Available online at www.sciencedirect.com
ScienceDirect
Chemometrics and Intelligent Laboratory Systems 85 (2007) 40–46
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

An adapted version of the element-wise weighted total least squares method for applications in chemometrics

M. Schuermans *, I. Markovsky, S. Van Huffel

KULeuven, ESAT-SCD, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium

Received 26 January 2006; received in revised form 31 March 2006; accepted 6 April 2006
Available online 27 June 2006

Abstract

The Maximum Likelihood PCA (MLPCA) method has been devised in chemometrics as a generalization of the well-known PCA method in order to derive consistent estimators in the presence of errors with known error distribution. For similar reasons, the Total Least Squares (TLS) method has been generalized in the field of computational mathematics and engineering to maintain consistency of the parameter estimates in linear models with measurement errors of known distribution. In a previous paper [M. Schuermans, I. Markovsky, P.D. Wentzell, S. Van Huffel, On the equivalance between total least squares and maximum likelihood PCA, Anal. Chim. Acta, 544 (2005), 254–267], the tight equivalences between MLPCA and Element-wise Weighted TLS (EW-TLS) have been explored. The purpose of this paper is to adapt the EW-TLS method in order to make it useful for problems in chemometrics. We will present a computationally efficient algorithm and compare this algorithm with the standard EW-TLS algorithm and the MLPCA algorithm in computation time and convergence behaviour on chemical data.
© 2006 Elsevier B.V. All rights reserved.

Keywords: EW-TLS; MLPCA; Rank reduction; Measurement errors

1. Introduction

This paper is an extension of paper [1]. In Ref. [1], it was shown that the Maximum Likelihood PCA (MLPCA) [2,3] method and the Element-wise Weighted Total Least Squares (EW-TLS) [4,5] method can be reduced to the same mathematical problem, i.e. finding the closest (in a certain sense) weighted low rank matrix approximation where the weight is derived from the distribution of the measurement errors in the given data. We will not repeat all the details here, but, in order to understand the rest of the paper, we will describe shortly this weighted low rank approximation problem to be solved. Mathematically, we will consider the following weighted low rank matrix approximation problem:

$$\min_{\hat{D}} \|D-\hat{D}\|_W \quad \text{s.t. } \operatorname{rank}(\hat{D}) \le r, \tag{1}$$

with $D \in \mathbb{R}^{m\times n}$, the noisy data matrix, $\operatorname{rank}(D)=k$, $r<k$, $\Delta\hat{D} = D-\hat{D}$ the estimated measurement noise, W the covariance matrix of $\operatorname{vec}(\Delta\hat{D})$ where $\operatorname{vec}(\Delta\hat{D})$ stands for the vectorized form of $\Delta\hat{D}$, i.e., a vector constructed by stacking the consecutive columns of $\Delta\hat{D}$ in one vector and $\|\cdot\|_W = \operatorname{vec}^\top(\cdot) W^{-1} \operatorname{vec}(\cdot)$. When the measurement noise is independently and identically distributed (i.i.d.), $W=I$, where I is the identity matrix, and the optimal closeness norm is the Frobenius norm, $\|\cdot\|_F$. This is used in the well-known TLS and PCA methods. Nevertheless, when the measurement errors are not i.i.d. the Frobenius norm is no longer optimal and a weighted norm is needed instead.

In the MLPCA approach, the rank constraint $\operatorname{rank}(\hat{D}) \le r$ is represented as

$$\hat{D} = TP^\top,$$

with $T \in \mathbb{R}^{m\times r}$ and $P \in \mathbb{R}^{n\times r}$. So, problem (1) can be rewritten as follows:

$$\min_{T}\left(\min_{P,\hat{D}} \operatorname{vec}^\top(D-\hat{D}) W^{-1} \operatorname{vec}(D-\hat{D}) \quad \text{s.t. } \hat{D} = TP^\top\right).$$

In the standard EW-TLS approach, the rank constraint is forced by rewriting $\operatorname{rank}(\hat{D}) \le r$ as

$$\hat{D}\begin{bmatrix} \hat{B} \\ -I_{n-r} \end{bmatrix} = 0, \tag{3}$$

* Corresponding author.
E-mail address: mieke.schuermans@esat.kuleuven.ac.be (M. Schuermans).

