

## Research Article

# Nonlinear Model for Condition Monitoring and Fault Detection Based on Nonlocal Kernel Orthogonal Preserving Embedding

Bo She , Fuqing Tian , Weige Liang, and Gang Zhang

*Department of Weaponry Engineering, Naval University of Engineering, Wuhan 430000, China*

Correspondence should be addressed to Fuqing Tian; [tianfq001@126.com](mailto:tianfq001@126.com)

Received 30 November 2017; Revised 28 April 2018; Accepted 5 May 2018; Published 11 June 2018

Academic Editor: Juan C. G. Prada

Copyright © 2018 Bo She et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The dimension reduction methods have been proved powerful and practical to extract latent features in the signal for process monitoring. A linear dimension reduction method called nonlocal orthogonal preserving embedding (NLOPE) and its nonlinear form named nonlocal kernel orthogonal preserving embedding (NLKOPE) are proposed and applied for condition monitoring and fault detection. Different from kernel orthogonal neighborhood preserving embedding (KONPE) and kernel principal component analysis (KPCA), the NLOPE and NLKOPE models aim at preserving global and local data structures simultaneously by constructing a dual-objective optimization function. In order to adjust the trade-off between global and local data structures, a weighted parameter is introduced to balance the objective function. Compared with KONPE and KPCA, NLKOPE combines both the advantages of KONPE and KPCA, and NLKOPE is also more powerful in extracting potential useful features in nonlinear data set than NLOPE. For the purpose of condition monitoring and fault detection, monitoring statistics are constructed in feature space. Finally, three case studies on the gearbox and bearing test rig are carried out to demonstrate the effectiveness of the proposed nonlinear fault detection method.

## 1. Introduction

Mechanical equipment is widely used in modern industrial production, but it often suffers from damage during the long time operation, such as the fracture of bearings and the broken tooth of gears; the defect of these parts may cause the performance of the machine to degrade, or even cause security accidents. Therefore, the fault detection of mechanical equipment is of great significance to ensure the safety of the industrial production process and the economic benefits. In recent years, the multivariate statistical process monitoring (MSPM) technique has been developed and used to detect the faults in industrial production process, such as principal component analysis (PCA) [1], partial least squares (PLS) [2], and independent component analysis (ICA) [3]. These classical monitoring methods perform dimension reduction on the process data and extract few components to construct monitoring statistics which can reflect the characteristics of the original data, at this point, the performance of dimension reduction will affect the monitoring effect.

Multivariate data-driven statistical PCA-based monitoring framework is the most frequently employed method in condition monitoring and fault detection field. To overcome the weakness that linear monitoring method may perform poorly in processing the nonlinear monitoring processes, KPCA-based monitoring method is widely investigated and used to detect faults [4, 5]. Although the improved PCA-based monitoring methods can retain latent features of raw data, they only capture the global structure of the data, and the local structure characteristics in the data have been ignored. However, the features extracted from the local structure of the data can also represent the different aspects of the data. The loss of the important information may have impact on dimension reduction and monitoring result [6].

As opposed to the global data structure preserving dimension reduction techniques, manifold learning methods have been developed to preserve the local data structure characteristics, represented by Laplacian eigenmap (LE) [7], local preserving projections (LPP) [8], locally linear embedding (LLE) [9], and neighborhood preserving embedding

(NPE) [10]. LPP and NPE both are linear projection methods that can process the testing data conveniently; manifold learning based monitoring methods can overcome some limits of the PCA-based monitoring method. However, these manifold learning methods only consider the neighborhood relationships to preserve local properties among samples and thus may also lose crucial information contained in the global data structure. In order to take both global and local data structure characteristics into account, the methods which unify LPP and PCA have been proposed, and the fault detection performances have proven to be better than LPP and PCA [11, 12]. But these approaches are still linear methods, as they are employed to process the nonlinear process data; these methods have limitations and may obtain a poor monitoring performance.

On the other hand, kernel function is usually investigated to extend linear methods to nonlinear methods, by mapping original data from input space into high dimension feature space, and then perform the linear method in feature space. For the purpose of taking full advantages of global and local data structure and processing the nonlinear monitoring problem efficiently, a kernel global-local preserving projections (KGLPP) method [13] based on KLPP and KPCA has been proposed, and the results show that it outperforms the linear global-local preserving projections (GLPP) method [14]. Orthogonal neighborhood preserving embedding (ONPE) is an orthogonal form of conventional NPE algorithm, which adds an additional orthogonal constraint on the projection vectors [15]; thus, ONPE not only inherits the local structure preserving feature, but also can avoid the distortion defects of NPE [16]. Moreover, the orthogonal property is also an advantage for fault detection and fault diagnosis. Firstly, orthogonal transformations can enhance the locality preserving power, which is effective in data reconstruction and computing reconstruction error; it is useful for fault detection. Secondly, the dimension reduction methods considering orthogonal constraint can improve the performance of identification, which is helpful to detect fault effectively [17].

In this paper, a new nonlinear dimension reduction method named nonlocal kernel orthogonal preserving embedding (NLKOPE) is proposed on the basis of a linear dimension reduction method named nonlocal orthogonal preserving embedding (NLOPE). NLOPE takes both advantages of ONPE and PCA into account; NLKOPE is a nonlinear extension of NLOPE. The exponentially weighted moving average (EWMA) statistic is built for condition monitoring and fault detection. To verify the effectiveness of our proposed methods, these methods are employed to detect the faults of gearbox and evaluate the performance degradation of bearing. In order to diagnose the fault type of the bearing, dual-tree complex wavelet packet transform (DTCWPT) is used for noise reduction, and Hilbert transform envelope algorithm is employed to extract the fault characteristic frequency.

The rest of the paper is organized as follows. KPCA, ONPE, and KONPE are reviewed and analyzed in Section 2. The proposed NLOPE-based monitoring method is developed in Section 3. The proposed NLKOPE-based monitoring

method is developed in Section 4. In Section 5, three cases are used to demonstrate the effectiveness of the proposed methods. Finally, conclusions are drawn in Section 6.

## 2. Background Techniques

*2.1. Kernel Principal Component Analysis.* As a multivariate method, PCA is widely used for process monitoring. However, for some complicated cases in industrial processes with nonlinear characteristics, PCA performs poorly as it takes the process data as linear, and some useful nonlinear features may be lost when the PCA is used to reduce the dimension and extract features. KPCA performs a nonlinear PCA that constructs a nonlinear mapping from the input space to the feature space through the kernel function. Given data set  $\{x_1, x_2, \dots, x_N\} \in R^m$ , where  $N$  is the number of samples and  $m$  is the number of variables, the samples in the input space are extended into the feature space by using a nonlinear mapping  $\Phi : R^m \rightarrow F$ , the covariance matrix in the feature space  $F$  can be expressed as

$$C^F = \frac{1}{N} \sum_{j=1}^N \Phi(x_j) \Phi(x_j)^T \quad (1)$$

where it is assumed that the data set  $\{\Phi(x_1), \Phi(x_2), \dots, \Phi(x_N)\}$  in feature space is centered and  $\sum_{j=1}^N \Phi(x_j) = 0$ . The principal component can be calculated by solving the eigenvalue problem in the feature space.

$$\lambda v = C^F v \quad (2)$$

$$\begin{aligned} C^F v &= \frac{1}{N} \sum_{j=1}^N \left( \Phi(x_j) \Phi(x_j)^T \right) v \\ &= \frac{1}{N} \sum_{j=1}^N \langle \Phi(x_j), v \rangle \Phi(x_j) \end{aligned} \quad (3)$$

where  $\langle \Phi(x_j), v \rangle$  denotes the dot product between  $\Phi(x_j)$  and  $v$  and  $\lambda$  denotes eigenvalue and  $v$  denotes eigenvector. For  $\lambda \neq 0$ , eigenvector  $v$  is regarded as a linear combination of  $\{\Phi(x_1), \Phi(x_2), \dots, \Phi(x_N)\}$  with the coefficients  $a_1, a_2, \dots, a_N$ ,  $v = \sum_{i=1}^N a_i \Phi(x_i)$ .

Multiplying  $\Phi(x_k)$  at both sides of (2),  $k = 1, 2, \dots, N$ , we obtain

$$\begin{aligned} &\lambda \sum_{i=1}^N a_i \langle \Phi(x_k), \Phi(x_i) \rangle \\ &= \frac{1}{N} \sum_{i=1}^N a_j \left\langle \Phi(x_k), \sum_{j=1}^N \Phi(x_j) \right\rangle \langle \Phi(x_i), \Phi(x_j) \rangle \end{aligned} \quad (4)$$

Defining a  $N \times N$  kernel matrix  $K$  with  $K_{ij} = \langle \Phi(x_i), \Phi(x_j) \rangle$ , (4) can be expressed as

$$\lambda \alpha = \frac{K}{N} \alpha \quad (5)$$

where  $\alpha = [a_1, a_2, \dots, a_N]^T$ , by solving the eigenvalue problem of (5), this yields eigenvectors  $a_1, a_2, \dots, a_N$  with

the eigenvalues  $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N$ . The coefficients  $a$  are normalized to satisfy  $\|a_k^2\| = 1/\lambda_k$  and to ensure  $\|v_k^2\| = 1$ . The projection of a test sample  $x$  is obtained as follows:

$$t_k = \langle v_k, \Phi(x) \rangle = \sum_{i=1}^N a_i^k \langle \Phi(x_i), \Phi(x) \rangle \quad (6)$$

