PCA Fault Feature Extraction in Complex Electric Power Systems

Yagang ZHANG, Zengping WANG, Jinfang ZHANG, Jing MA
Key Laboratory of Power System Protection and Dynamic Security Monitoring and Control under Ministry of Education, North China Electric Power University, Baoding, 071003, China
yagangzhang@ncepu.edu.cn

***Abstract*—Electric power system is one of the most complex artificial systems in the world. The complexity is determined by its characteristics about constitution, configuration, operation, organization, etc. The fault in electric power system cannot be completely avoided. When electric power system operates from normal state to failure or abnormal, its electric quantities (current, voltage and angles, etc.) may change significantly. Our researches indicate that the variable with the biggest coefficient in principal component usually corresponds to the fault. Therefore, utilizing real-time measurements of phasor measurement unit, based on principal components analysis technology, we have extracted successfully the distinct features of fault component. Of course, because of the complexity of different types of faults in electric power system, there still exists enormous problems need a close and intensive study.**

***Index Terms*— Complexity, Fault feature extraction, Principal components analysis, PCA, Phasor measurement unit, PMU, Electric power system.**

I. INTRODUCTION

The electric power system is one of the most complex artificial systems in this world and its safe, steady, economical and reliable operating plays a very important part in guaranteeing socioeconomic development, even in safeguarding social stability. In early 2008, the snow and ice disaster that occurred in south China had confirmed it again. The complexity of electric power system is determined by its characteristics: constitution, configuration, operation, organization, etc., that have caused many disastrous accidents, such as the large-scale blackout of America-Canada electric power system on August 14, 2003 and the large-scale blackout of Italy electric power system on September 28, 2003. To solve this complex and difficult problem, some methods and technologies reflecting modern science and technology level have been introduced, such as computer and communication technology, control technology, superconduct and new material technology, and so on. Obviously, no matter if we adopt, new analytical methods or technical means, we must have a distinct recognition of electric power system itself and its complexity, and continuously increase analysis, operation and control level. [1-3]

Principal components analysis (PCA) is a powerful technique that has been applied successfully in many fields such as face recognition and image compression, and generally it is used to reduce the dimension of the multi-dimensional data set and to extract features in the data set, highlight their similarities and differences, [4-6]. PCA is implemented to transform the original data into a smaller set of uncorrelated principal components, and the principal components are obtained by analyzing the covariance matrix of the data. Since patterns in data can be hard to find in data of high dimension, PCA is an outstanding tool for analyzing data. According to nonlinear complex systems, we have carried out large numbers of basic researches, [7-11]. In combination with graph theory and cluster analysis theory, the fault analysis in complex electric power systems has been amplified in detail, [12]. In this paper, we will discuss the fault feature extraction based on PCA.

The fault in electric power system cannot be completely avoided. When electric power system operates from normal state to failure or abnormal, its electric quantities (current, voltage, angles, etc.) may change significantly. In our researches, after some accidents, utilizing real-time measurements of phasor measurement unit (PMU), [13-16], based on principal components analysis technology, we will extract the distinct features of fault component.

II. PRINCIPLES OF PRINCIPAL COMPONENT ANALYSIS

Principal components analysis is well suited to data reduction, it is also used to identify underlying factors or processes that may explain the variance in a large data matrix. The analysis seeks to collect the common variance among many variables into one principal component. PCA has three main functions, [17]:

(1) Remove correlation between variables in the data set;
(2) Transform data into a new set of axes that preserves the distance between samples and where data variance can be observed;
(3) Reduce the number of variables necessary to describe most of the variance in the data.

Fig. 1 is presents the flow chart of principal component analysis. And the mathematical principle of PCA can be explained as follows:

Let,

$$F = a_1 X_1 + a_2 X_2 + \cdots + a_p X_p = a' X ,$$

hereinto, $a = (a_1, a_2, \cdots, a_p)'$, $X = (X_1, X_2, \cdots, X_p)'$.

The process to get the principal components is just seeking ask for the linear function $a'X$ of X , and making the variance as great as possible, that is,

$$\begin{aligned} Var(a'X) &= E(a'X - E(a'X))(a'X - E(a'X))' \\ &= a'E(X - E(X))(X - E(X))'a \\ &= a'\Sigma a \end{aligned}$$

reaches the maximum, and $a'a = 1$.

