JOURNAL OF CHEMOMETRICS, VOL. 11, 367–377 (1997)

COMMENTS ON MULTILINEAR PLS

AGE K. SMILDE
Laboratory for Analytical Chemistry, University of Amsterdam, Nieuwe Achtergracht 166, NL-1018 WV Amsterdam, Netherlands

SUMMARY

Recently, Bro published a paper on multilinear PLS (*J. Chemometrics*, **10**, 47–61 (1996)) in which he proposed a generalization of PLS to multiway situations, called multilinear PLS, which is a mixture of a trilinear model (PARAFAC) and PLS. However, Bro does not give the equations for the prediction step. In this paper these prediction equations are given in both their full and closed forms. The least squares properties of the proposed multilinear PLS are established and a more comprehensive notation is given. Using this notation, it is clear that some other multiway analysis methods such as PARAFAC and Tucker1 models can be combined with PLS. Multiway methods such as Tucker2 and Tucker3 need a different approach. A framework is given for general two-block multiway models. © 1997 John Wiley & Sons, Ltd.

Journal of Chemometrics, Vol. **11,** 367–377 (1997) (No. of Figures: 0 No. of Tables: 0 No. of References: 18)

KEY WORDS multilinear PLS; multiway calibration; PARAFAC; Tucker; PLS; principal covariates regression

INTRODUCTION

The prediction of a y-value using a multivariate **x**-vector as a generic problem has attracted a lot of attention in chemometrics. A very popular method to do this prediction is partial least squares (PLS).[1–3] In another area of chemometrics, multiway data analysis methods have gained attention.[4,5] These methods deal with data that can be arranged meaningfully in a cube or a four-way data matrix, etc. Most data analysis applications in chemometrics deal with two-way data, e.g. predicting a univariate **y** with a matrix **X**. This is the two-block two-way case in the terminology of Smilde.[5] Bro has extended the PLS method to multiway data; that is, predicting a **Y** using an **X**, where both **X** and **Y** can be three-way or even four-way, in general multiway.[6]

The extension of PLS to multiway data is fruitful, since many chemical problems generate data that are essentially multiway. The combination of three-way analysis methods such as PARAFAC[7] and Tucker1[8] with the PLS method is logical and possible, as Bro has already shown for the PARAFAC/PLS case. For other multiway methods (Tucker2 and Tucker3),[8] different methods are needed to obtain two-block solutions.

In the following, a brief review of Bro's method is given. Another notation is introduced and some properties of Bro's algorithms are established. A general framework is given for two-block multiway models.

In this paper, scalars are written as lowercase italic characters, vectors as boldface lowercase characters, matrices as boldface uppercase characters and three-way (and in general multiway) matrices as underlined boldface uppercase characters. The lowercase characters i, j and k will be used as running indices, where $i=1, \ldots, I$, $j=1, \ldots, J$ and $k=1, \ldots, K$.

CCC 0886–9383/97/050367–11 $17.50

Received 5 November 1996
Accepted 4 February 1997

MULTILINEAR PLS: ALGORITHM AND LEAST SQUARES PROPERTY

For convenience, two-way PLS (ordinary PLS) is repeated briefly, because details of PLS are needed in what follows. For simplicity, only PLS1 is considered. Suppose $\mathbf{X}$ ($I \times J$) and $\mathbf{y}$ ($I \times 1$) are column-centered (and scaled if necessary) matrices. The first PLS1 component which is calculated to predict $\mathbf{y}$ from $\mathbf{X}$ solves

$$\max_{\mathbf{w}} [\mathrm{cov}(\mathbf{t}, \mathbf{y}) | \mathbf{X}\mathbf{w} = \mathbf{t} \quad \text{and} \quad \|\mathbf{w}\| = 1] \tag{1}$$

Equation (1) can be written in summation form as

$$\max_{\mathbf{w}} \left[\sum_{i=1}^{I} t_i y_i \;\middle|\; t_i = \sum_{j=1}^{J} x_{ij} w_j \quad \text{and} \quad \|\mathbf{w}\| = 1 \right] \tag{2}$$

where t_i, y_i, x_{ij} and w_j are the typical elements of $\mathbf{t}$, $\mathbf{y}$, $\mathbf{X}$ and $\mathbf{w}$ respectively. For convenience, without changing the problem, the covariance is expressed in a summation without correction for the degrees of freedom. This can be rewritten as

$$\max_{\mathbf{w}} \left[\sum_{i=1}^{I} \sum_{j=1}^{J} y_i x_{ij} w_j \;\middle|\; \|\mathbf{w}\| = 1 \right] \tag{3}$$

and by defining $\mathbf{z} = \mathbf{X}^{\mathrm{T}}\mathbf{y}$ or, equivalently, $z_j = \Sigma_i y_i x_{ij}$, where $\mathbf{z} = (z_1, \ldots, z_J)^{\mathrm{T}}$, this is equivalent to

$$\max_{\mathbf{w}} \left[\sum_{j=1}^{J} z_j w_j \;\middle|\; \|\mathbf{w}\| = 1 \right] \tag{4}$$

This expression is clearly maximized if and only if

$$\mathbf{w} = \frac{\mathbf{z}}{\|\mathbf{z}\|} = \frac{\mathbf{X}^{\mathrm{T}}\mathbf{y}}{\|\mathbf{X}^{\mathrm{T}}\mathbf{y}\|} \tag{5}$$

or, stated differently, if $\mathbf{w}$ is the left singular vector of $\mathbf{X}^{\mathrm{T}}\mathbf{y}$ which is equal to the first eigenvector of $\mathbf{X}^{\mathrm{T}}\mathbf{y}\mathbf{y}^{\mathrm{T}}\mathbf{X}$. This is a well-known result.[2]

After finding the first component, $\mathbf{y}$ and $\mathbf{X}$ are deflated and the second component is found. Since this will be used in the sequel, the basic PLS1 algorithm is repeated for convenience. Some extra notation is needed: $\mathbf{X}^{(0)}$ and $\mathbf{y}^{(0)}$ are the original $\mathbf{X}$ and $\mathbf{y}$ respectively; $\mathbf{X}^{(a)}$ and $\mathbf{y}^{(a)}$ refer to the updated $\mathbf{X}$ and $\mathbf{y}$ respectively after the ath PLS component. The subscript a in $\mathbf{t}_a$, $\mathbf{w}_a$ and $\mathbf{p}_a$ refers to the ath component in the PLS1 model. Then PLS1 proceeds as follows:

1. $\mathbf{t}_1 = \mathbf{X}^{(0)}\mathbf{w}_1$
2. max cov $(\mathbf{t}_1, \mathbf{y}^{(0)}) \Rightarrow \mathbf{w}_1 = \mathbf{X}^{(0)\mathrm{T}}\mathbf{y}^{(0)} / \|\mathbf{X}^{(0)\mathrm{T}}\mathbf{y}^{(0)}\|$
3. $\mathbf{p}_1 = \mathbf{X}^{(0)\mathrm{T}}\mathbf{t}_1 / (\mathbf{t}_1^{\mathrm{T}}\mathbf{t}_1)$
4. $\mathbf{X}^{(1)} = \mathbf{X}^{(0)} - \mathbf{t}_1\mathbf{p}_1^{\mathrm{T}}$
5. $\mathbf{y}^{(1)} = \mathbf{y}^{(0)} - \mathbf{t}_1 b_1$; $b_1 = (\mathbf{t}_1^{\mathrm{T}}\mathbf{t}_1)^{-1}\mathbf{t}_1^{\mathrm{T}}\mathbf{y}^{(0)}$ (6)
6. $\mathbf{t}_2 = \mathbf{X}^{(1)}\mathbf{w}_2$
7. max cov $(\mathbf{t}_2, \mathbf{y}^{(1)}) \Rightarrow \mathbf{w}_2 = \mathbf{X}^{(1)\mathrm{T}}\mathbf{y}^{(1)} / \|\mathbf{X}^{(1)\mathrm{T}}\mathbf{y}^{(1)}\|$
8. $\mathbf{p}_2 = \mathbf{X}^{(1)\mathrm{T}}\mathbf{t}_2 / (\mathbf{t}_2^{\mathrm{T}}\mathbf{t}_2)$
9. $\mathbf{X}^{(2)} = \mathbf{X}^{(1)} - \mathbf{t}_2\mathbf{p}_2^{\mathrm{T}}$
10. $\mathbf{y}^{(2)} = \mathbf{y}^{(1)} - \mathbf{t}_2 b_2$; $b_2 = (\mathbf{t}_2^{\mathrm{T}}\mathbf{t}_2)^{-1}\mathbf{t}_2^{\mathrm{T}}\mathbf{y}^{(1)}$

. . .

until A components

and, by defining $\mathbf{T}=[\mathbf{t}_1, \ldots, \mathbf{t}_A]$, $\mathbf{W}=[\mathbf{w}_1, \ldots, \mathbf{w}_A]$ and $\mathbf{P}=[\mathbf{p}_1, \ldots, \mathbf{p}_A]$, results in the model[2,3]

$$\begin{aligned}&1.\ \mathbf{T}=\mathbf{X}\mathbf{W}(\mathbf{P}^{\mathrm{T}}\mathbf{W})^{-1}\\&2.\ \mathbf{X}=\mathbf{T}\mathbf{P}^{\mathrm{T}}+\mathbf{E}_{\mathbf{x}}\\&3.\ \mathbf{y}=\mathbf{T}\mathbf{b}+\mathbf{y}^{(A)}\\&4.\ \max \operatorname{cov}(\mathbf{t}_a, \mathbf{y}^{(a-1)});\ a=1, \ldots, A\end{aligned} \tag{7}$$

where $\mathbf{W}^{\mathrm{T}}\mathbf{W}=\mathbf{I}$, $\mathbf{T}^{\mathrm{T}}\mathbf{T}$ is diagonal and $\mathbf{w}_i^{\mathrm{T}}\mathbf{p}_j=0$ $(i<j)$. It can be shown that $\mathbf{P}$ solves the problem of minimizing $\|\mathbf{X}-\mathbf{T}\mathbf{P}^{\mathrm{T}}\|^2$ for given $\mathbf{T}$ using the orthogonality of the $\mathbf{t}$-vectors. Note that $\mathbf{P}$ and $\mathbf{T}$ do not solve the problem of mimimizing $\|\mathbf{X}-\mathbf{T}\mathbf{P}^{\mathrm{T}}\|^2$ for general $\mathbf{P}$ and $\mathbf{T}$, hence the name *partial* least squares. The above PLS1 model is Wold's version of PLS1.

There is another formulation of the PLS1 model due to Martens and Naes.[9] For convenience the same notation as in Wold's version is used, but the two versions generate different $\mathbf{t}$s, $\mathbf{b}$s and deflated $\mathbf{X}$s. In Martens' PLS1 the $\mathbf{t}$s are not forced to be orthogonal. Martens' PLS1 scheme is as follows:

$$\begin{aligned}&1.\ \mathbf{t}_1=\mathbf{X}^{(0)}\mathbf{w}_1\\&2.\ \max \operatorname{cov}(\mathbf{t}_1, \mathbf{y}^{(0)})\Rightarrow\mathbf{w}_1=\mathbf{X}^{(0)\mathrm{T}}\mathbf{y}^{(0)}/\|\mathbf{X}^{(0)\mathrm{T}}\mathbf{y}^{(0)}\|\\&3.\ \mathbf{X}^{(1)}=\mathbf{X}^{(0)}-\mathbf{t}_1\mathbf{w}_1^{\mathrm{T}}\\&4.\ \mathbf{y}^{(1)}=\mathbf{y}^{(0)}-\mathbf{t}_1 b_1;\ b_1=(\mathbf{t}_1^{\mathrm{T}}\mathbf{t}_1)^{-1}\mathbf{t}_1^{\mathrm{T}}\mathbf{y}^{(0)}\\&5.\ \mathbf{t}_2=\mathbf{X}^{(1)}\mathbf{w}_2\\&6.\ \max \operatorname{cov}(\mathbf{t}_2, \mathbf{y}^{(1)})\Rightarrow\mathbf{w}_2=\mathbf{X}^{(1)\mathrm{T}}\mathbf{y}^{(1)}/\|\mathbf{X}^{(1)\mathrm{T}}\mathbf{y}^{(1)}\|\\&7.\ \mathbf{X}^{(2)}=\mathbf{X}^{(1)}-\mathbf{t}_2\mathbf{w}_2^{\mathrm{T}}\\&8.\ \mathbf{y}^{(2)}=\mathbf{y}^{(0)}-\mathbf{T}\mathbf{b}_2;\ \mathbf{T}=[\mathbf{t}_1\quad\mathbf{t}_2];\ \mathbf{b}_2=(\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{y}^{(0)}\\&\ldots\\&\text{until } A \text{ components}\end{aligned} \tag{8}$$

where the extra regression step (calculating the $\mathbf{p}$-vectors) is omitted, which makes the algorithm more transparent. By defining again $\mathbf{T}=[\mathbf{t}_1, \ldots, \mathbf{t}_A]$ and $\mathbf{W}=[\mathbf{w}_1, \ldots, \mathbf{w}_A]$, Martens' PLS1 model after A components is

$$\begin{aligned}&1.\ \mathbf{T}=\mathbf{X}\mathbf{W}\\&2.\ \mathbf{X}=\mathbf{T}\mathbf{W}^{\mathrm{T}}+\mathbf{E}_{\mathbf{x}}\\&3.\ \mathbf{y}=\mathbf{T}\mathbf{b}_A+\mathbf{y}^{(A)}\\&4.\ \max \operatorname{cov}(\mathbf{t}_a, \mathbf{y}^{(a-1)});\ a=1, \ldots, A\end{aligned} \tag{9}$$