## 2.2. Kernel Orthogonal Neighborhood Preserving Embedding

**2.2.1. Orthogonal Neighborhood Preserving Embedding.** Given a high dimensional data set  $X = \{x_1, x_2, \dots, x_N\} \in R^m$ , as a linear dimension reduction method, ONPE is used to find a transformation matrix  $A = [a_1, a_2, \dots, a_d] \in R^{m \times d}$  ( $d < m$ ) that maps the high dimensional data set to the low dimensional data set  $Y = \{y_1, y_2, \dots, y_N\} \in R^d$ , that is,  $Y = A^T X$ . In NPE algorithm, in order to preserve the local geometric structure, the adjacency graph is built to reflect the relationship between samples; each sample can be reconstructed by a linear combination of its neighbors with the corresponding weight coefficients. The weight coefficients matrix  $W$  is computed by minimizing the following objective function:

$$\min \sum_i \left\| x_i - \sum_j W_{ij} x_j \right\|^2 \quad (7)$$

The reconstruction weight coefficients between  $x_i$  and its neighbors are preserved to reconstruct  $y_i$  by using its corresponding neighbors. Then, the low dimensional embedding  $Y$  is obtained by optimizing the error function:

$$\min \sum_i \left\| y_i - \sum_j W_{ij} y_j \right\|^2 \quad (8)$$

$$\text{s.t. } Y^T Y = A^T X X^T A = I \quad (9)$$

where  $\sum_{j=1}^k W_{ij} = 1$ ,  $i = 1, 2, \dots, N$ ,  $j = 1, 2, \dots, k$ , and  $k$  is the number of neighbors of  $x_i$ . If  $x_j$  is not the neighbor of  $x_i$ ,  $W_{ij} = 0$ . In ONPE, any high dimensional data can be mapped into the reduced space by the orthogonal projection matrix  $A$ , according to (7)-(8), matrix  $A$  is obtained by the following formulations:

$$a_1 = \arg \min_a \sum_i \left\| y_i - \sum_j W_{ij} y_j \right\|^2 \quad (10)$$

$$= \arg \min_a A^T X M X^T A$$

$$\text{s.t. } A^T X X^T A = I \quad (11)$$

$$a_k = \arg \min_a \sum_i \left\| y_i - \sum_j W_{ij} y_j \right\|^2 \quad (12)$$

$$= \arg \min_a A^T X M X^T A$$

$$\text{s.t. } v^T v_1 = v^T v_2 = \dots = v^T v_{k-1} = 0 \quad (13)$$

$$A^T X X^T A = I$$

where  $k = 2, 3, \dots, d$ ,  $M = (I - W)^T (I - W)$ . Using the Lagrange multiplier method, the orthogonal vectors  $A$  is calculated iteratively as follows:

(1)  $a_1$  is the eigenvector corresponding to the smallest eigenvalue of matrix  $(X X^T)^{-1} X M X^T$ .

(2)  $a_k$  is the eigenvector corresponding to the smallest eigenvalue of matrix  $Q^{(k)}$ :

$$Q^{(k)} = \left\{ I - (X X^T)^{-1} \cdot A^{(k-1)} \left[ (A^{(k-1)})^T (X X^T)^{-1} A^{(k-1)} \right]^{-1} (A^{(k-1)})^T \right\} \cdot (X X^T)^{-1} X M X^T \quad (14)$$

where  $A^{(k-1)} = [a_1, a_2, \dots, a_{k-1}]$ .

**2.2.2. Kernel Orthogonal Neighborhood Preserving Embedding.** ONPE algorithm has some nonlinear data processing ability, but it is essentially a linear method, the limitations are obvious as it comes to extract the nonlinear features of data. KONPE is a nonlinear extension of ONPE; the original data are mapped into the kernel space that it is feasible to use ONPE to obtain low dimensional nature characteristics.

Given a data set  $X = \{x_1, x_2, \dots, x_N\} \in R^m$ , the data are projected on to the high dimensional feature space  $\{\Phi(x_1), \Phi(x_2), \dots, \Phi(x_N)\}$  by using the nonlinear mapping function  $\Phi$ , assuming that the data is centered and satisfies  $\sum_{j=1}^N \Phi(x_j) = 0$ . Defining  $v = [v_1, v_2, \dots, v_d]$  be the linear mapping matrix to project the data from feature space  $\Phi(x)$  to the low dimensional space, the low dimensional embedding  $Y = \{y_1, y_2, \dots, y_N\} \in R^d$  are obtained by using matrix  $v$ , that is  $Y = v^T \Phi(x)$ . The matrix  $v$  can be expressed as a linear combination of  $\Phi(x)$  with the coefficients  $a_i^l$  ( $i = 1, 2, \dots, N$ ).

$$v_l = \sum_{i=1}^N a_i^l \Phi(x_i) \quad (15)$$

where  $l = 1, 2, \dots, d$ . According to  $K_{ij} = \Phi(x_i) \Phi(x_j)$ , the formulations for calculating the matrix  $v$  are as follows:

$$v_1 = \arg \min_v \sum_i \left\| v^T \Phi(x_i) - \sum_j W_{ij} v^T \Phi(x_j) \right\|^2 \quad (16)$$

$$= a^T K^T M K a$$

$$\text{s.t. } v^T \Phi(X) \Phi(X)^T v = 1 \quad (17)$$

$$v_k = \arg \min_v \sum_i \left\| v^T \Phi(x_i) - \sum_j W_{ij} v^T \Phi(x_j) \right\|^2 \quad (18)$$

$$= a^T K^T M K a$$

$$\text{s.t. } v^T v_1 = v^T v_2 = \dots = v^T v_{k-1} = 0 \quad (19)$$

$$v^T \Phi(X) \Phi(X)^T v = 1$$

where  $k = 2, 3, \dots, d$ ,  $M = (I - W)^T(I - W)$ . The problem of computing  $v$  can be converted to solve the coefficients vectors  $a$  based on the Lagrange multiplier method. According to (16) and (18), the specific steps are as follows:

- (1)  $a_1$  is the eigenvector corresponding to the smallest eigenvalue of matrix  $(KK^T)^{-1}KMK^T$ .
- (2)  $a_k$  is the eigenvector corresponding to the smallest eigenvalue of matrix  $Q^{(k)}$ :

$$Q^{(k)} = \left\{ I - (K^T K)^{-1} \cdot K^T a^{(k-1)} \left[ (a^{(k-1)})^T (K^T K)^{-1} K^T a^{(k-1)} \right]^{-1} \cdot (a^{(k-1)})^T \right\} (K^T K)^{-1} K^T M K \quad (20)$$

where  $a^{(k-1)} = [a_1, a_2, \dots, a_{k-1}]$ ; the kernel matrix  $K$  should be centered by

$$\bar{K} = K - EK - KE + EKE \quad (21)$$

where  $E_{ij} = 1/N$  ( $i, j = 1, 2, \dots, N$ ).

Given a test sample  $x_{new}$ , the  $k$ th variable of the low dimensional sample  $y_{new}$  is obtained by mapping  $\Phi(x_{new})$  into vector  $v_k$  in the feature space,  $k = 1, 2, \dots, d$ .

$$y_{new}^k = v_k \Phi(x_{new}) = \sum_{i=1}^N a_k^i (\Phi(x_i) \Phi(x_k)) = \sum_{i=1}^N a_k^i \bar{K}_{new}(x_i, x_{new}) \quad (22)$$

The test kernel matrix should also be centered as follows:

$$\bar{K}_{new} = K_{new} - eK - K_{new}E + eKE \quad (23)$$

where  $K_{new} = \Phi(x_{new})\Phi(x)$ ,  $e = 1/N[1, \dots, 1] \in R^{1 \times N}$ .

### 3. Nonlocal Orthogonal Preserving Embedding

**3.1. Algorithm Description.** In order to preserve both the local and global data structures, NLOPE algorithm is proposed to unify the advantages of PCA and ONPE. Given a data set  $x = \{x_1, x_2, \dots, x_N\} \in R^{m \times N}$ , the objective function of NLOPE is as follows:

$$J(a)_{NLOPE} = \eta J(a)_{Local} - (1 - \eta) J(a)_{Global} = \eta \min_a a^T x M x^T a - (1 - \eta) \max_a a^T C a = \min_a a^T (\eta x M x^T - (1 - \eta) C) a = \min_a a^T (\eta L' - (1 - \eta) C) a \quad (24)$$

$$\text{s.t. } a_k^T a_1 = a_k^T a_2 = \dots = a_k^T a_{k-1} = 0 \quad (25)$$

$$a^T [\eta x x^T + (1 - \eta) I] a = 1$$

where  $C = (1/N) \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T$ ,  $\bar{x} = (1/N) \sum_{i=1}^N x_i$ .

Using the Lagrange multiplier method, the projection vector  $a$  can be calculated by solving following eigenvector problems:

- (1)  $a_1$  is the eigenvector corresponding to the smallest eigenvalue of matrix  $S^{-1}D$ .
- (2)  $a_k$  is the eigenvector corresponding to the smallest eigenvalue of matrix  $Q^{(k)}$ :

$$Q^{(k)} = \left\{ I - (S)^{-1} a^{(k-1)} \left[ (a^{(k-1)})^T (S)^{-1} a^{(k-1)} \right]^{-1} (a^{(k-1)})^T \right\} \cdot S^{-1} D \quad (26)$$

where  $k = 2, 3, \dots, d$ ,  $d$  is the dimension of samples in NLOPE space,  $a^{(k-1)} = [a_1, a_2, \dots, a_{k-1}]$ ,  $S = \eta x x^T + (1 - \eta)I$ , and  $D = \eta x M x^T - (1 - \eta)C$ . A strict mathematical proof of the projection vector  $a$  is given in Appendix A.

**3.2. Selection of Parameter  $\eta$ .** The parameter  $\eta$  describes different roles of global and local data structure preserving in constructing the NLOPE model; it is important to choose an appropriate value of  $\eta$ , which will affect the extraction of latent variables. As we need to solve a dual-objective optimization problem, usually it is hard to find an absolutely optimal solution which simultaneously optimizes the two subobjectives. However, it is possible to obtain a relatively optimal solution by making balance between them. The parameter  $\eta$  is used to balance the matrix  $L'$  and matrix  $C$  in (24), and it can be regarded as balancing the energy variations of  $L'$  and  $C$ . Thus, we choose spectral radius of the matrix to estimate the value of  $\eta$ .