Digital Object Identifier 10.4316/AECE.2010.03017

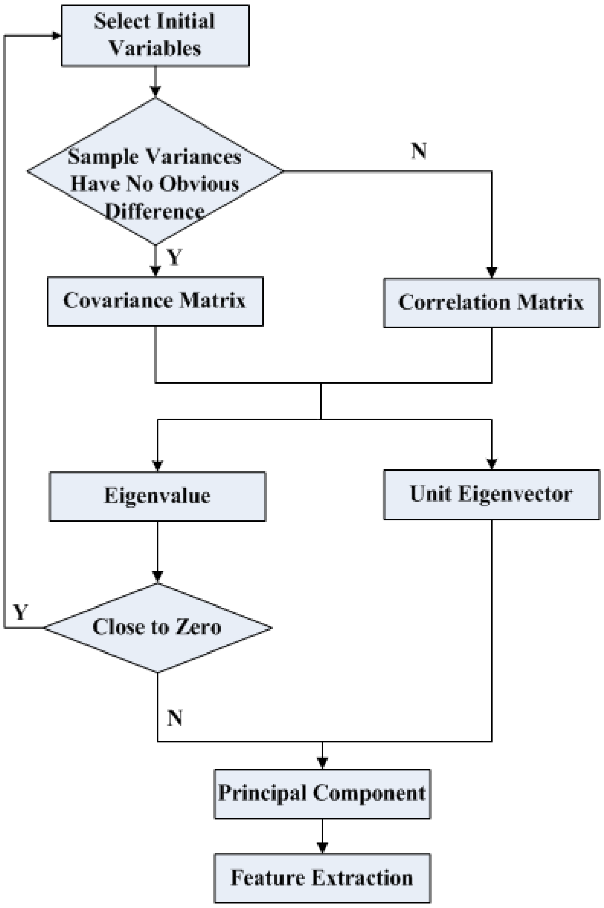

Figure 1. Flow chart of principal component analysis.

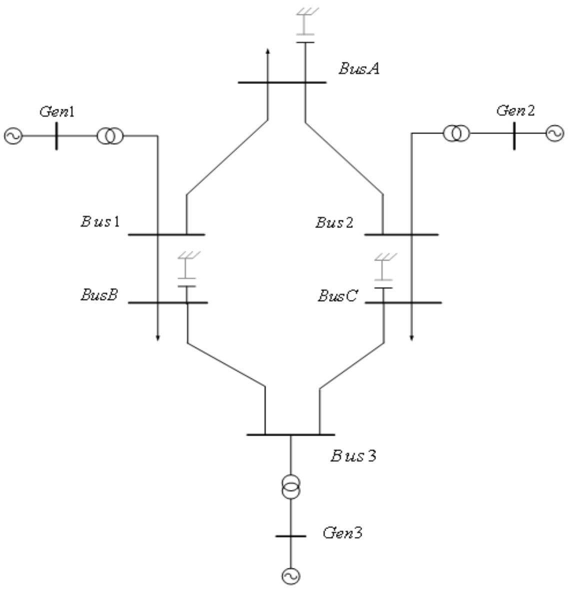

Figure 2. Electric diagram of IEEE 9-Bus system.

Let the eigenvalues of covariance matrix Σ be $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p > 0$, and their corresponding unit eigenvectors are $u_1, u_2, \cdots, u_p$, namely,

$$U_{(p\times p)} = (u_1, \cdots, u_p) = \begin{pmatrix} u_{11}, u_{12}, \cdots, u_{1p} \\ u_{21}, u_{22}, \cdots, u_{2p} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ u_{p1}, u_{p2}, \cdots, u_{pp} \end{pmatrix}.$$

Obviously $U'U = UU' = I$, and

$$\Sigma = U \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} U' = \sum_{i=1}^{p} \lambda_i u_i u_i' .$$

$$\because \quad a'\Sigma a = \sum_{i=1}^{p} \lambda_i a' u_i u_i' a$$

$$= \sum_{i=1}^{p} \lambda_i (a'u_i)(a'u_i)' = \sum_{i=1}^{p} \lambda_i (a'u_i)^2$$

$$\therefore \quad a'\Sigma a \le \lambda_1 \sum_{i=1}^{p} (a'u_i)^2 = \lambda_1 a'u(a'u)'$$

$$= \lambda_1 a'uu'a = \lambda_1 a'a = \lambda_1$$

Besides, if $a = u_1$, then

$$u_1'\Sigma u_1 = u_1'(\sum_{i=1}^{p} \lambda_i u_i u_i')u_1$$

$$= \sum_{i=1}^{p} \lambda_i u_1' u_i u_i' u_1 = \lambda_1 (u_1'u_1)^2 = \lambda_1$$

