Chemometrics and Intelligent Laboratory Systems 133 (2014) 70–83

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

An error-in-variable projection to latent structure framework for monitoring technical systems with orthogonal signal components

Zhe Li [a], Uwe Kruger [b,*], Xun Wang [c], Lei Xie [a,**]

[a] Institute of Cyber Systems and Control, Zhejiang University, 310018 Hangzhou, P.R. China
[b] Department of Mechanical and Industrial Engineering, Sultan Qaboos University, P.O. Box 33, Muscat, Oman
[c] Department of Mathematics, The Petroleum Institute, P.O. Box 2533, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Article history:
Received 7 August 2013
Accepted 5 February 2014
Available online 13 February 2014

Keywords:
Projection to latent structures
Maximum redundancy
Error-in-variables
Maximum likelihood estimation
Orthogonal signal components
Multivariate statistical process control

ABSTRACT

This article introduces an error-in-variable (EIV) data structure for process monitoring models, that assumes the presence of source variables that can be correlated and uncorrelated to the system response variables. To identify such models, the paper proposes a different objective function for the projection to latent structure (PLS) approach. Compared to existing work, this modified PLS formulation does not remove uncorrelated, or orthogonal, components from the predictor set prior to the identification of a PLS model nor identify an initial PLS model and carry out a subsequent extraction of various correlated and orthogonal components from the predictor set. The proposed PLS algorithms extract latent components to predict the system response variables as accurately as possible after estimating the error covariance matrices using a maximum likelihood algorithm that is also introduced in this article. A detailed analysis of this extended PLS framework yields that the objective function is an iterative maximum redundancy formulation. The article finally shows that the developed algorithms here are computationally more efficient than existing PLS-based approaches.

1. Introduction

Over the past few decades, principal component analysis (PCA) and PLS have gained significant attention for monitoring complex systems [1–6]. Both are modeling tools and a core part of the multivariate statistical process control (MSPC) methodology. Based on its conceptual simplicity, however, most reported MSPC applications utilize PCA. One problem is that PLS applications are not well understood. For example, it relies on the assumption that the predictor variables are noise free [6], which is a restriction of generality:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad \mathbf{y} = \mathbf{C}^{*}\mathbf{s} + \mathbf{f}. \tag{1}$$

Here, $\mathbf{x} \in \mathbb{R}^{n_x}$ and $\mathbf{y} \in \mathbb{R}^{n_y}$ represent the predictor and response sets, respectively, $\mathbf{f} \in \mathbb{R}^{n_y}$ is an error vector, $\mathbf{s} \in \mathbb{R}^{n}$ stores $n \le n_x$ source signals describing common cause variation and $\mathbf{A} \in \mathbb{R}^{n_x \times n}$ and $\mathbf{C}^{*} \in \mathbb{R}^{n_y \times n}$ are parameter matrices.

The assumption for the data structure in Eq. (1) includes $E\left\{f_i^2\right\} \ll E\left\{\left(\mathbf{c}_j^{*^T}\mathbf{s}\right)^2\right\}$ for all $1 \le i, j \le n_y$, where $\mathbf{c}_j^{*^T}$ is the jth row vectors of $\mathbf{C}^{*}$ and $E\{\cdot\}$ is the expectation operator. Furthermore, the random variables stored in $\mathbf{s}$ and $\mathbf{f}$ (i) are Gaussian distributed, (ii) possess no serial correlation, (iii) are statistically independent and (iv) $E\{\mathbf{s}\} = E\{\mathbf{f}\} = \mathbf{0}$. Offset terms $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ can be added to the predictor and response sets, that is $\mathbf{x}_M = \mathbf{x} + \bar{\mathbf{x}}$ and $\mathbf{y}_M = \mathbf{y} + \bar{\mathbf{y}}$, to describe the recorded variables, denoted by the subscript M. The first aim of this article is to remove the noise-free assumption for the predictor variables.

In contrast, PCA assumes that each process variable included in the analyzed data set is corrupted by an error term, that is, $\mathbf{z} = \mathbf{D}\mathbf{s} + \mathbf{g}$ [6–8]. Here, $\mathbf{z} \in \mathbb{R}^{n_z}$ is a data vector, $\mathbf{D} \in \mathbb{R}^{n_z \times n}$ is a parameter matrix and $\mathbf{g} \in \mathbb{R}^{n_z}$ is an error vector. It is important to note that assuming each of the recorded process variables is corrupted by an error term is practically more relevant than the restricted assumption that only the system response set is affected by errors.

For process monitoring, PLS extracts n sets of latent variables (LVs) that allow constructing a number of non-negative quadratic statistics [6]. Successful applications involving PLS include Refs. [9–13]. For PLS models, the response set typically includes product quality and safety critical variables, while the predictor set contains variables that are expected to directly affect the response set and other recorded variables describing common cause variation.

Besides the noise-free assumption of the predictor set, the data structure in Eq. (1) is practically too restrictive for a second reason. It does not cover cases where some of the predictor variables are uncorrelated to the response variables. An example of this is the presence of controller feedback, which maintains key process variables affecting product quality at predefined set-point values.