0169-7439/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2006.04.003

where $\hat{B} \in \mathbb{R}^{r\times(n-r)}$. Moreover, the weighting matrix W is assumed to be block diagonal

$$W = \begin{bmatrix} W_1 & & \\ & \ddots & \\ & & W_m \end{bmatrix},$$

where each block W_i is the covariance matrix of the errors in the i-th row of the data matrix D. So, for the EW-TLS approach, problem (1) can be rewritten as

$$\min_{\hat{B}} \left(\min_{\hat{D}} \sum_{i=1}^{m} (d_i - \hat{d}_i) W_i^{-1} (d_i - \hat{d}_i)^\top \quad \text{s.t. } \hat{D} \begin{bmatrix} \hat{B} \\ -I_{n-r} \end{bmatrix} = 0 \right), \quad (4)$$

with $d_i, \hat{d}_i \in \mathbb{R}^n$ the i-th row of D and $\hat{D}$, respectively, and W_i the i-th weighting matrix defined as the covariance matrix of the errors in d_i. Algorithms 3 and 5, described in Ref. [1], were designed to solve the standard EW-TLS problem (4) for the case when $m \geq n$ and when the measurement errors are only row-wise correlated. In chemometrics, however, the data matrix usually has size $m \times n$ with $m \leq n$, e.g., in problems of mixture analysis, curve resolution and data fusion. When the measurement errors are uncorrelated or column-wise correlated, the algorithms presented in Ref. [1], can still be applied to the transposed data matrix. For other cases of measurement error correlation, the algorithms need to be optimized by considering the left kernel of $\hat{D}$, i.e., the following modification of Eq. (3) should be used:

$$[\hat{B}_2^\top - I_{m-r}]\hat{D} = 0, \quad (5)$$

where $\hat{B}_2 \in \mathbb{R}^{r\times(m-r)}$. In Section 4 of the previous paper [1], we have shown through simulations that the EW-TLS method certainly has potential for problems when the data matrix has size $m \times n$ with $m \geq n$ and only row-wise correlated measurement errors. In that section, we have also shown that the standard EW-TLS approach is not the right method of choice for the case when $m \leq n$ and only row-wise correlated measurements and we have pointed out that the EW-TLS approach needed to be adapted for handling this case of row-wise correlated measurement errors in data sets where $m \leq n$. In this paper, an algorithm will be derived to solve the following adapted version of the EW-TLS problem:

$$\min_{\hat{B}_2} \left(\min_{\hat{D}} \sum_{i=1}^{m} (d_i - \hat{d}_i) W_i^{-1} (d_i - \hat{d}_i)^\top \quad \text{s.t. } [\hat{B}_2^T - I_{m-r}]\hat{D} = 0 \right), \quad (6)$$

with $d_i, \hat{d}_i \in \mathbb{R}^n$ the i-th row of D and $\hat{D}$, respectively, and W_i the i-th weighting matrix defined as the covariance matrix of the errors in the i-th row of the data matrix $D \in \mathbb{R}^{m\times n}$, with $m \leq n$ and only row-wise correlated measurement errors. The measurement errors among the columns are uncorrelated.

The paper is organized as follows. In Section 2, we will re-derive the standard EW-TLS problem in a different way than is usually done in the literature. In a symmetric way, a solution for the adapted EW-TLS problem, with modified constraint (5), will be derived. An algorithm to solve the adapted EW-TLS problem (6) will be presented in Section 3. In Section 4, we will compare the computation times of both EW-TLS algorithms, the standard and the adapted one, and the MLPCA algorithm on simulated chemical data and discuss their convergence behaviour. Conclusions are made in Section 5.

2. Derivation of the adapted EW-TLS problem

2.1. The standard EW-TLS problem

For a given noisy data matrix $D \in \mathbb{R}^{m\times n}$, with $m \geq n$, with only row-wise correlated measurement errors and given row error covariance matrices $W_i \in \mathbb{R}^{n\times n}$, for $i = 1, \ldots, m$, the standard EW-TLS problem can be formulated as follows:

$$\min_{\hat{D}\in\mathbb{R}^{m\times n}} \operatorname{vec}^\top (D^\top - \hat{D}^\top) W^{-1} \operatorname{vec}(D^\top - \hat{D}^\top) \quad \text{s.t. } \hat{D} \begin{bmatrix} \hat{B} \\ -I_{n-r} \end{bmatrix} = 0, \quad (7)$$

where the weighting matrix W is block diagonal, because the measurements are uncorrelated among the columns:

$$W = \begin{bmatrix} W_1 & & \\ & \ddots & \\ & & W_m \end{bmatrix}. \quad (8)$$

By defining $R := [\hat{B}^\top - I_{n-r}] \in \mathbb{R}^{(n-r)\times n}$, the rank constraint $\hat{D} \begin{bmatrix} \hat{B} \\ -I_{n-r} \end{bmatrix} = 0$ in problem (7) can be written as $\hat{D}R^\top = 0$ or $R\hat{D}^\top = 0$. So, problem (7) can be written as the following optimization problem:

$$\min_{\hat{B}} \left(\min_{\substack{\hat{D}\in\mathbb{R}^{m\times n} \\ R\hat{D}^\top = 0}} \operatorname{vec}^\top (D^\top - \hat{D}^\top) W^{-1} \operatorname{vec}(D^\top - \hat{D}^\top) \right). \quad (9)$$

