Chemometrics and Intelligent Laboratory Systems 146 (2015) 426–436

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Adaptive sparse principal component analysis for enhanced process monitoring and fault isolation

Kangling Liu [a], Zhengshun Fei [b], Boxuan Yue [a], Jun Liang [a,*], Hai Lin [c]

[a] State Key Lab of Industrial Control Technology, Institute of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, PR China
[b] School of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou, 310023, PR China
[c] Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

ARTICLE INFO

Article history:
Received 20 December 2014
Received in revised form 19 June 2015
Accepted 20 June 2015
Available online 26 June 2015

Keywords:
Principal component analysis
Adaptive
Sparse
Process monitoring
Fault isolation

ABSTRACT

Principal component analysis (PCA) has been widely applied for process monitoring and fault isolation. However, PCA lacks physical interpretation of principal components (PCs) since each PC is a linear combination of all variables, which makes the fault detection difficult. Moreover, since the PCA model is time invariant while all real world processes are time varying and subject to disturbances. This mismatch may cause a false alarm or missed detection. Due to these motivations, we propose an adaptive sparse PCA (ASPCA) for enhanced process monitoring and fault isolation. which obtains sparse loadings by imposing a sparsity constraint on PCA. ASPCA with sparse loadings improves the interpretation and then facilitates the isolation of faulty variables. Meanwhile, ASPCA enhances model adaptability by updating the loadings with the sparsity constraint modified with changes in operating conditions. Next, a process monitoring and fault isolation strategy is presented based on ASPCA. Qusi-T^2 and squared prediction error monitoring statistics are defined in the PC and residual subspaces, respectively. Nonzero variables in dominant PCs with most contributions to the fault are preferentially reconstructed. Case studies of TE process and waveform system demonstrate that the ASPCA method performs better in process monitoring and fault isolation compared to the PCA method.

1. Introduction

Recent decades have seen extensive efforts in the process monitoring and fault isolation driven by the growing demand for high efficiency, reliability and process safety [1–3]. Among many methods being proposed, principal component analysis (PCA) [4,5] has been widely adopted and successfully used in complex processes. The essential of PCA lies on the extraction of massive data into some less synthesize independent variables, which are termed as principal components (PCs). The extracted PCs sequentially capture the maximum variance of process data to guarantee minimum information loss. In detail, let an $n \times m$ matrix X, with m process variables and n observations, denote historical process data collected under normal operating conditions. Without loss of generality, assume that X is normalized to zero mean and unit variance. The optimization problem of PCA can be described as follows.

$$p_i^* = \arg \max \|Xp_i\|_2^2$$
$$s.t. \begin{cases} \|p_i\|_2^2 = 1 \\ p_i^T p_j^* = 0 \end{cases} (j = 1, 2, \ldots, i-1) \tag{1}$$

The ith PC is denoted as $t_i = Xp_i^* (i = 1, 2, \ldots, m)$, and its sample variance is $\lambda_i = p_i^{*\,T} \Sigma_X p_i^*$, where p_i^* is the loading vector and $\Sigma_X = X^T X / (n-1)$. The first l ($l < m$) PCs represent the PC subspace, and the remaining m–l PCs span the residual subspace. Define the loadings as $\hat{P} = [p_1^* \quad p_2^* \quad \cdots \quad p_l^*]$. The matrix X can be decomposed as $X = T\hat{P}^T + E$, where the residual matrix $E \in R^{n \times m}$ and the score matrix $T = X\hat{P} \in R^{n \times l}$. To select the optimal number of PCs, cumulative percent variance (CPV) with 85% of normal variability is widely adopted for its simplicity. With PCA model established based on historical normal data, process monitoring is achieved by comparing the monitoring statistics against the nominal model. Typically, the T^2 and squared prediction error (SPE) statistics are used in the PC and residual subspaces, respectively.