Therefore, $a = u_1$ makes $Var(a'X) = a'\Sigma a$ reach the maximum, and $Var(u_1'X) = u_1'\Sigma u_1 = \lambda_1$. And for the same reason, $Var(u_i'X) = \lambda_i \quad (i = 1, \cdots, p)$. Furthermore,

$$Cov(u_i'X, u_j'X) = \sum_{k=1}^{p} \lambda_k (u_i'u_k)(u_k'u_j) = 0 \quad (i \ne j).$$

III. FAULT FEATURE EXTRACTION BASED ON PRINCIPAL COMPONENT ANALYSIS

PCA analyzes the correlations between variables, and can reduce the dimensions of the data without losing much statistical information. Based on PCA, we can transform the data of mufti-dimensions from a correlated variable space into a less-correlated (or uncorrelated) variable space. A key characteristic of the PCA is that the first component accounts for the most variance in the original data, that is, the first principal component is considered a new variable that explains as much departure in all of the original variables as is possible; the second component explains as much of the remaining departure as is possible, and so on, each successive component accounts for a smaller amount of the departure than its predecessor. Thus principal components lead to a more essential description of the original data. In fact, normally only the first few components explain the significant amount of departure in the original data. The eigenvalue refers to the magnitude of the component vector, larger eigenvalues explain more departure.

Firstly, let us consider IEEE9-Bus system, Fig. 2 presents the IEEE 9-bus system electric diagram. In the structure of electric power system, Bus1 appears single-phase to ground fault. Through simulation experiments, using these actual measurement data of corresponding variables, we can carry through feature extraction of fault component.

A. Fault feature extraction of IEEE9-Bus system based on node positive sequence voltage

After computing IEEE9-Bus system, we can get node positive sequence voltages at $\boldsymbol{T_{-1}}$, $\boldsymbol{T_0}$(Fault) and $\boldsymbol{T_1}$ three times. Firstly, the covariance matrix of node positive sequence voltages can be calculated, see Table I.

In the covariance matrix, there is a remarkable characteristic, the covariance of Bus1 is 0.051393, which is the biggest. So, one can analyze preliminarily that the Bus1 is a probable fault component.

Now let's solve the eigenvalues of this covariance matrix, these results have been listed in Table II.

Finally, the first principal component can be obtained, its expression is,

$$\begin{aligned} F_1 = {} & 0.578699X_1 + 0.231257X_2 + 0.239876X_3 \\ & + 0.444402X_4 + 0.450557X_5 + 0.231363X_6 \\ & + 0.241800X_7 + 0.132581X_8 + 0.154979X_9 \end{aligned}.$$

Because the cumulative of the first principal component has reached 99.84%, in this place, we only need to extract the first principal component.

Based on a comprehensive analysis of these present results, a main conclusion has been reached as followings: From the feature of the first principal component, Bus1 corresponds with variable X_1, and the coefficient of X_1 is 0.578699, which is also the biggest, so, Bus1 is just the fault component. This conclusion is entirely identical with the fault set in advance.

B. Fault feature extraction of IEEE9-Bus system based on node negative sequence voltage

By simulation experiments, we can also get node negative sequence voltages at $\boldsymbol{T_{-1}}$, $\boldsymbol{T_0}$(Fault) and $\boldsymbol{T_1}$ three times, and the covariance matrix of node negative sequence voltages can be calculated, see Table III.

In the covariance matrix, there is a remarkable characteristic, the covariance of Bus1 is 0.037687, which is also the biggest. Therefore one can analyze preliminarily that the Bus1 is a probable fault component.

Let's further solve the eigenvalues of this covariance matrix, see Table IV.

Finally, the first principal component can be expressed as,

$$\begin{aligned} F_1 = {} & 0.634352X_1 + 0.196920X_2 + 0.217486X_3 \\ & + 0.425640X_4 + 0.450264X_5 + 0.191164X_6 \\ & + 0.247588X_7 + 0.103507X_8 + 0.138045X_9 \end{aligned}$$

Because the cumulative of the first principal component has reached 100%, in this place, we only need to extract the first principal component.