Solving the inner minimization of problem (9) via Lagrange multipliers gives:

$$\begin{aligned} \psi(L, \hat{D}) &= \operatorname{vec}^\top (D^\top - \hat{D}^\top) W^{-1} \operatorname{vec}(D^\top - \hat{D}^\top) - \operatorname{tr}(L^\top R\hat{D}^\top) \\ &= \operatorname{vec}^\top (D^\top - \hat{D}^\top) W^{-1} \operatorname{vec}(D^\top - \hat{D}^\top) - \operatorname{vec}^\top (L) \operatorname{vec}(R\hat{D}^\top) \\ &= \operatorname{vec}^\top (D^\top - \hat{D}^\top) W^{-1} \operatorname{vec}(D^\top - \hat{D}^\top) \\ &\quad - \operatorname{vec}^\top (L)(I_m \otimes R) \operatorname{vec}(\hat{D}^\top), \end{aligned}$$

where L is the Lagrange multiplier and we have used the following properties

$$\begin{aligned} \operatorname{tr}(A^\top C) &= \operatorname{vec}^\top (A) \operatorname{vec}(C) \\ \operatorname{vec}(AC) &= (C^\top \otimes I_q) \operatorname{vec}(A) \text{ with } \sharp \operatorname{row}(A) = q \\ &= (I_p \otimes A) \operatorname{vec}(C) \text{ with } \sharp \operatorname{col}(C) = p, \end{aligned}$$

where $\otimes$ denotes the Kronecker product. For more information about manipulations involving Kronecker products and the vec operator, we refer the interesting reader to Ref. [6]. Setting the partial derivatives of $\psi(L, \hat{D})$ equal to zero, gives:

$$\begin{aligned} \frac{\partial \psi}{\partial L} &= 0 \Leftrightarrow (I_m \otimes R) \operatorname{vec}(\hat{D}^\top) = 0. \\ \frac{\partial \psi}{\partial \hat{D}} &= 0 \Leftrightarrow 2W^{-1} \operatorname{vec}(\hat{D}^\top) - (I_m \otimes R^\top) \operatorname{vec}(L) = 2W^{-1} \operatorname{vec}(D^\top) \end{aligned}$$

In matrix form, this gives the following:

$$\begin{bmatrix} 2W^{-1} & -(I_m \otimes R^\top) \\ (I_m \otimes R) & 0 \end{bmatrix} \begin{bmatrix} \mathrm{vec}(\hat{D}^\top) \\ \mathrm{vec}(L) \end{bmatrix} = \begin{bmatrix} 2W^{-1}\mathrm{vec}(D^\top) \\ 0 \end{bmatrix} \tag{10}$$

Because of the specific form of the matrix in the left-hand side of Eq. (10), we can give a closed-form expression for the vectorized form of the minimizing $\hat{D}$:

$$\mathrm{vec}(\hat{D}^\top) = \mathrm{vec}(D^\top) - W(I_m \otimes R^\top)[(I_m \otimes R)W(I_m \otimes R^\top)]^{-1} \times (I_m \otimes R)\mathrm{vec}(D^\top) \tag{11}$$

By filling in this minimizing $\hat{D}$ in the cost function of problem (9), an expression for the cost function $f(\hat{B})$ to be minimized can be written as follows:

$$f(\hat{B}) = \mathrm{vec}^\top(D^\top)(I_m \otimes R^\top)[(I_m \otimes R)W(I_m \otimes R^\top)]^{-1} \times (I_m \otimes R)\mathrm{vec}(D^\top), \tag{12}$$

with $R = [\hat{B}^\top - I_{n-r}] \in \mathbb{R}^{(n-r)\times n}$.

Problem (9) can now be written as the following non-convex unconstrained optimization problem:

$$\min_{\hat{B}} f(\hat{B}). \tag{13}$$

2.2. The adapted EW-TLS problem

For a given noisy data matrix $D \in \mathbb{R}^{m\times n}$, with $m \le n$, with only row-wise correlated measurement errors and given row error covariance matrices $W_i \in \mathbb{R}^{n\times n}$, for $i = 1, \ldots, m$, the adapted EW-TLS problem can be formulated as follows:

$$\min_{\hat{D}\in\mathbb{R}^{m\times n}} \mathrm{vec}^\top(D^\top - \hat{D}^\top)W^{-1}\mathrm{vec}(D^\top - \hat{D}^\top)$$
$$\text{s.t. } [\hat{B}_2^\top - I_{m-r}]\hat{D} = 0, \tag{14}$$

where the weighting matrix W is block diagonal, because the measurements are uncorrelated among the columns:

$$W = \begin{bmatrix} W_1 & & \\ & \ddots & \\ & & W_m \end{bmatrix}. \tag{15}$$

The adapted EW-TLS problem formulation can be useful in chemometrics. In chemometrics, the data matrix usually has size $m \times n$ with $m \le n$. Moreover, it makes sense to study the case where correlations between the measurement errors exist only along the rows, because in calibration problems [2] for example, the rows of the data matrix D are formed by individual spectra.

