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.

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:

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