{w_o} = [\boldsymbol{E}_1|\ldots|\boldsymbol{E}_n] \begin{bmatrix} \boldsymbol{w}_{\boldsymbol{o},1} \\ \vdots \\ \boldsymbol{w}_{\boldsymbol{o},n} \end{bmatrix} = \sum_{i=1}^{n} \boldsymbol{E}_i \boldsymbol{w}_{\boldsymbol{o},i} = \sum_{i=1}^{n} \boldsymbol{t}_{\boldsymbol{o},i} \tag{15}$$

From this equivalence, and the fact that steps (2–11) and (21) in MB-OPLS constitute an MB-PLS iteration, we arrive at the equivalence between MB-OPLS and OPLS. Thus, orthogonality between the responses and orthogonal super scores $\boldsymbol{t_o}$ computed by MB-OPLS is also ensured. However, because the computation of orthogonal weights involves all blocks, the resulting orthogonal block scores $\boldsymbol{t}_{\boldsymbol{o},i}$ are not guaranteed to be orthogonal to the responses.

2.3.4. Computation from an OPLS Model

The equivalence between MB-OPLS super scores and OPLS scores may be leveraged to generate an MB-OPLS model from an existing OPLS model of a partitioned data matrix, saving computation time during cross-validated model training. The following algorithm details the extraction of MB-OPLS block scores and loadings from an OPLS model:

```
1. Initialize $a = 1, b = 1$
2. $\boldsymbol{t_o} \leftarrow [\boldsymbol{T_o}]_a$
3. $\boldsymbol{w}_{\boldsymbol{o},i} \leftarrow [\boldsymbol{W_o}]_i \ \forall \ i \in \{1, \ldots, n\}$
4. $\boldsymbol{t}_{\boldsymbol{o},i} \leftarrow \boldsymbol{X}_i \boldsymbol{w}_{\boldsymbol{o},i} \ \forall \ i \in \{1, \ldots, n\}$
5. $\boldsymbol{p}_{\boldsymbol{o},i} \leftarrow (\boldsymbol{X}_i{}^T \boldsymbol{t_o}) \cdot (\boldsymbol{t_o}^T \boldsymbol{t_o})^{-1} \ \forall i \in \{1, \ldots, n\}$
6. $\boldsymbol{T}_{\boldsymbol{o},i} \leftarrow [\boldsymbol{T}_{\boldsymbol{o},i}|\boldsymbol{t}_{\boldsymbol{o},i}] \ \forall \ i \in \{1, \ldots, n\}$
7. $\boldsymbol{P}_{\boldsymbol{o},i} \leftarrow [\boldsymbol{P}_{\boldsymbol{o},i}|\boldsymbol{p}_{\boldsymbol{o},i}] \ \forall \ i \in \{1, \ldots, n\}$
8. $\boldsymbol{W}_{\boldsymbol{o},i} \leftarrow [\boldsymbol{W}_{\boldsymbol{o},i}|\boldsymbol{w}_{\boldsymbol{o},i}] \ \forall \ i \in \{1, \ldots, n\}$
9. If another orthogonal component exists, increment $a$ and return to step (2).
   Otherwise, continue to step (10).
10. $\boldsymbol{X}_i \leftarrow \boldsymbol{X}_i - \boldsymbol{T_o} \boldsymbol{P}_{\boldsymbol{o},i}{}^T \ \forall i \in \{1, \ldots, n\}$
11. $\boldsymbol{u} \leftarrow [\boldsymbol{U}]_b$
12. $\boldsymbol{t} \leftarrow [\boldsymbol{T}]_b$
13. $\boldsymbol{w}_i \propto \boldsymbol{X}_i{}^T \boldsymbol{u} \ \forall i \in \{1, \ldots, n\}$
14. $\boldsymbol{t}_i \leftarrow \boldsymbol{X}_i \boldsymbol{w}_i \ \forall \ i \in \{1, \ldots, n\}$
15. $\boldsymbol{p}_i \leftarrow (\boldsymbol{X}_i{}^T \boldsymbol{t}) \cdot (\boldsymbol{t}^T \boldsymbol{t})^{-1} \ \forall i \in \{1, \ldots, n\}$
16. $\boldsymbol{T}_i \leftarrow [\boldsymbol{T}_i|\boldsymbol{t}_i] \ \forall \ i \in \{1, \ldots, n\}$
17. $\boldsymbol{p}_i \leftarrow [\boldsymbol{P}_i|\boldsymbol{p}_i] \ \forall \ i \in \{1, \ldots, n\}$.
18. $\boldsymbol{W}_i \leftarrow [\boldsymbol{W}_i|\boldsymbol{w}_i] \ \forall \ i \in \{1, \ldots, n\}$
19. $\boldsymbol{X}_i \leftarrow \boldsymbol{X}_i - \boldsymbol{t}\boldsymbol{p}_i{}^T \ \forall i \in \{1, \ldots, n\}$
20. If another predictive component exists, increment $b$ and return to step (1).Otherwise, end.
```

