PERFORMANCE MONITORING BASED ON CHARACTERISTIC SUBSPACE

Ming Guo Shuqing Wang

National Key Laboratory of Industrial Control Technology

Zhejiang University, Hangzhou, 310027, P.R.China

Abstract： In the operation and control of chemical process, automatic data logging systems produce large volumes of data. It is important for supervising daily operation that how to exploit the valuable information about normal and abnormal operation, significant disturbance and changes in operational and control strategies. In this paper, principal component analysis (PCA) is clarified its essence from the view of space, and every different subspace represents different operational mode and process performance. Based on that, distance between two subspaces is calculated to evaluate the difference between them. The method is illustrated by a case study of a fluid catalytic cracking unit (FCCU) reactor-regenerator system.

Keywords： Principal component analysis, statistical process control, space distance, process performance monitoring, FCCU.

1. INTRODUCTION

Advances in computer technology and application of advanced control theory have resulted in routine collection and storage of large volumes of data in chemical plant. Massive amounts of stored data can be used for analysis of the process operation and previous occurrences of abnormal situation. Principal component analysis (PCA) can extract valuable information from large historical database. Notable applications of PCA in chemical engineering have been in process monitoring (Nomikos and MacGregor, 1995; Kresta, *et al.,* 1991), disturbance detection (Ku and Storer, 1995), sensor fault diagnosis (Wang and Song, 2002) and process fault diagnosis (Kano, *et al.,*2001; Dunia and Qin,1998; Zhang, *et al.*,1996).

As far as process fault diagnosis is concerned, statistical process monitoring via PCA involves the use of Hotelling T^2 and Q (also known as Square Prediction Error or SPE) charts. Fault is identified with contributions of process variables to SPE. It is only valid for simple fault situation, and difficult to identify the root causes. Zhang and Martin (1996) proposed fault direction to identify different fault. Fault diagnosis is achieved by comparing the direction of the current on-line measurements with those of a database of known trajectories of identified faults. This method based on angle measurement does not make full use of principal component information of faults, and only the first loading vector is used. Dunia and Qin (1998) analyze the detectability, identifiability and reconstructability of faults using subspace approach. But they assume that the fault effect is not propagated into the other variables, which restricts its application. Kano, *et al.* (2001) proposed a novel statistical process monitoring method based on changes in the subspace which is spanned by several principal components. The method makes use of principal component information sufficiently, and has better monitoring performance than conventional PCA based on Hotelling T^2 and Q charts. In essence, the method proposed in this paper is similar to the one proposed by Kano, *et al.*. Their work is not dealt with fault identification, while our approach goes beyond the fault detection task. Once a fault is detected, we have proposed a method based on subspace distance to identity the type of fault.

The paper is structured as follows: the second section gives a more strict procedure of deduction for PCA based on subspace distance, and proposes a method of fault identification according to historical database. The third section presents an application of the approach to FCCU reactor-regenerator system. The final section summarizes the approach.

2. PCA BASED SUBSPACE

2.1 Spacial Signification of PCA

PCA decomposes a normalized sample vector into two portions,

$$\boldsymbol{x} = \hat{x} + \tilde{x}\,, \tag{1}$$

where $\boldsymbol{x} \in \Re^m$ is the sample vector normalized to zero mean and unit variance. The vector $\hat{\boldsymbol{x}}$ is the projection on the principal component subspace S:

$$\hat{\boldsymbol{x}} = \boldsymbol{P}\boldsymbol{P}^T\boldsymbol{x} = \boldsymbol{C}\boldsymbol{x} \tag{2}$$

where $\boldsymbol{P} \in \Re^{m\times k}$ is the PCA loading matrix, and $k \geqslant 1$ is the number of PCs retained in the PCA model. The matrix $\boldsymbol{C} = \boldsymbol{P}\boldsymbol{P}^T$ is projection operator on the principal component subspace S, $\hat{\boldsymbol{x}} \in S \subseteq \Re^m$, with $\dim(S)=k$. The columns of the loading matrix $\boldsymbol{P}$ are the eigenvectors of the correlation matrix associated with the k largest eigenvalues.