where $\mathbf{W}^{\mathrm{T}}\mathbf{W}=\mathbf{I}$ and $\mathbf{T}^{\mathrm{T}}\mathbf{T}$ is tridiagonal. Note that $\mathbf{T}=\mathbf{X}\mathbf{W}$ solves the problem of minimizing $\|\mathbf{X}-\mathbf{T}\mathbf{W}^{\mathrm{T}}\|^2$ for given $\mathbf{W}$ owing to the property $\mathbf{W}^{\mathrm{T}}\mathbf{W}=\mathbf{I}$. Nevertheless, $\mathbf{T}=\mathbf{X}\mathbf{W}$ does not solve the problem of minimizing $\|\mathbf{X}-\mathbf{T}\mathbf{W}^{\mathrm{T}}\|^2$ for general $\mathbf{T}$ and $\mathbf{W}$. Therefore Martens' PLS1 is also a *partial* least squares solution. It can be shown that both Martens' and Wold's PLS1 versions give the same predictions for $\mathbf{y}$.[10]

For the case of a univariate $\mathbf{y}$ $(I\times 1)$ and an $\underline{\mathbf{X}}$ dimensioned $I\times J\times K$, with typical elements y_i and x_{ijk} respectively, Bro generalizes equation (2) to

$$\max_{(\mathbf{w}^J, \mathbf{w}^K)}\left[\sum_{i=1}^{I} t_i y_i \;\middle|\; t_i=\sum_{j=1}^{J}\sum_{k=1}^{K} x_{ijk} w_j^J w_k^K \quad\text{and}\quad \|\mathbf{w}^J\|=\|\mathbf{w}^K\|=1\right] \tag{10}$$

where $\mathbf{w}^J$ $(J\times 1)$, with typical element w_j^J, and $\mathbf{w}^K$ $(K\times 1)$, with typical element w_k^K, are the weights for the second and third modes, respectively and assuming that $\underline{\mathbf{X}}$ (with typical element x_{ijk}) is centered in the direction of I, i.e. $\Sigma_{i=1}^{I} x_{ijk}=0$. These weighing vectors serve exactly the same goal as $\mathbf{w}$ in two-way PLS1. The vectors $\mathbf{w}^J$ and $\mathbf{w}^K$ can be solved in a very elegant way be defining $\mathbf{Z}$ as the matrix

with typical element $z_{jk}=\Sigma_i\, y_i x_{ijk}$ and using the singular value decomposition of $\mathbf{Z}$. It can be shown that $\mathbf{w}^J$ and $\mathbf{w}^K$ are equal to the first left and right singular vectors of $\mathbf{Z}$ respectively.

It is convenient to introduce another notation. Let $\mathbf{X}_i\,(J\times K)$ be the ith slice of $\underline{\mathbf{X}}\,(I\times J\times K)$, corresponding to the ith sample, then $\mathrm{Vec}(\mathbf{X}_i)=(x_{i11}, x_{i21}, x_{i31}, \ldots, x_{iJK})^{\mathrm{T}}$, and define

$$\mathbf{X}=\begin{bmatrix}\mathrm{Vec}(\mathbf{X}_1)^{\mathrm{T}}\\ \cdots\\ \mathrm{Vec}(\mathbf{X}_I)^{T}\end{bmatrix}$$

which is an $I\times JK$ matrix. In this notation, $\mathrm{Vec}(\mathbf{Z})=\mathbf{X}^{\mathrm{T}}\mathbf{y}$.
Problem (10) can be rewritten as

$$\max_{(\mathbf{w}^J,\,\mathbf{w}^K)}\,[\mathbf{t}^{\mathrm{T}}\mathbf{y}\,|\,\mathbf{t}=\mathbf{X}(\mathbf{w}^K\otimes\mathbf{w}^J)\quad\text{and}\quad\|\mathbf{w}^J\|=\|\mathbf{w}^K\|=1] \tag{11}$$

where $\otimes$ is the Kronecker product.

The question arises of whether the vectors $\mathbf{t}$, $\mathbf{w}^J$ and $\mathbf{w}^K$ have least squares properties. Consider the problem

$$\min_{\mathbf{t}}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(x_{ijk}-t_i w_j^J w_k^K)^2\quad\text{given }\mathbf{w}^J\text{ and }\mathbf{w}^K;\ \|\mathbf{w}^J\|=\|\mathbf{w}^K\|=1 \tag{12}$$

Using the Vec and Kronecker notation, this can be rewritten as

$$\min_{\mathbf{t}}\sum_{i=1}^{I}[\mathrm{Vec}(\mathbf{X}_i)-t_i(\mathbf{w}^K\otimes\mathbf{w}^J)]^2\quad\text{given }\mathbf{w}^J\text{ and }\mathbf{w}^K;\ \|\mathbf{w}^J\|=\|\mathbf{w}^K\|=1 \tag{13}$$

or

$$\min_{\mathbf{t}}\|\mathbf{X}-\mathbf{t}(\mathbf{w}^K\otimes\mathbf{w}^J)^{\mathrm{T}}\|^2\quad\text{given }\mathbf{w}^J\text{ and }\mathbf{w}^K;\ \|\mathbf{w}^J\|=\|\mathbf{w}^K\|=1 \tag{14}$$

The least squares solution of (14) is

$$\mathbf{t}=\mathbf{X}\mathbf{w}(\mathbf{w}^{\mathrm{T}}\mathbf{w})^{-1}\quad\text{where }\mathbf{w}=(\mathbf{w}^K\otimes\mathbf{w}^J) \tag{15}$$

Now the following holds:

$$\mathbf{w}^{\mathrm{T}}\mathbf{w}=(\mathbf{w}^K\otimes\mathbf{w}^J)^{\mathrm{T}}(\mathbf{w}^K\otimes\mathbf{w}^J)=[(\mathbf{w}^K)^{\mathrm{T}}\otimes(\mathbf{w}^J)^{\mathrm{T}}](\mathbf{w}^K\otimes\mathbf{w}^J)=[(\mathbf{w}^K)^{\mathrm{T}}(\mathbf{w}^K)\otimes(\mathbf{w}^J)^{\mathrm{T}}\mathbf{w}^J)=1 \tag{16}$$

Therefore

$$\mathbf{t}=\mathbf{X}(\mathbf{w}^K\otimes\mathbf{w}^J)\quad\text{or}\quad t_i=\sum_{j=1}^{J}\sum_{k=1}^{K}x_{ijk}w_j^J w_k^K \tag{17}$$

which shows that indeed, as in the two-way PLS case, the $\mathbf{t}$-vector as defined in (11) solves the least squares problem (12). Hence also multilinear PLS simultaneously gives a $\mathbf{t}$-vector maximally covarying with $\mathbf{y}$ and building a *partial* least squares model for $\underline{\mathbf{X}}$.