The keen observer will recognize the equivalence between steps (10–19) above and the procedure outlined by Westerhuis et al. for extracting MB-PLS block components from a PLS model [13]. By using the above algorithm to compute MB-OPLS models, the analyst avoids the unnecessary computation of block components during cross-validated model training. For example, a G-fold Monte Carlo cross-validation having R iterations requires the construction of RG models in order to yield R cross-validated response matrix estimates. In each of these RG models, MB-PLS requires $2Nn$ additional floating-point multiplications (per power iteration) over PLS. In addition, computation of multiblock components from single-block models ensures greater stability of super scores and loadings, especially in cases of missing data [13].

3. Datasets

Two datasets will be described to illustrate how MB-OPLS effectively integrates an OSC filter into an MB-PLS decomposition of a set of n matrices. The first synthetic dataset contrasts the mixing of predictive information in MB-PLS with its separation in MB-OPLS using a contrived three-block regression example similar to that introduced by Löfstedt and Trygg [16]. The second dataset, a joint set of nuclear magnetic resonance (NMR) and mass spectrometry (MS) observations [22,23], is used to demonstrate the enhanced interpretability of MB-OPLS models over MB-PLS in a real example of discriminant analysis. All modeling and validation were performed using routines available in the MVAPACK chemometrics toolbox (http://bionmr.unl.edu/mvapack.php) [24].

3.1. Synthetic example

In the first dataset, three matrices (all having 100 rows and 200 columns) were constructed to hold one $\boldsymbol{y}$-predictive component ($\boldsymbol{t}\boldsymbol{p}_i{}^T$) and one $\boldsymbol{y}$-orthogonal component ($\boldsymbol{t_o}\boldsymbol{p}_{\boldsymbol{o},i}{}^T$). The score vectors were non-overlapping (orthogonal) Gaussian density functions, and all block loading vectors were mutually overlapping Gaussian density or square step functions. The true synthetic block loadings are illustrated in Fig. 1A. A two-component MB-PLS-R regression model was trained on the synthetic three-block example dataset, as well as a 1 + 1 (one predictive, one orthogonal) component MB-OPLS-R regression model. Block loadings extracted by MB-PLS-R and MB-OPLS-R are shown in Figs. 1B and C, respectively.

3.2. Joint ^{1}H NMR and DI-ESI-MS datasets

The second dataset is a pair of processed and treated data matrices, collected on 29 samples of metabolite extracts from human dopaminergic neuroblastoma cells treated with various neurotoxic agents [23]. The first matrix, collected using ^{1}H NMR spectroscopy, contains 16,138 columns and the second, collected using direct injection electrospray ionization mass spectrometry (DI-ESI-MS), contains 2095 columns. Prior to all modeling, block weighting was applied after Pareto scaling to ensure equal contribution of each block to the models (fairness) [12].