Similarly, the residual $\tilde{\boldsymbol{x}}$ satisfies

$$\tilde{\boldsymbol{x}} = (\boldsymbol{I} - \boldsymbol{C})\boldsymbol{x} = \tilde{\boldsymbol{C}}\boldsymbol{x} \in \tilde{S} \subset \Re^m\,, \tag{3}$$

where $\tilde{\boldsymbol{C}}$ is projection operator on the residual subspace $\tilde{S}$, with $\dim(\tilde{S}) = m - k$. From the view of space, PCA divides the measurement space S_{m} ($\dim(S_{\mathrm{m}})=m$) into two orthogonal subspaces, a principal component subspace and a residual subspace. That is,

$$S_m = S \oplus \tilde{S} \tag{4}$$

Principal component subspace primarily characterizes the measurement subspace. When a change in variable correlation occurs, that is, space S_{m} has a change, the bases of principal component subspace also produce corresponding changes. We call principal component subspace S as characteristic subspace.

For a certain chemical process, we can define fault set $\{F_i\}_{i=1}^n$ according to the data recorded in historical database and technologic information. We denote S_i, S_j as the characteristic subspace of fault F_i, F_j respectively. They are spanned by the corresponding loading vectors, respectively, that is,

$$S_i = span(\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_r) \tag{5}$$

$$S_j = span(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_s) \tag{6}$$

where $\dim(S_i)=r$, $\dim(S_j)=s$. Without loss of generality, suppose $s \le r$. The dimensions of subspace S_i, S_j can be determined by the percent of contribution to the accumulative variances. The difference between F_i and F_j can be reflected by the difference of bases of their characteristic subspace. In order to identify different fault, the distance between two subspaces is used to measure the difference. Let matrix

$\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots \boldsymbol{u}_r]$, $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_s]$, with $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$, $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$. The projection operator from subspace S_j onto subspace S_i can be represented as

$$\boldsymbol{C} = \boldsymbol{U}\boldsymbol{U}^T \tag{7}$$

For any unit vector $\boldsymbol{y} \in S_j$, that is, $\|\boldsymbol{y}\|_2 = 1$, its projection on subspace S_i is written as $\hat{\boldsymbol{y}} = \boldsymbol{C}\boldsymbol{y}$. Now, the distance between two subspaces is defined as

$$d_{i,j} = \max_{y} \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2 \tag{8}$$

subject to $\|\boldsymbol{y}\|_2 = 1$ (9)

Since $\boldsymbol{y}$ is a unit vector in S_j, it can be represented as

$$\boldsymbol{y} = \boldsymbol{Vt} \tag{10}$$

where $\boldsymbol{t}$ is the coordinate coefficients vector correspond to bases $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_s$, with $\|\boldsymbol{t}\|_2 = 1$. According to Lagrange's method, we have

$$L(\boldsymbol{t}, \lambda) = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2^2 + \lambda(\|\boldsymbol{y}\|_2^2 - 1) \tag{11}$$

Let $\partial L/\partial \boldsymbol{t} = 0$ and $\partial L/\partial \lambda = 0$, with substitution of Eq.10 in Eq.11, we get the following expression,

$$\boldsymbol{At} = \lambda\ \boldsymbol{t} \tag{12}$$

where $\boldsymbol{A} = \boldsymbol{V}^T\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{V}$. The coordinate coefficients vector $\boldsymbol{t}$ is an eigenvector of the matrix $\mathbf{A}$, and λ is the corresponding eigenvalue. The distance between two subspaces is obtained by the substitution of Eq.12 into Eq.8,

$$d = \sqrt{1 - \lambda_{\min}(\boldsymbol{A})} \tag{13}$$

Now, we prove the distance $d \in [0,1]$.

Proof. $\because \boldsymbol{A} = \boldsymbol{V}^T\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{V} = (\boldsymbol{U}^T\boldsymbol{V})^T\boldsymbol{U}^T\boldsymbol{V}$,

$\therefore$ $\boldsymbol{A}$ is nonnegative definite , that is , $\boldsymbol{A} \geq 0$, $\lambda_{\min}(\boldsymbol{A}) \geq 0$.

Suppose the bases of residual subspace of fault F_i is $\tilde{\boldsymbol{U}} = [\tilde{\boldsymbol{u}}_{r+1}, \tilde{\boldsymbol{u}}_{r+2}, \cdots, \tilde{\boldsymbol{u}}_m]$. Let $\boldsymbol{E} = [\boldsymbol{U}, \tilde{\boldsymbol{U}}]$, then $\boldsymbol{E}$ is the bases of the measurement space S_{m}, with $\boldsymbol{E}^T\boldsymbol{E} = \boldsymbol{E}\boldsymbol{E}^T = \boldsymbol{I}$. $\because \boldsymbol{E}\boldsymbol{E}^T = \boldsymbol{U}\boldsymbol{U}^T + \tilde{\boldsymbol{U}}\tilde{\boldsymbol{U}}^T = \boldsymbol{I}$, and $\because \tilde{\boldsymbol{U}}\tilde{\boldsymbol{U}}^T \geq 0$, $\therefore \boldsymbol{U}\boldsymbol{U}^T \leq \boldsymbol{I}$. Thus, $\boldsymbol{A} = \boldsymbol{V}^T\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{V} \leq \boldsymbol{I}$, that is, $\lambda_{\min}(A) \leq 1$

Therefore, $0 \leq d \leq 1$, End.

