2017 13th IEEE International Conference on Control & Automation (ICCA)
July 3-6, 2017. Ohrid, Macedonia

A New PCA Similarity Factor Based on Elliptical Principal Angle

Renming Pang and Hao Ye

***Abstract*—Measuring the similarity between two different PCA models is a key step in pattern matching, monitoring of processes with multiple operating conditions and other applications. This paper first explains the limitations of several existing methods for measuring the similarity of PCA models through a comprehensive comparison based on simulation, then defines the elliptical principal angle, based on which, a new PCA similarity factor capable of reflecting both the gradual rotation and the gradual eigenvalue change of a PCA model sensitively and smoothly is proposed.**

I. INTRODUCTION

Principal component analysis (PCA) has been widely used in the process industry [1-5]. In the monitoring of processes with multiple operating modes or multi-stage/multi-phase batch processes, traditional PCA based methods may give rise to frequent false-alarms because they need to assume that the process has a constant operating region [14]. To deal with the problem, several PCA-based multiple-model methods [11-15] are proposed, in which an essential step is to calculate the similarity or distance between PCA models built on different dataset of a same process. Therefore, how to calculate PCA similarity may directly affect the performance of PCA-based multiple-model methods in process monitoring.

In pattern matching, measuring the similarity of PCA models is also a key step to analyze large amounts of historical data and locate similar conditions [6-10]. For example, the methods of pattern matching based on PCA similarity factor have been used for clustering data in industrial bio-processes [8], detecting faults in air handling unit system [6] and classification of batch processes in automotive metallic coatings [16].

Let X_F and X_G denote two datasets of a process. So far, the following similarity factors have been proposed to measure the similarity between the PCA models of X_F and X_G, which can be classified into three groups: (i) Krzanowski and W.J. [17] propose a PCA similarity factor (denoted as S_{PCA} in this paper) by using the angles between the PCA loading vectors of X_F and those of X_G. Johannesmeyer and Charles [18] further give a modified definition of S_{PCA} (denoted as S_{PCA}^{λ}) by introducing the eigenvalues as weights. While Yao and Gao [11] define the similarity factor (denoted as $S_{PCA}^{PC-pair}$) by measuring only the angles between each pair of the corresponding loading vectors (i.e. loading vectors with the same sequence number) in X_F and X_G, weighted by the corresponding eigenvalues; (ii) Zhao et al. [14] propose to use the principal angles [19] between the subspaces spanned by the retained PCA loading vectors of X_F and X_G in their definition for similarity factor (denoted as $S_{PCA}^{subspace}$), and further give the modified definition (denoted as $S_{PCA}^{\lambda,subspace}$) by introducing the eigenvalues as weights; (iii) Lu et al. [13] and Zhao et al. [12] give the measurement for the similarity (denoted as D_{PCA}) by calculating the Euclidean distance between the PCA loading matrices of X_F and X_G, weighted by the corresponding eigenvalues.

The above methods have been compared and evaluated in some literatures from different aspects. For example, Yao and Gao [11] compare D_{PCA} and $S_{PCA}^{PC-pair}$ by evaluating the performance of a batch process monitoring method, in which the two similarity factors are involved respectively, and similarly, Singhal et al. [9] compare S_{PCA} and S_{PCA}^{λ} by evaluating the performance of a pattern matching method. In addition, Yao et al. [11] point out that S_{PCA} will stay unchanged if the subspaces spanned by the retained PCA loading vectors of X_F and X_G cover the same space.

In this paper, we first give a comprehensive simulation comparison of all of the above mentioned methods, which has not been covered by any existing literatures. In addition, we propose to evaluate how effective a PCA similarity factor is by evaluating whether it can sensitively and smoothly reflect any changes in the loading vectors and eigenvalues of the PCA model built from a data matrix. The word "smoothly" means that the similarity factor shouldn't change abruptly when the process has just experienced a drift or a slow time varying change. This is obviously a more direct and essential approach for evaluation, compared with literatures [11] and [9], which evaluate the similarity factors according to the final performance of the applications such as process monitoring and pattern matching. Our comprehensive comparison indicates that none of the existing methods work in all simulation scenarios simultaneously in the above mentioned sense (Please refer to Sec. III for detail).

To deal with the problem, we give a new definition to PCA similarity factor by replacing the principal angles used in $S_{PCA}^{subspace}$ of [14] with the elliptical principal angles defined in this paper. For convenience, we call it $S_{PCA}^{elliptical}$. The merit of the proposed PCA similarity factor is that it reflects both the rotation and eigenvalue changes of a PCA model with sensitivity and smoothness, as demonstrated by all cases of the simulation, and also supported by physical interpretation.

*Research supported by National Natural Science Foundation of China under Grant 61290324.

