

Fault detection with principal component pursuit method

Yijun Pan, Chunjie Yang*, Youxian Sun, Ruqiao An and Lin Wang

Department of Control Science and Engineering, University of Zhejiang, Hangzhou Zhejiang, China

E-mail: cjyang@iipc.zju.edu.cn

Abstract. Data-driven approaches are widely applied for fault detection in industrial process. Recently, a new method for fault detection called principal component pursuit(PCP) is introduced. PCP is not only robust to outliers, but also can accomplish the objectives of model building, fault detection, fault isolation and process reconstruction simultaneously. PCP divides the data matrix into two parts: a fault-free low rank matrix and a sparse matrix with sensor noise and process fault. The statistics presented in this paper fully utilize the information in data matrix. Since the low rank matrix in PCP is similar to principal components matrix in PCA, a T^2 statistic is proposed for fault detection in low rank matrix. And this statistic can illustrate that PCP is more sensitive to small variations in variables than PCA. In addition, in sparse matrix, a new monitored statistic performing the online fault detection with PCP-based method is introduced. This statistic uses the mean and the correlation coefficient of variables. Monte Carlo simulation and Tennessee Eastman (TE) benchmark process are provided to illustrate the effectiveness of monitored statistics.

1. Introduction

Nowadays, the industrial process control system tends to be large-scale and complicated. In order to maintain the efficiency and reliability of the system, it's important to develop a fault detection method, which can monitor the process and detect the fault promptly. The fault detection method can be divided into three categories: model-based method, knowledge-based method and data-driven method [1]. Compared with model-based method and knowledge-based method, data-driven method can detect faults without the exactly models. Compared to other two methods, the model-based method is based on exact process models. The knowledge-based methods are based on the available knowledge of the process behavior and the experience of expert plant operators. The data-based process monitoring methods have no requirement of the process model and the associated expert knowledge. It has become more and more popular in recent years, especially in those complex industrial processes whose models and expert knowledge are difficult to build and obtain in practice. This method is suitable to complicated industrial process. The common applied data-driven methods are principal component analysis(PCA), partial least squares(PLS), support vector machine(SVM), artificial neural network(ANN) and some improved methods based on these basic methods. Process monitoring are divided into four steps: model building, fault detection, fault isolation and process reconstruction. An advantage of principal component pursuit(PCP) is that it can simultaneously accomplish the four steps of process monitoring. Besides, it's robust to outliers.



Candes first proposed the PCP method in detail in 2011[2]. A low rank matrix and a sparse matrix could be recovered from a data matrix by using a convex optimization algorithm. This method was able to realize the detection of objects in a cluttered background. Bouwmans reviewed the recent developments in the field of RPCA solved via PCP, and investigated how these methods were solved[3]. Then Isom indicated that PCA had several shortcomings compared with PCP, and PCP was robust to outliers[4]. In addition, PCP could simultaneously accomplish the objectives of model building, fault detection, fault isolation, and process reconstruction. This is the first time that the PCP-based method is used for process monitoring. Cheng proposed a new scaling preprocessing step that improved quality of data matrix in the sense of low coherence[5]. Moreover, a residual generator and a post-filter suitable for PCP generated process models were provided.

However, as far as we know, there are few appropriate monitored statistics to perform process monitoring with PCP-based method. The PCP is a useful matrix decomposition technique, and it can handle the data collected during some time. So its hard to perform online process monitoring in theory. In PCA-based method, a T^2 statistic is used for fault detection in principal component subspace. The principal component subspace is basically fault-free, which is similar to low rank matrix in PCP. In this paper, a T^2 statistic is introduced to be used for fault detection in low rank matrix. The process monitoring performance is better in PCP-based method than PCA-based method. PCA determines the number of principal components by calculating the cumulative variance proportion. But it's not sure that the total important information is put into the principal component subspace. In addition, the number of principal components must be an integer number, while the parameter λ can be any real number in PCP-based method. It can be adjusted little by little. This means that the adjustable range is flexible in PCP-based method. In PCA-based method, the important information is divided into principal component subspace and residual space. However, in PCP-based method, almost total important information is emerged in the low rank matrix. The low rank matrix is used for building the model. By using this monitored statistic, the fact that PCP is more sensitive to small variations in variables than PCA is illustrated. The reason why we compares the PCP with PCA is that they are both decomposition method. The PCA-based method divides the matrix into a principal component matrix and a residual matrix. The PCP-based method divides the matrix into a low rank matrix and a sparse matrix. Both principal component matrix and low rank matrix are fault-free.