To balance the global and local structure of the data,  $\eta$  can be selected as follows [6]:

$$\eta S_{Local} = (1 - \eta) S_{Global} \quad (27)$$

where  $S_{Global} = \rho(C)$  and  $S_{Local} = \rho(L')$  denote the energy variations of  $J(a)_{Global}$  and  $J(a)_{Local}$ .  $\rho(\cdot)$  is the spectral radius of the matrix.  $L'$  and  $C$  are defined in (24). Thus,  $\eta$  is computed by

$$\eta = \frac{\rho(C)}{\rho(L') + \rho(C)} \quad (28)$$

**3.3. Monitoring Model.** In PCA-based monitoring method, Hotelling's  $T^2$  statistic and the squared prediction error  $SPE$  statistic are often used for fault detection. Similarly, the two monitoring statistics are applied in NLOPE-based model. Hotelling's  $T^2$  is used to measure the variation in latent variable space and detects a new sample if the variation in the latent variables is greater than the variation explained by the model, which is computed as

$$T^2 = y^T \Lambda^{-1} y \quad (29)$$

where  $\Lambda = yy^T/(N - 1)$  is the covariance matrix of the projection vectors of training samples in the NLOPE subspace.

The squared prediction error  $SPE$  statistic is a measurement of the variation in residual space and is used to measure the goodness of fit of the new sample to the model; it is defined as follows [20]:

$$SPE = \langle x, x \rangle - \langle y, y \rangle \quad (30)$$

where  $y$  is the embedding of input sample  $x$  in NLOPE subspace.

As it is hard to estimate the condition of the machine only by the original vibration signal, some features need to be constructed. Time-domain and frequency-domain features can be generated from vibration data and are also widely used to characterize the state of machinery, time-domain features such as kurtosis, crest factor, and impulse factor, i.e., are sensitive to impulsive oscillation, and frequency-domain features can reveal some information that can not be characterized by time-domain features. In this study, 11 time-domain features and 13 frequency-domain features [21] were extracted from each sample to construct the high dimensional feature sample. For the purpose of condition monitoring and fault detection, it is critical to extract the most useful information hidden in current machine state. Therefore, the dimension reduction methods can be employed to extract latent features effectively.

In order to detect the incipient fault of mechanical equipment more accurately and reliability, the exponentially weighted moving average (EWMA) statistic based on a combined index of  $T^2$  and  $SPE$  statistics is developed to detect the fault of the mechanical equipment. The combined index  $U$  is a summation of  $T^2$  and  $SPE$  statistics as follows:

$$U = \frac{T^2}{L_{T^2}} + \frac{SPE}{L_{SPE}} \quad (31)$$

where  $L_{T^2}$  and  $L_{SPE}$  are the control limits of  $T^2$  and  $SPE$  statistics, they can be computed by kernel density estimation (KDE) algorithm, the values of  $T^2/L_{T^2}$  statistic should be normalized between 0 and 1 by using the maximal value and minimum value of  $T^2/L_{T^2}$ , and the values of  $SPE/L_{SPE}$  statistic should be normalized, too.

The EWMA statistic is computed as follows:

$$W_t = (1 - \gamma)W_{t-1} + \gamma U_t \quad (32)$$

where  $W_t$  is calculated by the average of preliminary data and  $\gamma$  is a smoothing constant between 0 and 1. While  $\gamma$  is large, the value of  $W_t$  puts more weight on the current statistic  $U_t$  than on the historic statistic. Calculate the control limit  $L_{EWMA}$  for EWMA statistic by KDE method, too. In this study, the value of  $\gamma$  is set to 0.2.

The offline modeling procedure is listed as follows:

- (1) The healthy samples are used to be the training samples, convert each original training sample into the high dimensional feature sample, and then normalize the high dimensional feature samples  $x$  to zero mean and unit variance.

- (2) Use (26) to calculate the projection matrix  $a$ , and calculate the projected vectors of the training samples in the NLOPE subspace.
- (3) Compute  $T^2$  and  $SPE$  statistics of all training samples, and calculate the control limits  $L_{T^2}$  and  $L_{SPE}$ , and then obtain the EWMA statistics and the control limit  $L_{EWMA}$ .

The online monitoring procedure is listed as follows:

- (1) Convert each testing sample into the high dimensional feature sample, and normalize the high dimensional feature samples  $x_{new}$  with the mean and variance of the training feature samples  $x$ .
- (2) Calculate the projected vectors of testing samples as  $y_{new} = a^T x_{new}$ .
- (3) Compute EWMA statistics associated with  $x_{new}$ , and monitor if they exceed the control limit  $L_{EWMA}$ .

## 4. Nonlocal Kernel Orthogonal Preserving Embedding

**4.1. Algorithm Description.** NLKOPE algorithm performs a nonlinear NLOPE by using the kernel trick. Given a data set  $x = \{x_1, x_2, \dots, x_N\} \in R^{m \times N}$ , the nonlinear mapping  $\Phi$  is used to project  $x$  onto the feature space, the data set  $\Phi(x) = \{\Phi(x_1), \Phi(x_2), \dots, \Phi(x_N)\}$  in the feature space is assumed to be centered, and  $\sum_{j=1}^N \Phi(x_j) = 0$ . The objective function of NLKOPE algorithm is computed as follows:

$$\begin{aligned} J(a)_{NLKOPE} &= \eta J(a)_{Local} - (1 - \eta) J(a)_{Global} \\ &= \eta \min_a a^T \bar{K}^T M K a \\ &\quad - (1 - \eta) \max_a a^T \frac{\bar{K}^T \bar{K}}{N} a \\ &= \min_a a^T \left( \eta \bar{K}^T M \bar{K} - (1 - \eta) \frac{\bar{K}^T \bar{K}}{N} \right) a \\ &= \min_a a^T (\eta L' - (1 - \eta) L) a \end{aligned} \quad (33)$$

$$\begin{aligned} \text{s.t. } v^T v_1 = \dots = v^T v_{k-1} &= a^T \bar{K}^T a_{k-1} = 0 \\ v^T \Phi(X) \Phi(X)^T v &= a^T \bar{K}^T K a = 1 \\ v^T v &= a^T \bar{K} a = 1 \end{aligned} \quad (34)$$

According to (33), computing  $v$  is reduced to obtain the coefficients vectors  $a$ , using the Lagrange multiplier method, the formulation is converted as follows:

$$\begin{aligned} a^T \left[ \eta \bar{K}^T M \bar{K} - (1 - \eta) \frac{\bar{K}^T \bar{K}}{N} \right] a \\ - \lambda \left\{ a^T \left[ \eta (\bar{K}^T \bar{K} + \bar{K}) + (1 - \eta) I \right] a - 1 \right\} \\ - \sum_{i=1}^k u_i a_i^T \bar{K}^T a_i = 0 \end{aligned} \quad (35)$$

The coefficients vectors  $a$  are obtained as

- (1)  $a_1$  is the eigenvector corresponding to the smallest eigenvalue of matrix  $Q^{(1)}$ :

$$Q^{(1)} = S^{-1}D \quad (36)$$

- (2)  $a_k$  is the eigenvector corresponding to the smallest eigenvalue of matrix  $Q^{(k)}$ :

$$Q^{(k)} = \left\{ I - S^{-1}\bar{K}^T a^{(k-1)} \left[ (a^{(k-1)})^T S^{-1}\bar{K}^T a^{(k-1)} \right]^{-1} \cdot (a^{(k-1)})^T \right\} S^{-1}D \quad (37)$$

where  $k = 2, 3, \dots, d$ ,  $d$  is the dimension of samples in NLKOPE space,  $a^{(k-1)} = [a_1, a_2, \dots, a_{k-1}]$ ,  $S = \eta(\bar{K}^T \bar{K} + \bar{K}) + (1-\eta)I$ ,  $D = \eta\bar{K}^T M\bar{K} - (1-\eta)\bar{K}^T \bar{K}/N$ , and  $\bar{K}$  is the centered kernel matrix of  $x$  obtained by (21). A strict mathematical proof of the coefficients vectors  $a$  is given in Appendix B.

4.2. *Selection of Parameter  $\eta$* . The method to choose the parameter  $\eta$  in NLKOPE model is same as in the NLOPE model, while the values of spectral radius are different, and  $\eta$  is set as

$$\eta = \frac{\rho(L)}{\rho(L') + \rho(L)} \quad (38)$$

where  $L = \bar{K}^T \bar{K}/N$ ,  $L' = \bar{K}^T M\bar{K}$ .

4.3. *Monitoring Model*. Hotelling's  $T^2$  statistic and the squared prediction error  $SPE$  statistic are also used in NLKOPE-based model to monitor the abnormal variations. The  $T^2$  statistic defines as

$$T^2 = y^T \Lambda^{-1} y \quad (39)$$

where  $\Lambda = yy^T/(N-1)$  is the covariance matrix of the projection vectors in the training samples.

The  $SPE$  statistic defines as [22]

$$SPE = \langle \Phi(x), \Phi(x) \rangle - \langle y, y \rangle \\ = k(x, x) - \frac{2}{N} \sum_{i=1}^N k(x_i, x) + \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N k(x_i, x_j) - y^T y \quad (40)$$

$$= 1 - \frac{2}{N} \sum_{i=1}^N k(x_i, x) + \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N k(x_i, x_j) - y^T y$$

$$y_{new} = a^T \bar{K}_{new}(x, x_{new}) \quad (41)$$

where  $\bar{K}_{new}(x, x_{new})$  is the centered kernel vector of  $x_{new}$  obtained via (23).

The offline modeling procedure is listed as follows:

- (1) The healthy samples are used to be the training samples, convert each original training sample into the high dimensional feature sample, and then normalize the high dimensional feature samples  $x$  to zero mean and unit variance.
- (2) Compute the kernel matrix  $K$  by selecting a kernel function, and center the kernel matrix  $K$  via (21).
- (3) Obtain the projection matrix  $v$  by calculating the eigenvector problem of (36)-(37).
- (4) Compute  $T^2$  and  $SPE$  statistics of all training samples, and calculate the control limits  $L_{T^2}$  and  $L_{SPE}$ , and then obtain the EWMA statistics and the control limit  $L_{EWMA}$ .