R.M. Pang is with the Dept. of Automation, Tsinghua University, Beijing, 100084, China (e-mail: prm14@foxmail.com; phone: (86)150-1004-6219)
C.A. H. Ye is with the Dept. of Automation, Tsinghua University, Beijing, 100084, China (e-mail: haoye@tsinghua.edu.cn).

II. PRELIMINARIES

A. Principal Component Analysis

For a data matrix $X \in \Re^{n\times m}$ with n sample points and m observational variables, singular value decomposition can be performed on its covariance matrix as following [1]:

$$\frac{1}{n-1} X^T X = P \Lambda P^T \tag{1}$$

where the orthogonal loading matrix $P = [p_1, p_2, \cdots, p_m]$ is composed of the m eigenvectors (which is also called loading vectors below), the eigenvalue matrix $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_m)$ is composed of the m corresponding eigenvalues, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$. Then X can be expressed as [4]:

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_r p_r^T + E \tag{2}$$

where $E = t_{r+1} p_{r+1}^T + \cdots + t_m p_m^T$ represents the residual, t_i and p_i, for $i = 1, 2, \cdots, m$, are called principal component vectors and loading vectors respectively, and r is the number of retained principal components.

Notation: Throughout the paper, let X_F and X_G denote the two data matrices for which a PCA similarity will be used. Then for a data matrix $*$, where $* = F$ or G, (i) Let $P_{*_k} = [p_{*_1}, p_{*_2}, \cdots, p_{*_k}]$ represent the corresponding loading matrix and loading vectors of its PCA model with $k \leq m$; (ii) Similarly, let $\Lambda_{*_k} = diag(\lambda_1^*, \lambda_2^*, \cdots, \lambda_k^*)$ denote the eigenvalue matrix and eigenvalues of the PCA model of data matrix $*$; (iii) Let θ_{ij} denote the angle between p_{F_i} and p_{G_j}, for $i = 1, \cdots, m$ and $j = 1, \cdots, m$; (iv) Let $\theta_k^{F,G}$ denote the k-th principal angle [19] between the subspaces Ψ_F and Ψ_G defined in (3) and (4).

B. Principal Angle

Let Ψ_F and Ψ_G be two given subspaces of real space $\Re^m$ that spanned by P_{F_r} and P_{G_q} respectively and assume that $r = \dim(\Psi_F) \geq \dim(\Psi_G) = q$. Then the principal angles between Ψ_F and Ψ_G can be recursively defined as [19]

$$cos\theta_1^{F,G} = \max_{u \in \Psi_F \cap v \in \Psi_G} (u^T v) = u_1^T v_1 \tag{3}$$

subject to $\|u\|_2 = \|v\|_2 = 1$, and

$$cos\theta_k^{F,G} = \max_{u \in \Psi_F \cap v \in \Psi_G} (u^T v) = u_k^T v_k \tag{4}$$

subject to $\|u\|_2 = \|v\|_2 = 1,\ u_i^T u = 0,\ v_i^T v = 0,\ i = 1, 2, \cdots, k-1$, where $k = 1, 2, \cdots, q$.

Literature [19] further indicates that $cos\theta_k^{F,G}$ in (3) and (4) can be equivalently computed as following [19].

$$cos\theta_k^{F,G} = SVD_k\left(P_{F_r}{}^T P_{G_q}\right) \tag{5}$$

where $SVD_k(*)$ denotes the k-th singular value of matrix $*$.

C. A Brief Introduction to Existing PCA Similarity Factors

In the following, a brief introduction to the existing PCA similarity factors (denoted as S_{PCA}, S_{PCA}^{λ}, $S_{PCA}^{PC-pair}$, $S_{PCA}^{subspace}$, $S_{PCA}^{\lambda,subspace}$, D_{PCA} in this paper) will be given.

Krzanowski and W.J. [17] first define a PCA similarity factor as:

$$S_{PCA} = \frac{1}{r}\sum_i^r \sum_j^r \cos^2(\theta_{ij}) = \frac{1}{r} trace\left(P_{G_r}{}^T P_{F_r} P_{F_r}{}^T P_{G_r}\right) \tag{6}$$