Thus, we have the following three special cases:

(i) if the subspace $S_{\mathrm{i}} = S_{\mathrm{j}}$, that is , the two subspace are identical, then $\boldsymbol{U} = \boldsymbol{VQ}$, where $\boldsymbol{Q}$ is nonsingular orthogonal matrix. With Eq.13, we can get $\lambda_{\min}(\boldsymbol{A}) = 1$, that is, d=0. In the case of this, the fault F_i, F_j can be considered as the same fault.

(ii) If the subspace $S_i \subset S_j$, that is, the subspace spanned by the model for the fault F_j contains the subspace spanned by the model for fault F_i, we can also get d=0. It means that the fault F_i is masked by fault F_j, and they can not be distinguished from each other. In fact, they are mistaken for the same fault.

(iii) If the subspace $S_i = S_j^{\perp}$, that is, they are orthogonal, then $\boldsymbol{U}^T\boldsymbol{V} = 0$. Thus, we can get $\lambda_{\min}(\mathbf{A}) = 0$, that is , d=1. It means that the fault F_i, F_j can be distinguished from each other to the most extent.

2.2 Fault Diagnosis Based on Distance

From above description, the distance between subspaces can be used to identify the different faults. We define a match function as follows,

$$\begin{aligned} p_{i,j} &= (1 - d_{i,j}) \times 100\% \\ &= 1 - \sqrt{1 - \lambda_{\min}(A_{i,j})} \times 100\% \end{aligned} \tag{14}$$

When a fault occurs, the loading vectors of the fault data are calculated through PCA and used to represent the bases of characteristic subspace of the fault. On the basis of that, the library of characteristic subspace of faults can be formed and represented as follows,

$$S_F = [S_1, S_2, \cdots, S_n] \tag{15}$$

Where S_i is the characteristic subspace corresponding to the fault F_i, and S_F the subspace set, and n the number of faults.

The currently monitored process measurements can then be analyzed using PCA. The calculated loading vectors form the bases of the subspace corresponding to the current observations. Denoting the current characteristic subspace by S_{cur}, the matching degree between S_{cur} and the every subspace in S_F can be measured by Eq.14 respectively after a fault is detected. If the matching degree between S_{cur} and some subspace (for example, S_i) is very close to 100%, then the current abnormal occurrence may be probably ascribed to fault F_i. On the contrary, if the matching degree is close to zero, it may be least ascribed to fault F_i. Thus, fault identification can be performed by calculating the matching degree between the characteristic subspace of the current data and the library of subspace of known faults. As already discussed, some faults may be masked, so domain knowledge is further needed in that case to analyze the results and determine which fault has occurred on earth.

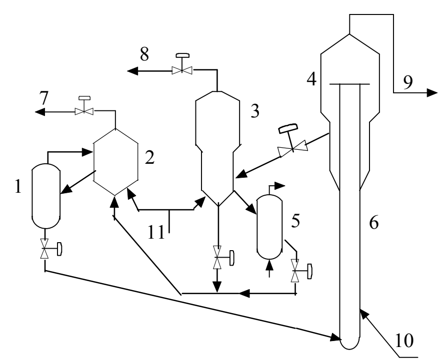
In practical application, a diagnostic threshold is required to be defined in advance. The maximum of matching degree between the current data and the faults in the library should be larger than the diagnostic threshold. Otherwise, if the maximum of matching degree is less than the diagnostic threshold, that is, the current data subspace is not well matched with any fault characteristic subspace in the library, then it is likely that a novel fault has occurred. Once the occurrence of a novel fault is confirmed, the bases of the current data subspace can be stored in the library. Through this method, diagnostic knowledge about novel faults is progressively learnt and the library updated.