In order to show the behavior of multilinear PLS for more than one component, some extra notation is needed. Define $\mathbf{t}_a$ to be the ath score vector of $\underline{\mathbf{X}}$ (a=1, . . . , A); $\mathbf{w}_a^J$ and $\mathbf{w}_a^K$ the weight vectors of the

ath dimension for mode J and mode K, respectively, $\mathbf{w}_a=(\mathbf{w}_a^K \otimes \mathbf{w}_a^J)$; $\mathbf{X}^{(0)}$ and $\mathbf{y}^{(0)}$ the original $\mathbf{X}$ (rearranged as

$$\mathbf{X}=\begin{bmatrix} \mathrm{Vec}(\mathbf{X}_1)^{\mathrm{T}} \\ \cdots \\ \mathrm{Vec}(\mathbf{X}_I)^{T} \end{bmatrix}$$

with $\mathrm{Vec}(\mathbf{X}_i)=(x_{i11}, x_{i21}, x_{i31}, \ldots, x_{iJK})^{\mathrm{T}}$; see above) and original $\mathbf{y}$ respectively; $\mathbf{X}^{(a)}$ and $\mathbf{y}^{(a)}$ the deflated $\mathbf{X}$ and deflated $\mathbf{y}$ after the ath dimension respectively.

The first multilinear PLS step can be written as

$$\begin{aligned} &\mathbf{t}_1=\mathbf{X}^{(0)}\mathbf{w}_1 \\ &\mathbf{X}^{(1)}=\mathbf{X}^{(0)}-\mathbf{t}_1\mathbf{w}_1^{\mathrm{T}} \\ &b_1=(\mathbf{t}_1^{\mathrm{T}}\mathbf{t}_1)^{-1}\mathbf{t}_1^{\mathrm{T}}\mathbf{y}^{(0)} \\ &\mathbf{y}^{(1)}=\mathbf{y}^{(0)}-b_1\mathbf{t}_1 \end{aligned} \tag{18}$$

where b_1 is a scalar. The second step in multilinear PLS becomes

$$\begin{aligned} &\mathbf{t}_2=\mathbf{X}^{(1)}\mathbf{w}_2 \\ &\mathbf{X}^{(2)}=\mathbf{X}^{(1)}-\mathbf{t}_2\mathbf{w}_2^{\mathrm{T}} \\ &\mathbf{b}_2=(\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{y}^{(0)} \\ &\mathbf{y}^{(2)}=\mathbf{y}^{(0)}-\mathbf{T}\mathbf{b}_2 \end{aligned} \tag{19}$$

where $\mathbf{b}_2$ is a 2×1 vector and $\mathbf{T}=[\mathbf{t}_1 \quad \mathbf{t}_2]$. Several comments are appropriate at this point: all these comments are related to the property that the successive $\mathbf{t}$-vectors calculated in the multilinear PLS method are not orthogonal. First, in the second step the $\mathbf{b}$-vector has to be calculated with all previous $\mathbf{t}$-vectors. Secondly, the deflating of $\mathbf{y}$ has to be done with all $\mathbf{t}$-vectors and the appropriate $\mathbf{b}$. Thirdly, the third dimension is calculated analogously to the second one: $\mathbf{t}_3$ is calculated, a new $\mathbf{T}=[\mathbf{t}_1 \quad \mathbf{t}_2 \quad \mathbf{t}_3]$ is formed, and so on. This is exactly equal to the steps in Marten's PLS1.

After A components the multilinear PLS algorithm has generated $\mathbf{T}=[\mathbf{t}_1 \quad \ldots \quad \mathbf{t}_A]$, $\mathbf{b}_A$, $\mathbf{w}_1$ to $\mathbf{w}_A$ and all the deflated $\mathbf{y}$- and $\mathbf{X}$-matrices. Examining equations (18) and (19) more closely, it can be seen that after A components, multilinear PLS can be written formally as

$$\begin{aligned} &1.\ \mathbf{T}=\mathbf{X}\mathbf{V} \\ &2.\ \mathbf{X}=\mathbf{T}\mathbf{W}^{\mathrm{T}}+\mathbf{E}_{\mathbf{X}};\ \mathbf{W}=[\mathbf{w}_1^K \otimes \mathbf{w}_1^J | \ldots | \mathbf{w}_A^K \otimes \mathbf{w}_A^J] \\ &3.\ \mathbf{y}=\mathbf{T}\mathbf{b}_A+\mathbf{y}^{(A)} \\ &4.\ \max \mathrm{cov}(\mathbf{t}_a, \mathbf{y}^{(a-1)});\ a=1, \ldots, A \end{aligned} \tag{20}$$

where $\mathbf{V}$ is a matrix of weighing coefficients which can be written in terms of $\mathbf{w}_1$ to $\mathbf{w}_A$. This will be shown in the next section.

Note that multilinear PLS in the formulation of equation (20) is very similar to Martens' PLS1 version.[9] There is, however, a noticeable difference: whereas Martens' PLS1 gives orthogonal $\mathbf{w}$-vectors ($\mathbf{W}^{\mathrm{T}}\mathbf{W}=\mathbf{I}$), Bro's multilinear PLS does not have this property.[11] Therefore, in multilinear PLS, $\mathbf{T}=\mathbf{X}\mathbf{W}$ is not the solution of the problem of minimizing $\|\mathbf{X}-\mathbf{T}\mathbf{W}^{\mathrm{T}}\|^2$ for fixed $\mathbf{W}$. Multilinear PLS generates non-orthogonal $\mathbf{t}_a$-vectors, non-orthognal $\mathbf{w}_a$-vectors and componentwise least squares solutions for fixed $\mathbf{w}_a$. The model for $\mathbf{X}$ (see the second line in (20)) is a trilinear model, hence the name multilinear PLS.

In Bro's paper, generalizations to four and higher orders are given. For a fourth-order array $\underline{\mathbf{X}}$ ($I\times J\times K\times L$) a formula similar to equation (10) with an extra weight vector $\mathbf{w}^L$ results in an analogous formulation of the problem as in equation (11), after properly rearranging $\underline{\mathbf{X}}$ using the Vec notation. It can be shown along the same lines as above that for the fourth-order array a $\mathbf{t}$-vector is

found that maximally covaries with $\mathbf{y}$. Moreover, $\mathbf{t}$, $\mathbf{w}^J$, $\mathbf{w}^K$ and $\mathbf{w}^L$ are a (*partial*) least squares approximation of $\underline{\mathbf{X}}$. For proving this *partial* least squares property, an equation similar to (16) is found via the multiplication and transposition properties of Kronecker products. This can easily be extended to higher orders. Moreover, for generalizations to $\underline{\mathbf{X}}$-arrays of order four and higher, multilinear PLS can be formulated equivalent to (20) and the comments following that equation are also valid for this case.