In sparse matrix, the Q statistic in PCA is not suitable to fault detection. An online PCP-based process monitoring statistic that fully uses the information of noise and fault in sparse matrix is presented. The mean and correlation coefficient of variables in training matrix are used for building monitor statistic. In normal condition, the sparse matrix contains the noise. In abnormal condition, the sparse matrix contains sensor noises and process fault. Building a statistic by comparing the mean of variables in two sparse matrices. That can eliminate the influence of noises. So this statistic can increase the fault detection rate. This is the reason that this statistic is suitable to fault detection in sparse matrix. Both T^2 and new mean-coefficient statistics fully consider the advantage of PCP. The T^2 statistic uses the important process information, while the mean-coefficient statistic uses the noise and fault information. The total information in data matrix is used for fault detection. So these PCP-based statistics are appropriate to fault detection. The PCP-based method divides the data matrix into two parts including a low rank matrix and a sparse matrix. In actual industrial process, the variables are relevant. This means that the matrix is low rank only in the condition without sensor noise and process fault in process. This is the reason that low rank matrix is without fault data. The sparse matrix contains sensor noise and process fault. Finally, the monitored statistics are used for fault detection on numerical simulation and Tennessee Eastman(TE) process to prove the efficiency.

In section 2, a brief description of PCP is given. The proposed monitor statistic based on

PCP for process monitoring will be presented in section 3. Section 4 provides the simulation results on numerical simulation and TE process respectively. Section 5 gives the conclusion.

2. Principal component pursuit

Given a data matrix $X \in R^{n \times m}$ with n observations and m measurement variables. X can be decomposed by PCP as follows

$$X = A + E \quad (1)$$

where A is a low rank matrix and E is a sparse matrix. And $E_{ij} = 0$ means that the i th observation of j th variable is without sensor noise and process fault. The essence of PCP is to solve the optimization problem

$$\begin{aligned} \min \quad & \|A\|_* + \lambda \|E\|_1 \\ \text{s.t.} \quad & X = A + E. \end{aligned} \quad (2)$$

where $\|A\|_*$ is the nuclear norm of matrix A , and it is equal to the sum of singular values. $\|E\|_1$ is the l_1 norm of matrix E , and it is equal to the sum of magnitudes of all entries[2]. λ is a parameter used for optimizing algorithm. The value of parameter λ represents the importance of $\|A\|_*$ and $\|E\|_1$ in (2). The larger value of parameter λ is, the more sensor noise and process fault the sparse matrix contains. But if the value of parameter λ is too large, the relationship of data in low rank matrix A will be broken. On the contrary, the sparse matrix has little sensor noise and process fault, which leads to a low fault detection rate. So choosing an appropriate λ is important. In this paper, the parameter λ is chosen by trial. The model in (2) is a simple convex optimization problem. In this work, we solve it by singular value thresholding method[6].

3. Fault detection based on PCP

3.1. A monitored statistic for fault detection in low rank matrix

As mentioned above, the low rank matrix contains the information without sensor noise and process fault. This is similar to the principal components space in PCA-based method. So the T^2 statistic is used for fault detection in low rank matrix in PCP-based method. And with this monitored statistic, the PCP-based method is illustrated to be sensitive to small variations in variables by simulation.