Johannesmeyer and Charles [18] further introduce the eigenvalues as weights into (6) as following

$$\begin{aligned} S_{PCA}^{\lambda} &= \frac{\sum_{i=1}^r \sum_{j=1}^r \left(\lambda_i^F \lambda_j^G \cos^2\theta_{ij}\right)}{\sum_{i=1}^r \lambda_i^F \lambda_i^G} \\ &= \frac{\sum_{k=1}^r SVD_k\left(\Lambda_{G_r}^{1/2} P_{G_r}{}^T P_{F_r} \Lambda_{F_r} P_{F_r}{}^T P_{G_r} \Lambda_{G_r}^{1/2}\right)}{\sum_{i=1}^r \lambda_i^F \lambda_i^G} \\ &= \frac{trace\left(\Lambda_{G_r}^{1/2} P_{G_r}{}^T P_{F_r} \Lambda_{F_r} P_{F_r}{}^T P_{G_r} \Lambda_{G_r}^{1/2}\right)}{trace(\Lambda_{F_r} \Lambda_{G_r})} \end{aligned} \tag{7}$$

Different from (6) and (7), which consider the angles θ_{ij} for all possible combinations of $i \leq r$ and $j \leq r$, Yao et al. [11] only consider the angles θ_{ii} between each pair of the corresponding loading vectors p_{F_i} and p_{G_i}, for $i = 1, \cdots, r$, in the similarity factor definition, i.e.

$$S_{PCA}^{PC-pair} = \frac{\sum_{i=1}^r \left(\lambda_i^F \lambda_i^G \cos^2\theta_{ii}\right)}{\sum_{i=1}^r \lambda_i^F \lambda_i^G} \tag{8}$$

Instead of using the angles between loading vectors, Zhao et al. [14] propose to use the principal angles $\theta_k^{F,G}$ [19] between Ψ_F and Ψ_G spanned by the loading vectors P_{F_r} and P_{G_q} (Please refer to (3)-(5) for the definition of $\theta_k^{F,G}$) in the similarity factor, i.e.

$$S_{PCA}^{subspace} = \frac{1}{q}\sum_{k=1}^q \cos^2\left(\theta_k^{F,G}\right) = \frac{1}{q} trace\left(P_{G_q}{}^T P_{F_r} P_{F_r}{}^T P_{G_q}\right) \tag{9}$$

It is worth noting that Zhao et al. [20] indicate that S_{PCA} is identical to (9), as long as q is equal to r.

Again, Zhao et al. [14] further introduce eigenvalues as weights into (9), and then gives the following similarity factor

$$S_{PCA}^{\lambda,subspace} = \frac{1}{q} trace\left(\Lambda_{G_q}^{1/2} P_{G_q}{}^T P_{F_r} \Lambda_{F_r} P_{F_r}{}^T P_{G_q} \Lambda_{G_q}^{1/2}\right) \tag{10}$$

For multistage batch processes, a stage-based sub-PCA model [13] and a soft-transition multiple PCA [12] are proposed, which both assess the dissimilarity by comparing the Euclidean distance between two weighted loading matrices as following.

$$D_{PCA}=\sum_{i}^{m}\left\|\frac{\lambda_i^F}{\sum_{j=1}^{m}\lambda_j^F}p_{F_i}-\frac{\lambda_i^G}{\sum_{j=1}^{m}\lambda_j^G}p_{G_i}\right\| \tag{11}$$

III. A Comprehensive Comparison of the Existing Methods

A. A Description of the Simulation Scenario

According to (1), any difference between the data matrices X_F and X_G can be represented as a difference between their eigenvalues and/or that between their loading vectors, and the later one actually represents a rotation between the two data matrices. So a more direct approach to evaluate a PCA similarity factor is to observe how it changes with the gradual rotation and the gradual eigenvalue change of the PCA model of a data matrix respectively.

To this end, in the scenario of rotation, let data length $n=1000$, $m=3$, $r=2$ (where m and r represent number of columns of X_F and the number of retained principal components respectively, please refer to Sec. II.A) and generate the data matrix $X_F\sim N\left(\vec{0},\ \mathrm{diag}(16,\ 4,\ 0.1)\right)$. Then define the rotation matrix $A_*(\omega)$ as in TABLE I, where $*=x,y,z$ denotes around which axis the rotation is made and ω denotes the rotation angle. Suppose we directly generate X_G^* by rotating X_F according to $X_G^*=X_F A_*(\omega)$, its covariance matrix satisfies

TABLE I. The interpretation of rotation around different axis

axis	affine matrix	The meaning of rotation
x	$A_x(\omega)=\begin{pmatrix}1 & 0 & 0\\ 0 & \cos(\omega) & \sin(\omega)\\ 0 & -\sin(\omega) & \cos(\omega)\end{pmatrix}$	rotating the second principal component to the residual subspace.
y	$A_y(\omega)=\begin{pmatrix}\cos(\omega) & 0 & \sin(\omega)\\ 0 & 1 & 0\\ -\sin(\omega) & 0 & \cos(\omega)\end{pmatrix}$	rotating the first principal component to the residual subspace.
z	$A_z(\omega)=\begin{pmatrix}\cos(\omega) & \sin(\omega) & 0\\ -\sin(\omega) & \cos(\omega) & 0\\ 0 & 0 & 1\end{pmatrix}$	rotating the principal components within the principal subspace.

