Automatica 80 (2017) 313–327

Contents lists available at ScienceDirect

Automatica

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

Recursive transformed component statistical analysis for incipient fault detection☆

Jun Shang[a], Maoyin Chen[a], Hongquan Ji[a], Donghua Zhou[b,a]

[a] Department of Automation, TNList, Tsinghua University, Beijing 100084, China

[b] College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China

ARTICLE INFO

Article history:
Received 28 January 2016
Received in revised form
24 January 2017
Accepted 31 January 2017

Keywords:
Recursive transformed component statistical analysis
Process monitoring
Incipient fault detection
Ordered eigenvalues

ABSTRACT

This paper presents a new data-driven process monitoring method called recursive transformed component statistical analysis (RTCSA) for the purpose of incipient fault detection. Without space partition, RTCSA processes data in sliding windows to obtain orthogonal transformed components (TCs) recursively using rank-one modification. The statistical information of TCs can reveal some important process features, implying that faults can be detected by monitoring the statistics of TCs. With second-order statistics, the detection index reduces to relative changes of ordered eigenvalues of the sample covariance matrix. Fault detectability is analyzed in a statistical sense, leading to the analysis of the eigenvalues of stochastic matrices, including the closed-form expressions for the probability distribution function of the arbitrary lth largest eigenvalue of a class of real uncorrelated Wishart matrices. It indicates that a scaled ordered eigenvalue is sensitive to small changes. The structure of the detection index ensures that RTCSA is sensitive to incipient faults. Compared with existing multivariate statistical process monitoring approaches such as principal component analysis (PCA) and its variants, the superior detectability of RTCSA is illustrated by a numerical example and the Tennessee Eastman process.

1. Introduction

Recent years have witnessed an increase in the demand for safety and reliability of modern industrial processes. Along these lines, data-driven process monitoring has attracted considerable interest owing to the merit that neither system models nor a priori fault information is required (Ge, Song, & Gao, 2013; Qin, 2012; Yin, Ding, Xie, & Luo, 2014; Yin, Li, Gao, & Kaynak, 2015). As an important branch of data-driven process monitoring techniques, multivariate statistical process monitoring (MSPM) has been successfully applied in various industrial processes, including chemicals, polymers, and microelectronics manufacturing (Qin, 2003). Principal component analysis (PCA), which is an important method of multivariate analysis, has been widely used in various fields such as data compression, feature extraction, pattern recognition, and process monitoring (Ding, 2014; Duan, Sun, & Shi, 2012; Lloyd, Mohseni, & Rebentrost, 2014; Price et al., 2006; Ringnér, 2008; Sun, Zhang, Jiang, & Zhang, 2008). As a basic technique of MSPM, PCA plays an important role in numerous industrial processes for fault detection and diagnosis (Chiang, Russell, & Braatz, 2000; Kruger, Kumar, & Littler, 2007). Its variants such as recursive PCA (RPCA) (Li, Yue, Valle-Cervantes, & Qin, 2000), dynamic PCA (DPCA) (Ku, Storer, & Georgakis, 1995; Russell, Chiang, & Braatz, 2000), and kernel PCA (Choi, Lee, Lee, Park, & Lee, 2005) are usually used for advanced process monitoring such as adaptive process monitoring, dynamic process monitoring, and nonlinear process monitoring.

In practical cases, numerous abnormal conditions gradually evolve from incipient faults (Watanabe, Matsuura, Abe, Kubota, & Himmelblau, 1989). This implies that, if faults are detected in their incipient stages, abnormal conditions may be effectively avoided. However, compared with serious faults, incipient faults are easily masked by normal variation or measurement noise owing to their small magnitudes; as a result, incipient fault detection is a challenging task. Recently, some approaches have been proposed in the literature to address the problems associated with incipient fault detection. Kiasi, Prakash, and Shah

☆ This work was supported by National Natural Science Foundation of China under Grants 61490701, 61210012, 61290324, 61473164, Tsinghua University Initiative Scientific Research Program under Grant 20131089240, and Research Fund for the Taishan Scholar Project of Shandong Province of China under Grant LZB2015-162. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Antonio Vicino under the direction of Editor Torsten Söderström.