The steps of PCP-based method for fault detection are as follows:

- Normalization
The normal training data matrix X is normalized by mean and variance of the corresponding variables.
- Matrix decomposition
The PCP is used for dividing the normalized training data matrix X^* into a low rank matrix A and a sparse matrix E by solving the convex optimization problem.
- Eigenvalue decomposition
Get the singular values diagonal matrix Σ and loading vector P via eigenvalue decomposition of low rank matrix A .
- T^2 statistic threshold of normal condition
The threshold of T^2 statistic in training data matrix can be determined by (3).

$$T_\alpha^2 = \frac{(n^2 - 1)m}{n(n - m)} \times [F_\alpha(m, n - m)] \quad (3)$$

where n, m are the dimension of matrix X^* and $F_\alpha(m, n - m)$ is the value of F-distribution at significance level with m and $(n - m)$ degrees of freedom [7].

- T^2 statistic of testing data matrix

The T^2 statistic of testing data matrix can be computed by (4).

$$T^2 = \mathbf{x}^T P \left(\sum \right)^{-2} P^T \mathbf{x} \quad (4)$$

where \mathbf{x} is an observation vector.

- Fault detection

If the T^2 statistic exceeds the normal threshold value, then a fault will happen.

3.2. A new monitored statistic for fault detection in sparse matrix

The data matrix is divided into two parts by PCP-based method: a low rank matrix and a sparse matrix. In training data matrix, the sparse matrix E_1 contains the sensor noise, while in testing data matrix, the sparse matrix E_2 contains the sensor noise and process fault. For the purpose of fault detection, the following assumptions is made. In the same operation condition, the magnitude and direction of noises are nearly the same in both training matrix and testing matrix. This statistic compares the mean of variables of two sparse matrices. This way eliminates the influence of noises, which improves the performance of fault detection. To get a normal condition threshold, the correlation coefficient of variables is calculated. This way realizes the online fault monitoring. Then the steps of building a new monitored statistic are as follows:

- Calculate the low rank matrix A_1 and sparse matrix E_1 of training matrix X_1 by PCP-based method.
- Calculate the correlation coefficient $c_i, i \in 1, 2, \dots, m-1$ of every variables with the first variable in training matrix X .
- Calculate the proportion $p_i, i \in 1, 2, \dots, m-1$ of every variables in correlation coefficient.

$$p_{i+1} = \frac{c_i}{1 + \sum_1^{m-1} c_i}, p_1 = \frac{1}{1 + \sum_1^{m-1} c_i} \quad (5)$$

- Calculate the normal threshold M . Multiply the proportion p_i by the mean of variables $d_i, i \in 1, 2, \dots, m$ in training matrix.

$$M = p_i \times d_i \quad (6)$$

- Use the last $(n-i) \times m, i \in 1, 2, \dots, n$ row data in training matrix and the first $i \times m$ row data in testing matrix to make up a new matrix D .
- Calculate the low rank matrix A_2 and sparse matrix E_2 of matrix D by PCP-based method.
- Calculate the monitored statistic G in testing matrix. Multiply the proportion p_i by sparse matrix E_2 .

$$G = p_i \times E_2 \quad (7)$$

- Get the last value in G to make up the new monitored statistic MG .

$$MG(i) = G(n) \quad (8)$$

- Repeat n times to get the monitored statistic MG
- If the statistic MG exceeds the threshold value M , then a fault will happen.

4. Simulation result

In this section, the simulation results of PCP-based method and PCA-based method are discussed. The new monitored statistics are used for fault detection on numerical simulation and TE process.

4.1. Numerical simulation

The purpose of this Monte Carlo simulation is to verify that T^2 statistic and mean-coefficient statistic are suitable to the fault detection. Besides, PCP is more sensitive to small change in variables. Monte Carlo simulation results are provided for the case of modest fault magnitudes by randomly assigning fault sensors and fault magnitudes. In 2009, Alcalá used the Monte Carlo simulation to build a numerical simulation example, which is suitable to test statistics[8]. The process model to be used is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} -0.2310 & -0.0816 & -0.2662 \\ -0.3241 & 0.7055 & -0.2158 \\ -0.217 & -0.3056 & -0.5207 \\ -0.4089 & -0.3442 & -0.4501 \\ -0.6408 & 0.2105 & -0.2372 \\ -0.5655 & -0.433 & -0.5938 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} + noise \quad (9)$$

where t_1 , t_2 and t_3 are zero-mean random variables with standard deviations of 1, 0.8 and 0.6, respectively. The noise included in the process is zero-mean with standard deviation of 0.2 and is normally distributed. The simulated testing data matrix is of the form

$$x_{test} = x^* + \varepsilon_i f \quad (10)$$

where x^* is generated according to the model given above and the fault magnitude f is a random number uniformly distributed between 0 and 0.9. The chosen fault magnitude is small. It not only can illustrate the effectiveness of T^2 statistic, but also illustrate that PCP is more sensitive to small variations. Also, the direction ε_i is uniformly random out of the six possible variable directions. In this section, total six variables are used for building the model.