The online monitoring procedure is listed as follows:

- (1) Convert each testing sample into the high dimensional feature sample, and normalize the high dimensional feature samples  $x_{new}$  with the mean and variance of the training feature samples  $x$ .
- (2) Compute the kernel vector  $K_{new}(x, x_{new})$  and center it to get  $\bar{K}_{new}(x, x_{new})$  via (23).
- (3) Calculate the projected vectors of testing samples  $y_{new}$  via (41).
- (4) Compute EWMA statistics associated with  $x_{new}$ , and judge whether they exceed the control limit  $L_{EWMA}$ .

The procedure of condition monitoring and fault detection by the method of NLKOPE is shown in Figure 1. The healthy vibration signals are collected to implement NLKOPE and construct the offline model, and then the model will be employed to implement online condition monitoring, fault detection, and performance degradation assessment.

## 5. Case Studies and Result Analysis

5.1. *Fault Detection of Gearboxes*. The 2009 PHM gearbox fault data [18] is a representative of generic industrial gearbox data; we use it to evaluate the proposed methods. The gearbox contains 4 gears, 6 bearings, and 3 shafts, the measured signals consist of two accelerometer signals and a tachometer signal with a sampling frequency of 66.67 kHz, and the schematic and overview of the gearbox are shown in Figure 2. In this study, 3 different health conditions of the helical gearbox under low load and 30 Hz speed are used to test the effect of fault detection, and the detailed description of the data and pattern is shown in Table 1. In the pattern of health, all the mechanical elements in the gearbox are normal. In the pattern of fault 1, the gear with 24 teeth on the idler shaft is chipped. In the pattern of fault 2, the gear with 24 teeth on the idler shaft is broken, and the bearing at output side of the idler shaft also has inner race defect.

In this case, 1024 sampling points are selected as a sample, and we extract 30 samples for each pattern. The first 30 samples from the pattern of health are used as training samples, and the remaining 60 samples from pattern of fault 1

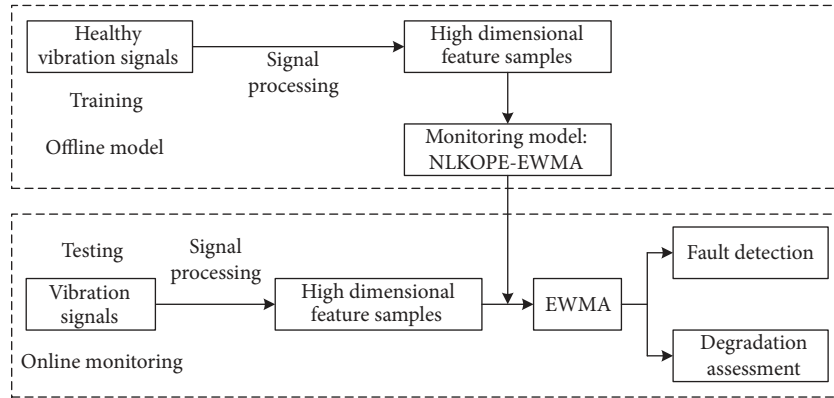


FIGURE 1: Procedure of condition monitoring and fault detection.

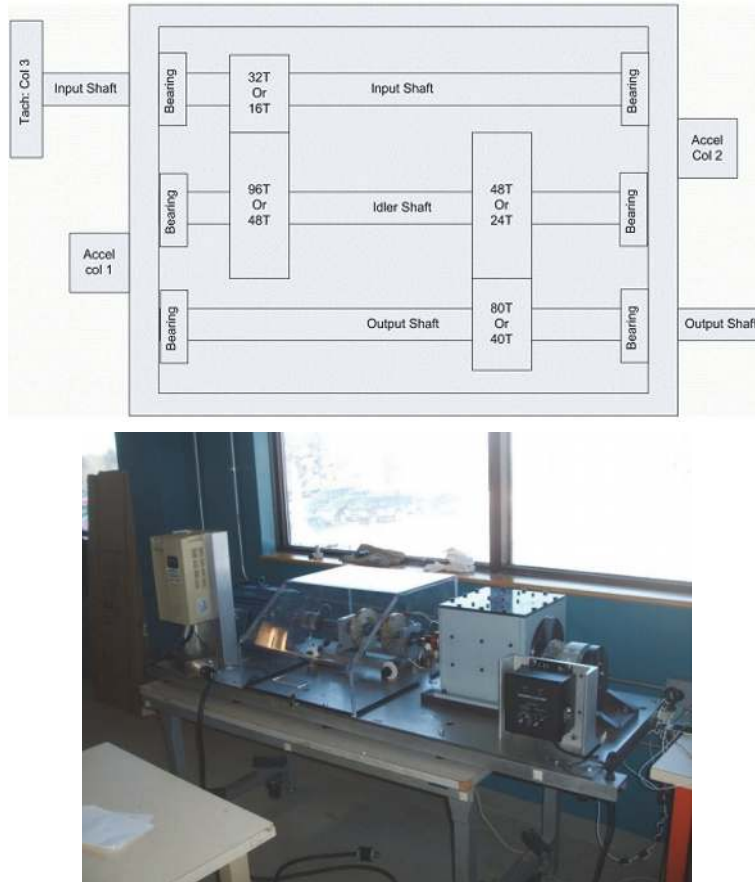


FIGURE 2: Schematic and overview of the gearbox used in PHM 2009 Challenge Data [18].

TABLE 1: Pattern description of the gearbox: IS=input shaft; IS=input side; ID=idler shaft; OS=output Side; OS=output shaft.

Pattern	Gear				Bearing						Shaft	
	16T	48T	24T	40T	IS:IS	ID:IS	OS:IS	IS:OS	ID:OS	OS:OS	Input	Output
Health	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good
Fault 1	Good	Good	Chipped	Good	Good	Good	Good	Good	Good	Good	Good	Good
Fault 2	Good	Good	Broken	Good	Good	Good	Good	Good	Inner	Good	Good	Good

TABLE 2: Fault detection rates of five methods.

	KPCA	KONPE	NLOPE	KGLPP [13]	NLKOPE
FDR(%)	90	96.67	98.33	98.33	100

and fault 2 are collected as testing samples. In other words, we use these 90 samples to detect whether the gearbox is faulty, and actually the gearbox starts to fail at 31st sample.

For the purpose of comparison, five monitoring methods based on KPCA, KONPE, NLOPE, KGLPP [13], and NLKOPE are presented to detect the fault of gearbox respectively. The embedding dimension in each model is set to 3, and the number of nearest neighbors is set to 20 in KONPE, NLOPE, KGLPP, and NLKOPE models. The 99% confidence limit is used for  $T^2$ ,  $SPE$ , and EWMA statistics. In order to compare the results more clearly, the indicator of fault detection rate (FDR) is applied in this case.

Monitoring charts of five methods are shown in Figure 3, and the detailed fault detection results of these monitoring methods are listed in Table 2. Obviously, gearbox starts to fail at 33rd sample that detected by KPCA monitoring model as shown in Figure 3(a), and 35th, 36th, 37th, 38th samples are all under the control limit, that means the failure of detection at these samples. Figure 3(b) illustrates that gearbox starts to fail at 33rd sample that detected by KONPE, that means the failure of detection at 31st and 32nd samples. As shown in Figures 3(c)-3(d), gearbox starts to fail at 32nd sample that detected by NLOPE and KGLPP, but in fact, gearbox starts to fail at 31st sample. The detection result of NLKOPE monitoring model is shown in Figure 3(e); the EWMA statistic can work well and detect the fault of gearbox accurately. Besides, as shown in Table 2, the fault detection rates of NLOPE, KGLPP, and NLKOPE are higher than KPCA and KONPE which only consider the global or local data structure, although NLOPE has considered the global-local data information; the ability to process nonlinear data is not prominent when compared to NLKOPE. The results indicate that the NLKOPE-based monitoring method outperforms KPCA, KONPE, NLOPE, and KGLPP-based monitoring method.

**5.2. Dimension Reduction Performance Assessment.** In this case, the experimental data from Case Western Reserve University [23] are used to evaluate the dimension reduction performance of the proposed methods. The bearings used at the drive end are the deep groove ball bearing 6205-2RS JEM SKF. Data was collected with 12kHz sampling frequency at the rotating speed of 1797 rpm and 0HP load. The sample sets include 7 different severity conditions, i.e., health, inner race faults with faulty sizes 0.007, 0.014, 0.021, and 0.028, respectively, outer race, and ball fault with faulty size 0.014, respectively. We select 1024 sampling points as a sample, and extract 70 samples for each severity condition. Furthermore, the first 35 samples of each severity condition are collected as training samples, then the remaining 35 samples are used as testing samples.

The purpose of dimension reduction is to make the intraclass low dimensional samples clustering and interclass

TABLE 3: The clustering degree of different reduction algorithms.

	KPCA	KONPE	NLOPE	KGLPP [13]	NLKOPE
J	0.0198	0.0182	0.0185	0.0102	0.0088

separation, which will be helpful to improve the performance of fault classification. Thus, the clustering degree is used as a quantification index to evaluate the dimension reduction performance; it defines as follows:

$$J = \frac{S_w}{S_b} \quad (42)$$

$$S_w = \frac{1}{n_i} \sum_{i=1}^C \sum_{y \in A_i} (y - m_i)(y - m_i)^T \quad (43)$$

$$S_b = \sum_{i=1}^C (m_i - m)(m_i - m)^T$$

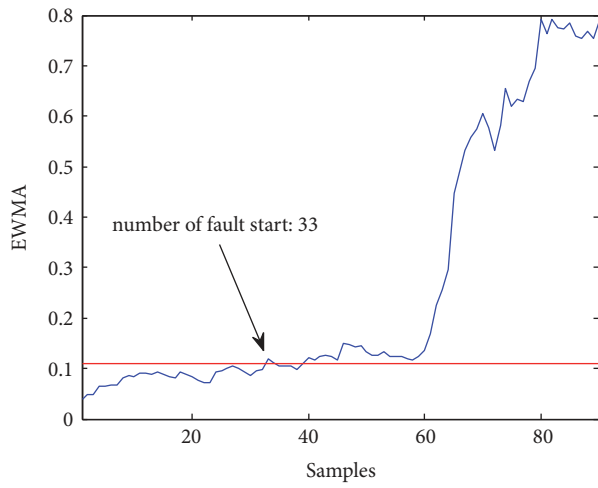
where  $C$  is the number of fault types,  $n_i$  is the sample size of the  $i$ th fault type,  $y$  is the low dimension embedded coordinate,  $A_i = (y_i^1, \dots, y_i^{n_i})$ ,  $m_i$  is the mean value of embedded coordinates of the  $i$ th fault type, and  $m$  is the mean value of all low dimension embedded coordinates.