$$\Sigma_G^*=\frac{\left(X_G^*\right)^T X_G^*}{n-1}=A_*^T(\omega)\Sigma_F A_*(\omega) \tag{12}$$

So to generate a data matrix X_G^* which is not only a rotation compared with X_F but also independent to X_F, we can first determine Σ_G^* according to $\Sigma_G^*=A_*^T(\omega)\Sigma_F A_*(\omega)$, and then generate X_G^* according to $X_G^*\sim N\left(0,\Sigma_G^*\right)$, for each axis, i.e. for $*=x,y,z$, and for $\omega=0^\circ\sim 45^\circ$, with an interval of 4.5°. Then we can observe how a PCA similarity factor changes with the gradual change of ω in each axis.

Similarly, for the scenario of eigenvalue change, let $n=1000$, $m=4$, $r=3$ and generate the data matrix $X_F\sim N\left(\vec{0},\ \mathrm{diag}(16,\ 9,\ 4,\ 0.1)\right)$. Define the scale matrix $B_i(\alpha)$, for $i=1,2,3$ as $B_1(\alpha)=diag(\alpha,1,1,1)$, $B_2(\alpha)=$

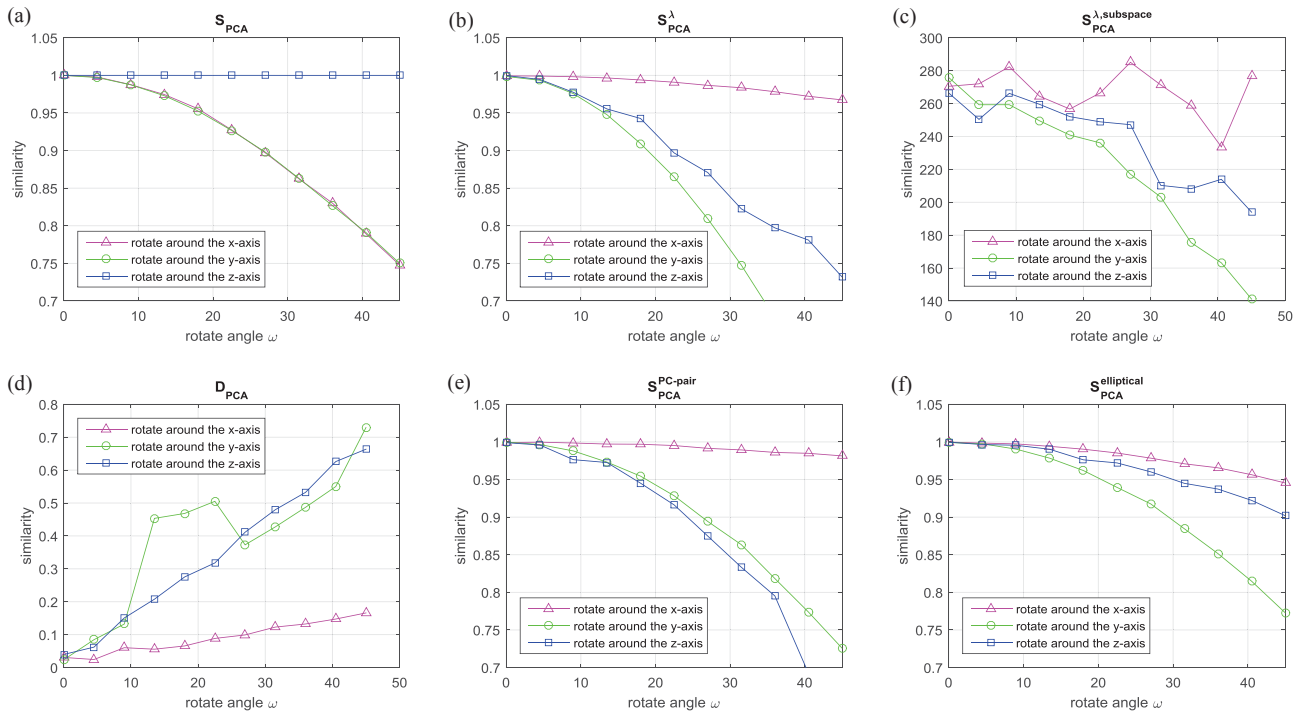

Figure 1. Different versions of PCA similarity factors with different rotation angle ω in different directions

$diag(1,\alpha,1,1)$ and $B_3(\alpha) = diag(1,1,\alpha,1)$ respectively, with the scale factor $\alpha = 0.3 \sim 3$, with an interval of 0.05. Then calculate Σ_G^i through $\Sigma_G^i = B_i^T(\alpha)\Sigma_F B_i(\alpha)$ and generate X_G^i according to $X_G^i \sim N(0, \Sigma_G^i)$. With the above definition, it is intuitive that the eigenvalues of the PCA model of X_G^i are obtained by changing those of X_F gradually. Then again, we can observe how a PCA similarity factor changes with the gradual change of α for each eigenvalue λ_i^G.