4.1.1. The T^2 statistic for fault detection in low rank matrix The fault detection results of PCA-based method and PCP-based method for this process model are shown in Fig.1- Fig.2. The fault detection rate (FDR), fault alarm rate (FAR) of T^2 statistic are shown in Table 1. In PCA-based method, the T^2 and Q statistics are used for fault detection.

In this simulation, a small change in the variables which works as a process fault is introduced in testing data matrix. The testing data matrix consists of 100 normal data and 500 fault data. In PCA-based method, the number of principal components is three by calculating the accumulative variance contribution rate. The T^2 statistic threshold of normal condition is equal to 0.3468, and the Q statistic threshold of normal condition is equal to 8.4063×10^{-30} . In PCP-based method the parameter λ is equal 0.05. The T^2 statistic threshold of normal condition is equal to 0.6986.

From Fig.1-Fig.2 and Table 1, the PCP-based method gets higher FDR and lower FAR than PCA-based method. This shows that the T^2 statistic is suitable to fault detection in low rank matrix based on PCP. And PCP-based method is more sensitive to small variations.

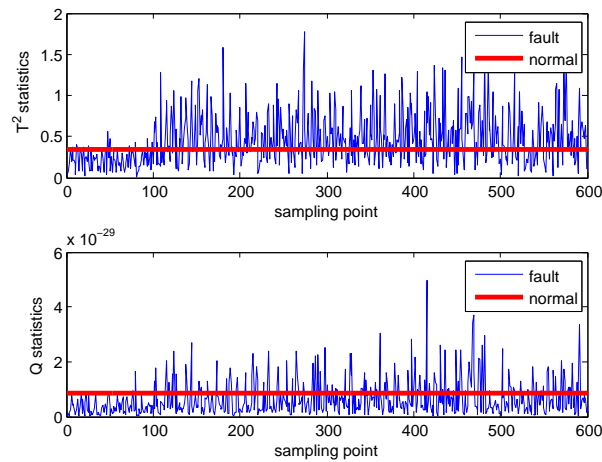


Figure 1. The T^2 statistic of PCA-based method for fault detection in numerical simulation

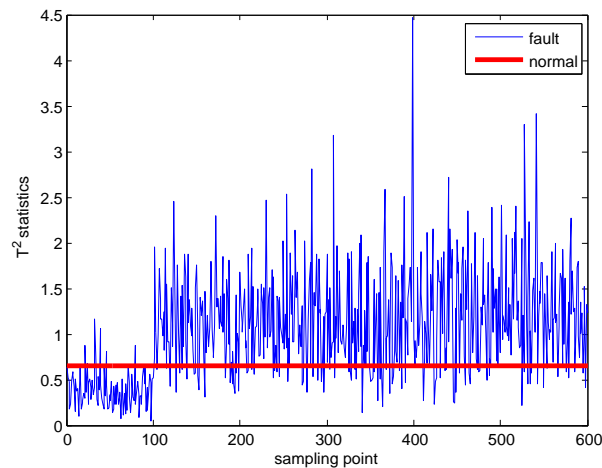


Figure 2. The T^2 statistic of PCP-based method for fault detection in numerical simulation

Table 1. The FAR and FDR of PCP and PCA.