PCA has a wide application for process monitoring and fault detection, but it still suffers from the interpretation and adaptability problems especially when applied to complex processes that are often time varying and of large-scale. The interpretation problem means that each PC is difficult to be interpreted since elements of its corresponding loadings are typically nonzero. Approaches to address this problem can be classified into two types. On the one hand, multi-block method [6–9] divides the variables into conceptually meaningful blocks. The block division requires prior process knowledge, unfortunately, which is often unavailable or deficient in complex processes. On the other hand, the sparse method [10–12] imposes an L_1-norm constraint on

* Corresponding author at: Zhejiang University, No.38 Zheda Road, Hangzhou, 310027, PR China. Tel.: +86 138 0574 9569.
E-mail address: jliang@iipc.zju.edu.cn (J. Liang).

http://dx.doi.org/10.1016/j.chemolab.2015.06.014
0169-7439/

the optimization problem of PCA to control the sparsity of loadings and an additional tuning parameter is studied to balance between the variance of PCs and the number of nonzero elements in the loadings. PCA and its extension methods mentioned above estimate a fixed model and the application of a fixed model to time varying processes may cause model adaptability problem. For the adaptability problem, one solution is to describe pre-defined sets of faults by using supervised learning method [13–15]. Obviously, this method is not complete since it does not guarantee that all the possible faults are pre-defined. A more reasonable solution is to use process measurements to update the model dynamically [16–20].

To improve both the interpretation and the adaptability of PCA, this work proposes an adaptive sparse PCA (ASPCA). Firstly, ASPCA obtains sparse loadings with a small number of nonzero elements by introducing a sparsity constraint, which is defined based on actual correlation structure of process. Since each PC is a linear combination of several variables, the interpretation of model is improved. Secondly, to solve the adaptability problem, ASPCA unitizes the current correlation structure to update the sparsity constraint, which is carried out entirely based on the process data and is independent of known fault data. It is of particular interest when the process knowledge or the information of fault data set is deficient. Thirdly, a process monitoring and fault isolation strategy is presented based on ASPCA. Qusi-T^2 and SPE monitoring statistics are defined in the PC and residual subspaces, respectively. The capability of identifying faulty variables is enhanced by narrowing down the faulty variables to nonzero variables in dominant PCs with most contributions to the fault.

The rest of the work is organized as follows. In Section 2, the ASPCA method is detailed, including the development of ASPCA model and a Bayesian information criterion (BIC) for selection of the number of PCs, in which, ASPCA based process monitoring and fault isolation schemes are also given. This is followed by the case studies of application of ASPCA in the TE process and the waveform system, as demonstrated in Sections 3 and 4. Finally, Section 5 provides a concluding summary.

2. Adaptive sparse principal component analysis

In Section 2.1, the formulation of adaptive sparse principal component analysis (ASPCA) is detailed, and a BIC-type criterion is introduced to select the number of PCs of ASPCA model. ASPCA based schemes for process monitoring and fault isolation are proposed in Sections 2.2 and 2.3.

2.1. ASPCA model

PCA focuses on global correlation structure among all process variables. Process changes or faults affect the structure. Fig. 1 shows a motivation example for the concept of ASPCA. Fig. 1a illustrates limitations of PCA in two operating conditions. (*i*) In the normal operating condition, variables 1 and 2 are independent. With PCA model, the elements in the loadings p_1, p_2 are nonzero, then each PC is difficult to be interpreted physically. (*ii*) Since the model built in the normal operating condition is time-invariant, PCA cannot detect the change of operating condition which occurs within its confidence limit. To update the model in compliance with current operating condition, we consider the current correlation structure to construct the ASPCA model, as shown in Fig. 1b. This method is detailed as follows.