Based on a comprehensive analysis of these present results, one can conclude as follows: From the feature of the first principal component, Bus1 corresponds with variable X_1, and the coefficient of X_1 is 0.634352, which is the biggest. Consequently, Bus1 is just the fault component. This conclusion is also entirely identical with the fault set in advance.

Now let us further consider IEEE 39-Bus system. The IEEE39-Bus system is well known as 10-machine 39-bus New-England Power System, and Fig. 3 presents its electric diagram. In the structure of electricity grid, Bus-18 appears three-phase short-circuit fault. According to the results of the simulation experiments, using these actual measurement data of corresponding variables, we can carry through feature extraction of fault component.

C. Fault feature extraction of IEEE 39-Bus system based on node positive sequence voltage

Likewise, we calculate the node positive sequence voltage at $\boldsymbol{T_{-1}}$, $\boldsymbol{T_0}$(Fault) and $\boldsymbol{T_1}$ three times. Firstly, the covariance matrix of node positive sequence voltages has been calculated. In this place, we only intercept Bus18 section, see Table V.

In Table V, a remarkable characteristic is in esse, the covariance of Bus18 is 0.293031, which is not only the biggest one in Table V (only intercept Bus18 section), but also the biggest one in the complete covariance matrix based on node positive sequence voltage of IEEE 39-Bus system. So, one can analyze preliminarily that the Bus18 is a probable fault component.

Let's further solve the eigenvalues of this covariance matrix, see Table VI.

Finally, the first principal component is obtained, which can be expressed as,

$$\begin{aligned} F_1 = {} & 0.080303X_1 + 0.171509X_2 + 0.266442X_3 \\ & + 0.165149X_4 + 0.124600X_5 + 0.119829X_6 \\ & + 0.112038X_7 + 0.109471X_8 + 0.059017X_9 \\ & + 0.124305X_{10} + 0.122919X_{11} + 0.125781X_{12} \\ & + 0.133050X_{13} + 0.155540X_{14} + 0.186957X_{15} \\ & + 0.205857X_{16} + 0.306195X_{17} + 0.425979X_{18} \\ & + 0.129325X_{19} + 0.097090X_{20} + 0.165830X_{21} \\ & + 0.131574X_{22} + 0.129598X_{23} + 0.195589X_{24} \\ & + 0.161128X_{25} + 0.200178X_{26} + 0.247223X_{27} \\ & + 0.148679X_{28} + 0.132369X_{29} + 0.105041X_{30} \\ & + 0.075669X_{31} + 0.079530X_{32} + 0.088549X_{33} \\ & + 0.081280X_{34} + 0.096113X_{35} + 0.078531X_{36} \\ & + 0.108539X_{37} + 0.095568X_{38} + 0.022421X_{39} \end{aligned}$$

Because the cumulative of the first principal component has reached 100%, in this place, we only need to extract the first principal component.

Based on a comprehensive analysis of these present results, one can conclude as follows: From the feature of the first principal component, Bus18 corresponds with variable X_{18}, and the coefficient of X_{18} is 0.425979, which is the biggest. Consequently, Bus18 is just the fault component. This conclusion is also entirely identical with the fault set in advance.

These instances have fully proven that the fault feature extraction can be performed successfully by principal component analysis and calculation. The results are accurate and reliable.

TABLE I. THE COVARIANCE MATRIX BASED ON NODE POSITIVE SEQUENCE VOLTAGE

Bus	*Bus*1	*Bus*2	*Bus*3	*Bus*A	*Bus*B	*Bus*C	*Gen*1	*Gen*2	*Gen*3
*Bus*1	**0.051393**	**0.020496**	**0.021251**	**0.039452**	**0.039995**	**0.020501**	**0.021416**	**0.011727**	**0.013698**
*Bus*2	0.020496	0.008212	0.008523	0.015747	0.015967	0.008218	0.008594	0.00472	0.005523
*Bus*3	0.021251	0.008523	0.008848	0.01633	0.016559	0.00853	0.008923	0.004904	0.005741
*Bus*A	0.039452	0.015747	0.01633	0.03029	0.030708	0.015752	0.016458	0.009017	0.010536
*Bus*B	0.039995	0.015967	0.016559	0.030708	0.031132	0.015973	0.016689	0.009145	0.010686
*Bus*C	0.020501	0.008218	0.00853	0.015752	0.015973	0.008224	0.008602	0.004726	0.005531
*Gen*1	0.021416	0.008594	0.008923	0.016458	0.016689	0.008602	0.008999	0.004948	0.005793
*Gen*2	0.011727	0.00472	0.004904	0.009017	0.009145	0.004726	0.004948	0.002725	0.003195
*Gen*3	0.013698	0.005523	0.005741	0.010536	0.010686	0.005531	0.005793	0.003195	0.003747