Item	FDR	FAR
PCA	FDR	FAR
PCP	81.20	8.00

4.1.2. The new monitored statistic in sparse matrix In this part, an online monitored statistic is presented. The fault detection results with mean-coefficient monitored statistic are shown in Fig.3, and fault detection rate (FDR), fault alarm rate(FAR) are shown in Table 2. In this part, the fault magnitude f is a random number uniformly distributed between 0 and 5, which is the same as the original condition in (8). And the parameter λ is equal to 1.5. The threshold of normal condition is equal to 5.3434×10^{-4} . This new statistic can deal with online fault detection, which overcomes the main shortcoming of PCP-based method. However, this new statistic can't simultaneously deal with the four steps of process monitoring.

From Fig.3 and Table 2, we can obtain that the new statistic in sparse matrix performances

better than the T^2 statistic in low rank matrix in numerical simulation. This means that the new PCP-based statistic is suitable to fault detection.

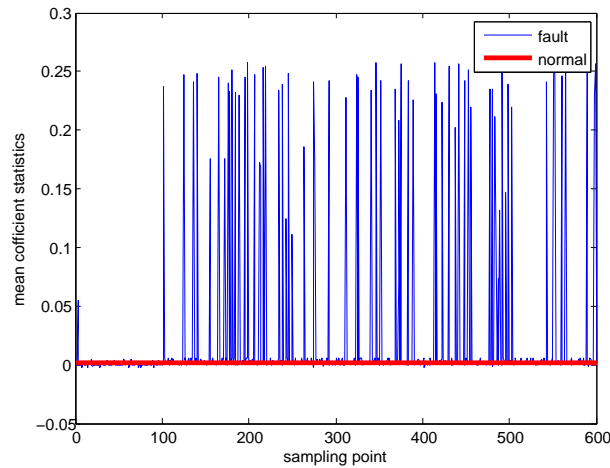


Figure 3. The mean-coefficient statistic of PCP-based method for fault detection in numerical simulation.

Table 2. The FAR and FDR of PCP and PCA.

Item	PCP
FDR	98.60
FAR	2.30

4.2. TE process simulation

4.2.1. TE process In this section, due to the actual industrial process has few fault data, TE process simulated the actual industrial process is used. This process contains 41 measured variables and 12 manipulated variables. In these 53 variables, reactor agitator speed remains constant. Therefore, reactor agitator speed is removed from the variables, remaining 52 variables.

There are 21 process faults in the TE process, and Chiang described it in detail[9]. The 21 faults in TE are introduced in Table 3. Two sets of data matrices are collected from the TE process - training and testing data matrix. To collect the data matrix, 22 running condition (1 normal condition and 21 abnormal conditions) are performed. In data matrix, the fault is introduced at 8th hour with a sampling interval of 3 minutes. Hence, the data matrix is composed of 22 separate data matrices containing 1 normal condition data matrix and 21 fault data matrices, each having dimension of 960×52 (160 normal data/800 fault data). In this part, the total 52 measured and manipulated variables in TE process are used for building the model in PCA-based and PCP-based method.

Table 3. Description of Faults in the TE Process.

no.	description	type
1	A/C feed ratio, B composition constant(stream 4)	step
2	D feed temperature (stream 2)	step
3	B composition, A/C ratio constant (stream 4)	step
4	reactor cooling water inlet temperature	step
5	condenser cooling water inlet temperature	step
6	A feed loss (stream 1)	step
7	C header pressure loss?reduced availability(stream 4)	step
8	A, B, and C feed composition (stream 4)	random variation
9	D feed temperature (stream 2)	random variation
10	C feed temperature (stream 4)	random variation
11	reactor cooling water inlet temperature	random variation
12	condenser cooling water inlet temperature	random variation
13	reaction kinetics	slow drift
14	reactor cooling water valve	sticking
15	condenser cooling water valve	sticking
16-20	unknown	unknown
21	valve for stream 4 was fixed at steady-state position	constant position

4.2.2. Simulation results in low rank matrix The fault detection results of PCA and PCP-based method for fault 1, 5 are shown in Fig.4-Fig.7. The parameters of simulation are shown below. In Fig.4, the T^2 statistic threshold of normal condition is equal to 4.8430. The parameter λ is equal to 0.05. In Fig.5, the T^2 statistic threshold of normal condition is equal to 2.1172, and the Q statistic threshold of normal condition is equal to 17.3456. The number of principal components is 24. In Fig.6, the T^2 statistic threshold of normal condition is equal to 7.8914. The parameter λ is equal to 0.05. In Fig.7, the T^2 statistic threshold of normal condition is equal to 2.1172, and the Q statistic threshold of normal condition is equal to 17.3456. The number of principal components is 24.