Note that the rank constraint $rank(\hat{D}) \le r$ is written differently in this subsection. By using the same equation as in the previous subsection,

$$\hat{D}\begin{bmatrix} \hat{B} \\ -I_{n-r} \end{bmatrix} = 0, \tag{16}$$

we would overparameterize the constraint because of $m \le n$. By defining $R := [\hat{B}_2^\top - I_{m-r}] \in \mathbb{R}^{(m-r)\times m}$, the rank constraint $[\hat{B}_2^\top - I_{m-r}]\hat{D} = 0$ can now be written as $R\hat{D} = 0$ or $\hat{D}^\top R^\top = 0$. So, problem (14) can be written as the following optimization problem:

$$\min_{\hat{B}} \left(\min_{\substack{\hat{D}\in\mathbb{R}^{m\times n} \\ R\hat{D}^\top = 0}} \mathrm{vec}^\top(D^\top - \hat{D}^\top)W^{-1}\mathrm{vec}(D^\top - \hat{D}^\top) \right). \tag{17}$$

Note that the blocks W_i itself, for $i = 1, \ldots, m$, have larger dimensions than the non-zero blocks of the matrix W in the previous subsection. Solving the inner minimization of problem (17) via Lagrange multipliers gives:

$$\psi(L, \hat{D}) = \mathrm{vec}^\top(D^\top - \hat{D}^\top)W^{-1}\mathrm{vec}(D^\top - \hat{D}^\top) - \mathrm{tr}(L^\top \hat{D}^\top R^\top),$$

where L is the Lagrange multiplier. Analogously as in the previous subsection, a closed-form expression for the vectorized form of the minimizing $\hat{D}$ is given by

$$\mathrm{vec}(\hat{D}^\top) = \mathrm{vec}(D^\top) - W(R^\top \otimes I_n)[(R \otimes I_n)W(R^\top \otimes I_n)]^{-1} \times (R \otimes I_n)\mathrm{vec}(D^\top). \tag{18}$$

Now, an expression for the cost function $g(\hat{B}_2)$ of the inner minimization of (17) can be found and denoted as follows:

$$g(\hat{B}_2) = \mathrm{vec}^\top(D^\top)(R^\top \otimes I_n)[(R \otimes I_n)W(R^\top \otimes I_n)]^{-1} \times (R \otimes I_n)\mathrm{vec}(D^\top), \tag{19}$$

with $R = [\hat{B}_2^\top - I_{m-r}] \in \mathbb{R}^{(m-r)\times m}$.

Problem (17) can now be written as the following non-convex unconstrained optimization problem:

$$\min_{\hat{B}_2} g(\hat{B}_2). \tag{20}$$

Algorithm 1. Adapted EW-TLS algorithm.

1. Input: the data matrix $D \in \mathbb{R}^{m\times n}$, the covariance matrices W_i, $i = 1, \ldots, m$, a rank specification r, and a convergence tolerance ε.
2. Initial approximation $B_2^{(0)}$: compute the TLS solution $B_2^{(0)} = -U_{12}U_{22}^{-1}$; where $D = U\Sigma V^\top$ is the Singular Value Decomposition (SVD) of D and the matrix U is partitioned as follows:

$$U = \begin{matrix} \\ r \\ m-r \end{matrix} \begin{matrix} \begin{matrix} r & m-r \end{matrix} \\ \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} \end{matrix}.$$

3. Apply a standard optimization algorithm, e.g., the BFGS (Broyden, Fletcher, Goldfarb, and Shanno) quasi-Newton method, for the minimization of g over $\hat{B}_2$ with initial approximation $B_2^{(0)}$ and with cost function evaluation performed via implementation of expression (23). Let $\hat{B}_2*$ be the approximation found by the optimization algorithm upon convergence.

4. Compute the vectorized form of the low rank r approximation matrix $\hat{D}$ by filling in $\hat{B}_2*$ in expression (18).
5. Reconstruct $\hat{D}$ from vec($\hat{D}$).
6. Output: $\hat{D}$.