In previously performed work, a two-component, two-class (vector-$\boldsymbol{y}$) multiblock discriminant analysis (MB-PLS-DA) model was trained on the dataset in order to discriminate between untreated and neurotoxin-treated cell samples. To highlight the improved interpretability of MB-OPLS over MB-PLS, a 1 + 1 MB-OPLS-DA model was trained on the data using an identical vector of class labels. Block components were extracted from an OPLS-DA model of the concatenated matrix $\boldsymbol{X} = [\boldsymbol{X}_{\text{NMR}}|\boldsymbol{X}_{\text{MS}}]$ using the above algorithm. For both models, fifty rounds of Monte Carlo seven-fold cross-validation [25,26] were performed to compute per-component Q^2 statistics [3], in addition to the R^2 statistics

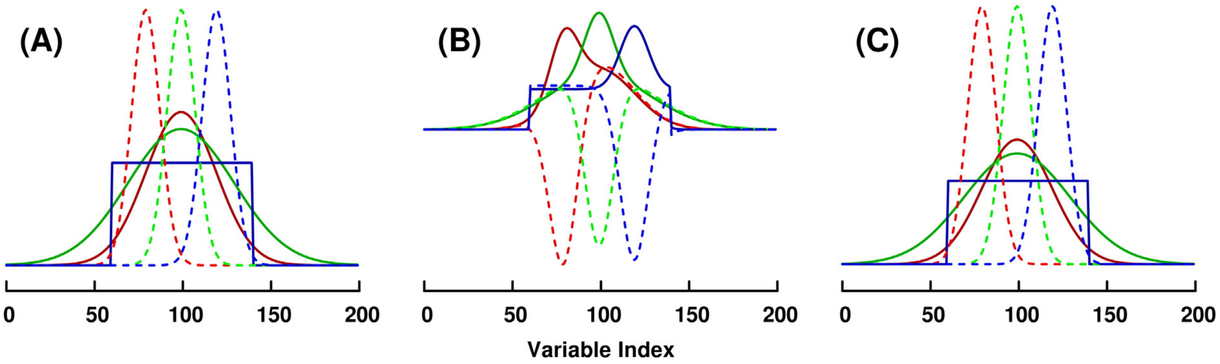

Fig. 1. Block loadings in the synthetic multiblock example dataset. (A) True predictive loadings (solid) and orthogonal loadings (dashed) used to construct the three-block dataset. First, second and third block loadings are colored in red, green and blue, respectively. (B) First component (solid) and second component (dashed) loadings identified by MB-PLS modeling of the three data blocks. (C) Predictive (solid) and orthogonal (dashed) block loadings identified by MB-OPLS, illustrating the separation of ***y***-uncorrelated variation accomplished by the integrated OSC filter. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

available from model training. CV-ANOVA significance testing was also applied to further assess model reliability [27].

4. Results and discussion

In both the contrived dataset and the real spectroscopic dataset, the interpretative advantage offered by MB-OPLS over MB-PLS is strikingly apparent. In the synthetic example, MB-OPLS capably identifies the true predictive and orthogonal loadings in the presence of ***y***-orthogonal variation that clouds the interpretation of MB-PLS loadings (Fig. 1). By design, this comparison between MB-OPLS and MB-PLS is highly similar to the first example presented by Löfstedt and Trygg to compare nPLS and OnPLS for general data discovery [16]. However, as is evidenced by the differences between Eqs. (3) and (7) above, MB-OPLS solves an inherently distinct problem from OnPLS: the identification of consensus variation in multiple blocks of data that predicts a single set of responses.

The ability of MB-OPLS to separate predictive and orthogonal variation from multiple data matrices is further exemplified in the discriminant analysis of the real spectroscopic dataset. From the rotated discrimination axis in the MB-PLS-DA scores (Fig. 2A), it is clear that predictive and orthogonal variation have become mixed in the corresponding block loadings (Fig. 3). Integration of an OSC filter into the multiblock model in the form of MB-OPLS-DA achieves the expected rotation of super scores to place more predictive variation into the first component (Fig. 2B). As a consequence of this rotation, spectral information that separates paraquat treatment from other neurotoxin treatments is also moved into the orthogonal component. For example, strong loadings from citrate in the ^{1}H NMR MB-PLS block loadings (Fig. 3A, 2.6 ppm) are substantially diminished in the predictive block loadings from MB-OPLS (Fig. 4), as separation between paraquat and other treatments has been isolated along the orthogonal component in super scores. Inspection of the orthogonal block loadings from MB-OPLS (Supplementary Fig. S-4) will reveal, as expected, that citrate contributes more to separation between neurotoxin treatments than to separation between treatments and controls. Similar patterns were observed in the DI-ESI-MS block loadings at m/z 203.058 and 233.067, which were assigned via accurate mass and tandem MS measurements as sodium adducts of hexose and heptose, respectively [23]. These results agree with detailed prior analyses of pairwise MB-PLS-DA models between each drug treatment and untreated cells, which indicate that paraquat treatment uniquely alters metabolic flux through glycolysis and the pentose phosphate pathway [22]. In contrast to the multiple MB-PLS-DA models employed by Lei et al. to arrive at this conclusion [22], the MB-OPLS-DA model has provided the same set of core results in a single, substantially more interpretable model.