Assume N new measurements with m variables $Z=[z_1, z_2, \ldots, z_N]^T \in R^{N\times m}$ is normalized using the preprocessing information of normal dataset X. Covariance matrix of $Z_{1:\tau}=[z_1, z_2, \ldots, z_\tau]^T$ can be written as $\Sigma_{Z_\tau}=Z_{1:\tau}{}^T Z_{1:\tau}/(\tau-1)=(\tau-1)\Sigma_{Z_{\tau-1}}/\tau+z_\tau/\tau$, where $\Sigma_{Z_{\tau-1}}$ denotes the covariance matrix of $Z_{1:\tau-1}=[z_1, z_2, \ldots, z_{\tau-1}]^T$. The t-test [21] is then applied to Σ_{Z_τ}, resulting in a matrix $S^{z_\tau}=[s^{z_\tau}_{jk}]\in R^{m\times m}$. Each element $s^{z_\tau}_{jk}$, ranging from 0 and 1, denotes the probability of getting a correlation of variables k and j when the true correlation is zero. A smaller $s^{z_\tau}_{jk}$-value means that the correlation of corresponding variables j and k is more significant. With the τth new measurement $z_\tau(\tau=1, 2, \ldots, N)$, the following ASPCA optimization problem is individually performed for each sparse loading vector $p_i^\tau \in R^{m\times 1}$ $(i=1, 2, \ldots, m)$, where $\beta>0$ is to tune the sparsity and the variance, and empirical evidence indicates that the choice of $\beta=0.1$ is appropriate in this work.

$$p_i^\tau=\min\ -p_i{}^T\bullet\Sigma_X\bullet p_i+\beta\sum_{j=1}^{m}\sum_{k=1}^{m}\left|p_{i,k}p_{i,j}s^{z_\tau}_{jk}\right| \\ s.t.\quad \begin{cases}\|p_i\|_2^2=1 \\ p_i^T p_j^\tau=0\end{cases}(j=1,2,\ldots,i-1) \tag{2}$$

where $p_{i,k}$, $p_{i,j}$ are the kth and jth elements of loading vector p_i, respectively. To minimize the second term, each $p_{i,k}p_{i,j}s^{z_\tau}_{jk}$-value should be equal to zero or near zero. In other words, if $s^{z_\tau}_{jk}$-value is sufficiently large, at least one of $p_{i,k}$ and $p_{i,j}$ is shrunken to zero via a trade-off of variance in the first term. Otherwise if $s^{z_\tau}_{jk}$-value is closer to zero, elements $p_{i,k}$ and $p_{i,j}$ are less subject to the constraint. Thus, with ASPCA model, elements in the loadings are zeros if corresponding variables are independent, and values of the elements are increased with the degree of correlation between its corresponding variables. Each loading vector grasps intrinsic local information of nonzero variables. As a result, the whole process is divided into a diversity of different local correlation sub-structures. In each sub-structure, the selected subset of variables is relative strongly correlated. Note that variables may overlap in

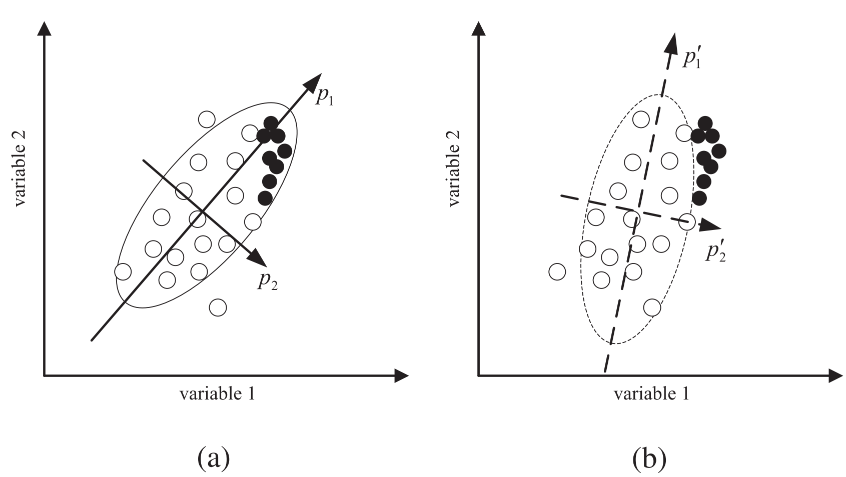

Fig. 1. PCA (a) and ASPCA (b) on data with variables 1 and 2 in two operating conditions. The variables are independent in the normal operating condition (○), but correlated in the current operating condition (●).