Fault Diagnosis of Multi-level Principal Component Space Based on High-Dimensional Feature Representation

Caixia Zhang¹,², Shaolin Hu³, Jing Guo¹,² and Xiangdong Wang¹

¹ Foshan University, Foshan, China
² Guangdong Province Smart City Infrastructure Health Monitoring and Evaluation Engineering Technology Research Center, Foshan, China
³ Guangdong University of Petrochemical Technology, Maoming, China
Email: 380351447@qq.com

Abstract. Due to the influence of space equivalent representation and pattern composition, traditional principal component analysis (PCA) is faced with challenges in identifying small faults. This paper proposes a fault diagnosis method is proposed based on multi-levels principal component space based on high-dimensional feature representation. Firstly, based on the original PCA, the dimensions are expanded first and then reduced, so that the information that cannot be expressed in the original space is fully expressed. This method further divides the principal element space based on the eigenvalues, and projects the occurrence of the event and the size of the identification abnormality according to the projection to different sub-spaces. Compared with the usual fault detection methods, simulations verify that this algorithm can effectively measure the level of faults and improve the correct detection rate of minor faults.

1. Introduction

Fault diagnosis is very important to improve the safety and reliability of complex industrial systems. Fault separation and identification is the goal of fault diagnosis [1-2]. The key to deal with faults with corresponding measures is how to quickly detect faults, increase the detection rate of faults and estimate the size of the faults.

Principal component analysis [3-5] is a commonly means to deal with fault diagnosis. The multivariate projection technique is used to disintegrate the sample space into residual space and low-dimensional latent variable space. Statistics are built in these two subspaces and the faults are diagnosed. Since the process operation data is analysed after being transformed from the measurement space to the feature space, then a serious mode compound effect will arise, which making it hard to interpret fault modes that brings about the abnormal system. Moreover, the dimension reduction is divided into residual space and principal component space, and the ability to distinguish insignificant patterns is also affected.

For the problem of model compounding effect of PCA, Ref. [6] proposes the concept of specified meta-analysis designated component analysis (DCA), in which the projection data is projected onto a specified pattern to determine the significance of each specified element and judge whether the corresponding fault has occurred. Zhou et al. [7] proposed a stepwise DCA analysis algorithm and solved the orthogonal diagnosis among DCA groups in a specified mode, but it is still non-orthogonal in the group. In order to overcome the shortcomings of traditional PCA due to different dimensions and standardization processes, Wen et al. [8] introduced the relative importance of different variables into the system and established a method called RPCA. From the perspective of subspace projection,
this method extracts subspaces that can reflect both the importance of variables and the main changes in the sample. Zhao [9] applied the RPCA method to the fault diagnosis of the flight control system to realize the accurate identification and locating of the stuck fault of the actuator. Hu et al. [10] proposed a method which calculate the proportion factor based on important contribution rate, sensitivity coefficient and fault sensitivity. The ability of fault detection of important variables can be significantly improved by this way.

The limitation of the mode compound effect of PCA in the diagnosis of multi-level micro faults, Ref. [11] proposed a minor fault diagnosis in the multi-stage projection PCA framework. In consideration of small deviation between the small fault and the normal state, Hu [12] subdivided the residual subspace via determining the principle of subspace partition based on the size of eigenvalues and eigenvector and proposed a multi-space detection method based on PCA, which can improve the fault’s significance in the residual subspace. Ref. [13] is based on the summary of data-driven micro-fault diagnosis methods, and looks forward the following three perspective: adding new information, mining un-utilized implicit information and adopting new mathematical tools.

These solutions all improve the ability to monitor faults based on the PCA method to some extent. Aiming at the problem of fault pattern recognition and dimension reduction caused by mode complex effect of traditional PCA, which can make small faults unable to be fully expressed, this paper proposes fault diagnosis of multi-level principal element space. By this method, the dimension is expanded and then is reduced, and then divides the main variable space after dimension reduction into different levels of subspace and residual space. Our method expands the research of fault diagnosis from two aspects: Fault detection capability could be enhanced; Detection performances are improved based on the ability to project low-level faults in the main element space to high-level space.