The partial correlation of both predictive and orthogonal block scores in MB-OPLS is readily observed in the comparison of block scores from MB-PLS and MB-OPLS (Supplementary Figs. S-2 and S-3). While the super scores in Fig. 2B are rotated to separate predictive and orthogonal variation, block scores in Figs. S-2B and S-3B have rotated back into alignment with the MB-PLS block scores. This partial correlation and re-

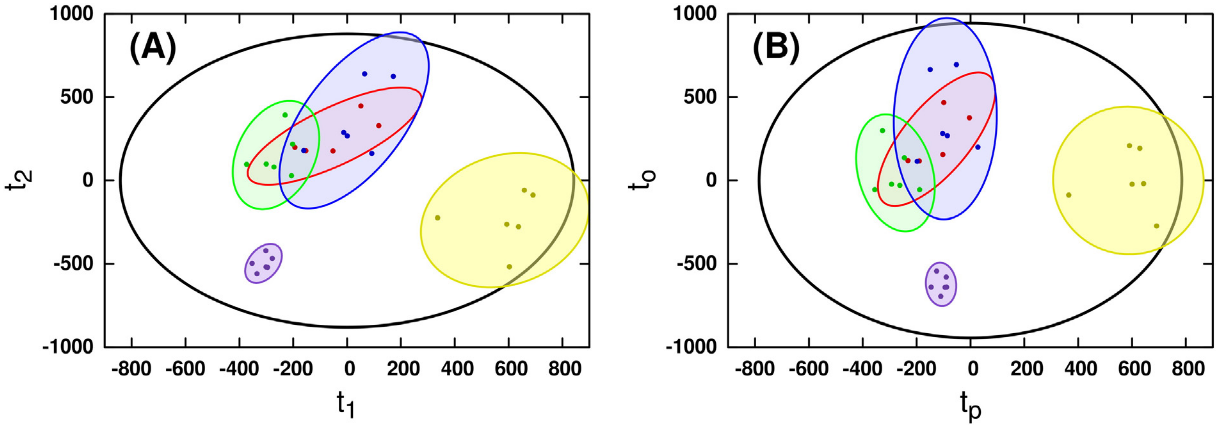

Fig. 2. Super scores identified by (A) MB-PLS and (B) MB-OPLS modeling of the joint ^{1}H NMR and DI-ESI-MS data matrices. Extraction of ***y***-orthogonal variation from the first PLS component is clear in the MB-OPLS scores. Ellipses represent the 95% confidence regions for each sub-class of observations, assuming normal distributions. Colors indicate membership to the untreated (yellow), 6-hydroxydopamine (red), 1-methyl-4-phenylpyridinium (green) and paraquat (violet) sub-classes. Cross-validated super scores for each model are shown in Supplementary Figure S-1. Block scores for each model are shown in Supplementary Figs. S-2 and S-3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