11 time-domain features and 13 frequency-domain features [21] are extracted from each sample to be the variations and make up the high dimensional sample, and in order to visualize clearly, the embedding dimension is set to 3. For the purpose of comparison, five methods including KPCA, KONPE, NLOPE, KGLPP [13], and NLKOPE are presented to obtain the dimension reduction results on the training and testing samples, respectively; scatter plots of three features are as shown in Figure 4. Furthermore, Health, Fault1, Fault2, Fault3, Fault4, Fault5, and Fault6 in Figure 4 represent 7 different fault types which contain health, four inner race faults with faulty sizes 0.007, 0.014, 0.021, and 0.028 in, outer race, and ball fault with faulty size 0.014 in, respectively, and  $x$ ,  $y$ ,  $z$  indicate three-dimensional representation based on the three features extracted from the training and testing samples by the proposed methods. Figure 4 illustrates the classification abilities of five methods for the 3D-clusters samples, where samples in the same fault types are marked in the same color.

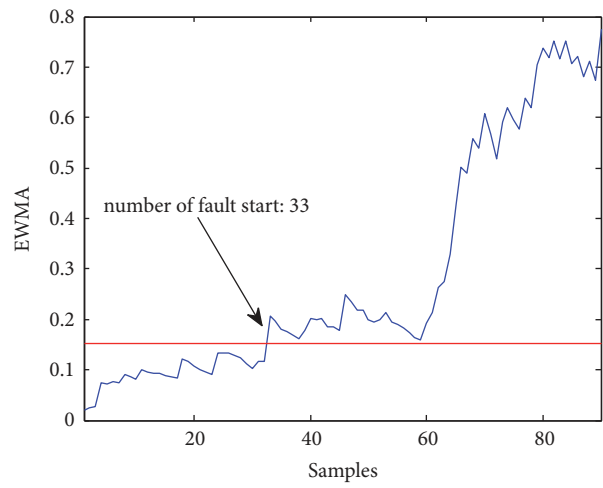
The distribution of the same fault type of samples is dispersed in Figure 4(a), and the different fault types of samples gather together as shown in Figure 4(b), both of these two situations may increase the probability of misclassification. The clustering degree is calculated as shown in Table 3, the clustering degree value of KGLPP is close to NLKOPE, and the result of dimension reduction based on NLKOPE has the minimum clustering degree, which is beneficial to improve the accuracy of fault classification.

**5.3. Condition Monitoring and Performance Degradation Assessment of Bearing.** In this case, the aim is to implement condition monitoring and evaluate the performance degradation of bearing, and the degradation index is important to assess the state of bearing. Thus, we hope to identify the

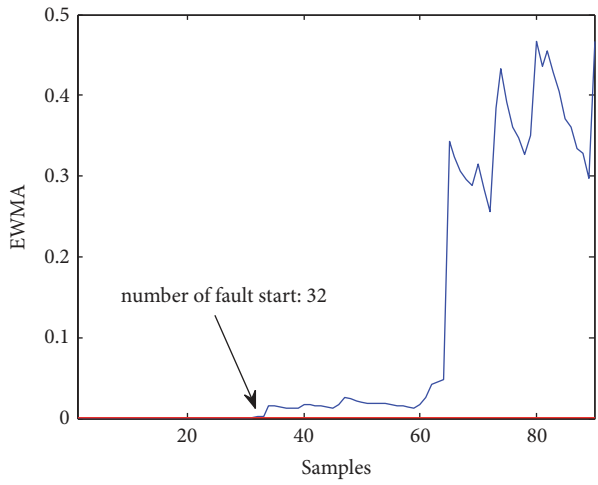




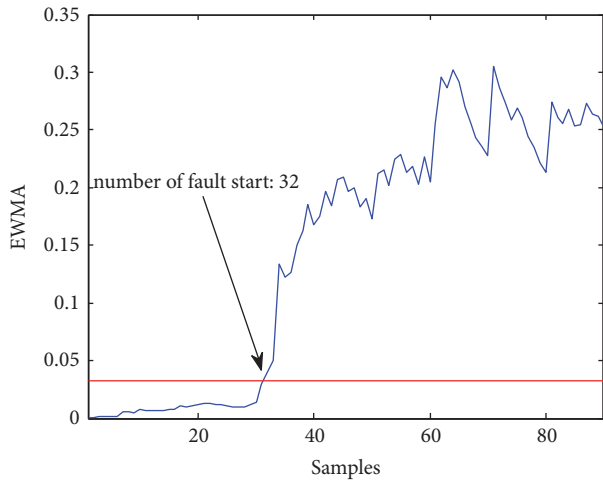
(a) Fault detection result using KPCA



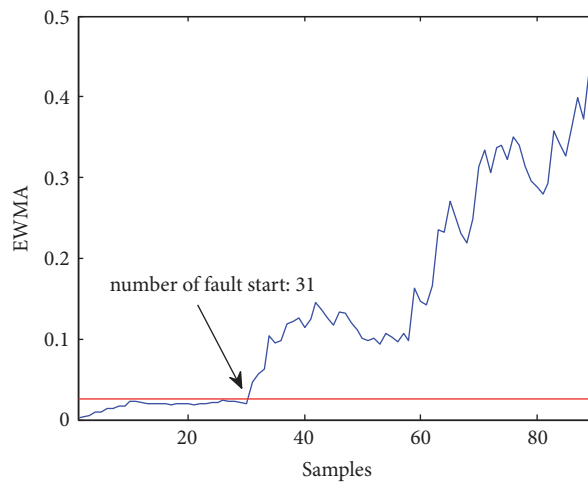
(b) Fault detection result using KONPE



(c) Fault detection result using NLOPE



(d) Fault detection result using KGLPP



(e) Fault detection result using NLKOPE

FIGURE 3: Monitoring charts: (a) KPCA, (b) KONPE, (c) NLOPE, (d) KGLPP, and (e) NLKOPE.

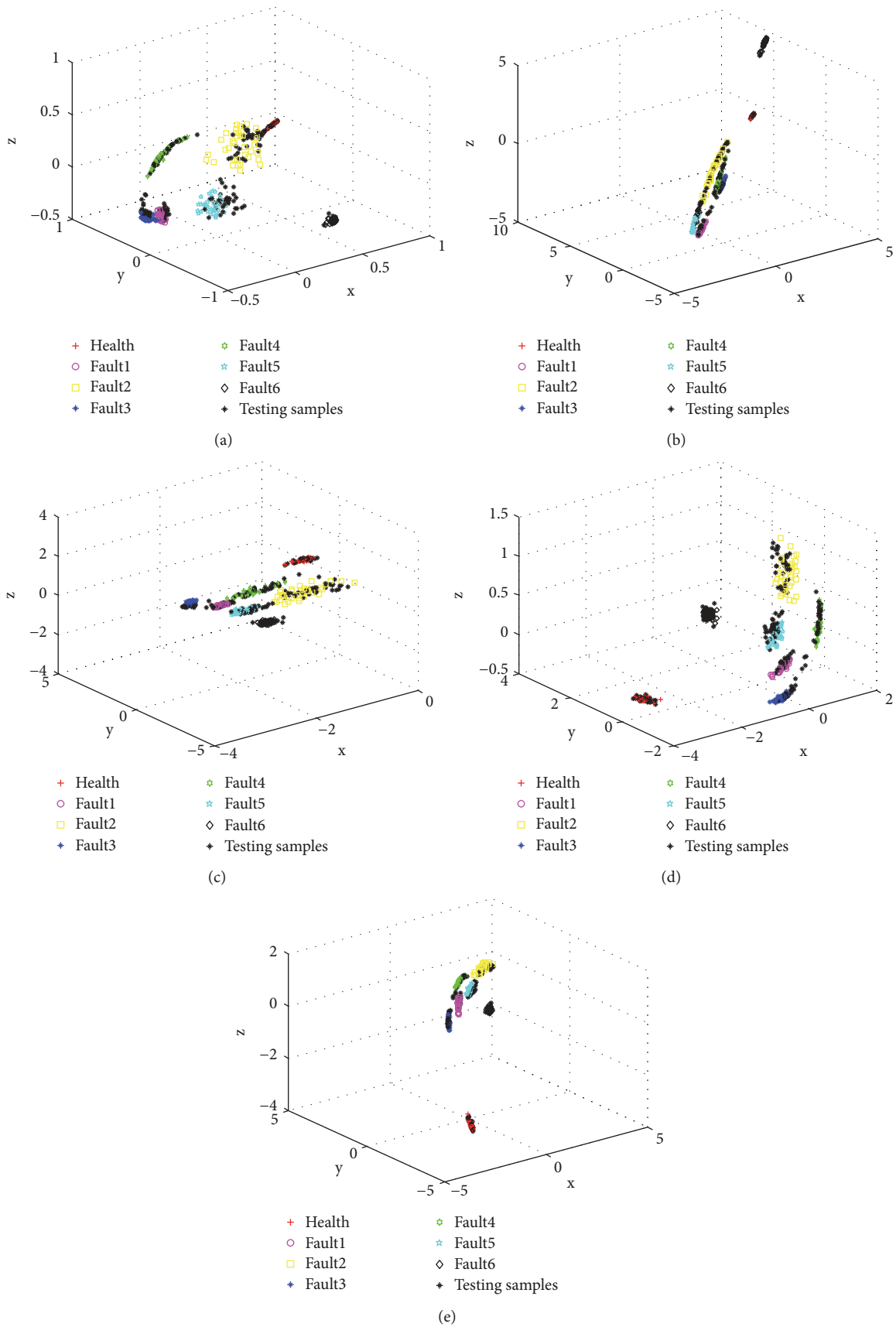


FIGURE 4: Three-dimensional representation based on the three features extracted from the training and testing samples by (a) KPCA, (b) KONPE, (c) NLOPE, (d) KGLPP, and (e) NLKOPE.