TABLE II. THE EIGENVALUES OF COVARIANCE MATRIX BASED ON NODE POSITIVE SEQUENCE VOLTAGE

No.	Eigenvalues	Proportion	Cumulative
1	0.15332717	0.9984	0.9984
2	0.00024435	0.0016	1.0000

TABLE III. THE COVARIANCE MATRIX BASED ON NODE NEGATIVE SEQUENCE VOLTAGE

Bus	*Bus*1	*Bus*2	*Bus*3	*Bus*A	*Bus*B	*Bus*C	*Gen*1	*Gen*2	*Gen*3
*Bus*1	**0.037687**	**0.011699**	**0.012921**	**0.025288**	**0.02675**	**0.011357**	**0.014709**	**0.006149**	**0.008201**
*Bus*2	0.011699	0.003632	0.004011	0.00785	0.008304	0.003526	0.004566	0.001909	0.002546
*Bus*3	0.012921	0.004011	0.00443	0.00867	0.009171	0.003894	0.005043	0.002108	0.002812
*Bus*A	0.025288	0.00785	0.00867	0.016968	0.017949	0.00762	0.00987	0.004126	0.005503
*Bus*B	0.02675	0.008304	0.009171	0.017949	0.018988	0.008061	0.010441	0.004365	0.005821
*Bus*C	0.011357	0.003526	0.003894	0.00762	0.008061	0.003423	0.004433	0.001853	0.002472
*Gen*1	0.014709	0.004566	0.005043	0.00987	0.010441	0.004433	0.005741	0.0024	0.003201
*Gen*2	0.006149	0.001909	0.002108	0.004126	0.004365	0.001853	0.0024	0.001003	0.001338
*Gen*3	0.008201	0.002546	0.002812	0.005503	0.005821	0.002472	0.003201	0.001338	0.001785

TABLE IV. THE EIGENVALUES OF COVARIANCE MATRIX BASED ON NODE NEGATIVE SEQUENCE VOLTAGE

No.	Eigenvalues	Proportion	Cumulative
1	0.09365568	1.0000	1.0000
2	0	0	1.0000

TABLE V. THE SIMPLE COVARIANCE MATRIX BASED ON NODE POSITIVE SEQUENCE VOLTAGE OF IEEE 39-BUS SYSTEM (ONLY INTERCEPT BUS18 SECTION)

Bus	*Bus*1	*Bus*2	*Bus*3	*Bus*4	*Bus*5	*Bus*6	*Bus*7	*Bus*8
*Bus*18	**0.05524**	**0.117981**	**0.183285**	**0.113606**	**0.085712**	**0.082431**	**0.077071**	**0.075305**
Bus	*Bus*9	*Bus*10	*Bus*11	*Bus*12	*Bus*13	*Bus*14	*Bus*15	*Bus*16
*Bus*18	**0.040598**	**0.085509**	**0.084556**	**0.086525**	**0.091525**	**0.106996**	**0.128607**	**0.141609**
Bus	*Bus*17	*Bus*18	*Bus*19	*Bus*20	*Bus*21	*Bus*22	*Bus*23	*Bus*24
*Bus*18	**0.210632**	**0.293031**	**0.088963**	**0.066788**	**0.114075**	**0.09051**	**0.08915**	**0.134546**
Bus	*Bus*25	*Bus*26	*Bus*27	*Bus*28	*Bus*29	*Bus*30	*Bus*31	*Bus*32
*Bus*18	**0.11084**	**0.137702**	**0.170065**	**0.102277**	**0.091057**	**0.072258**	**0.052052**	**0.054709**
Bus	*Bus*33	*Bus*34	*Bus*35	*Bus*36	*Bus*37	*Bus*38	*Bus*39	
*Bus*18	**0.060913**	**0.055912**	**0.066116**	**0.054021**	**0.074664**	**0.065741**	**0.015424**	