3. CASE STUDY

3.1 Process Description

Fluid catalytic cracking unit (FCCU) is considered as one of the most important unit in the refinery. A simplified flow diagram is shown in Fig.1. Briefly, the fresh feed and recycle sludge oil are preheated, mixed, and then enter into the riser reactor where they contact regenerated catalyst and start the cracking reactions. The spent catalyst passes to the steam stripping section and enters the regenerator where the coke on the catalyst is burnt off with air. The heat released by the combustion of coke is supplied to the endothermic cracking reactions. The extra heat than what is required by cracking reactions is taken away by the heat exchanger outside the regenerator. The FCCU reactor–regenerator model which is used in this paper can be referred to the work done by Yang, *et al.* (1997).

To generate an instance, the simulation is running at normal mode. When all parameters become stable, a disturbance or fault is introduced and at the same time, data recording is started. Fourteen variables are chosen to be recorded, including temperature of feed preheated, flow rate of recycle oil, sludge oil, slurry oil, distillate oil and assembled feed, flow rate of air of the first regenerator and the second regenerator, outlet temperature of riser, the heat of exchanger, carbon dioxide content in the flue gas from the first regenerator, oxygen content in the flue gas from the second regenerator, the temperature in the high density bed of the first and the second regenerator. During the simulation, random noise was added to the measurement and controller outputs. Altogether ten data instances have been generated and summarized in Table 1. Sample time is 4 minutes. Each instance is simulated for 1000 minutes, and the database is a 250×14×10 matrix.



1：deaerator 2: 2nd regenerator 3: 1st regenerator
4: settler 5: heat exchanger 6: riser
7: 2nd regenerator flue gas 8: 1st regenerator flue gas 9: product 10: feed 11: air

Fig.1 FCCU Reactor-regenerator Flow Sheet

3.2 Data Analysis and Fault Identification

When the measurement data are obtained, data reconciliation is performed to validate the sensor data. Then they are normalized to zero mean and unit variance before the data of each instance is analyzed by PCA. The distance between two corresponding subspaces is calculated with Eq.13. Table 2 is the calculated results. From Table 2, it can be seen that the distance between subspaces of case 2, 3, 4 is relatively small compared with other distance, which indicates that their difference between them is relatively small. This is because they are all the flow disturbance of fresh feed, only different in the magnitude, and they have similar effect on the correlation of data. The distance between case 2,3,4 and case 5 is relatively large compared with the one between case 2,3,4. That is because case 5 has an adverse disturbance direction. The distance between other cases is large. So we can use the distance to identify different fault or disturbance.

For online monitoring, the data matrix representing the current operating conditions is updated by moving the time-window step by step as proposed by Kano *et al* (2001). PCA is applied to the data matrix, and the distance between the subspace of current data and the one of normal operation data is calculated with Eq.13 at each step. If the distance goes beyond the given control limit, the process is judged to be out of normal operation condition. And then, the distances between the subspace of current data and the one of known faults in the library are calculated respectively, and the match degrees between them are also obtained with Eq.14. If the maximum of matching degree is larger than the diagnostic threshold, then the fault in the library corresponding to the maximum has probably occurred.

In order to verify the method for fault identification, the preheated temperature of fresh feed decreased 6K at 100 minute in the simulation. The monitoring results are shown in Fig.2, and the 95% warning limit which can be determined by statistical method is also shown.

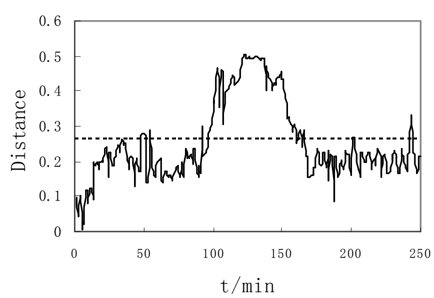


Fig 2 Monitoring results for FCCU reactor-regenerator system

When the distance is out of the control limit, the matching degrees between the current data and the faults in the historical database are calculated, and the results are shown in Fig.3. The diagnostic threshold is predefined as 0.80. It can be seen that the current case matches well with case 8 at 86.32% matching degree which is above the diagnostic threshold, and the fault is successfully identified.