In the presence of measured or unmeasured disturbances, some manipulated variables may not directly affect product quality, which

* Corresponding author. Tel.: +968 24 142 549; fax: +968 241 41 316.
** Corresponding author. Tel.: +86 130 7184 7893.
E-mail addresses: uwekruger@squ.edu.om (U. Kruger), leix@csc.zju.edu.cn (L. Xie).

http://dx.doi.org/10.1016/j.chemolab.2014.02.001
0169-7439/

follows from the aim of the control system to isolate the effect of disturbances upon product quality. Consequently, the common cause variation captured in the predictor variables may be divided into variation that is correlated and orthogonal to the response set. The second aim of this article is to introduce an error-in-variable EIV data structure that includes both types of source signals.

Addressing both aims together requires the introduction of the following and more generic data structure for PLS models:

$$\mathbf{x} = \mathbf{A}\mathbf{s}_1 + \mathbf{B}\mathbf{s}_2 + \mathbf{e} \quad \mathbf{y}=\mathbf{C}\mathbf{s}_1+\mathbf{f}. \tag{2}$$

Here, $\mathbf{s}_1 \in \mathbb{R}^n$ and $\mathbf{s}_2 \in \mathbb{R}^m$ are vectors storing source signals that are correlated and uncorrelated, or orthogonal, to the response variables, respectively, $\mathbf{e}$ is an error term and $n + m \le n_x$. If $n + m = n_x$ Eq. (2) presents the well-known EIV formulation for least squares problems [14]. It is important to note that PCA cannot extract latent components which discriminate $\mathbf{s}_1$ from $\mathbf{s}_2$. It is also important to note that the data structure in Eq. (2) can be rewritten to reduce to the PCA data structure, which Section 2 of this article shows.

On the basis of the assumption that $\mathbf{e} = \mathbf{0}$ and for the application of near-infrared spectra, Wold et al. [15] developed an orthogonalization routine based on the iterative NIPALS algorithm to extract components that are orthogonal to the response set. Removing these components from the predictor set then allows the application of standard PLS algorithms [16,17]. To guarantee that the extracted orthogonal components have a maximum variance, Fearn [18] introduced a slightly different filtering algorithm.

Trygg and Wold [19] augmented the PLS algorithm to determine score variables that are either correlated or orthogonal to the response set. The advantages of this orthogonal PLS (O-PLS) algorithm lies in its ability (i) to utilize cross-validation to prevent overfitting and (ii) to exclude specific variation of the predictor variable set for predicting the response set. Thus, the work in Refs. [15,18,19] can extract $\mathbf{s}_2$ up to a similarity transformation if $\mathbf{e} = \mathbf{0}$.

An alternative approach has been proposed in Zhao et al. [20]. This approach relates to a post-processing of the PLS components to separate those associated with the predictor set into components that are correlated and orthogonal to the response variable set. A detailed analysis of this approach in this paper, however, shows that the separation is not conducted in an optimal fashion.

Different to PCA and standard PLS, the techniques in Refs. [15,18–20] can discriminate between $\mathbf{s}_1$ and $\mathbf{s}_2$ under the restriction that $\mathbf{e} = \mathbf{0}$. Hence, a modeling technique that can separately extract $\mathbf{s}_1$ and $\mathbf{s}_2$ for $\mathbf{e} \neq \mathbf{0}$ has not been introduced in the literature. For process monitoring, however, such a technique is important, as it allows monitoring and assessing whether (i) an abnormal event compromises product quality, (ii) this event is safety critical, (iii) operator intervention is required or (iv) the process can continue to operate at present.

Compared to the work in Refs. [15,18,19], this article proposes a different way to addresses the second aim. Instead of extracting the orthogonal components $\mathbf{s}_2$ first, the paper introduces an augmented objective function for PLS to extract the correlated components $\mathbf{s}_1$ instead. The deflated predictor sets then describes $\mathbf{s}_2$. In a similar fashion to the work in Ref. [20], PCA can be applied to the deflated predictor set in case some of the source signals in $\mathbf{s}_2$ have a small variance. The paper shows that resultant algorithms for extracting $\mathbf{s}_1$ prior to $\mathbf{s}_2$ are computationally more efficient than those extracting $\mathbf{s}_2$ first.

To address the first aim of this paper, this article develops a maximum likelihood (ML) formulation for PLS models that estimates the error covariance matrices for $\mathbf{e}$ and $\mathbf{f}$. This ML formulation is similar to the ML extension for PCA models that has recently been discussed in the literature [21,8]. For addressing the second aim, besides introducing a number of algorithms for computing a solution to the augmented PLS objective function, a further contribution of this article is the development of computationally improved versions of the algorithms in Refs. [19,20].