TABLE VI. THE EIGENVALUES OF COVARIANCE MATRIX BASED ON NODE POSITIVE SEQUENCE VOLTAGE OF IEEE 39-BUS SYSTEM

No.	Eigenvalues	Proportion	Cumulative
1	1.61486656	1.0000	1.0000
2	0.00000006	0	1.0000

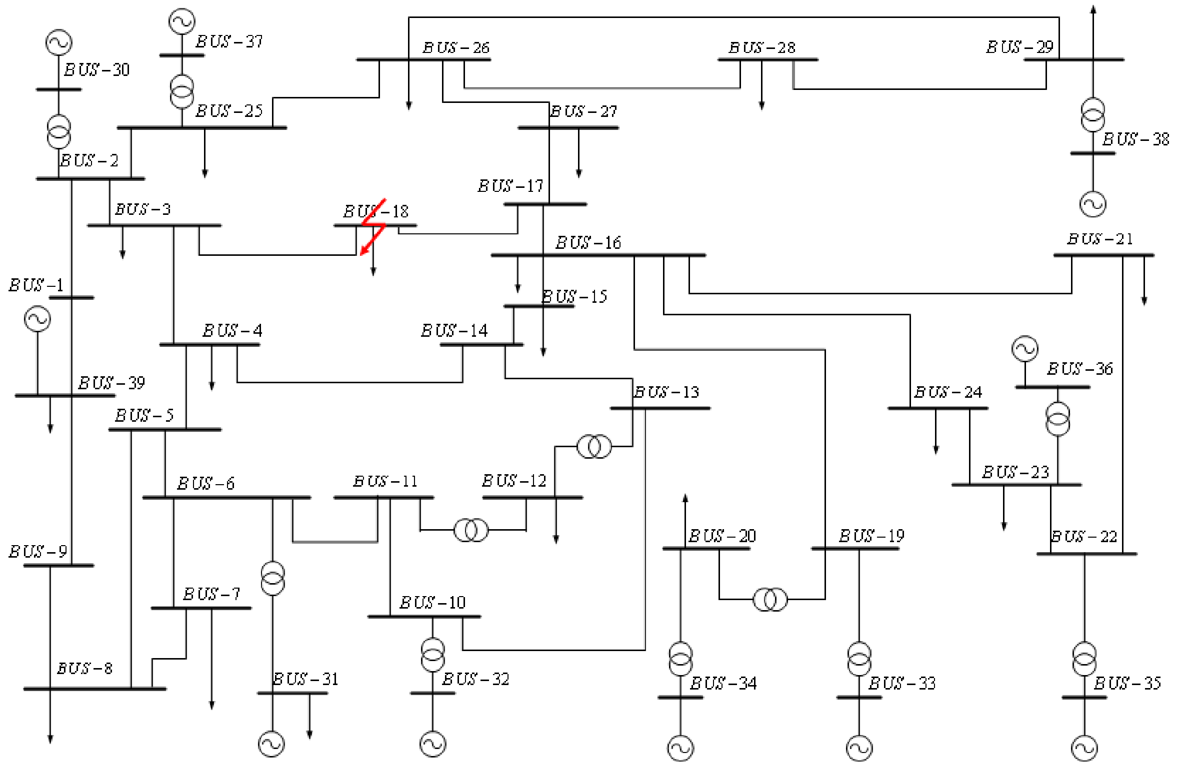

Figure 3. Electric diagram of IEEE 39-Bus system.

IV. CONCLUSION AND DISCUSSION

Electric power system is one of the most complex artificial systems in this world, which safe, steady, economical and reliable operation plays a very important part in guaranteeing socioeconomic development, even in safeguarding social stability. The complexity of electric power system is determined by its characteristics about constitution, configuration, operation, organization, etc. However, no matter what we adopt new analytical method or technical means, we must have a distinct recognition of electric power system itself and its complexity, and increase continuously analysis, operation and control level. The fault in electric power system cannot be completely avoided. When electric power system operates from normal state to failure or abnormal operates, its electric quantities (current, voltage, angles, etc.) may change significantly. Our researches have indicated that the variable with the biggest coefficient in principal component usually corresponds to the fault. So, utilizing real-time measurements of PMU, based on principal components analysis technology, we have extracted successfully the distinct features of fault component.