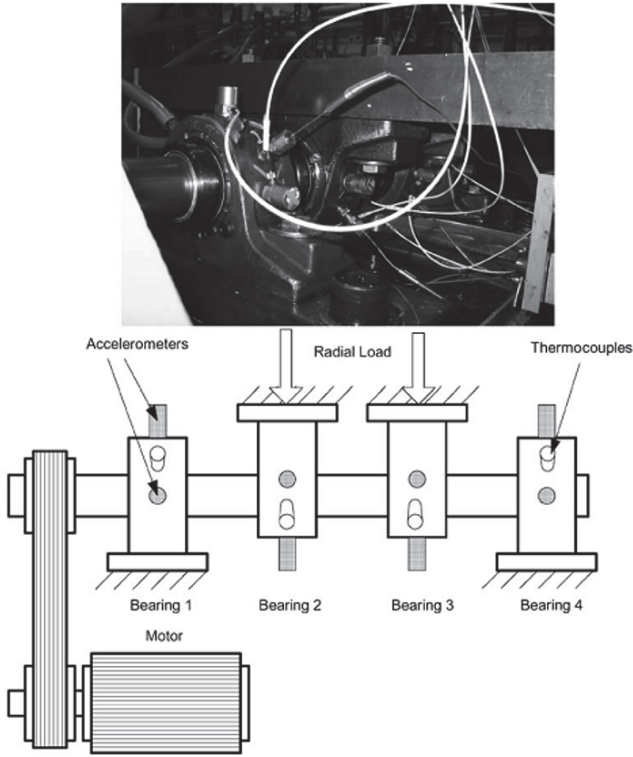


FIGURE 5: Bearing test rig [19]. Figure 5 is reproduced from H. Qiu et al. (2006).

degradation at an early stage to avoid continuous deterioration of the state and minimize machine downtime. The bearing experimental data were generated from the run-to-failure test [19, 24]. Figure 5 illustrates the bearing test rig. The rotation speed was kept constant at 2000 rpm, and each sample consists of 20480 points with the sampling rate set at 20kHz. The structural parameters and kinematical parameters (shaft frequency  $f_r$ , inner-race fault frequency  $f_i$ , rolling element fault frequency  $f_b$ , and outer-race fault frequency  $f_o$ ) of the experiment bearing are listed in Table 4, and the detailed information about the experiments have been introduced in the literature [19]. One bearing (i.e., the bearing 3 of testing 1) with inner race defect is used to verify the performance of the proposed algorithm. We extract 2100 sets of test-to-fail samples recorded for the bearing 3, the first 500 samples are used as the training samples, and the rest are generated as the testing samples.

For the purpose of comparison, five monitoring methods based on KPCA, KONPE, NLOPE, KGLPP [13], and NLKOPE are presented to explain the bearing performance state, respectively. The 99% confidence limit is used for  $T^2$ ,  $SPE$ , and EWMA statistics. In this case, we extracted 2100 test-to-fail samples, and the 1790th sample was regarded as the initial weak degradation point based on the research in literature [25].

As shown in Figures 6 and 7, the EWMA statistic has presented the state of the bearing, and the 1797th sample is considered to be the initial weak degradation point where the performance of the bearing begins to degrade. As the

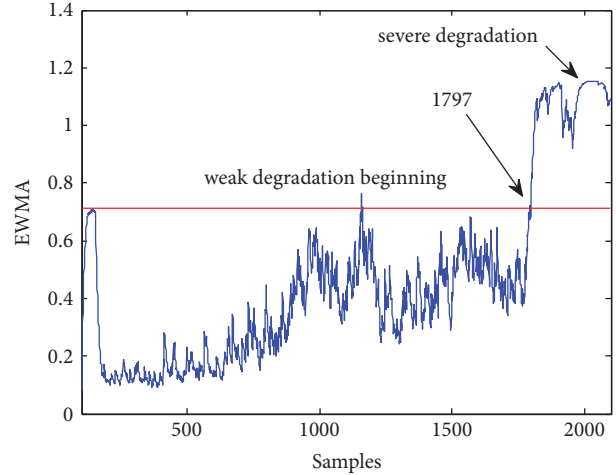


FIGURE 6: Monitoring chart by KPCA.

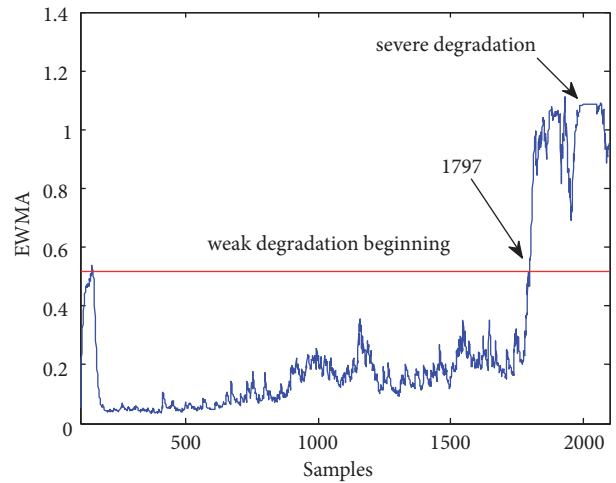


FIGURE 7: Monitoring chart by KONPE.

samples were recorded every ten minutes, it is 70 minutes late to detect the failure by KPCA or KONPE-based monitoring method when compared with the result in literature [25], and the KPCA-EWMA statistic with large fluctuations is not suitable for condition monitoring. Figures 8–10 illustrate the detection results of NLOPE, KGLPP, and NLKOPE-based monitoring methods, and they all obtain the initial weak degradation point of the bearing at the 1789th sample, which is 10 minutes earlier than the result in literature [25], and the statistics after 1789th sample all exceed the control limits, but LPP-EWMA statistics [25] between 1950th sample and 2150th sample are below the control limit, that means the failure of detection in these interval. Though the fault detection accuracy of NLOPE-based monitoring method outperforms KPCA and KONPE, this advantage of NLOPE is not prominent, since the EWMA statistic after the initial weak degradation point has a relative big fluctuation as shown in Figure 8, which is not conducive to evaluating the bearing performance state. The performance degradation assessment of KGLPP-based monitoring method is slightly

TABLE 4: Structural parameters and kinematical parameters of the experiment bearing.

Bearing designation	Number of rolling elements	Diameter of the pitch (in.)	Contact angle	Diameter of the rolling element (in.)
	16	2.815	15.17°	0.331
Characteristic frequency	$f_r$ (Hz)	$f_i$ (Hz)	$f_b$ (Hz)	$f_o$ (Hz)
	33	296.9	279.8	236.4

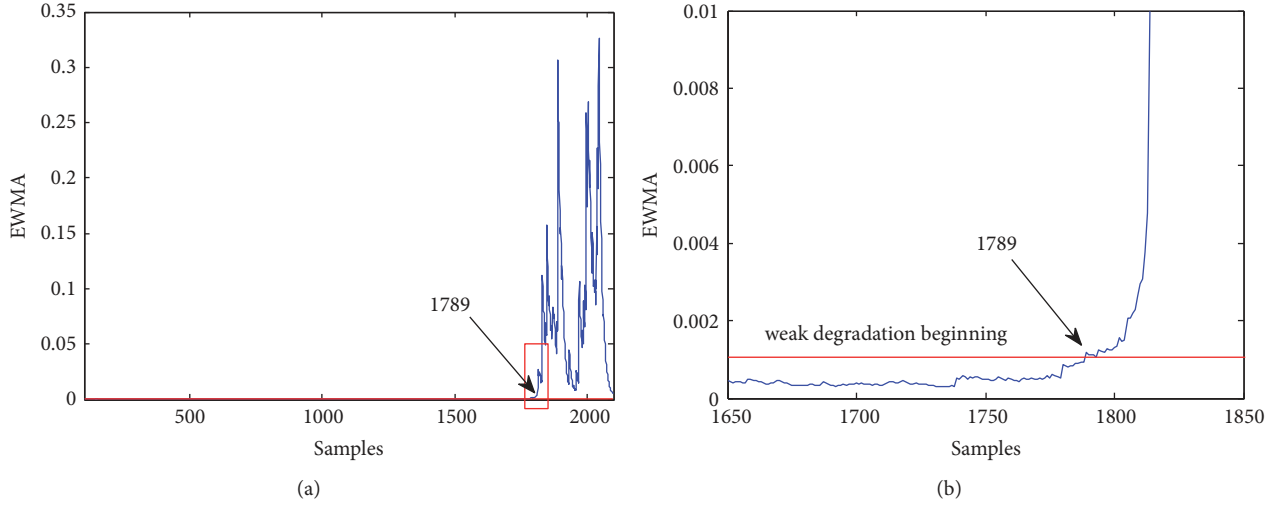


FIGURE 8: Monitoring charts by NLOPE: (a) EWMA statistic; (b) local enlargement of (a).

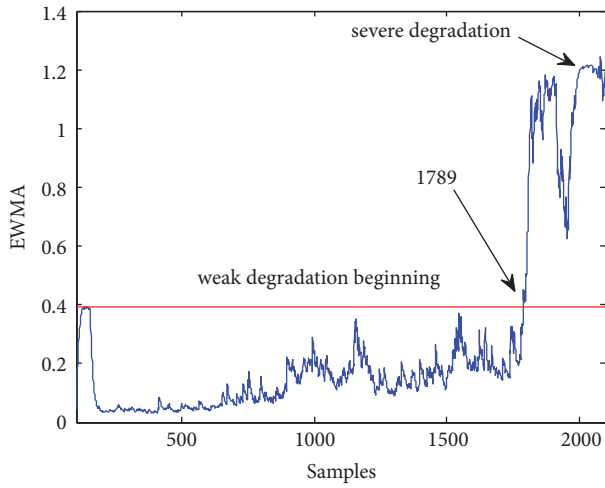


FIGURE 9: Monitoring chart by KGLPP.