3. An algorithm to solve the adapted EW-TLS problem

The cost functions in the optimization problems (13) and (20) are non-convex. Due to the non-convexity of these problems, we consider a standard method for *local* optimization: the Levenberg–Marquardt algorithm (Matlab's lsqnonlin), which is a nonlinear least squares optimization algorithm. By applying such an optimization method, the efficiency of the cost function evaluation is of great importance. In order to derive an efficient algorithm to solve problems (9) and (17), the cost functions $f(\hat{B})$ and $g(\hat{B}_2)$ in problems (13) and (20), respectively, need to be studied in further detail. The expression (12) for the cost function $f(\hat{B})$ that needs to be minimized in the standard EW-TLS case, can be rewritten in a simpler way. It is not difficult to see that, instead of using Kronecker products, the cost function $f(\hat{B})$ can be written in terms of a summation over the rows of the data matrix D:

$$f(\hat{B}) = \sum_{i=1}^{m} d_i R^{\top} (R W_i R^{\top})^{-1} R d_i^{\top}, \tag{21}$$

where $R = [\hat{B}^{\top} - I_{n-r}] \in \mathbb{R}^{(n-r)\times n}$ and d_i denotes the i-th row of matrix D. By implementing these sums instead of the Kronecker products and minimizing over the simplified expression (21) of the cost function $f(\hat{B})$, an efficient algorithm to solve the standard EW-TLS problem was described in Ref. [1, Algorithm 5]. For the adapted EW-TLS problem, however, it is not so straightforward to rewrite the expression (19) for the cost function $g(\hat{B}_2)$ in a simpler way. Still, computational savings can be achieved as follows. By denoting R_i as the i-th column of matrix $R = [\hat{B}_2^{\top} - I_{m-r}] \in \mathbb{R}^{(m-r)\times m}$, expression $(R \otimes I_n) W (R^{\mathrm{T}} \otimes I_n)$ in Eq. (19) can be rewritten as

$$[R_1 \otimes I_n \ldots R_m \otimes I_n] \begin{bmatrix} W_1 & & \\ & \ddots & \\ & & W_m \end{bmatrix} \begin{bmatrix} R_1^{\top} \otimes I_n \\ \vdots \\ R_m^{\top} \otimes I_n \end{bmatrix} = \sum_{i=1}^{m} R_i R_i^{\top} \otimes W_i. \tag{22}$$

Hence, by exploiting the structure of the weighting matrix W, a simplified expression for the cost function $g(\hat{B}_2)$ is given as follows:

$$g(\hat{B}_2) = \left(\sum_{i=1}^{m} R_i^{\top} \otimes d_i \right) \left(\sum_{i=1}^{m} R_i R_i^{\top} \otimes W_i \right)^{-1} \times \left(\sum_{i=1}^{m} R_i \otimes d_i^{\top} \right). \tag{23}$$

So, the evaluation of the efficient cost function g in (23) is a matter of numerical implementation of the involved operations. The proposed algorithm, based on a classical optimization method, to solve problem (20) is outlined in Algorithm 1.

In the next section, we will compare the standard EW-TLS method (minimizing the simplified expression (21) of cost function f) with the derived adapted EW-TLS method (minimizing the simplified expression (23) of cost function g) on simulated chemical data with only row-wise correlated measurement errors. We expect that in the case of $m \leq n$ the number of iterations needed to compute the adapted EW-TLS solution will be less than the number of iterations needed to compute the standard EW-TLS solution, and the other way around.

4. Performance comparison between EW-TLS and MLPCA

For the discussion of the performance of the EW-TLS algorithms and the MLPCA algorithm three simulated data sets are used: two Monte-Carlo simulations are used and a simulated data set from chemical measurements is used which was previously described in Ref. [1, Example 3]. The presented results are obtained by implementing the different algorithms in Matlab (version 6.1) on a PC i686 with 800 MHz and 256 MB memory.

Example 1. The simulated data set contained matrices $D \in \mathbb{R}^{10\times n}$, for $n=6, 7, \ldots, 15$. The noise-free data matrix D_0 of rank 2 was calculated by multiplying an arbitrary 10×2 matrix by an arbitrary $2\times n$ matrix. For every n, 100 different noise realizations have been added to D_0 in order to construct a full rank noisy matrix $D = D_0 + \Delta D$. ΔD is a noise matrix with correlations only within the rows.

Both, the standard EW-TLS method (defined with the right kernel (16)) and the adapted EW-TLS method (14) are applied to D, described in Example 1, in order to find the best low rank $r=2$ approximation matrix $\hat{D}$ of D. For every n, the mean value of the number of iterations over the 100 runs is of interest and is visualized in Fig. 1. We expected that in the case of $m \geq n$ the number of iterations needed to compute the standard EW-TLS