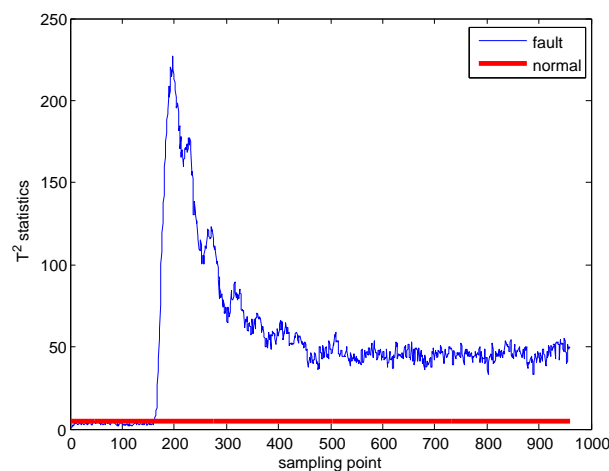


Figure 4. The T^2 statistic in low rank matrix of PCP-based method for fault detection of fault 1 in TE process.

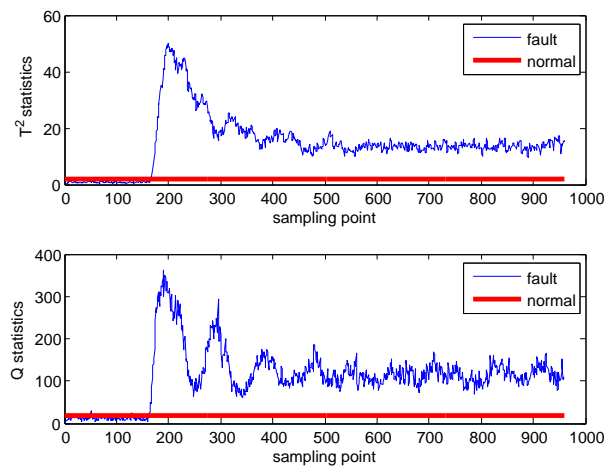


Figure 5. The T^2 and Q statistic of PCA-based method for fault detection of fault 1 in TE process.

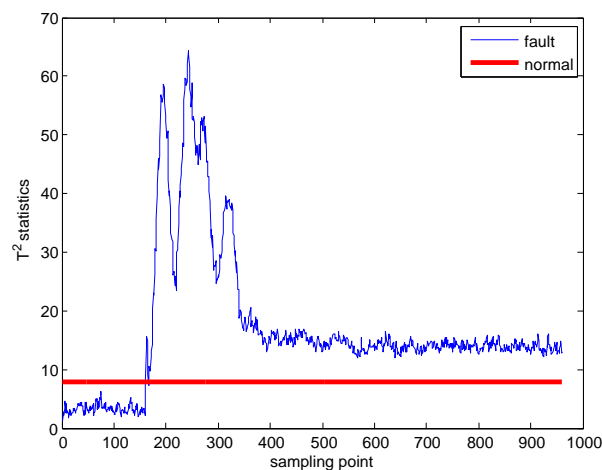


Figure 6. The T^2 statistic in low rank matrix of PCP-based method for fault detection of fault 5 in TE process.

Fault 5 is associated with step change in the condenser cooling water inlet temperature. The PCA-based method for fault 5 in Fig.7 only detects the fault from 160th to 350th. Since the control loops compensated some of the changes in the variables and only small variations remained, the PCA-based method fails to detect the small variations. PCP-based method for fault 5 in Fig.6 can detect the fault from 161th to 960th, which means that the PCP method is more sensitive to small variations in industrial process. And the T^2 statistic is suitable to fault detection with PCP-based method. The FDR and FAR of faults by using T^2 statistic in TE process are shown in Table 4. In most faults, the PCP-based method gets better results than PCA-based method. The average values of FAR and FDR of two methods also illustrate that PCP-based method is better.