This article elaborates on the following four sections. Firstly, the apply about PCA algorithm in fault diagnosis is introduced. In second part, how to implement the high-dimensional feature frame based on the basic PCA algorithm is completed. In the third part, the classification of principal component space is proposed under the high-dimensional feature representation. The fourth part gives the experiment simulation and the effectiveness of the algorithm is verified.

2. Principal Component Analysis (PCA)
PCA method usually treats the data of the original variable space through zero mean unit variance, and the covariance matrix $S$ obtained is as follows:

$$S = \text{cov}(Z) \approx \frac{1}{m-1} \bar{Z}^T \bar{Z} = V_0^T \Lambda V_0$$

(1)

We calculate the eigenvectors $V$ and eigenvalues $\gamma$ of the covariance matrix, the eigenvalues are arranged in descending order of $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_m \geq 0$, and the eigenvectors are arranged in order, too.

Then $V = (V_1, V_2, V_3, \cdots V_n)$ constitutes a set of standard orthonormal basis sets in the sample equivalent projection space.

The original variable can be decomposed as follows:

$$Z = \bar{Z} + \bar{Z} = TV^T + \bar{Z} = ZV^T + \bar{Z}$$

(2)

$t$ is a hidden variable:

$$t = zV$$

When it is used for fault detection, Hotelling T2 and Square Prediction Error (SPE) are used to detect if an abnormality has occurred in the process.

$$z^T (\Sigma)^{-1} z = \frac{t_1^T t_1}{\gamma_1} + \frac{t_2^T t_2}{\gamma_2} + \cdots + \frac{t_A^T t_A}{\gamma_A} + \frac{t_{A+1}^T t_{A+1}}{\gamma_{A+1}} + \cdots + \frac{t_m^T t_m}{\gamma_m}$$

(3)
Among them,

\[ T^2 = \frac{t_1^T t_1 + t_2^T t_2 + \cdots + t_A^T t_A}{\gamma_1 + \gamma_2 + \cdots + \gamma_A} \]  

(4)

The variation of variables in the principal component space is measured by \( T^2 \), obeying \( \chi^2 \) distribution, the degree of freedom of the distribution is \( h \), and the confidence level is \( \alpha \).

And because the residual space statistics are

\[ SPE = \tilde{z}^T \tilde{z} = z^T (I - PP^T)(I - PP^T)z = z^T (I - PP^T)z = \sum_{i=1}^{m} t_i^T t_i \]  

(5)

The new test data are projected to each subspace, and statistics are obtained. Compared the statistics with the corresponding control limits, then we can judge whether the process has an abnormality.

3. Formation of High Dimensional Feature Representation

3.1. High-Dimensional Feature Frame

When the traditional algorithm of PCA is used for fault detection, a fault with a small sign that is submerged in noise or a change of a large normal process may appear inapplicable. The information is fully expressed by expanding the dimension, and adding the projections to the original projection. Thus, \( i \)-level projection frame of high-dimensional feature representation is formed, and a small fault detection research is performed. Based on the eigenvalues obtained in (1) and the corresponding eigenvectors, a vector is inserted in the middle of the linear connection lines between two adjacent base vectors to obtain a first-level feature representation frame:

\[ \tilde{V}_1 = (V_{01}, \alpha_0 V_{01} + \beta_0 V_{02}, V_{02}, \cdots, V_{(n-1)}, \alpha_{0(n-1)} V_{0(n-1)} + \beta_{0(n-1)} V_{0n}, V_{0n}) \]

Insert two vectors in the middle of the linear connection between two adjacent base vectors to obtain the second-level feature representation frame:

\[ \tilde{V}_2 = (V_i, \alpha_i V_i + \beta_i V_{i+1}, \alpha_{2_i} V_{i+1} + \beta_{2_i} V_{i+2}, V_{i+2}, \cdots, V_{n}, \alpha_{n-1} V_{n-1} + \beta_{n-1} V_n, V_{n}) \]

Similarly, according to the first and two projection frame generation method, the standard orthogonal sorting base generated based on the sample set is connected end to end, and the following vectors are inserted in the middle of the linear connecting line between the two adjacent base vectors, which form the first, second, ..., to \( i \)-level projection frame as \( V_i = (V_{01}, V_{02}, \cdots, V_{(i+1)(n-1)}) \).