Summarizing, the general scheme of Wold's PLS1 Martens' PLS1 and Bro's multilinear PLS1 is

$$\begin{aligned}
&1.\ \mathbf{T}=\mathbf{XV}\\
&2.\ \mathbf{X}=\mathbf{TP}^{\mathrm{T}}+\mathbf{E}_{\mathbf{X}}\\
&3.\ \mathbf{y}=\mathbf{Tb}+\mathbf{e}_{\mathbf{y}}\\
&4.\ \max \operatorname{cov}(\mathbf{t}_a, \mathbf{y}^{(a-1)}) \text{ with } \mathbf{t}_a=\mathbf{X}^{(a-1)}\mathbf{w}_a \text{ and } \|\mathbf{w}_a\|=1;\ a=1, \ldots, A
\end{aligned} \tag{21}$$

where, depending on the nature of $\mathbf{X}$ (two-way or rearranged multiway) and restrictions put on $\mathbf{T}$, $\mathbf{V}$, $\mathbf{P}$ and $\mathbf{W}=[\mathbf{w}_1, \ldots, \mathbf{w}_A]$, the result is one of the three above-mentioned models. All three PLS1 models clearly work in a componentwise fashion, as exemplified by step 4 in (21).

MULTILINEAR PLS: PREDICTION EQUATIONS

The prediction equations can be derived as follows. Suppose a new sample is available, denoted by $\mathbf{X}_{I+1}^{(0)}$ $(J\times K)$. Then the full prediction becomes

$$\begin{aligned}
&\mathbf{t}_{I+1,\,1}=\mathrm{Vec}(\mathbf{X}_{I+1}^{(0)})^{\mathrm{T}}\mathbf{w}_1\\
&\mathrm{Vec}(\mathbf{X}_{I+1}^{(1)})=\mathrm{Vec}(\mathbf{X}_{I+1}^{(0)})-t_{I+1,\,1}\mathbf{w}_1\\
&t_{I+1,\,2}=\mathrm{Vec}(\mathbf{X}_{I+1}^{(1)})^{T}\mathbf{w}_2\\
&\cdots\\
&t_{I+1,\,A}=\mathrm{Vec}(\mathbf{X}_{I+1}^{(A-1)})^{\mathrm{T}}\mathbf{w}_A\\
&\mathrm{Vec}(\mathbf{X}_{I+1}^{(A)})=\mathrm{Vec}(\mathbf{X}_{I+1}^{(A-1)})-t_{I+1,\,A}\mathbf{w}_A=\mathrm{Vec}(\mathbf{E}_{I+1})\\
&\mathbf{t}_{I+1}=(t_{I+1,\,1}, \ldots, t_{I+1,\,A})^{\mathrm{T}}\\
&\hat{y}_{I+1}=\mathbf{t}_{I+1}^{\mathrm{T}}\mathbf{b}_A
\end{aligned} \tag{22}$$

and it is clear that with this full prediction scheme the residual matrix $\mathbf{E}_{I+1}$ $(J\times K)$ is available for checking whether the new sample is typical or not.

The closed form of the prediction equations can be derived as follows. The general idea is to get a direct relationship between $\mathbf{X}^{(0)}$ and $\hat{\mathbf{y}}$:

$$\hat{\mathbf{y}}=\mathbf{Tb}_A=\mathbf{X}^{(0)}\mathbf{b}_{\mathrm{PLS}} \tag{23}$$

Using the deflation equations for $\mathbf{X}$ (see (18) and (19)), it holds that

$$\begin{aligned}
&\mathbf{t}_1=\mathbf{X}^{(0)}\mathbf{w}_1\\
&\mathbf{t}_2=\mathbf{X}^{(1)}\mathbf{w}_2=(\mathbf{X}^{(0)}-\mathbf{t}_1\mathbf{w}_1^{\mathrm{T}})\mathbf{w}_2=(\mathbf{X}^{(0)}-\mathbf{X}^{(0)}\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}})\mathbf{w}_2=\mathbf{X}^{(0)}(\mathbf{I}-\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}})\mathbf{w}_2\\
&\cdots\\
&\mathbf{t}_A=\mathbf{X}^{(0)}(\mathbf{I}-\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}})\ldots(\mathbf{I}-\mathbf{w}_{A-1}\mathbf{w}_{A-1}^{\mathrm{T}})\mathbf{w}_A
\end{aligned} \tag{24}$$

and using $\mathbf{T}=[\mathbf{t}_1, \ldots, \mathbf{t}_A]$, it holds that

$$\mathbf{T}=\mathbf{X}^{(0)}[\mathbf{w}_1\,|\,(\mathbf{I}-\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}})\mathbf{w}_2\,|\ldots|(\mathbf{I}-\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}})(\mathbf{I}-\mathbf{w}_2\mathbf{w}_2^{\mathrm{T}})\ldots(\mathbf{I}-\mathbf{w}_{A-1}\mathbf{w}_{A-1}^{\mathrm{T}})\mathbf{w}_A]=\mathbf{X}^{(0)}\mathbf{V} \tag{25}$$

Then

$$\mathbf{Tb}_A=\mathbf{X}^{(0)}[\mathbf{w}_1\,|\,(\mathbf{I}-\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}})\mathbf{w}_2\,|\ldots|(\mathbf{I}-\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}})(\mathbf{I}-\mathbf{w}_2\mathbf{w}_2^{\mathrm{T}})\ldots(\mathbf{I}-\mathbf{w}_{A-1}\mathbf{w}_{A-1}^{\mathrm{T}})\mathbf{w}_A]\mathbf{b}_A \tag{26}$$

Combining (26) and (23) gives

$$\mathbf{b}_{\mathrm{PLS}}=[\mathbf{w}_1|(\mathbf{I}-\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}})\mathbf{w}_2|\ldots|(\mathbf{I}-\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}})(\mathbf{I}-\mathbf{w}_2\mathbf{w}_2^{\mathrm{T}})\ldots(\mathbf{I}-\mathbf{w}_{A-1}\mathbf{w}_{A-1}^{\mathrm{T}})\mathbf{w}_A]\mathbf{b}_A \tag{27}$$

and $\mathbf{b}_{\mathrm{PLS}}$ can be used for predicting a new sample using

$$\hat{y}_{I+1}=\mathrm{Vec}(\mathbf{X}_{I+1}^{(0)})^{\mathrm{T}}\mathbf{b}_{\mathrm{PLS}} \tag{28}$$

If $\mathbf{w}_i^{\mathrm{T}}\mathbf{w}_j=0$ $(i\neq j)$, then (27) reduces to

$$\mathbf{b}_{\mathrm{PLS}}=[\mathbf{w}_1 \quad \mathbf{w}_2 \quad \ldots \quad \mathbf{w}_A]\mathbf{b}_A=\mathbf{W}\mathbf{b}_A \tag{29}$$

This is, however, not the case for multilinear PLS. Note that $\mathbf{w}_i^{\mathrm{T}}\mathbf{w}_j=0$ $(i\neq j)$ holds for Martens' PLS algorithm[9] and (29) indeed coincides with the closed prediction equation for that algorithm.