B. Rotation

Subfigures (a)-(e) of Figure 1 show the results of the five existing PCA similarities between X_F and X_G^* generated in Sec. A. It is worth noting that Subfigure (f) is for the result of the ellipsoid based PCA similarity proposed in this paper, and will be discussed in Sec. IV.D.

From Figure 1, the following conclusions can be made:

1. S_{PCA} keeps unchanged when the rotation is around z-axis. That is because the retained loading vectors in the PCA model of X_G^z cover the same subspace in spite of the rotation, as pointed out by Yao et al. [11]. In addition, although S_{PCA} changes with the rotation both around x-axis and around y-axis, it is still not appropriate as pointed out by Singhal et al. [10], because S_{PCA} have the same sensitivity to these two rotations, but the corresponding eigenvalues in the rotation direction vary significantly.
2. $S_{PCA}^{\lambda,subspace}$ in all cases and D_{PCA} for the rotation around y-axis is not smooth enough (i.e. fluctuate seriously), which has been verified in several repeated experiments but still need further investigation to give a physical interpretation.
3. Both S_{PCA}^{λ} and $S_{PCA}^{PC-pair}$ are not sensitive to the rotation around x-axis. This is because when they calculate the product of eigenvalues between F and G, according to (7) and (8), the biggest eigenvalues are paid too much attention.

The simulation results indicate that none of the existing methods can correctly reflect the gradual rotation in all three cases simultaneously.

C. Eigenvalue Change

It should be noted that a befitting similarity factor should reach the maximum (i.e. 1) when $\alpha = 1$, because in that case $B_i(\alpha)$ becomes a unitary matrix and X_G^i follows the same distribution as X_F. While as α increase or decrease from 1, the similarity factor should get smaller because X_G^i follows a distribution different from X_F.

Subfigures (a)-(e) of Figure 2 show the result of the five existing PCA similarities. Subfigure (f) is for the result of the ellipsoid based PCA similarity proposed in this paper, and will be discussed in Sec. IV.D.

From Figure 2, the following conclusions can be made:

1. S_{PCA} keeps unchanged in all cases because it does not consider the eigenvalues in the definition.
2. In all of the three cases, S_{PCA}^{λ} keeps unchanged when α varies around 1, which means that it cannot reflect the eigenvalues change. That is because for those values of α, although the eigenvalues λ_i^G changes with α,