According to the numerical simulation and TE process simulation, the PCP-based method gets higher FDR and lower FAR than PCA-based method. This means that the T^2 statistic is suitable to fault detection in PCP-based method than PCA-based method. PCA-based method determines the number of principal components by calculating the cumulative variance

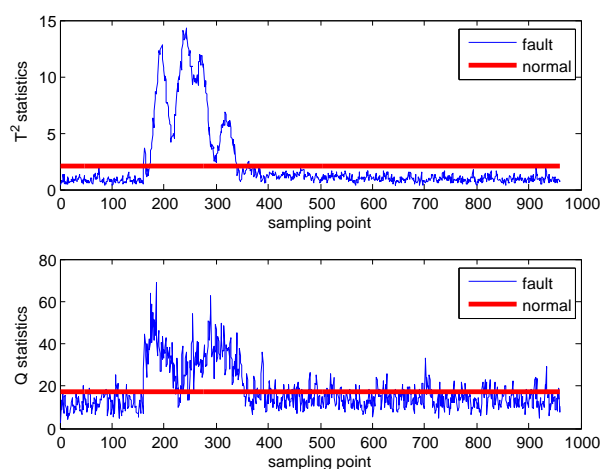


Figure 7. The T^2 and Q statistic of PCA-based method for fault detection of fault 5 in TE process.

proportion. The number of principal components must be an integer number. This means that the adjustable range is inflexible. So the important information is put into principal component subspace and residual space. However, in PCP-based method the parameter λ is a real number. It can be adjusted little by little, so the noises and faults are almost isolated from low rank data matrix.

Table 4. The FAR and FDR of PCP and PCA

Fault	FAR-PCP	FAR-PCA	FDR-PCP	FDR-PCA
1	0.00	0.00	98.75	99.25
2	1.25	0.00	98.38	98.25
3	6.25	97.25	16.75	11.85
4	1.25	2.50	91.25	56.50
5	0.00	2.50	99.50	33.00
6	0.00	1.87	99.50	99.12
7	5.63	0.00	86.50	99.75
8	10.00	2.50	87.88	97.88
9	23.13	15.00	16.63	11.13
10	3.13	1.87	83.13	50.75
11	5.63	11.25	76.38	70.50
12	5.63	16.25	99.00	99.25
13	1.97	4.37	94.50	95.87
14	5.63	8.75	99.88	99.75
15	10.63	6.25	23.13	26.12
16	5.00	41.25	85.12	48.25
17	6.88	6.88	84.25	88.50
18	5.00	8.13	89.88	91.87
19	11.87	3.76	51.38	36.25
20	5.00	4.37	77.50	60.75
21	11.87	16.25	53.13	50.75

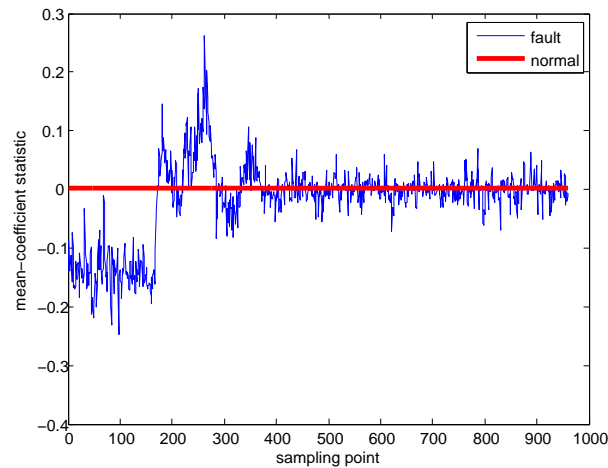


Figure 8. The mean-coefficient statistic in sparse matrix of PCP-based method for fault detection of fault 1 in TE process.