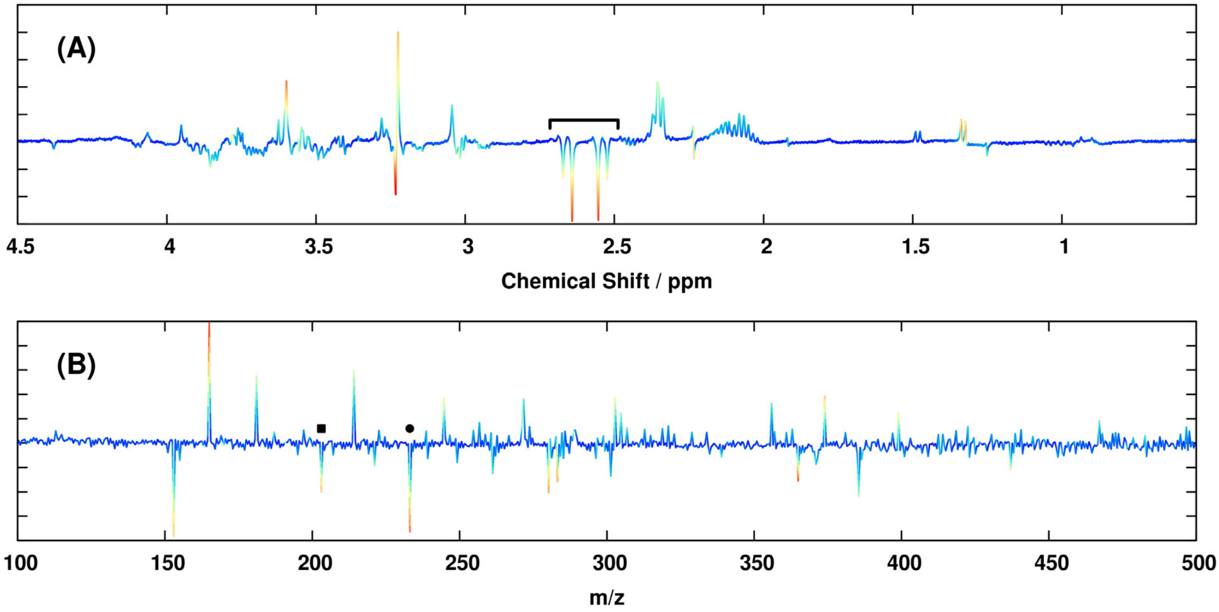

Fig. 3. Backscaled first-component block loadings from the MB-PLS model of the (A) ^{1}H NMR and (B) DI-ESI-MS data matrices. Coloring of each loading vector ranges from blue to red based on the amount of point-wise weighting applied during Pareto scaling. It is important to note that a second PLS component exists in the MB-PLS model that is not shown. Spectral contributions from citrate in the ^{1}H NMR MB-PLS block loadings (2.6 ppm) are indicated by a bracket, and contributions from hexose and heptose in the DI-ESI-MS loadings are indicated by black squares and circles, respectively.

mixing of predictive and orthogonal variation in MB-OPLS block scores is a consequence of the use of super score deflation in the proposed algorithm. When all matrices contain similar patterns of orthogonal variation, their MB-OPLS block scores will reflect this by retaining the OSC-induced rotation captured at the consensus level by the super scores. However, because the interpretative advantage of MB-OPLS over MB-PLS lies in the relationship between super scores and block loadings, the fact that orthogonal block scores have partial $\boldsymbol{y}$-correlation is relatively benign.

Because the MB-OPLS-DA model of the real spectral data matrices was trained using the single-block OPLS routine already present in MVAPACK, all readily available cross-validation metrics were available in the model without further computational expenditure. Monte Carlo cross-validation of the MB-PLS model produced cumulative $R^2{}_Y$ and Q^2 statistics of 0.903 and 0.819 ± 0.024, respectively, and validation of the MB-OPLS model resulted in statistics of 0.903 and 0.736 ± 0.021, respectively. As expected, the MB-OPLS model captured the same fraction of response variation ($R^2{}_Y$) as MB-PLS, reaffirming the fact that the two methods have the same predictive ability. In addition, MB-OPLS modeling yielded $R^2{}_{Xp}$ and $R^2{}_{Xo}$ statistics of 0.378 and 0.245 for the first block, and 0.236 and 0.083 for the second block. Monte Carlo cross-validated super scores from MB-PLS and MB-OPLS are depicted in Supplementary Fig. S-1. Compared to MB-PLS scores in Fig. S-1A, MB-OPLS scores (Fig. S-1B) exhibit an increased uncertainty during cross-validation due to the coupled nature of predictive and orthogonal components in OPLS models. Further validation of the MB-OPLS-DA model via CV-ANOVA produced a p value equal to 2.88×10^{-6}, indicating a sufficiently reliable model.