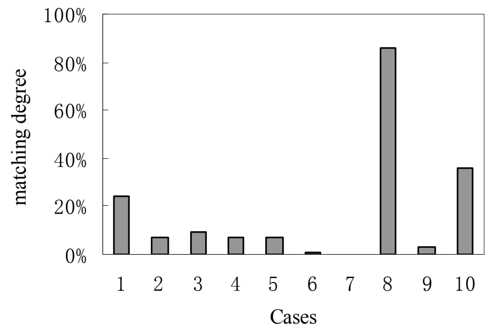


Fig 3 Matching degree of the current case with the cases in the historical database

4. CONCLUSIONS

The diagnosis of abnormal operation can be greatly facilitated if similar system performance has been recorded in the historical database. Principal component analysis is among the most popular methods for extracting information from data. Through PCA, features associated with different faults can be identified and used in fault diagnosis. The features are the characteristic subspace spanned by several loading vectors. Fault diagnosis can then be performed by calculating the distance between the subspace of current data matrix and the one of known faults in the library. It can also deal with novel faults and learn diagnosis knowledge about novel faults. This method is applied to monitoring the FCCU reactor-regenerator system. The results have shown that the method can successfully identify different faults, because it makes full use of information about several principal components. It is important to note that although the approach is well founded, there are problems to be solved in real industrial application. It is advisable to combine domain knowledge with data mining method to diagnosis fault.

5. ACKNOWLEDGEMENTS

The support from the National High Technology (863) Project (No: 2001AA413110) is gratefully acknowledged.

6. REFERRENCES

Dunia, R and Qin, S.J(1998). Subspace approach to multidimensional fault identification and reconstruction. *AIChE.J.*,**44**(8),1813-1831.

Kano, M., Hasebe,S. and Hashimoto, I.(2001). A new multivariate statistical process monitoring method using principal component analysis. *Comp.Chem.Eng*., **25**,1103-1113.

Kresta, J., MacGregor, J.F. and Marlin,T.E.(1991). Multivariate statistical monitoring of process operating performance. *Can.J.Chem.Eng*., **69**(1),35-47.

Ku, W., Storer, R.H., and Georgakis, C.(1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemical Intellectual Laboratory System*, **30**,179-196.

Nomikos,P. and MacGregor, J.F.(1995). Multivariate SPC charts for monitoring batch processes. *Technometrics*, **37**(1), 41-59.

Wang, H.Q., Song, Z.H. and Wang, H.(2002). Statistical process monitoring using improved PCA with optimized sensor locations. *J. Proc. Cont.,* 6, 735-744.

Yang, M.Y, Rong, G. and Wang, S.Q.(1997). FCCU reactor-regenerator advanced control. *International Symposium on Advanced Control of Chemical Processes*, Banff, Canada, 653-657.

Zhang, J., Martin, E.B. and Morris, A.J(1996). Fault detection and diagnosis using multivariate statistical techniques. *Chemical Engineering Research and Design*, **74**(1), 89-96.

Table 1 Summary of the 10 simulated cases

Cases	Description of cases
1	normal operation
2	a step increase of 10% in fresh feed flow rate
3	a step increase of 20% in fresh feed flow rate
4	a step increase of 40% in fresh feed flow rate
5	a step decrease of 30% in fresh feed flow rate
6	a step increase of 10% in air flow rate of 1st regenerator
7	a step increase of 10% in heat of heat removal system
8	decrease of 3K in preheated temperature of fresh feed
9	increase of 3K in outlet temperature of riser reactor
10	increase of 3K in high density bed temperature of 1st regenerator

Table 2 The distance between ten different cases

d	case 1	case 2	case 3	case 4	case5	case 6	case 7	case 8	case 9	case 10
case 1	0	0.3896	0.5799	0.6619	0.5953	0.3406	0.4191	0.7596	0.5754	0.5355
case 2	0.3896	0	0.0678	0.2054	0.6251	0.9915	0.9427	0.9454	0.9998	0.9881
case 3	0.5799	0.0678	0	0.1300	0.6535	0.9991	0.9622	0.9227	0.9527	0.9555
case 4	0.6619	0.2054	0.1300	0	0.6990	0.9998	0.9653	0.9469	0.9360	0.9609
case 5	0.5953	0.6251	0.6535	0.6990	0	0.9769	0.9958	0.9359	0.9566	0.9510
case 6	0.3406	0.9915	0.9991	0.9998	0.9769	0	0.8441	0.9960	0.8684	0.9926
case 7	0.4191	0.9427	0.9622	0.9653	0.9958	0.8441	0	0.9991	0.9618	0.9998
case 8	0.7596	0.9454	0.9227	0.9469	0.9359	0.9960	0.9991	0	0.9747	0.6478
case 9	0.5754	0.9998	0.9527	0.9360	0.9566	0.8684	0.9618	0.9747	0	0.9230
case 10	0.5355	0.9881	0.9555	0.9609	0.9510	0.9926	0.9998	0.6478	0.9230	0