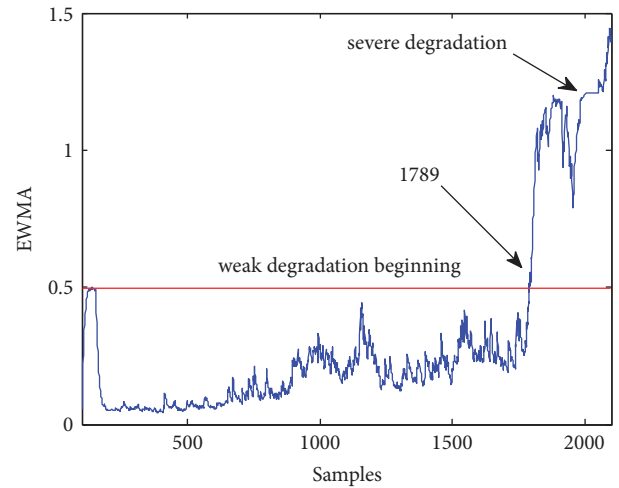


FIGURE 10: Monitoring chart by NLKOPE.

inferior to the performance of NLKOPE-based monitoring method, because NLKOPE-EWMA statistic can reflect the damage degree of the bearing from the severe degradation occurrence of incipient defect to final failure, as shown in Figure 10, the EWMA statistic continues to grow from severe degradation to final failure stage, which is consistent with actual bearing degradation. Thus we can draw the conclusion that the NLKOPE-based monitoring model which considers the global and local data structure together will obtain better monitoring performance than the model considering only the global structure or local structure of data.

The above results have shown that the proposed method can be effectively used for the task of bearing fault detection. The next step is to diagnose the fault type of the bearing. We extract the 1789th sample to analysis, the signal is complex and messy that contains lots of noise as shown in Figure 11, and thus it is hard to diagnose whether the bearing is faulty only by the time waveform of the vibration signal, as the features have been submerged by the strong noise. In order to extract useful features for diagnosis, it is necessary to eliminate the noise in the original vibration signal. Dual-tree complex wavelet packet transform (DTCWPT) is a

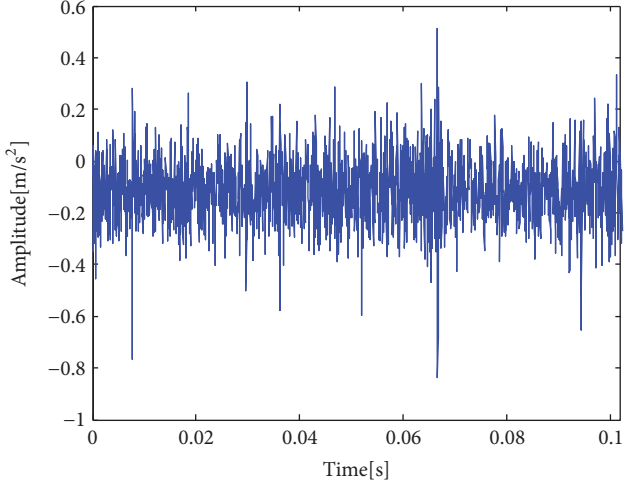


FIGURE 11: Time waveform of vibration signal.

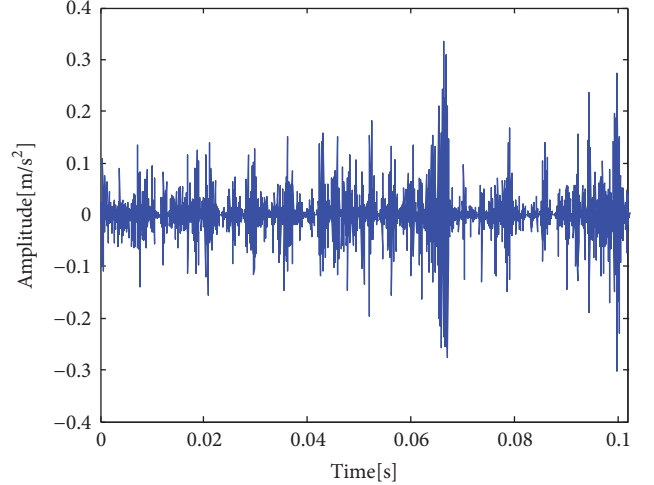


FIGURE 12: Time waveform of denoised vibration signal.

multiscale method with such attractive properties as nearly shift-invariance and reduced aliasing, which has been widely used in signal processing [26]. In this study, DTCWPT is employed to denoise the original vibration signal combined with the threshold method, and Hilbert transform envelope algorithm is applied to extract the fault characteristic frequency. As shown in Figure 12, the noise in vibration signal has been greatly reduced, and transient periodicity can be found because of the impacts produced by the bearing defect. The envelop spectrum of denoised vibration signal is presented in Figure 13, we can find the shaft frequency  $f_r$  and its harmonics, the fault characteristic frequency  $f_i$  and its harmonics are all quite effectively extracted, and there are also side bands  $f_1$  and  $f_2$  on both sides of the fault characteristic frequency  $f_i$ . Therefore, the bearing inner race can be judged to be faulty, which is also in line with the actual condition of the bearing.

## 6. Conclusions

In this paper, a linear dimension reduction method called nonlocal orthogonal preserving embedding is proposed, and the nonlinear form of NLOPE named nonlocal kernel orthogonal preserving embedding is also presented. In order to retain the geometric of the latent manifold, NLOPE and NLKOPE both take global and local data structures into account, and a tradeoff parameter is introduced to balance the global preserving and local preserving. Hence, compared to KPCA and KONPE, NLKOPE is more general and flexible, and it is also more powerful to extract latent information from nonlinear data than NLOPE. Based on the results of three cases, the dimension reduction performance of NLKOPE is the best, which is beneficial to improve the accuracy of fault classification, and NLKOPE-based monitoring method has higher fault detection rate, it is also more sensitive and effective to evaluate the performance degradation of bearing in comparison with KPCA, KONPE, and NLOPE-based monitoring method.

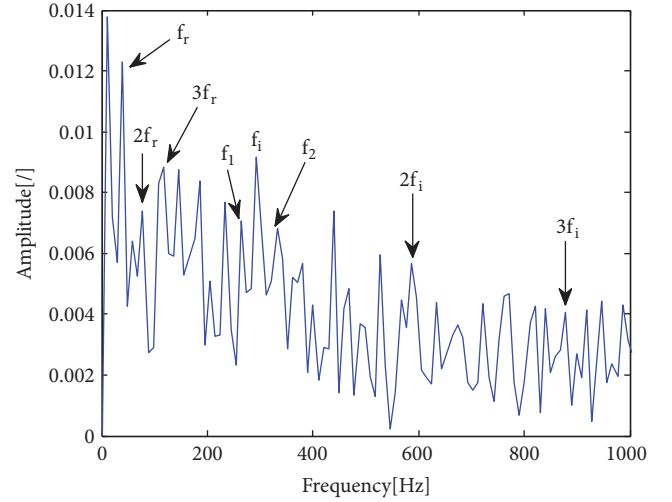


FIGURE 13: Envelop spectrum of denoised vibration signal.

## Appendix

### A.

To obtain the result of  $a_1$ , we construct the Lagrange function of  $L(a_1)$  based on (24)

$$L(a_1) = a_1^T [\eta x M x^T - (1 - \eta) C] a_1 - \lambda \{ a_1^T [\eta x x^T + (1 - \eta) I] a_1 - 1 \} \quad (\text{A.1})$$

Let  $S = \eta x x^T + (1 - \eta) I$ ,  $D = \eta x M x^T - (1 - \eta) C$ , set the partial derivative of  $L(a_1)$  with respect to  $a_1$  to be zero, and we get

$$\frac{\partial L(a_1)}{\partial a_1} = 2D a_1 - 2\lambda S a_1 = 0 \quad (\text{A.2})$$

Thus,  $a_1$  is the eigenvector corresponding to the smallest eigenvalue of matrix  $S^{-1}D$ .

To obtain the result of  $a_k$ , we construct the Lagrange function of  $L(a_k)$  based on (24)

$$L(a_k) = a_k^T [\eta x M x^T - (1 - \eta) C] a_k - \lambda \{ a_k^T [\eta x x^T + (1 - \eta) I] a_k - 1 \} - \sum_{i=1}^{k-1} u_i a_k^T a_i \quad (\text{A.3})$$

Substituting  $S$  and  $D$  into (A.3), we get

$$L(a_k) = a_k^T D a_k - \lambda (a_k^T S a_k - 1) - \sum_{i=1}^{k-1} u_i a_k^T a_i \quad (\text{A.4})$$

Set the partial derivative of  $L(a_k)$  with respect to  $a_k$  to be zero

$$\frac{\partial L(a_k)}{\partial a_k} = 2D a_k - 2\lambda S a_k - \sum_{i=1}^{k-1} u_i a_i = 0 \quad (\text{A.5})$$

Multiplying the left side of (A.4) by  $a_i^T S^{-1}$ , we have

$$a_i^T S^{-1} \sum_{i=1}^{k-1} u_i a_i = 2a_i^T S^{-1} D a_k \quad (\text{A.6})$$

$i = 1, 2, \dots, k-1$ , (A.6) can be represented as

$$\begin{bmatrix} a_1^T \\ \vdots \\ a_{k-1}^T \end{bmatrix} S^{-1} [a_1, \dots, a_{k-1}] \begin{bmatrix} u_1 \\ \vdots \\ u_{k-1} \end{bmatrix} = 2 \begin{bmatrix} a_1^T \\ \vdots \\ a_{k-1}^T \end{bmatrix} S^{-1} D a_k \quad (\text{A.7})$$

Let  $u^{(k-1)} = [u_1, u_2, \dots, u_{k-1}]^T$ ,  $a^{(k-1)} = [a_1, a_2, \dots, a_{k-1}]$ , we obtain

$$u^{(k-1)} = 2 \left[ (a^{(k-1)})^T S^{-1} a^{(k-1)} \right]^{-1} (a^{(k-1)})^T S^{-1} D a_k \quad (\text{A.8})$$