It is worthy of final mention that the objective solved by MB-OPLS is but a single member of a superfamily of multiblock methods introduced in detail by Hanafi and Kiers [17]. In the first family, nPLS and OnPLS maximally capture the between-matrix covariances before and after orthogonal signal correction, respectively, and thus serve to regress a set of matrices against each other. Methods in the second family capture *both* within-matrix variances and between-matrix covariances of a set of matrices (CPCA-W), a set of response-weighted matrices (MB-PLS), and a set of response-weighted OSC-filtered matrices (MB-OPLS). By casting these methods in the light of MAXDIFF and MAXBET, we obtain an informative picture of their characteristics, commonalities, and differences. For example, nPLS and OnPLS force an equal contribution of each matrix to the solution through the constraint $\|\boldsymbol{w}_i\| = 1$, while CPCA-W, MB-PLS and MB-OPLS allow contributions to float based on the "importance" of each matrix to the modeling problem at hand. This super weight approach necessitates a block scaling procedure to avoid highly weighting any given matrix due to size alone [12,13].

5. Conclusions

The MB-OPLS method proposed here is a versatile extension of MB-PLS to include an OSC filter, and belongs to a family of MAXBET optimizers that share an equivalence with their single-block factorizations (Supplementary Fig. S-5). By removing consensus response-uncorrelated variation from a set of n data matrices, MB-OPLS expands the scope and benefits of OPLS to cases where blocking information is available. The ability of MB-OPLS to separate predictive and orthogonal variation from multiple blocks of data has been demonstrated on both synthetic and real spectral data, both in cases of vector-$\boldsymbol{y}$ regression and discriminant analysis. Of course, while both examples were interpreted in the light of spectroscopic datasets like those used in metabolomics [22,23], MB-OPLS is a fully general algorithm that admits any multiblock dataset for the purposes of regression or discriminant analysis. For example, recent applications of MB-PLS for investigating food spoilage [28], iron-ore content [29], chemical toxicity [30], the evolution of human anatomy [31], and the assessment of cortical and muscle activity in Parkinson's disease patients [32] would benefit from our MB-OPLS algorithm. The presented algorithm admits either a vector or a matrix as responses, and is implemented in the latest version of the open-source MVAPACK chemometrics toolbox [24].

Notes

The authors declare no competing financial interest.