E-mail addresses: shangj13@mails.tsinghua.edu.cn (J. Shang), mychen@tsinghua.edu.cn (M. Chen), jihq12@mails.tsinghua.edu.cn (H. Ji), zdh@mail.tsinghua.edu.cn (D. Zhou).

http://dx.doi.org/10.1016/j.automatica.2017.02.028
0005-1098/

(2015) presented a modified GLR-based approach to detect and diagnose incipient sensor faults of an LTI system. For a class of nonlinear distributed processes with incipient component and actuator faults, Armaou and Demetriou (2008) presented a robust detection and accommodation scheme. Alwi, Edwards, and Tan (2009) proposed sliding mode estimation schemes for incipient sensor faults. Ge, Wang, Zhou, Wu, and Jin (2015) proposed a two-step incipient fault detection method combining wavelet analysis with residual evaluation. Harmouche, Delpha, and Diallo (2014) presented an incipient fault detection method based on Kullback–Leibler divergence using PCA.

For most MSPM methods, process variables are jointly monitored to detect faults. Based on correlations, statistical models are built based on sufficient training data, leading to the decomposition of the original measurement space (Qin, 2003). During online monitoring, sample vectors are directly projected onto corresponding subspaces in sequence. This implies that the latest sample is projected separately without considering statistical information among samples. When detecting incipient faults, samples belonging to normal and abnormal conditions usually overlap to a large extent owing to their small fault magnitudes. As a result, conventional sample-wise MSPM methods easily lead to a significant number of missed detections.

One possible solution to reduce the missed detection rate is to utilize statistical information among measurements. Window-based monitoring methods can partially alleviate data overlap. He and Wang (2011) proposed statistics pattern analysis (SPA) to address the challenges encountered in semiconductor processes, which was also extended to continuous process monitoring (Wang & He, 2010). Instead of monitoring process variables, SPA monitors the statistics of process variables in sliding windows, demonstrating a superior performance over PCA and DPCA. However, SPA may not effectively detect some incipient faults with small magnitudes. Kano, Hasebe, Hashimoto, and Ohno (2002) proposed DISSIM to monitor the dissimilarity of process data. DISSIM monitors data distribution in sliding windows, and uses the dissimilarity index to differentiate between normal and abnormal conditions. It is sensitive to incipient faults occurring in some processes but may lack the portability for others.

Considering the problem of incipient fault detection, we propose a new MSPM method called recursive transformed component statistical analysis (RTCSA). It obtains orthogonal vectors called transformed components (TCs) by transforming the axes in the original measurement space. This transformation represents a rigid rotation of axes such that the scores in the transformed coordinates are orthogonal with maximum sample variance under constraints. TCs extracted in sliding windows are linear combinations of normalized process measurement vectors. The statistical information of TCs can reveal some important process features, which implies that condition changes can be detected by monitoring the statistics of TCs. We also use rank-one modification to update the sample covariance matrix and its eigenpairs recursively to improve the algorithm efficiency.

The main contributions of this paper are summarized as follows. (i) A new MSPM method RTCSA is proposed. Different from existing methods such as PCA and SPA, RTCSA extracts orthogonal TCs without space partition. Statistical information including higher-order statistics of TCs is extracted for process monitoring. (ii) The detection index is well-designed to ensure that RTCSA is sensitive to incipient faults. With second-order statistics, the detection index reduces to relative changes of ordered eigenvalues of the sample covariance matrix. Its structure ensures a wide spectrum of fault detection, because a scaled ordered eigenvalue is sensitive to faults with small magnitudes. (iii) The fault detectability of RTCSA is analyzed in a statistical sense for a general multivariate process (Alcala & Qin, 2009). For multivariate Gaussian distribution, the sample covariance matrix is decomposed into five parts considering the small magnitude of the incipient fault. This leads to the analysis of the eigenvalues of stochastic matrices, including the closed-form expressions for the probability distribution function (p.d.f.) of the arbitrary lth largest eigenvalue of a class of real uncorrelated Wishart matrices. (iv) A numerical example and the benchmark Tennessee Eastman process both illustrate the superior detectability of RTCSA, compared with the existing MSPM methods, such as PCA, RPCA, DPCA, SPA, and DISSIM.