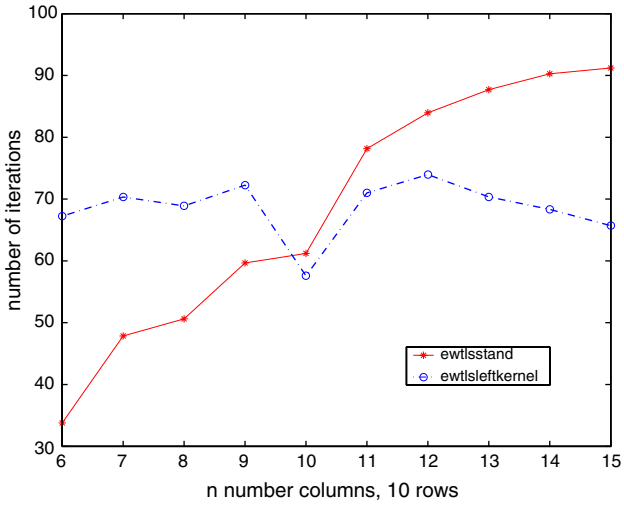

Fig. 1. Number of iterations needed to compute the standard EW-TLS solution and the adapted EW-TLS solution of the problem described in Example 1.

solution would be less than the number of iterations needed to compute the adapted EW-TLS solution, and the other way around. The reason for this is that representations (3) and (5) lead to a different number of parameters. Fewer optimization variables result in less computations per iteration as well as in fewer iteration steps. Thus, by solving problem (1) it is important to use a representation of the rank constraint $\text{rank}(\hat{D}) \leq r$ that leads to as few as possible optimization variables. From Fig. 1, it is clear that a smaller number of iterations is needed to compute the adapted EW-TLS solution for $m<n$. Indeed, the average number of iterations depends on the number of parameters in the optimization problem to solve.

Example 2. The simulated data set contained spectra from 10 samples of three-component mixtures. The concentration of each component in each of the 10 mixtures had a value between 0 and 1 from a uniform random number distribution. The spectral profiles of the three components were Gaussian with a standard deviation of 20 nm and maximum molar absorptivities at 480 nm, 500 nm and 520 nm, respectively. Pure spectral vectors were generated between 400 nm and 600 nm at 20 nm intervals. The noise-free data matrix D_0 was calculated by multiplying the 10×3 matrix of concentrations by the 3×11 matrix of pure component spectra. To add a noise matrix ΔD of correlated measurement errors to the noise-free matrix D_0, first a 10×11 matrix D' of uncorrelated measurement errors was generated. The matrix of the measurement standard deviations corresponding to this 10×11 matrix is determined by generating a 10×11 matrix of uniformly distributed random numbers between 0 and 0.01. This ensures that there is no pattern in the standard deviation matrix. Now, the 10×11 matrix of uncorrelated measurement errors D' is generated by taking a 10×11 random matrix with normally distributed elements (with Matlab's randn) and multiplying this matrix, element-wise, by the standard deviation matrix. To introduce correlations among the errors within the rows, the rows of matrix D' were filtered using a 1×5 moving average digital filter (see Ref. [2, Eqs. (34) (35) (36)] for the definition) to construct the correlated error matrix ΔD. This error matrix ΔD was added to the noise-free part D_0 in order to complete the noisy data matrix $D=D_0+\Delta D$ of size 10×11.

To the simulated data set described in Example 2 we have applied the MLPCA algorithm and the EW-TLS algorithms. All the algorithms start from the same initial approximation, the truncated SVD solution. After the algorithms reached their local minimum, we compared the final cost which is defined by $\text{vec}^\top(D-\hat{D})W^{-1}\text{vec}(D-\hat{D})$ for each algorithm and the number of iterations needed to converge to their local minimum. The results are presented in Table 1. From the table it is clear that the adapted EW-TLS algorithm needs the fewest number of iterations for the case when the data matrix has size $m\times n$, $m\leq n$, and only row-wise correlated measurement errors.

Table 1
MLPCA and EW-TLS applied to the chemical data set described in Example 2

Approach	Cost	Iterations
MLPCA	43.40462099863502	37
Standard EW-TLS	43.75752145651096	84
Adapted EW-TLS	43.40462102116915	25

In order to draw general conclusions about the right choice of algorithms to solve specific problems in chemometrics and other application fields, we created a simulation example, Example 3, which contains data matrices of size $m\times n$ with $m<n$, $m>n$ or $m=n$. All the data matrices have measurement errors which are only correlated among the rows.