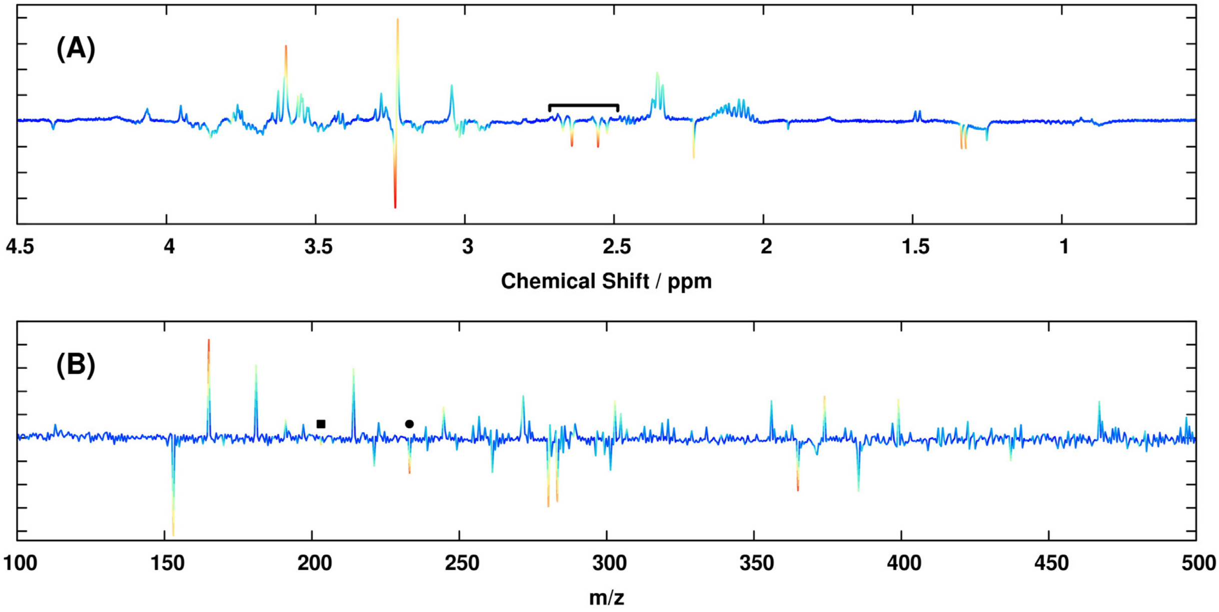

Fig. 4. Backscaled first-component block loadings from the MB-OPLS model of the (A) ^{1}H NMR and (B) DI-ESI-MS data matrices. Coloring of each loading vector is identical to that of Fig. 3. Spectral contributions from citrate in the ^{1}H NMR MB-OPLS block loadings (2.6 ppm) are indicated by a bracket, and contributions from hexose and heptose in the DI-ESI-MS loadings are indicated by black squares and circles, respectively. Unlike the two-component MB-PLS model, the single predictive MB-OPLS component here fully separates observations between the classes under discrimination. Backscaled orthogonal block loadings from the same model are shown in Supplementary Fig. S-4.

Acknowledgments

This manuscript was supported, in part, by funds from grants R01 CA163649, P01 AI083211-06 and P30 GM103335 from the National Institutes of Health. The research was performed in facilities renovated with support from the National Institutes of Health (RR015468-01).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.chemolab.2015.10.018.