Further, a polyhedron is formed based on adjacent base vectors, and interpolation of different densities is performed inside the polyhedron to form a corresponding high-dimensional feature frame.

3.2. Control Limits under High-Dimensional Feature Representation

After the original data \( X \) is projected to the corresponding high-dimensional feature frame, new data is obtained.

\[ T^{2}\langle i \rangle = t_i^T \Lambda^{-1} t_i = z_i^T V_i^T \Lambda^{-1} V_i^* z_i \]

\[ T^{2*}\langle i \rangle = t_i^* T \Lambda^{-1} t_i = z_i^* T V_i^* \Lambda^{-1} V_i^* z_i \]  

(6)

Frame statistics \( T^{2}\langle i \rangle \) of new projection are computed as follows:

\[ T^{2*}\langle i \rangle = t_i^T \Lambda^{-1} t_i = z_i^T V_i^T \Lambda^{-1} V_i^* z_i \]  

(7)

The statistic \( SPE^*(i) \) of the residual subspace is:
are the thresholds for the standard normal distribution, that is,

\[ z_i = z_i\left( t_{1-\alpha} \right) \]

and the

\[ t_{1-\alpha} \]

where

\[ \alpha \]

and the length of each load vector is normalized, that is

\[ \|V_i \|^2 = \sum_{j=1}^{b_i} \lambda_j \]

The two control limits in the new projection coordinate system are:

One control limit is \( T_\alpha^{i2} (i) \):

\[ T_\alpha^{i2} (i) = \frac{b^*(n-1)}{n-b^*} F_{b^*,n-b^*,\alpha} \]

where \( F_{b^*,n-b^*,\alpha} \) is the Critical value of F distribution, and the degree of freedom is \( b^*,n-b^* \), confidence degree is \( \alpha \).

Other control limit is \( Q_\alpha (i) \):

\[ Q_\alpha (i) = \theta \left( \frac{2 \theta_i h_0}{\theta_i^2} + 1 + \frac{\theta_i h_0 (h_0 - 1)}{\theta_i^2} \right)^{1/h_0} \]

where \( \theta_i = \sum_{j=b+1}^{m} \gamma_j^2 \), \( h_0 = 1 - 2 \theta_i / 3 \theta_i^2 \), \( C_\alpha \) are the thresholds for the standard normal distribution at the confidence coefficient \( \alpha \).

4. Multi-level Principal Element Space under High-Dimensional Feature Representation

The original data is represented to a high-dimensional feature by a data matrix \( Z_i \) after the projection frame is projected. Each column of the matrix is standardized by zero mean and unit variance. Main element subspace is subdivided according to the cumulative variance contribution rate, and \( k \) is the main element space number. \( V_i \) is a load matrix of the \( i \)-th principal element subspace, which is composed of \( b_{ij}+1 \)-th to \( b_j \)-th eigenvectors of the covariance matrix \( S \) of the sample \( Z_i \) after normalized projection data of the high-dimensional projection frame. \( b_i \) is the sum of the number of principal elements in the first \( i \) primary meta-subspaces, and the number of each subspace principal elements is determined by cumulative variance contribution method [14]. The columns of \( V = [v_1, v_2, \cdots, v_r] \) and \( u \) are orthogonal to each other, and the length of each load vector is normalized, that is

\[ V_i^TV_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}; V_j \in R^{m \times (b_{ij}+1)} \]