Example 3. The simulated data set contained $m\times n$ matrices D, for $m=6, 7,\ldots, 13$ and $n=20-m$. The noise-free data matrix D_0 of rank r was calculated by multiplying an arbitrary $m\times r$ matrix by an arbitrary $r\times n$ matrix. The desired rank r was varied from 1 to 4. For each combination of m and r, 50 independently generated data matrices D were generated as follows. First, an $m\times n$ matrix D' of uncorrelated measurement errors was generated. The matrix of the measurement standard deviations corresponding to this $m\times n$ matrix is determined by generating an $m\times n$ matrix of uniformly distributed random numbers between 0 and 0.01. This ensures that there is no pattern in the standard deviation matrix. Now, the $m\times n$ matrix of uncorrelated measurement errors D' is generated by taking an $m\times n$ random matrix with normally distributed elements (with Matlab's randn) and multiplying this matrix, element-wise, by the standard deviation matrix. To introduce correlations among the errors within the rows, the rows of matrix D' were filtered using a 1×5 moving average digital filter (see Ref. [2, Eqs. (34) (35) (36)] for the definition) to construct the correlated error matrix ΔD. This error matrix ΔD was added to the noise-free part D_0 in order to complete the noisy data matrix $D=D_0+\Delta D$ of size $m\times n$.

We have applied the MLPCA algorithm and the two EW-TLS algorithms to each of the data matrices D of the data set described in Example 3. The three algorithms were run from equivalent initial approximations obtained via the SVDs and the stopping criteria were set to the same tolerance. In all the runs, the same solution was found. As expected, the average number of iterations depends on the number of optimization parameters: the fewer the optimization variables, the fewer the average number of iterations for convergence. Numerical results are shown in Table 2. The table shows the number of iterations for each algorithm and for the different m, n and r.

From the table it is clear that the adapted EW-TLS algorithm needs the fewest number of iterations for the case when the data matrix has size $m\times n$ with $m<n$ and the measurement errors are only row-wise correlated. When $m>n$, the standard EW-TLS algorithm converges to the right solution within the fewest number of iterations. For square matrices with $m=n$, the MLPCA algorithm seems to converge in the fewest number of iterations.

Nevertheless, in order to draw general conclusions about the right method of choice for a specific problem in chemometrics, besides the number of iterations, we also need to take into account the number of floating point operations (flops) per iteration for each of the algorithms discussed. The MLPCA

Table 2
The number of iterations of the MLPCA and the EW-TLS algorithms by applying them to the chemical data set described in Example 3

$m\times n$		6×14	7×13	8×12	9×11	10×10	11×9	12×8	13×7
$r=1$	MLPCA	16	16	16	17	**11**	17	17	18
	Standard EW-TLS	15	16	16	13	14	**10**	**9**	8
	Adapted EW-TLS	**7**	8	**11**	11	23	13	14	17
$r=2$	MLPCA	27	28	30	32	**13**	35	31	36
	Standard EW-TLS	56	58	46	38	48	**30**	**30**	25
	Adapted EW-TLS	**13**	19	**19**	21	28	34	34	33
$r=3$	MLPCA	37	41	46	50	**18**	51	49	53
	Standard EW-TLS	88	84	72	67	66	53	**48**	35
	Adapted EW-TLS	**15**	19	**30**	36	41	**50**	63	66
$r=4$	MLPCA	46	50	53	62	**24**	64	63	61
	Standard EW-TLS	88	89	83	78	77	**63**	**49**	31
	Adapted EW-TLS	**16**	28	**42**	46	57	65	69	72

algorithm is dominated by an SVD, whose computational cost is of the order of $4m^2n+8mn^2+9n^3$ for an $m\times n$ matrix. We obtained the following theoretical number of flops for minimizing the cost function (step 3) in the adapted EW-TLS algorithm, Algorithm 1: $2mn(m-r)+(m-r)^2+mn^2(m-r)^2+2/3n^3(m-r)^3+n^2(m-r)^2+n(m-r)$. The theoretical number of flops for minimizing the cost function f, expressed in (21), in the standard EW-TLS algorithm is of the order $m(2n(n-r)+n^2(n-r)+n(n-r)^2+2/3(n-r)^3+(n-r)^2+(n-r))$. Given the iterative nature of the algorithms, the total number of flops is a multiple of the flops necessary to execute one iteration. Based on the number of iterations given in Table 2 and the theoretical number of flops per iteration, we computed the total number of flops for each algorithm and for the different sizes of m, n and r. The results are shown in Table 3. By underlining the number of iterations in the table, we emphasize which algorithm has the fewest computational load for the specific choices of m, n and r. We clearly see the computational advantage of the adapted EW-TLS algorithm for the cases of $m\ll n$ and $r>2$ and the computational advantage of the standard EW-TLS algorithm for the case when $m\gg n$. The standard EW-TLS algorithm seems to behave better. The reason for this is that we could not avoid the Kronecker product in the adapted EW-TLS algorithm. The EW-TLS algorithms are only developed for cases of row-wise correlated measurement errors. So, for more general cases of measurement correlations the MLPCA algorithm should still be the method of choice.

Based on this experiment, we can conclude that the EW-TLS-like algorithms can indeed be nice alternatives to the MLPCA algorithm for the specific cases of row-wise correlated measurement errors and data matrices which are far from squared. Moreover, in paper [1] we showed that for uncorrelated measurement errors with equal row variances the GTLS algorithm performs best and that the standard EW-TLS algorithm is also useful for cases where the measurement errors have unequal row variances. We refer to paper [1] for further explanation.