The method of principal components is primarily a data analytic technique that obtains linear transformations of correlated variables such that certain optimal conditions are achieved. The most important of these conditions is that the transformed variables are uncorrelated, [18]. The goal of PCA is just to clarify the correlations among the components of a multivariate variable. PCA has been successfully used in many applications as a feature reduction technique.

Our researches have sufficiently demonstrated that the fault feature extraction of complex electric power systems can be explored successfully by principal components analysis. Of course, because of the complexity of different types of faults in electric power systems, there still exists enormous problems need a close and intensive study.

ACKNOWLEDGMENT

This research was supported partly by the National Natural Science Foundation of China (50837002, 50907021) and Chinese Universities Science Foundation.

REFERENCES

[1] J.X. Yuan, Wide area protection and emergency control to prevent large scale blackout, China Electric Power Press, Beijing, 2007.

[2] L. Ye, "Study on sustainable development strategy of electric power in China in 2020", Electric Power, vol. 36, pp.1-7, Oct. 2003.

[3] Y.S. Xue, "Interactions between power market stability and power system stability", Automation of Electric Power Systems, vol. 26, pp.1-6, Nov. 2002.

[4] M.J. Baxter, "Standardization and transformation in principal components analysis, with applications to archaeometry", Applied Statistics, vol. 44, pp. 513-527, Oct. 1995.

[5] M.D. Glascock, E.B. Geoffrey and H.C. Robert, A systematic approach to obsidian source characterization, Plenum Publishing Co., New York, 1998.

[6] X.L. Yu and X.S. Ren, Multivariate statistical analysis, China Statistic Press, 1998.

[7] Y.G. Zhang, C.J. Wang and Z. Zhou, "Inherent randomicity in 4-symbolic dynamics", Chaos, Solitons and Fractals, vol. 28, pp. 236-243, Apr. 2006.

[8] Y.G. Zhang and C.J. Wang, "Multiformity of inherent randomicity and visitation density in *n*-symbolic dynamics", Chaos, Solitons and Fractals, vol. 33, pp. 685-694, Jul. 2007.

[9] Y.G. Zhang and Z.P. Wang, "Knot theory based on the minimal braid in Lorenz system", International Journal of Theoretical Physics, vol. 47, pp. 873-880, Apr. 2008.

[10] Y.G. Zhang, Y. Xu and Z.P. Wang, "Dynamical randomicity and predictive analysis in cubic chaotic system", Nonlinear Dynamics, vol.61, pp. 241-249, Jul. 2010. http://www.springerlink.com/content/50745j737g3t1x54/fulltext.pdf

[11] Y.G. Zhang, J.F. Zhang, Q. Ma, J. Ma and Z.P. Wang, "Statistical description and forecasting analysis of life system", International

Journal of Nonlinear Sciences and Numerical Simulation, vol.11, pp. 157-163, Mar. 2010.

[12] Z.P. Wang, Y.G. Zhang, J.F. Zhang and J. Ma, “Recent research progress in fault analysis of complex electric power systems”, Advances in Electrical and Computer Engineering, vol. 10, pp.28-33, Feb. 2010.

[13] A.G. Phadke and J.S. Thorp, Synchronized phasor measurements and their applications, Springer Verlag, 2008.

[14] C. Wang, C.X. Dou, X.B. Li and Q.Q. Jia, “A WAMS/PMU-based fault location technique”, Electric Power Systems Research, vol. 77, pp. 936-945, Jun. 2007.

[15] C. Rakpenthai, S. Premrudeepreechacharn, S. Uatrongjit and N.R. Watson, “Measurement placement for power system state estimation using decomposition technique”, Electric Power Systems Research, vol. 75, pp.41-49, Jul. 2005.

[16] J.N. Peng, Y.Z. Sun and H.F. Wang, “Optimal PMU placement for full network observability using Tabu search algorithm”, International Journal of Electrical Power & Energy Systems, vol. 28, pp. 223-231, May. 2006.

[17] M.D. Glascock, Characterization of archaeological ceramics at MURR by neutron activation analysis and multivariate statistics, Prehistory Press, Madison, 1992.

[18] J.E. Jackson, A user’s guide to principal components, John Wiley and Sons, New York, 1991.