\( b_{ij} \) is the principal element number of \( i \)-th principal element subspace, and also score matrix column. The principal element score matrix as \( T \in R^{m \times b_i} = [t_1, t_2, \cdots, t_{b_i}] \). The columns of which are orthogonal to each other, that is, \( i \neq j \) , \( t_i^Tv_j = 0 \). \( E \) is a residual information. Covariance matrix of principal component is \( \Lambda = diag(\gamma_1, \gamma_2, \cdots, \gamma_{b_i}) \), where \( \gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_{b_i} \) denotes the first \( b_i \) large eigenvalues of \( S \).

\[
\begin{bmatrix}
Z = \sum_{i=1}^{k} T V_i^T + E \\
T_i = Z V_i
\end{bmatrix}
\]


The conventional PCA can detect whether an abnormality occurs in the process according to the data of each space exceeds the control limit or not, but it cannot obtain the abnormal size information. In the hierarchical principal element space, when a small fault is projected into a space with large data changes, it cannot be detected due to pattern recombination, but the larger fault can be detected in each hierarchical space. Based on this performance, we can detect the size of the fault based on spatial anomaly. At the same time, this classification is a multi-level principal element space under the representation of high-dimensional features, and can also improve the inspection rate of minor faults.

5. Simulation Research

5.1. Offline Training

5.1.1. Data Pre-processing. Zero-means and unit-variance normalization for each column of the obtained normal history data matrix.

5.1.2. The Eigenvector Matrix V. The covariance matrix of process variables is transformed into the characteristic matrix V with diagonal form by similarity transformation.

5.1.3. Expansion. \(V_i\): A new eigenvector matrix \(V_i\) is formed by adding \(i\) vectors between two adjacent columns of eigenvector matrix.

5.1.4. New Projection Data. The original data is projected to the column vector of the new eigenvector matrix \(V^R\) as coordinates to obtain a new set of projected data \(Z_i\).

5.1.5. Data reprocessing. After the projection data matrix \(Z_i\) is obtained, the zero mean value and unit variance of each column of the matrix are standardized.

5.1.6. A Multi-level Principal Component Analysis Model Is Established for the Standardized Data. According to the contribution rate of the eigenvalues of the covariance matrix, the variable space is divided into \(k\) principal subspaces and a residual subspace, which is used as a multi-level principal component analysis projection model.

5.2. Simulation

In order to verify the effectiveness of the method, the following system variables are constructed considering the combination of small faults under normal noise or normal variation:

\[
\begin{align*}
    z_1(k) &= 10 + 0.2 \times \text{randn}(1,1) \\
    z_2(k) &= 20 + 0.05z_1(k) + 0.8 \times \text{randn}(1,1) \\
    z_3(k) &= 10 + 5 \times z_2(k) - 2 \times z_1(k) + 0.2 \times \text{randn}(1,1) \\
    z_4(k) &= 20 + 0.3 \times \text{randn}(1,1) \\
    z_5(k) &= 10 - 1.3 \times z_1(k) + 0.2 \times z_2(k) + 0.8 \times z_3(k)
\end{align*}
\] (13)

Matlab generates a random number of 1 row and 1 column such as randn (1,1). PCA model is established by 4000 normal data. Expand model dimension according to the actual demands, and establish the principal element model after the dimension expansion. Then a multi-level principal element model is set up, control limits of each space are calculated separately. Then 500 samples are selected as test data, and constant amplitude deviation faults of a certain amplitude are introduced at times 201-300, 301-400, and 501-600 of the variable \(z_1, z_3, z_5\) of the test data. The multi-level principal component space model under the high-dimensional frame was used for detection, and the results were compared with the results of the model without high-dimensional frame projection.
Figures 1 and 2 show that the principal component subspace and residual subspace at all levels can detect anomalies when the fault level is high. Figures 3 and 4 show the faults in the first principal element space when the fault is in the medium level. Although some sampling points can be detected after the dimension expansion (figure 4), the false negative rate is very large. However, the anomalies could be founded in 2nd and 3rd principal element and residual subspaces. Figure 5 shows that there is no high-dimensional feature representation when there is a relatively small fault, and it cannot be detected in all principal element spaces and residual subspaces. Figure 6 shows the high-dimensional feature, and corresponding control limits are projected and calculated. The simulation results show that the fault can be detected better in residual subspace than the traditional PCA method, and the detection rate is improved.