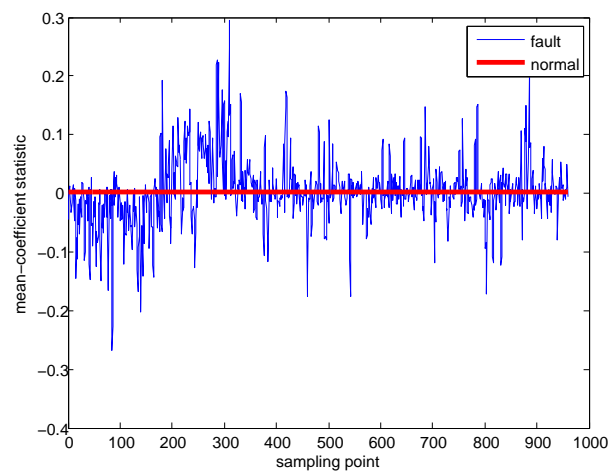


Figure 9. The mean-coefficient statistic in sparse matrix of PCP-based method for fault detection of fault 2 in TE process.

4.2.3. Simulation results in sparse matrix In this part, the fault detection results with new monitored statistic in sparse matrix for fault 1 and 2 in TE process are also presented in Fig.7-Fig.8. In Fig.7-Fig.8, the threshold of normal condition is equal to 0.0014. The parameter λ is equal to 0.5. The fault detection performance in low rank matrix is better than that in sparse matrix. Since the performance of fault detection is not well and the amount of calculation is huge, the FAR and FDR are not presented. There are two reasons to explain this poor performance. First, the values of parameter λ is not appropriate. Second, the assumption made in part 'A monitored statistic in sparse matrix' is hard to realize. Because in actual process, the noise is uncontrollable. Although the fault detection performance is not well, this method realizes the online monitoring in sparse matrix, which is more important.

5. Conclusion

In this paper, the monitored statistics based on PCP used for fault detection are introduced. In low rank matrix, a T^2 statistic is introduced. This statistic gets good performance. And by this statistic, the fact that PCP-based method is more sensitive to small variations is illustrated. In sparse matrix, an online mean-coefficient statistic is presented. The two statistics fully use the information in data matrix. The numerical simulation and TE process illustrate the effectiveness of these two monitored statistics. And there still exists large room to improve. While using the PCP to get two matrices, the parameter λ is chosen by trial. The value of λ is important to simulation results. In further work, the algorithm of how to get the optimal parameter will be discussed. And how to improve fault detection performance in sparse matrix is should also be considered.

References

- [1] Z. Q. Ge, Z. H. Song, and F. R. Gao, Review of recent research on data-based process monitoring, *Industrial and Engineering Chemistry Research*, vol. 52, no. 10, pp. 3543-3562, 2013
- [2] E. J. Candes, X. Li, Y. Ma, and J. Wright, Robust principal component analysis? *Journal of ACM*, vol. 58, no. 3, 2011.
- [3] T. Bouwmans and E. H. Zahzah, Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance, *Computer Vision and Image Understanding*, vol. 122, pp. 22-34, 2014.
- [4] J. D. Isom and R. E. LaBarre, Process fault detection, isolation, and reconstruction by principal component pursuit, in *2011 American Control Conference*, San Francisco, CA, USA, 2011, pp. 238-243.
- [5] Y. Cheng and T. W. Chen, Application of principal component pursuit to process fault detection and diagnosis, in *2013 American Control Conference*, Washington, DC, USA, 2013, pp. 3535-3540.
- [6] J. F. Cai, E. J. Candes, and Z. W. Shen, A singular value thresholding algorithm for matrix completion, *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956-1982, 2010.
- [7] H. L. Chiang, L. E. Russell, and D. R. Braatz, Fault diagnosis in chemical processes using fisher discriminant analysis, discriminant partial least squares, and principal component analysis, *Chemometrics and Intelligent Laboratory Systems*, vol. 50, pp. 243-252, 2000.
- [8] C. Alcalá and S. Qin, Reconstruction-based contribution for process monitoring, *Automatica*, vol. 45, no. 7, pp. 1593-1600, 2009.
- [9] L. Chiang, E. Russell, and R. Braatz, *Fault detection and diagnosis in industrial systems*. Springer Verlag, 2001.