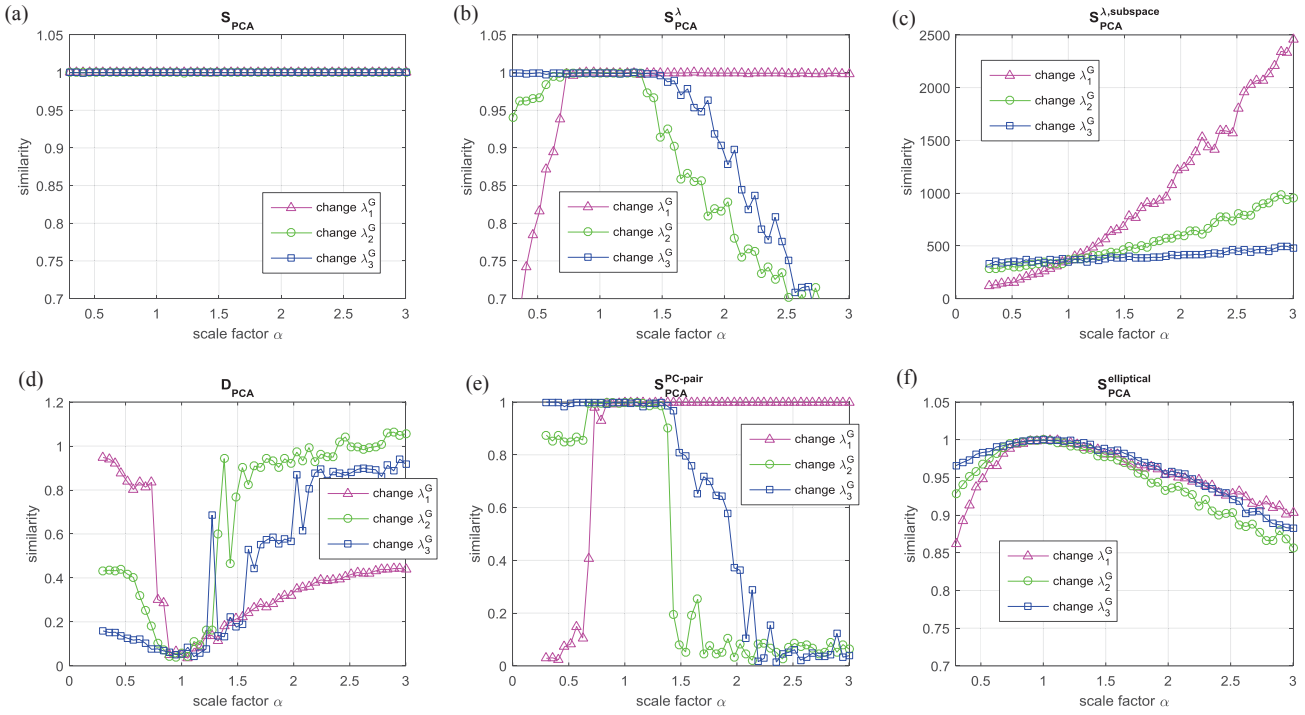

Figure 2. Different versions of PCA similarity factors with different scale factors ω of different eigenvalues

$P_{F_r} = P_{G_r}$ still holds, which leads to $S_{PCA}^{\lambda} = 1$ according to (7). Actually, S_{PCA}^{λ} will not change until the change of eigenvalues becomes great enough to cause the change of sequence of the principal components.

3. $S_{PCA}^{\lambda,subspace}$ keeps increasing as the eigenvalues of X_G^i grows, which is due to the lacking of normalization and is obviously inappropriate because the maximum value of a similarity factor should be reached when $\alpha = 1$.

4. Both D_{PCA} and $S_{PCA}^{PC-pair}$ sensitively rely on the sequence of the principal components, because they compare the loading vectors and eigenvalues by pair. Hence, they may suffer from abrupt changes when the sequence of principal components changes, as shown in subfigure (d) and (e). Especially, if similar eigenvalues exist in a PCA model, D_{PCA} and $S_{PCA}^{PC-pair}$ may fluctuate frequently because the sequence of the principal components may change frequently due to the influence of noise, drift or time varying change of the system.

The simulation results indicate that none of the existing methods can correctly reflect the eigenvalue change in all three cases simultaneously.

IV. An Ellipsoid Based PCA Similarity Factor

In this section, we first give a new definition for principal angle, which is called elliptical principal angle. Then based on it, propose a new PCA similarity factor by replacing the principal angles used in $S_{PCA}^{\lambda,subspace}$ of [14] with it.

A. Definition of Elliptical Principal Angle

In this paper, elliptical principal angles are defined recursively as following, which is a modification to that of principal angles defined in Sec. II.B [19],

$$cos\tilde{\theta}_1^{F,G} = \max_{u\in\Psi_F \cap v\in\Psi_G} \left(u^T v\right) = u_1^T v_1 \tag{13}$$

subject to

$$\begin{cases} u = P_{F_r}\tilde{\Lambda}_{F_r}^{1/2} y \\ v = P_{G_q}\tilde{\Lambda}_{G_q}^{1/2} z \\ \|y\|_2 = \|z\|_2 = 1 \end{cases} \tag{14}$$

and

$$cos\tilde{\theta}_k^{F,G} = \max_{u\in\Psi_F \cap v\in\Psi_G} \left(u^T v\right) = u_k^T v_k \tag{15}$$

subject to

$$\begin{cases} u = P_{F_r}\tilde{\Lambda}_{F_r}^{1/2} y \\ v = P_{G_q}\tilde{\Lambda}_{G_q}^{1/2} z \\ \|y\|_2 = \|z\|_2 = 1 \\ y_i^T y = 0,\ z_i^T z = 0,\ i = 1,2,\cdots,k-1 \end{cases} \tag{16}$$

where $\tilde{\Lambda}_{F_r} = \Lambda_{F_r} \Big/ trace\left(\Lambda_{F_r}\right), \tilde{\Lambda}_{G_q} = \Lambda_{G_q} \Big/ trace\left(\Lambda_{G_q}\right)$ and $k = 1,2,\cdots,q$.