Multiplying (A.4) by  $S^{-1}$  and substituting (A.8), we obtain

$$\left\{ I - S^{-1} a^{(k-1)} \left[ (a^{(k-1)})^T S^{-1} a^{(k-1)} \right]^{-1} (a^{(k-1)})^T \right\} \cdot S^{-1} D a_k = \lambda a_k \quad (\text{A.9})$$

Thus,  $a_k$  is the eigenvector corresponding to the smallest eigenvalue of matrix  $Q^{(k)}$ :

$$Q^{(k)} = \left\{ I - (S)^{-1} a^{(k-1)} \left[ (a^{(k-1)})^T (S)^{-1} a^{(k-1)} \right]^{-1} \cdot (a^{(k-1)})^T \right\} S^{-1} D \quad (\text{A.10})$$

## B.

To obtain the result of  $a_1$ , we construct the Lagrange function of  $L(a_1)$  based on (31)

$$L(a_1) = a_1^T \left[ \eta \bar{K}^T M \bar{K} - (1 - \eta) \frac{\bar{K}^T \bar{K}}{N} \right] a_1 - \lambda \left\{ a_1^T \left[ \eta (\bar{K}^T \bar{K} + \bar{K}) + (1 - \eta) I \right] a_1 - 1 \right\} \quad (\text{B.1})$$

Let  $S = \eta (\bar{K}^T \bar{K} + \bar{K}) + (1 - \eta) I$ ,  $D = \eta \bar{K}^T M \bar{K} - (1 - \eta) \bar{K}^T \bar{K} / N$ , set the partial derivative of  $L(a_1)$  with respect to  $a_1$  to be zero, we get

$$\frac{\partial L(a_1)}{\partial a_1} = 2D a_1 - 2\lambda S a_1 = 0 \quad (\text{B.2})$$

Thus,  $a_1$  is the eigenvector corresponding to the smallest eigenvalue of matrix  $S^{-1} D$ .

To obtain the result of  $a_k$ , we construct the Lagrange function of  $L(a_k)$  based on (31)

$$L(a_k) = a_k^T \left[ \eta \bar{K}^T M \bar{K} - (1 - \eta) \frac{\bar{K}^T \bar{K}}{N} \right] a_k - \lambda \left\{ a_k^T \left[ \eta (\bar{K}^T \bar{K} + \bar{K}) + (1 - \eta) I \right] a_k - 1 \right\} - \sum_{i=1}^{k-1} u_i a_k^T \bar{K}^T a_i \quad (\text{B.3})$$

Substituting  $S$  and  $D$  into (B.3) and setting the partial derivative of  $L(a_k)$  with respect to  $a_k$  to be zero, we obtain

$$L(a_k) = a_k^T D a_k - \lambda (a_k^T S a_k - 1) - \sum_{i=1}^{k-1} u_i a_k^T \bar{K}^T a_i \quad (\text{B.4})$$

$$\frac{\partial L(a_k)}{\partial a_k} = 2D a_k - 2\lambda S a_k - \sum_{i=1}^{k-1} u_i \bar{K}^T a_i = 0 \quad (\text{B.5})$$

Multiplying the left side of (B.4) by  $a_i^T S^{-1}$ , we have

$$a_i^T S^{-1} \sum_{i=1}^{k-1} u_i \bar{K}^T a_i = 2a_i^T S^{-1} D a_k \quad (\text{B.6})$$

$i = 1, 2, \dots, k-1$ , (B.6) can be represented as

$$\begin{bmatrix} a_1^T \\ \vdots \\ a_{k-1}^T \end{bmatrix} S^{-1} \bar{K}^T [a_1, \dots, a_{k-1}] \begin{bmatrix} u_1 \\ \vdots \\ u_{k-1} \end{bmatrix} = 2 \begin{bmatrix} a_1^T \\ \vdots \\ a_{k-1}^T \end{bmatrix} S^{-1} D a_k \quad (\text{B.7})$$

Let  $u^{(k-1)} = [u_1, u_2, \dots, u_{k-1}]^T$ ,  $a^{(k-1)} = [a_1, a_2, \dots, a_{k-1}]$ , and we obtain

$$u^{(k-1)} = 2 \left[ \left( a^{(k-1)} \right)^T S^{-1} \bar{K}^T a^{(k-1)} \right]^{-1} \left( a^{(k-1)} \right)^T S^{-1} D a_k \quad (\text{B.8})$$

Multiplying (B.4) by  $S^{-1}$  and substituting (B.8), we obtain

$$\left\{ I - S^{-1} \bar{K}^T a^{(k-1)} \left[ \left( a^{(k-1)} \right)^T S^{-1} \bar{K}^T a^{(k-1)} \right]^{-1} \cdot \left( a^{(k-1)} \right)^T \right\} S^{-1} D a_k = \lambda a_k \quad (\text{B.9})$$

Thus,  $a_k$  is the eigenvector corresponding to the smallest eigenvalue of matrix  $Q^{(k)}$ :

$$Q^{(k)} = \left\{ I - S^{-1} \bar{K}^T a^{(k-1)} \left[ \left( a^{(k-1)} \right)^T S^{-1} \bar{K}^T a^{(k-1)} \right]^{-1} \cdot \left( a^{(k-1)} \right)^T \right\} S^{-1} D \quad (\text{B.10})$$

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research is supported by National Nature Science Foundation of China (nos. 61640308 and 61573364) and Nature Science Foundation of Naval University of Engineering (no. 20161579). The author is also grateful to the 2009 PHM Challenge Competition, Case Western Reserve University, and NSF I/UCR Center for Intelligent Maintenance System, University of Cincinnati, USA, for providing the experimental data.

## References

- [1] Q. Jiang, X. Yan, and B. Huang, "Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and bayesian inference," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 1, pp. 377–386, 2016.
- [2] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annual Reviews in Control*, vol. 36, no. 2, pp. 220–234, 2012.
- [3] Z. Ge, "Review on data-driven modeling and monitoring for plant-wide industrial processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 171, pp. 16–25, 2017.
- [4] Z. Ge, C. Yang, and Z. Song, "Improved kernel PCA-based monitoring approach for nonlinear processes," *Chemical Engineering Science*, vol. 64, no. 9, pp. 2245–2255, 2009.
- [5] M. Yao and H. Wang, "On-line monitoring of batch processes using generalized additive kernel principal component analysis," *Journal of Process Control*, vol. 28, pp. 56–72, 2015.
- [6] M. Zhang, Z. Ge, Z. Song, and R. Fu, "Global-Local Structure Analysis Model and Its Application for Fault Detection and Identification," *Industrial & Engineering Chemistry Research*, vol. 50, no. 11, pp. 6837–6848, 2011.
- [7] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [8] X. He and P. Niyogi, "Locality preserving projections," *Advances in Neural Information Processing Systems*, vol. 16, pp. 153–160, 2004.
- [9] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [10] X. He, D. Cai, S. Yan, and H. Zhang, "Neighborhood preserving embedding," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, pp. 1208–1213, Beijing, China, October 2005.
- [11] J. Yu, "Local and global principal component analysis for process monitoring," *Journal of Process Control*, vol. 22, no. 7, pp. 1358–1373, 2012.
- [12] J. Wang, J. Feng, and Z. Y. Han, "Locally preserving PCA method based on manifold learning and its application in fault detection," *Control and Decision*, vol. 22, no. 5, pp. 683–687, 2013.
- [13] L. Luo, S. Bao, J. Mao, and D. Tang, "Nonlinear process monitoring based on kernel global-local preserving projections," *Journal of Process Control*, vol. 38, pp. 11–21, 2016.
- [14] L. J. Luo, "Process Monitoring with Global-Local Preserving Projections," *American Chemical Society*, vol. 53, no. 18, pp. 7696–7705, 2014.
- [15] X. Liu, J. Yin, Z. Feng, J. Dong, and L. Wang, "Orthogonal neighborhood preserving embedding for face recognition," in *Proceedings of the 14th IEEE International Conference on Image Processing, ICIP 2007*, pp. 1133–1136, USA, September 2007.
- [16] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal laplacianfaces for face recognition," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3608–3614, 2006.
- [17] J.-D. Shao, G. Rong, and J. M. Lee, "Generalized orthogonal locality preserving projections for nonlinear fault detection and diagnosis," *Chemometrics and Intelligent Laboratory Systems*, vol. 96, no. 1, pp. 75–83, 2009.
- [18] Phm data challenge 2009, <http://www.phmsociety.org/competition/PHM/09>.
- [19] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *Journal of Sound and Vibration*, vol. 289, no. 4-5, pp. 1066–1090, 2006.
- [20] F. He and J. Xu, "A novel process monitoring and fault detection approach based on statistics locality preserving projections," *Journal of Process Control*, vol. 37, pp. 46–57, 2016.
- [21] Y. Lei, Z. He, and Y. Zi, "A new approach to intelligent fault diagnosis of rotating machinery," *Expert Systems with Applications*, vol. 35, no. 4, pp. 1593–1600, 2008.
- [22] J.-H. Cho, J.-M. Lee, S. W. Choi, D. Lee, and I.-B. Lee, "Fault identification for process monitoring using kernel principal component analysis," *Chemical Engineering Science*, vol. 60, no. 1, pp. 279–288, 2005.
- [23] K. A. Loparo, *Bearings Vibration Data Set*, Case Western Reserve University, 2003, <http://csegroups.case.edu/bearingdatacenter/pages/apparatus-procedures>.
- [24] J. Lee, H. Qiu, G. Yu, and J. Lin, *Bearing Data Set. IMS*, University of Cincinnati. NASA Ames Prognostics Data Repository, NASA Ames, Moffett Field, CA, USA, <http://data-acoustics.com/measurements/bearing-faults/bearing-4>.

- [25] J.-B. Yu, "Bearing performance degradation assessment using locality preserving projections," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7440–7450, 2011.
- [26] J. Qu, Z. Zhang, and T. Gong, "A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion," *Neurocomputing*, vol. 171, no. 1, pp. 837–853, 2016.





**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