GENERAL FRAMEWORK FOR TWO-BLOCK MULTIWAY METHODS

In this section a framework is given for two-block multiway methods. Such methods try to derive a relationship between two blocks of data $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$ with at least one of them being multiway. These two blocks have at least one mode in common: the object mode. That is, the same objects play a role in both blocks of data. Therefore it is reasonable to build a relationship between the blocks. For the two-block two-way case this is very often done with PLS models. For a more general description of different multiblock multiway situations see Reference 5.

There are different ways to build models describing the relationship between two blocks of data. This can be done in a componentwise or a simultaneous fashion. Traditionally, two-way PLS models are built componentwise. There are, however, models available that build such relations simultaneously, e.g. principal covariate regression.[12] For describing the framework of two-block multiway methods, it is useful to make a distinction between componentwise solutions and simultaneous solutions. This section starts with a brief description of the multiway models considered. Then the generalizations to componentwise and simultaneous solutions are given. For convenience and without loss of generality it is assumed that $\underline{\mathbf{X}}$ is a properly centered three-way array and $\mathbf{y}$ is a centered vector which has to be predicted using $\underline{\mathbf{X}}$.

Multiway models

Define Define $\mathbf{A}=[\mathbf{a}_1, \ldots, \mathbf{a}_R]$, $\mathbf{B}=[\mathbf{b}_1, \ldots, \mathbf{b}_R]$ and $\mathbf{C}=[\mathbf{c}_1, \ldots, \mathbf{c}_R]$, where $\mathbf{a}_r$ $(r=1, \ldots, R)$ are $I\times 1$ vectors, $\mathbf{b}_r$ $(r=1, \ldots, R)$ are $J\times 1$ vectors and $\mathbf{c}_r$ $(r=1, \ldots, R)$ are $K\times 1$ vectors. Define $\mathbf{X}_k$ $(I\times J)$ as the kth slice of $\underline{\mathbf{X}}$ $(I\times J\times K)$ and $\mathbf{X}=[\mathbf{X}_1|\ldots|\mathbf{X}_K]$ $(I\times JK)$, which is the same as before. Then PARAFAC with R components models $\underline{\mathbf{X}}$ $(I\times J\times K)$ with $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ by solving the problem

$$\min_{(\mathbf{A},\mathbf{B},\mathbf{C})} \|\mathbf{X}-\mathbf{A}\mathbf{D}(\mathbf{I}\otimes\mathbf{B}^{\mathrm{T}})\|^2 \quad \text{where } \mathbf{D}=[\mathbf{D}_1|\ldots|\mathbf{D}_K] \tag{30}$$

and $\mathbf{D}_k$ is diagonal with diag$(\mathbf{D}_k)=k$th row of $\mathbf{C}$

When $R=1$, equation (30) reduces to

$$\min_{(\mathbf{a},\mathbf{b},\mathbf{c})} \|\mathbf{X}-\mathbf{a}(\mathbf{b}\otimes\mathbf{c})^{\mathrm{T}}\|^2 \tag{31}$$

which is very similar to equation (14), except that in (31) all loading vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ can be chosen freely. In other words, (14) is a restricted version of (31) and therefore the sum of squared residuals

of (14) is never smaller than the sum of squared residuals of (31). This is a restatement of the *partial* least squares property of Bro's multilinear PLS.

The Tucker3 model of $\underline{\mathbf{X}}$ ($I\times J\times K$) is found by solving the problem

$$\min_{(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{G})} \|\mathbf{X}-\mathbf{AG}(\mathbf{C}^{\mathrm{T}}\otimes\mathbf{B}^{\mathrm{T}})\|^2 \tag{32}$$

where $\mathbf{A}$ ($I\times R_1$), $\mathbf{B}$ ($J\times R_2$) and $\mathbf{C}$ ($K\times R_3$) are loading matrices, $\mathbf{G}$ is the properly concatenated three-way core matrix $\underline{\mathbf{G}}$ ($R_1\times R_2\times R_3$) and $\mathbf{X}$ is as before.[13] Note that $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are not the same as for PARAFAC; not only has the Tucker3 model a different number of factors in the three directions, but even if R_1=R_2=R_3, the PARAFAC and Tucker3 solutions differ. This is due to the core array in the Tucker3 model, which makes the Tucker3 model more general than the PARAFAC model. In fact, if the core array is chosen to have only non-zeros on the superdiagonal (assuming that R_1=R_2=R_3), then Tucker3 equals PARAFAC.

The Tucker2 model can be formalized as

$$\min_{(\mathbf{A},\mathbf{B},\mathbf{G})} \|\mathbf{X}-\mathbf{AG}(\mathbf{I}\otimes\mathbf{B}^{\mathrm{T}})\|^2 \tag{33}$$

where $\mathbf{X}$ is as before, $\mathbf{I}$ is the $K\times K$ identity matrix, $\mathbf{G}$ ($R_1\times KR_2$) is the properly concatenated extended core array $\underline{\mathbf{G}}$ ($R_1\times R_2\times K$) and $\mathbf{A}$ ($I\times R_1$) and $\mathbf{B}$ ($J\times R_2$) are loading matrices of the first and second modes respectively.[13] Note that the Tucker2 model is a special case of Tucker3 with R_3=K, which means that in the third mode no reduction to a small number of components is sought. Likewise, there are Tucker2 models corresponding to R_1=I and R_2=J.

The Tucker1 model can formalized as

$$\min_{(\mathbf{A},\mathbf{G})} \|\mathbf{X}-\mathbf{AG}\|^2 \tag{34}$$

where $\mathbf{X}$ is the same as before, $\mathbf{G}$ is an $R\times JK$ matrix and $\mathbf{A}$ ($I\times R$) is the loading matrix of the first mode. Note that there are three different formulations of the Tucker1 model, depending on which mode is being reduced.

The Tucker1 model is known in the chemometrics literature as 'multiway PCA'.[14] The popularity of this method is mainly due to the easy way to solve problem (34): unfold $\underline{\mathbf{X}}$ ($I\times J\times K$) to $\mathbf{X}$ ($I\times JK$) and use standard SVD routines. It is clear that there exist many more multiway PCA models than the Tucker1 model.