5. Conclusion

In order to make the EW-TLS method useful for problems in chemometrics, we have derived an adapted version of this EW-TLS method. For the case when the data matrix D has much more columns than rows and only correlated measurement errors within the rows, we have presented an algorithm to compute the best low rank matrix approximation of D. Moreover, we have compared this new algorithm with the standard EW-TLS algorithm and the MLPCA algorithm in convergence behaviour on chemical data. It can be concluded that the EW-TLS-like algorithms are nice alternatives to the MLPCA algorithm for the specific cases of row-wise correlated measurement errors and data matrices which are far from squared. For more general cases of correlations between the measurements the MLPCA algorithm should still be the method of choice.

Acknowledgements

Prof. Dr. Sabine Van Huffel is a full professor, Dr. Ivan Markovsky is a postdoctoral researcher and Mieke Schuermans is a research assistant at the Katholieke Universiteit Leuven, Belgium. Our research is supported by

Table 3
The total number of flops of the MLPCA and the EW-TLS algorithms by applying them to the chemical data set described in Example 3

$m\times n$		6×14	7×13	8×12	9×11	10×10	11×9	12×8	13×7
$r=1$	MLPCA	577,920	508,560	445,440	412,335	231,000	306,765	261,120	232,470
	Standard EW-TLS	623,220	618,240	552,870	386,880	345,240	195,950	133,560	84,864
	Adapted EW-TLS	1,847,300	2,930,000	5,061,287	5,868,400	13,272,633	7,609,810	7,802,300	8,437,644
$r=2$	MLPCA	975,240	889,980	835,200	776,160	273,000	631,575	476,160	464,940
	Standard EW-TLS	2,020,032	1,923,400	1,345,700	941,868	966,400	468,160	343,440	196,080
	Adapted EW-TLS	1,817,100	4,139,800	5,638,212	7,668,304	11,577,000	14,788,980	14,506,000	12,855,524
$r=3$	MLPCA	1,336,440	1,303,185	1,280,640	1,212,750	378,000	920,295	752,640	684,495
	Standard EW-TLS	2,733,632	2,367,680	1,762,560	1,363,584	1,065,680	643,632	410,880	194,130
	Adapted EW-TLS	934,425	2,207,200	5,323,350	8,514,500	11,648,000	15,638,000	20,040,615	19,751,000
$r=4$	MLPCA	1,661,520	1,589,250	1,475,520	1,503,810	504,000	1,154,880	967,680	787,815
	Standard EW-TLS	2,332,000	2,108,232	1,676,800	1,280,916	974,820	577500	300,272	113,646
	Adapted EW-TLS	327,850	1464,600	4,002,432	6,542,900	10,539,072	14,028,560	15,847,552	16,131,312

the Research Council KUL: GOA-AMBioRICS, several PhD/postdoc and fellow grants;
the Flemish Government: FWO: PhD/postdoc grants, projects, G.0270.02 (nonlinear Lp approximation), G.0360.05 (EEG, Epileptic), research communities (ICCoS, ANMMM);
the IWT: PhD Grants,
the Belgian Federal Government: IUAP V-22 (2002–2006): Dynamical Systems and Control: Computation, Identification and Modeling;
the EU: BIOPATTERN (FP6-2002-IST 508803), ETUMOUR (FP6-2002-LIFESCIHEALTH 503094).

References

[1] M. Schuermans, I. Markovsky, P.D. Wentzell, S. Van Huffel, On the equivalence between total least squares and maximum likelihood PCA, Anal. Chim. Acta 544 (2005) 254–267.

[2] P.D. Wentzell, D.T. Andrews, D.C. Hamilton, K. Faber, B.R. Kowalski, Maximum likelihood principal component analysis, J. Chemom. 11 (1997) 339–366.

[3] P.D. Wentzell, M.T. Lohnes, Maximum likelihood principal component analysis with correlated measurement errors: theoretical and practical considerations, Chemometr. Intell. Lab. Syst. 45 (1999) 65–85.

[4] A. Premoli, M.L. Rastello, The parametric quadratic form method for solving problems with element-wise weighting, in: S. Van Huffel, P. Lemmerling (Eds.), Total least Squares and Errors-In-Variables Modeling: Analysis, Algorithms and Applications, Kluwer, 2002, pp. 67–76.

[5] I. Markovsky, M.L. Rastello, A. Premoli, A. Kukush, S. Van Huffel, The element-wise weighted total least squares problem, Comp. Stat. Data Anal. 50 (1) (2005) 181–209.

[6] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, John Wiley & Sons, 1988.