References

[1] M. Andersson, A comparison of nine PLS1 algorithms, J. Chemom. 23 (2009) 518–529.
[2] A. Hoskuldsson, PLS regression methods, J. Chemom. 2 (1988) 18.
[3] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. 58 (2001) 109–130.
[4] G.H. Golub, C.F. Van Loan, Matrix Computations, 3rd ed. Johns Hopkins University Press, Baltimore, 1996.
[5] C.L. Lawson, R.J. Hanson, D.R. Kincaid, F.T. Krogh, Basic linear algebra subprograms for fortran usage, ACM Trans. Math. Softw. 5 (1979) 308–323.
[6] J. Trygg, S. Wold, Orthogonal projections to latent structures (O-PLS), J. Chemom. 16 (2002) 119–128.
[7] J.C. Boulet, J.M. Roger, Pretreatments by means of orthogonal projections, Chemom. Intell. Lab. 117 (2012) 61–69.
[8] T. Verron, R. Sabatier, R. Joffre, Some theoretical properties of the O-PLS method, J. Chemom. 18 (2004) 62–68.
[9] S. Wiklund, E. Johansson, L. Sjoestroem, E.J. Mellerowicz, U. Edlund, J.P. Shockcor, J. Gottfries, T. Moritz, J. Trygg, Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models, Anal. Chem. 80 (2008) 115–122.
[10] D.S. Wishart, Metabolomics: applications to food science and nutrition research, Trends Food Sci. Technol. 19 (2008) 482–493.
[11] N. Souihi, A. Lindegren, L. Eriksson, J. Trygg, OPLS in batch monitoring — opens up new opportunities, Anal. Chim. Acta 857 (2015) 28–38.
[12] A.K. Smilde, J.A. Westerhuis, S. de Jong, A framework for sequential multiblock component methods, J. Chemom. 17 (2003) 323–337.
[13] J.A. Westerhuis, T. Kourti, J.F. MacGregor, Analysis of multiblock and hierarchical PCA and PLS models, J. Chemom. 12 (1998) 301–321.
[14] J.A. Westerhuis, P.M.J. Coenegracht, Multivariate modelling of the pharmaceutical two-step process of wet granulation and tableting with multiblock partial least squares, J. Chemom. 11 (1997) 379–392.
[15] J.M.F. Tenberge, Generalized approaches to the maxbet problem and the maxdiff problem, with applications to canonical correlations, Psychometrika 53 (1988) 487–494.
[16] T. Lofstedt, J. Trygg, OnPLS-a novel multiblock method for the modelling of predictive and orthogonal variation, J. Chemom. 25 (2011) 441–455.
[17] M. Hanafi, H.A.L. Kiers, Analysis of K sets of data, with differential emphasis on agreement between and within sets, Comput. Stat. Data Anal. 51 (2006) 1491–1508.
[18] T. Löfstedt, OnPLS: orthogonal projections to latent structures in multiblock and path model data analysis, Chemistry, Umea University, Umea, Sweden 2012, p. 76.
[19] T. Lofstedt, L. Eriksson, G. Wormbs, J. Trygg, Bi-modal OnPLS, J. Chemom. 26 (2012) 236–245.
[20] T. Lofstedt, M. Hanafi, G. Mazerolles, J. Trygg, OnPLS path modelling, Chemom. Intell. Lab. 118 (2012) 139–149.
[21] T. Lostedt, D. Hoffman, J. Trygg, Global, local and unique decompositions in OnPLS for multiblock data analysis, Anal. Chim. Acta 791 (2013) 13–24.
[22] S. Lei, L. Zavala-Flores, A. Garcia-Garcia, R. Nandakumar, Y. Huang, N. Madayiputhiya, R.C. Stanton, E.D. Dodds, R. Powers, R. Franco, Alterations in energy/redox metabolism induced by mitochondrial and environmental toxins: a specific role for glucose-6-phosphate-dehydrogenase and the pentose phosphate pathway in paraquat toxicity, ACS Chem. Biol. 9 (2014) 2032–2048.
[23] D.D. Marshall, S.L. Lei, B. Worley, Y.T. Huang, A. Garcia-Garcia, R. Franco, E.D. Dodds, R. Powers, Combining DI-ESI-MS and NMR datasets for metabolic profiling, Metabolomics 11 (2015) 391–402.
[24] B. Worley, R. Powers, MVAPACK: a complete data handling package for NMR metabolomics, ACS Chem. Biol. 9 (2014) 1138–1144.
[25] J. Shao, Linear-model selection by cross-validation, J. Am. Stat. Assoc. 88 (1993) 486–494.
[26] Q.S. Xu, Y.Z. Liang, Monte Carlo cross validation, Chemom. Intell. Lab. 56 (2001) 1–11.
[27] L. Eriksson, J. Trygg, S. Wold, CV-ANOVA for significance testing of PLS and OPLS (R) models, J. Chemom. 22 (2008) 594–600.
[28] Y. Xu, E. Correa, R. Goodacre, Integrating multiple analytical platforms and chemometrics for comprehensive metabolic profiling: application to meat spoilage detection, Anal. Bioanal. Chem. 405 (2013) 5063–5074.
[29] P. Yaroshchyk, D.L. Death, S.J. Spencer, Comparison of principal components regression, partial least squares regression, multi-block partial least squares regression, and serial partial least squares regression algorithms for the analysis of Fe in iron ore using LIBS, J. Anal. At. Spectrom. 27 (2012) 92–98.
[30] J. Zhao, S. Yu, Quantitative structure–activity relationship of organophosphate compounds based on molecular interaction fields descriptors, Environ. Toxicol. Pharmacol. 35 (2013) 228–234.
[31] M. Coquerelle, J.C. Prados-Frutos, S. Benazzi, F.L. Bookstein, S. Senck, P. Mitteroecker, G.W. Weber, Infant growth patterns of the mandible in modern humans: a closer exploration of the developmental interactions between the symphyseal bone, the teeth, and the suprahyoid and tongue muscle insertion sites, J. Anat. 222 (2013) 178–192.
[32] J. Chiang, Z.J. Wang, M.J. McKeown, A multiblock PLS model of cortico-cortical and corticomuscular interactions in Parkinson's disease, Neuroimage 63 (2012) 1498–1509.