Summarizing, all multiway models for a three-way array $\underline{\mathbf{X}}$ can be written as

$$\mathbf{X}=\mathbf{AW}+\mathbf{E} \tag{35}$$

where $\mathbf{X}$ is the same as before, $\mathbf{E}$ is an error matrix of the same size as $\mathbf{X}$ of which the Frobenius norm is minimized, $\mathbf{A}$ is an $I\times R_1$ loading matrix of the first mode and the structure of $\mathbf{W}$ ($R_1\times JK$) depends on the specific model which is superimposed on $\underline{\mathbf{X}}$. In the Tucker models, $\mathbf{A}$ can be chosen to be column-orthogonal without loss of generality. This is due to the rotational freedom in those models. For PARAFAC models this is not the case: there is no rotational freedom in a PARAFAC model and hence constraining $\mathbf{A}$ to be column-orthogonal generates another model. Equation (35) is also mentioned by Kiers.[15]

Componentwise two-block multiway models

The general scheme of a componentwise solution is as follows.

1. Calculate component 1.

2. Deflate $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$.
3. Calculate component 2.
4. Etc.

This holds both for two-way data and for multiway data. The two-block componentwise Tucker1 model can be formalized as

$$\begin{aligned}&1.\ \mathbf{T}=\mathbf{XV}\\&2.\ \mathbf{X}=\mathbf{TP}^{\mathrm{T}}+\mathbf{E}_{\mathbf{X}}\\&3.\ \mathbf{y}=\mathbf{Tb}+\mathbf{y}^{(A)}\\&4.\ \max \operatorname{cov}(\mathbf{t}_a, \mathbf{y}^{(a-1)}) \text{ with } \mathbf{t}_a=\mathbf{X}^{(a-1)}\mathbf{w}_a \text{ and } \|\mathbf{w}_a\|=1;\ a=1, \ldots, A\end{aligned} \tag{36}$$

where $\mathbf{X}$ ($I\times JK$) is the properly rearranged three-way array $\underline{\mathbf{X}}$ and the matrix $\mathbf{T}$ ($I\times A$) has columns $\mathbf{t}_a$ (a=1, . . . , A) which are chosen in such a way that componentwise the covariance between $\mathbf{t}_a$ and the unexplained part of $\mathbf{y}$ is maximized (which is symbolized in line 4 of equation (36)). The residuals in $\mathbf{X}$ ($\mathbf{E}_{\mathbf{X}}$) and $\mathbf{y}$ ($\mathbf{y}^{(a)}$) are *partially* minimized in the sense described previously. Clearly, the model in (36) can be estimated with Wold's PLS1 if the restriction of orthogonal $\mathbf{t}$s is imposed (Wold's multiway PLS[14]). If this restriction is not imposed, then Martens' PLS1 algorithm can be used.

The two-block componentwise PARAFAC model can be formalized as

$$\begin{aligned}&1.\ \mathbf{T}=\mathbf{XV} \text{ with } \mathbf{V}=[\mathbf{v}_1, \ldots, \mathbf{v}_A]\\&2.\ \mathbf{X}=\mathbf{TP}^{\mathrm{T}}+\mathbf{E}_{\mathbf{X}} \text{ with } \mathbf{P}=[\mathbf{p}_1, \ldots, \mathbf{p}_A] \text{ and } \mathbf{p}_a=\mathbf{p}_a^K \otimes \mathbf{p}_a^J\\&3.\ \mathbf{y}=\mathbf{Tb}+\mathbf{y}^{(A)}\\&4.\ \max \operatorname{cov}(\mathbf{t}_a, \mathbf{y}^{(a-1)}) \text{ with } \mathbf{t}_a=\mathbf{X}^{(a-1)}\mathbf{w}_a \text{ and } \|\mathbf{w}_a\|=1;\ a=1, \ldots, A\end{aligned} \tag{37}$$

where $\mathbf{X}$ ($I\times JK$) is the properly rearranged three-way array $\underline{\mathbf{X}}$ and the matrix $\mathbf{T}$ ($I\times A$) has columns $\mathbf{t}_a$ (a=1, . . . , A) where the demand on $\mathbf{P}$ in line 2 of equation (37) assures that $\underline{\mathbf{X}}$ is fitted with a trilinear (PARAFAC-type) model. Problem (37) can be solved with Bro's multilinear PLS1.

In order to obtain a trilinear (*partial*) least squares model for $\underline{\mathbf{X}}$, only the special structure of $\mathbf{P}$ (line 2 of equation (37)) has to be demanded. This points to the existence of alternatives for Bro's multilinear PLS1, depending on which restrictions are imposed. To point out some of these alternatives, consider forcing the $\mathbf{t}$-vectors to be orthogonal; then another algorithm is needed and another model is estimated. Does this give the same results in terms of predicted $\mathbf{y}$-values as multilinear PLS (which was the case for Wold's and Martens' algorithms for two-way PLS)?

For two-block Tucker3 and Tucker 2 models no componentwise solutions are possible. To explain this, consider a Tucker3 model with R_1=2, R_2=3 and R_3=4. What is 'a component' in this context? Even if R_1=R_2=R_3=R, problems occur. Suppose R=2 and consider the (1, 2, 1)th element in the core array of this specific Tucker3 model. This element estimates the interaction between the first component in the first mode, the second component in the second mode and the first component in the third mode. How should this be done componentwise? No clear solutions are available.

Note that equations (36) and (37) can be extended for multivariate $\mathbf{Y}$ and even multiway $\underline{\mathbf{Y}}$. Hence componentwise Tucker1 and PARAFAC models can be built for the general problem of connecting two blocks $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$ of any order.

Simultaneous two-block multiway models

Principal covariate regression (PCovR) provides a framework for discussing simultaneous two-block multiway models and this will be taken as an example. The general formulation of PCovR for two-way data (again for simplicity a univariate $\mathbf{y}$ is taken) is

$$\begin{aligned}&1.\ \mathbf{T}=\mathbf{XW}\\&2.\ \mathbf{X}=\mathbf{TP}^{\mathrm{T}}+\mathbf{E}_{\mathbf{X}}\\&3.\ \mathbf{y}=\mathbf{Tb}+\mathbf{e}_{\mathbf{y}}\\&4.\ \min[\alpha\|\mathbf{X}-\mathbf{XWP}^{\mathrm{T}}\|^2+(1-\alpha)\|\mathbf{y}-\mathbf{XWb}\|^2]\end{aligned}\qquad(38)$$

where $\mathbf{X}$ ($I\times J$) is a (centered) matrix, $\mathbf{y}$ ($I\times 1$) is a (centered) vector, $\mathbf{E}_{\mathbf{X}}$ and $\mathbf{e}_{\mathbf{y}}$ are the residuals in $\mathbf{X}$ and $\mathbf{y}$ respectively and α is a fixed number between zero and one accounting for the weight given on modeling $\mathbf{X}$ or $\mathbf{y}$ in (38). Note that the first term in the minimization expression of line 4 in equation (38) corresponds to the unexplained part of $\mathbf{X}$ and the second term corresponds to the unexplained part of $\mathbf{y}$. PCovR balances the explanation of $\mathbf{X}$ and $\mathbf{y}$ by the parameter α. As explained by de Jong and Kiers,[12] α can be selected by cross-validation. Note that (38) is easily generalized to multivariate $\mathbf{Y}$.