The remainder of this paper is organized as follows. In Section 2, the algorithm of RTCSA is introduced in detail, including transformed components, statistical analysis, recursive computation, and the corresponding analysis of computational complexity. In Section 3, the detection indices of RTCSA and window width selection are demonstrated. The detectability of RTCSA for additive sensor fault and process fault is analyzed in Section 4, including the closed-form expectations for the arbitrary lth largest eigenvalue of a class of real uncorrelated Wishart matrices and the lower bounds on the expectations of eigenvalues of other stochastic matrices. In Section 5, both a numerical example and the Tennessee Eastman process are used to examine the detectability of RTCSA. Conclusions are given in Section 6.

2. Methodology

2.1. Transformed components (TCs)

Consider the original measurement matrix $\boldsymbol{X} \in \mathbb{R}^{n\times m}$, where n and m denote the number of samples and measured variables, respectively. We construct a one-step sliding window to stack process measurements:

$$\boldsymbol{X}_k = \begin{bmatrix} x_{k-w+1,1} & x_{k-w+1,2} & \cdots & x_{k-w+1,m} \\ x_{k-w+2,1} & x_{k-w+2,2} & \cdots & x_{k-w+2,m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k,1} & x_{k,2} & \cdots & x_{k,m} \end{bmatrix} \tag{1}$$

where k is the time index of the latest sample in the sliding window, and w denotes the window width. The original measurements are normalized as

$$\bar{\boldsymbol{X}}_k = (\boldsymbol{X}_k - \boldsymbol{1}\boldsymbol{\mu}_0^T)\boldsymbol{\Sigma}_0^{-1} \tag{2}$$

where $\boldsymbol{\mu}_0 \in \mathbb{R}^m$ denotes the reference mean, and the diagonal matrix $\boldsymbol{\Sigma}_0 \in \mathbb{R}^{m\times m}$ denotes the reference standard deviation, both of which are obtained using sufficient historical datasets under normal conditions. Then, the covariance matrix of $\bar{\boldsymbol{X}}_k$ can be approximated as follows:

$$\mathbf{cov}(\bar{\boldsymbol{X}}_k) \approx \boldsymbol{C}_k = \frac{1}{w}\bar{\boldsymbol{X}}_k^T\bar{\boldsymbol{X}}_k. \tag{3}$$

Furthermore, $\boldsymbol{C}_k = \boldsymbol{P}_k\boldsymbol{\Lambda}_k\boldsymbol{P}_k^T$, where the diagonal matrix $\boldsymbol{\Lambda}_k \in \mathbb{R}^{m\times m}$ denotes the eigenvalues of $\boldsymbol{C}_k$ in descending order, and $\boldsymbol{P}_k \in \mathbb{R}^{m\times m}$ denotes the corresponding eigenvectors. Then, the data matrix $\bar{\boldsymbol{X}}_k$ can be transformed into $\boldsymbol{T}_k = \bar{\boldsymbol{X}}_k\boldsymbol{P}_k$, where $\boldsymbol{P}_k$ denotes the loading matrix, and $\boldsymbol{T}_k \in \mathbb{R}^{w\times m}$ is the score matrix. Each column of $\boldsymbol{T}_k$ represents a corresponding TC, i.e., the linear combination of normalized measurement vectors. The statistical properties of TCs can reflect some important process features. Note that

$$\frac{1}{w}\boldsymbol{T}_k^T\boldsymbol{T}_k = \mathrm{diag}\{\lambda_{1,k}, \ldots, \lambda_{m,k}\} \tag{4}$$

where $\lambda_{i,k}$ denotes the ith largest eigenvalue of $\boldsymbol{C}_k$. This implies two remarkable properties of TCs for process monitoring: (i) TCs are orthogonal, which makes it more convenient for monitoring TCs than monitoring original measurement vectors; (ii) the sample variances of TCs are equivalent to the eigenvalues of the sample covariance matrix of the normalized data matrix $\bar{\boldsymbol{X}}$.