B. Physical Interpretation

For convenience, let $y' = \Lambda_{F_r}^{1/2} y$ and $\sigma_{F,i}^2 = \tilde{\Lambda}_{F_r}[i,i]$ where $\tilde{\Lambda}_{F_r}[i,i]$ denotes the element of $\tilde{\Lambda}_{F_r}$ in column i and row i, then there is $y_i' = \sigma_{F,i} y_i$. So we can get

$$\frac{y_1'^2}{\sigma_{F,1}^2} + \frac{y_2'^2}{\sigma_{F,2}^2} + \ldots + \frac{y_r'^2}{\sigma_{F,r}^2} = \sum_{i=1}^{r} y_i^2 = \|y\|^2 = 1 \tag{17}$$

Because $u = P_{F_r}\tilde{\Lambda}_{F_r}^{1/2} y = P_{F_r} y'$ according to (16), y' can be seen as a coordinate in the subspace of Ψ_F rotated by the loading matrix P_F from original coordination system. According to (17), y' is a point on the surface of a r-dimension ellipsoid whose semi-axes are $\sigma_{F,i}$, $i = 1,\cdots,r$ and u is the original coordinate of y' according to $u = P_{F_r} y'$. Due to the normalized weight $\tilde{\Lambda}_{F_r}$ and $\tilde{\Lambda}_{G_q}$, the semi-axes satisfy

$$\sum_{i=1}^{r} \alpha_{F,i}^2 = trace\left(\tilde{\Lambda}_{F_r}\right) = 1 \tag{18}$$

Hence, the normalized weights on principal directions can guarantee that the sum of the square of the semi-axes of ellipsoid is 1.

It is worth noting that the angles $\tilde{\theta}_k^{F,G}$ defined above do not have the meaning of traditional angles any more, and actually measure the similarity of two normalized multidimensional ellipsoid by comparing their directions and shape. The directions with larger semi-axes will have more weights to describe the similarity.

C. An Ellipsoid Based PCA Similarity Factor

Similar to (10) and by replacing principal angles with elliptical principal angles, we can define the ellipsoid based PCA similarity factor as following

$$S_{PCA}^{elliptical} = \sum_{k=1}^{q} \cos\tilde{\theta}_k^{F,G} = \sum_{k=1}^{q} SVD_k\left(\tilde{\Lambda}_{G_q}^{1/2} P_{G_q}{}^T P_{F_r} \tilde{\Lambda}_{F_r}^{1/2}\right) \tag{19}$$

where $\tilde{\theta}_k^{F,G}$ is defined in (13) and (15). Since $S_{PCA}^{elliptical}$ in (19) involves both the directions and the shape of the normalized ellipsoid according to (16) and (17), which actually correspond to the loading matrix and eigenvalue matrix of a PCA model respectively, any change of the loading vectors or the eigenvalues will result in the change of the directions or the shape of its corresponding ellipsoid and finally affect the $S_{PCA}^{elliptical}$. When $\tilde{\Lambda}_{F_r} = \tilde{\Lambda}_{G_q}$ and $P_{G_q} = P_{F_r}$, we can get

$$S_{PCA}^{elliptical} = \sum_{k=1}^{q} SVD_k\left(\tilde{\Lambda}_{G_q}^{1/2}\tilde{\Lambda}_{F_r}^{1/2}\right) = trace\left(\tilde{\Lambda}_{F_r}\right) = 1 \tag{20}$$

Therefore, the normalized weight $\tilde{\Lambda}_{F_r}$ and $\tilde{\Lambda}_{G_q}$ also guarantee that the defined similarity factor $S_{PCA}^{elliptical}$ is normalized to a real number changing from 0 to 1.

Remark

Equation (19) can be equivalently rewrite as

$$S_{PCA}^{elliptical} = \frac{\sum_{k=1}^{q} SVD_k\left(\Lambda_{G_q}^{1/2} P_{G_q}{}^T P_{F_r} \Lambda_{F_r}^{1/2}\right)}{\sqrt{\left(\sum_{k=1}^{q}\lambda_k^G\right)\left(\sum_{k=1}^{r}\lambda_k^F\right)}} \tag{21}$$

According to (21) and (7), there are two differences between $S_{PCA}^{elliptical}$ and S_{PCA}^{λ}: (i) the former calculates the sum of the cosines while the latter calculates the sum of the square of the cosines; (ii) different denominators are used to normalize the values. Specifically, $S_{PCA}^{elliptical}=1$ if $\tilde{\Lambda}_{F_r}=\tilde{\Lambda}_{G_q}$ and $P_{G_q}=P_{F_r}$ are satisfied, while $S_{PCA}^{\lambda}=1$ as long as $P_{G_q}=P_{F_r}$ is satisfied.

D. Results of Simulation Experiments

From subfigure (f) of Figure 1, the following conclusions can be made:

(i) Compared with S_{PCA}, $S_{PCA}^{elliptical}$ can reflect the rotation around different axis with different sensitivity.

(ii) Compared with $S_{PCA}^{\lambda,subspace}$ and D_{PCA}, $S_{PCA}^{elliptical}$ changes more smoothly as the rotation angle changes.