The paper is organized as follows. Section 2 analyzes the techniques in Refs. [15,18–20] and motivates the rational for introducing the proposed PLS framework. Assuming initially that $\mathbf{e} = \mathbf{0}$, Sections 3–5 introduce the augmented PLS objective function, shows that this objective function is an iterative formulation of a maximum redundancy index, describes the properties of the resulting algorithm and show that it can determine each of the latent variable sets simultaneously.

For $\mathbf{e} \neq \mathbf{0}$, Section 6 proposes an ML estimation of the error covariance matrices to consistently estimate the column space of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. Section 7 discusses the implementation of the algorithms in Refs. [19,20] and Sections 8 and 9 contrast the performance of the various algorithms and finally, Section 10 presents a concluding summary of this article.

2. Preliminaries

Following a brief analysis of the PLS technique, Sections 2.2 and 2.3 examine recently proposed techniques for extracting the $\mathbf{s}_2$ components and for subsequently extract $\mathbf{s}_1$ and $\mathbf{s}_2$ from an existing PLS model, respectively. Based on this analysis, Section 2.4 motivates the proposed PLS framework.

2.1. Analysis of PLS modeling

The need for introducing orthogonal signal correction as a preprocessing prior to the application of PLS [15,22,18] or to incorporate an orthogonal projections into the PLS algorithm [19] results from the PLS objective function for determining the weight vectors. Defining $\mathbf{w}_i$ and $\mathbf{v}_i$ as the ith pair of weight vectors for $\mathbf{x}$ and $\mathbf{y}$, respectively, this objective function is as follows:

$$\begin{pmatrix}\mathbf{w}_i\\ \mathbf{v}_i\end{pmatrix} = \arg\max_{\mathbf{w},\mathbf{v}} E\left\{\mathbf{w}^T\mathbf{x}_i\mathbf{y}^T\mathbf{v}\right\} - \frac{1}{2}\lambda\left(\mathbf{w}^T\mathbf{w}-1\right) - \frac{1}{2}\lambda\left(\mathbf{v}^T\mathbf{v}-1\right), \tag{3}$$

where $\mathbf{x}_i$ is predictor set after applying the deflation procedure $i - 1$ times and λ is a Lagrangian multiplier. Applying the solution to Eq. (3) for $i = 1$ determines linear combinations of $\mathbf{x}$ and $\mathbf{y}$ which represent the score variables t_1 and u_1:

$$t_1 = \mathbf{w}_1^T\mathbf{A}\mathbf{s}_1 + \mathbf{w}_1^T\mathbf{B}\mathbf{s}_2 \quad u_1 = \mathbf{v}_1^T\mathbf{C}\mathbf{s}_1 + \mathbf{v}_1^T\mathbf{f}\,. \tag{4}$$

Substituting Eq. (4) into the PLS objective function gives rise to:

$$\begin{pmatrix}\mathbf{w}_i\\ \mathbf{v}_i\end{pmatrix} = \arg\max_{\mathbf{w},\mathbf{v}} E\{t_1u_1\} - \frac{1}{2}\lambda\left(\mathbf{w}^T\mathbf{w}-1\right) - \frac{1}{2}\lambda\left(\mathbf{v}^T\mathbf{v}-1\right). \tag{5}$$

A more detailed analysis of the expectation yields:

$$E\{t_1u_1\} = r_{t_1u_1}\sigma_{t_1}\sigma_{u_1} = r_{t_1u_1}\sqrt{E\left\{\left(\mathbf{w}_1^T\mathbf{A}\mathbf{s}_1\right)^2 + \left(\mathbf{w}_1^T\mathbf{B}\mathbf{s}_2\right)^2\right\}}\sqrt{E\left\{\left(\mathbf{v}_1^T\mathbf{C}\mathbf{s}_1\right)^2 + \left(\mathbf{v}_1^T\mathbf{f}\right)^2\right\}}, \tag{6}$$

where $r_{t_1u_1}$ is the correlation coefficient, $\sigma_{t_1}^2$ and $\sigma_{u_1}^2$ are the variances of the score variables. Eq. (6) dictates a condition for discriminating between components that are correlated and orthogonal to $\mathbf{y}$. If $\mathbf{w}_1{}^T\mathbf{B} = \mathbf{0}$ and $\mathbf{w}_1^T\mathbf{A} = \mathbf{0}$, t_1 is correlated and orthogonal to $\mathbf{y}$, respectively. These conditions, however, are generally not be assumed. Moreover, incorporating the deflation procedure for $\mathbf{x}$ allows demonstrating that PLS is generally unable to extract t-score variables that discriminate between $\mathbf{s}_1$ and $\mathbf{s}_2$.

2.2. Extracting components from $\mathbf{x}$ that are orthogonal to $\mathbf{y}$

The approach by Wold et al. [15] is a modified NIPALS algorithm to extract score vectors, $\mathbf{t}_i \in \mathbb{R}^K$, $\mathbf{t}_i = \mathbf{X}_i\mathbf{w}_i$. The weight vector, $\mathbf{w}_i \in \mathbb{R}^{n_x}$, is determined such that $\mathbf{t}_i$ is orthogonal to the column space of the