Figure 1. Larger faults.

Figure 2. Large faults (After expanding the dimension).

Figure 3. Medium faults.
Figure 4. Medium faults (After expanding the dimension).

Figure 5. Small faults.

Figure 6. Minor faults (After expanding the dimension).

6. Conclusion
In this paper, a preliminary exploration is made on the problems of identification of fault size and the diagnosis of minor faults for traditional PCA. A new method is proposed to expand the redundancy of dimension to represent the system features, and then to reduce the dimension, so that the information that can not be expressed in the raw space can be fully expressed, and detection rate of small faults can be improved, and the method of further dividing the principal element space based on the eigenvalues is proposed. Online detection data is projected to different subspaces step by step. By comparing the control limits with different subspaces, it is possible to detect whether the monitoring range is faulty and which space a fault is in and then identify the level of the fault, so appropriate measures can be taken. Experimental results show that the algorithm can detect small faults that cannot be detected in one-dimensional space, and improve the ability of fault detection.
Acknowledgements
This work is supported by The Project of Department of Education of Guangdong Province (2019KTSCX192), Guangdong basic and Applied Basic Research Fund project (2020B1515310003), Foshan Core Technology Research Project (1920001001367), Guangdong Quality Safety of Grain and Oils Big Data Engineering Technology Research Center, National Natural Science Foundation of China Youth Fund (61803087).

References
[1] Zhou D H, Li G and Li Y 2011 Data-Driven Industrial Process Fault Diagnosis: Principal Component Analysis and Partial Least Squares (Science Press).
[2] Zhou D, Li G and Qin S J 2010 Total projection to latent structures for process monitoring AIChe E Journal 56 (1) 168 -178.
[3] Shen Y and Ding S X 2010 PCA-based fault diagnosis techniques SysTol 2010 (Nice, France) pp 179-184.
[4] Zhang M, Wang T and Tang T 2018 A fault detection method of current machine based on fractal and PCA of BP network Electric Machines and Control 22 (02) 79-88.
[5] Tang Y and Xiong Y 2018 Transformer fault diagnosis based on feature extraction of relative transformation principal component analysis Journal of System Simulation 30 (03) 1127-1133+1144.
[6] Tang Y and Xiong Y 2018 Transformer fault diagnosis based on feature extraction of relative transformation principal component analysis Journal of System Simulation.
[7] Zhou F N 2009 DCA based multiple faults diagnosis method Acta Automatica Sinica 35 (7) 971-982.
[8] Wen C L 2008 Relative PCA with applications of data compression and fault diagnosis Acta Automatica Sinica 34 (9) 1128-1139.
[9] Zhao D, Wang S and Wu Y J 2010 Fault diagnosis of flight control system based on relative principal component analysis Journal of Sichuan Armeries 31 (11) 29-31.
[10] Hu J, Wen C L and Li P 2013 A method for determining specific gravity factors in relative principal component analysis CAC2013 (Miami, Florida).
[11] Zhou F N, T H Tang and C L Wen 2008 DCA based multi-level small fault diagnosis ICCT 2008 (Hangzhou, China) pp 486-489.
[12] Hu J, Wen C and Li P, et al. 2014 Multi-space PCA with its application in fault diagnosis CCC 2014 pp 3311-3316.
[13] Wen C L, Lv F Y, Zhe J B, et al. 2016 A review of data driven-based incipient fault diagnosis Acta Automatica Sinica 42 (9) 1285-1299.
[14] Hu J 2015 Fault Diagnosis and Quality Monitoring Based on Multivariate Statistical Analysis (Zhejiang University).