The general two-block three-way PCovR problem is

$$\begin{aligned}&1.\ \mathbf{T}=\mathbf{XW}\\&2.\ \mathbf{X}=\mathbf{TP}^{\mathrm{T}}+\mathbf{E}_{\mathbf{X}}\\&3.\ \mathbf{y}=\mathbf{Tb}+\mathbf{e}_{\mathbf{y}}\\&4.\ \min[\alpha\|\mathbf{X}-\mathbf{XWP}^{\mathrm{T}}\|^2+(1-\alpha)\|\mathbf{y}-\mathbf{XWb}\|^2]\end{aligned}\qquad(39)$$

where $\mathbf{X}$ ($I\times JK$) is the properly rearranged $\underline{\mathbf{X}}$ ($I\times J\times K$) as in equation (35) and all other matrices/vectors are defined as before. Depending on the structure imposed on $\mathbf{P}$, the following models are obtained.

1. Simultaneous two-block Tucker3: $\mathbf{P}^{\mathrm{T}}=\mathbf{G}(\mathbf{C}^{\mathrm{T}}\otimes\mathbf{B}^{\mathrm{T}})$.
2. Simultaneous two-block Tucker2: $\mathbf{P}^{\mathrm{T}}=\mathbf{G}(\mathbf{I}\otimes\mathbf{B}^{\mathrm{T}})$.
3. Simultaneous two-block Tucker1: $\mathbf{P}$ is general.
4. Simultaneous two-block PARAFAC: $\mathbf{P}^{\mathrm{T}}=\mathbf{D}(\mathbf{I}\otimes\mathbf{B}^{\mathrm{T}})$, with $\mathbf{D}$ as in equation (30).

In the original PCovR formulation, $\mathbf{T}$ was restricted to have orthogonal columns. For the simultaneous two-block Tucker models the restriction of column orthogonality of $\mathbf{T}$ can be imposed without loss of generality because of the rotational freedom in the Tucker1, Tucker2 and Tucker3 models. For the simultaneous two-block PARAFAC model, imposing a constraint on $\mathbf{T}$ changes the solution.

Simultaneous two-block Tucker1 models can be solved easily by unfolding $\underline{\mathbf{X}}$ properly and using PCovR on $\mathbf{y}$ and the unfolded $\underline{\mathbf{X}}$. For all other simultaneous two-block models mentioned, no algorithms exist yet.

Several remarks are appropriate. First, the different versions of simultaneous two-block multiway models are obtained by putting a special structure on $\mathbf{P}$. It is not known what the consequences of this imposed structure are for the properties of $\mathbf{T}$ and $\mathbf{W}$. Secondly, principal covariates regression gives a framework for simultaneous two-block multiway models, but there are others, e.g. joint continuum regression.[16] Thirdly, the generic problem as stated in (39) can be generalized to multivariate $\mathbf{Y}$, multiway $\underline{\mathbf{Y}}$ (of any order) and multiway $\underline{\mathbf{X}}$ (of order higher than three), which opens up possibilities to have hybrid forms, e.g. imposing a Tucker3 structure on $\underline{\mathbf{X}}$ and a PARAFAC structure on $\underline{\mathbf{Y}}$. Fourthly, there are still other simultaneous two-block multiway models possible, e.g. those based on restricted or constrained Tucker models,[17,18] but these can be treated as special cases of the Tucker3-type models.

Clearly, developing theory and algorithms to solve the problem of equation (39) with proper constraints on $\mathbf{T}$, $\mathbf{P}$ and $\mathbf{W}$ is a challenge in the area of multiway data analysis.

ACKNOWLEDGEMENTS

Thanks go to Rasmus Bro, Henk Kiers and Sijmen de Jong for careful reading of the manuscript, pointing out some errors and making suggestions that improved the manuscript.

REFERENCES

1. S. Wold, A. Ruhe, H. Wold and W. J. Dunn III, 'The collinearity problem in linear regression, the partial least squares (PLS) appoach to generalized inverses', *SIAM J. Sci. Stat. Comput.* **5**, 735–743 (1984).
2. A. Höskuldsson, 'PLS regression methods', *J. Chemometrics* **2**, 211–228 (1988).
3. S. de Jong, 'SIMPLS: an alternative approach to partial least squares regression', *Chemometrics Intell. Lab. Syst.* **18**, 251–263 (1993).
4. P. Geladi, 'Analysis of multi-way (multi-mode) data', *Chemometrics Intell. Lab. Syst.* **7**, 11–30 (1989).
5. A. K. Smilde, 'Three-way analysis: problems and prospects', *Chemometrics Intell. Lab. Syst.* **15**, 143–157 (1992).
6. R. Bro, 'Multiway calibration. Multilinear PLS', *J. Chemometrics* **10**, 47–61 (1996).
7. R. A. Harshman, 'Foundations of the PARAFAC procedure: models and conditions for an exploratory multimodel factor analysis', *UCLA Working Papers on Phonetics*, **16**, 1–84 (1970).
8. L. Tucker, 'Implications of factor analysis of three-way matrices for measurement of change', *Problems of Measuring Change*, ed. by C. Harris, pp. 122–137, University of Wisconsin Press, Madison, WI (1963).
9. H. Martens and T. Naes, *Multivariate Calibration*, Wiley, Chichester (1989).
10. I. S. Helland, 'On the structure of partial least squares regression', *Commun. Stat. B: Simul.* **17**, 581–607 (1988).
11. J. Nilsson, H. Wikström and A. K. Smilde, 'Multiway calibration in 3D QSAR. Part 1: Calibration and validation', *Internal Report*, University Centre for Pharmacy, University of Groningen (1996).
12. S. de Jong and H. A. L. Kiers, 'Principal covariates regression. Part I. Theory', *Chemometrics Intell. Lab. Syst.* **14**, 155–164 (1992).
13. P. M. Kroonenberg, *Three-Mode Principal Component Analysis*, DSWO Press, Leiden (1983).
14. S. Wold, P. Geladi, K. Esbensen and J. Ohman, 'Multi-way principal components and PLS-analysis', *J. Chemometrics* **1**, 41–56 (1987).
15. H. A. L. Kiers, 'Hierarchical relations among three-way methods', *Psychometrika*, **56**, 449–470 (1991).
16. R. Brooks and M. Stone, 'Joint continuum regression for multiple predictands', *J. Am. Stat. Assoc.* **89**, 1374–1377 (1994).
17. H. A. L. Kiers, 'Tuckals core rotations and constrained Tuckals modeling', *Stat. Appl.* **4**, 659–667 (1992).
18. A. K. Smilde, R. Tauler, J. M. Henshaw, L. W. Burgess and B. R. Kowalski, 'Multicomponent determination of chlorinated hydrocarbons using a reaction-based chemical sensor. 2. Medium-rank second order calibration with restricted Tucker models', *Anal. Chem.* **66**, 3345–3351 (1994).