(iii) Compared with S_{PCA}^{λ} and $S_{PCA}^{PC-pair}$, $S_{PCA}^{elliptical}$ is more sensitive to the rotation around x-axis.

From subfigure (f) of Figure 2, the following conclusions can be made:

(i) Compared with S_{PCA} and S_{PCA}^{λ}, $S_{PCA}^{elliptical}$ is sensitive to eigenvalue changes in all cases.

(ii) Compared with D_{PCA} and $S_{PCA}^{PC-pair}$, $S_{PCA}^{elliptical}$ can reflect the eigenvalue changes smoothly.

V. Conclusions

In this paper, we propose to evaluate how effective a PCA similarity factor is by observing whether it can reflect the gradual rotation and gradual eigenvalue change of a PCA model with sensitivity and smoothness, and gives a comprehensive comparison of several existing PCA similarity factors in this sense through simulation, which indicates some limitations of the existing methods. Then we define the elliptical principal angle, and based on it, propose a new PCA similarity factor which is satisfying in the above mentioned sense.

References

[1] Yin, S., et al., A review on basic data-driven approaches for industrial process monitoring. IEEE Transactions on Industrial Electronics, 2014. 61(11): p. 6418-6428.

[2] Ge, Z., Z. Song and F. Gao, Review of recent research on data-based process monitoring. Industrial & Engineering Chemistry Research, 2013. 52(10): p. 3543-3562.

[3] Yin, S., et al., A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. Journal of Process Control, 2012. 22(9): p. 1567-1581.

[4] Jeng, J., Adaptive process monitoring using efficient recursive PCA and moving window PCA algorithms. Journal of the Taiwan Institute of Chemical Engineers, 2010. 41(4): p. 475-481.

[5] Vanhatalo, E., Multivariate process monitoring of an experimental blast furnace. Quality and Reliability Engineering International, 2010. 26(5): p. 495-508.

[6] Li, S. and J. Wen, Application of pattern matching method for detecting faults in air handling unit system. Automation in Construction, 2014. 43: p. 49-58.

[7] Deng, X. and X. Tian, Nonlinear process fault pattern recognition using statistics kernel PCA similarity factor. Neurocomputing, 2013. 121: p. 298-308.

[8] Gunther, J.C., et al., Pattern matching in batch bioprocesses—comparisons across multiple products and operating conditions. Computers & Chemical Engineering, 2009. 33(1): p. 88-96.

[9] Singhal, A. and D.E. Seborg, Evaluation of a pattern matching method for the Tennessee Eastman challenge process. Journal of Process Control, 2006. 16(6): p. 601-613.

[10] Singhal, A. and D.E. Seborg, Pattern matching in historical batch data using PCA. IEEE control systems, 2002. 22(5): p. 53-63.

[11] Yao, Y. and F. Gao, Phase and transition based batch process modeling and online monitoring. Journal of Process Control, 2009. 19(5): p. 816-826.

[12] Zhao, C., et al., Stage-based soft-transition multiple PCA modeling and on-line monitoring strategy for batch processes. Journal of Process Control, 2007. 17(9): p. 728-741.

[13] Lu, N., F. Gao and F. Wang, Sub-PCA modeling and on-line monitoring strategy for batch processes. AIChE Journal, 2004. 50(1): p. 255-259.

[14] Zhao, S.J., J. Zhang and Y.M. Xu, Monitoring of processes with multiple operating modes through multiple principle component analysis models. Industrial & engineering chemistry research, 2004. 43(22): p. 7025-7035.

[15] Ge, Z. and Z. Song, Semiconductor manufacturing process monitoring based on adaptive substatistical PCA. IEEE Transactions on Semiconductor Manufacturing, 2010. 23(1): p. 99-108.

[16] Medina, J.M. and J.A. Díaz, Classification of batch processes in automotive metallic coatings using principal component analysis similarity factors from reflectance spectra. Progress in Organic Coatings, 2015. 88: p. 75-83.

[17] Krzanowski, W.J., Between-groups comparison of principal components. Journal of the American Statistical Association, 1979. 74(367): p. 703-707.

[18] Johannesmeyer, M.C., Abnormal situation analysis using pattern recognition techniques and historical data. 1999.

[19] Rck, K.B. and G.H. Golub, Numerical Methods for Computing Angles Between Linear Subspaces. Mathematics of computation, 1973. 27(123): p. 579 - 594.

[20] Zhao, S.J., J. Zhang and Y.M. Xu, Performance monitoring of processes with multiple operating modes through multiple PLS models. Journal of Process Control, 2006. 16(7): p. 763-772.

[21] Hwang, D. and C. Han, Real-time monitoring for a process with multiple operating modes. Control Engineering Practice, 1999. 7(7): p. 891-902.