KSME

Springer

Journal of Mechanical Science and Technology 35 (6) 2021

**Original Article**
DOI 10.1007/s12206-021-0507-2

**Keywords:**
· Kernel principal component analysis (KPCA)
· Fisher criterion (FC)
· Planetary gearbox
· Wear fault
· Fault identification
· Particle swarm optimization (PSO)

**Correspondence to:**
Linzheng Ye
lz09020141@163.com

**Citation:**
He, Y., Ye, L., Zhu, X., Wang, Z. (2021). Feature extraction based on PSO-FC optimizing KPCA and wear fault identification of planetary gear. Journal of Mechanical Science and Technology 35 (6) (2021) 2347~2357.
http://doi.org/10.1007/s12206-021-0507-2

Received October 22nd, 2020
Revised December 30th, 2020
Accepted March 9th, 2021

† Recommended by Editor No-cheol Park



# Feature extraction based on PSO-FC optimizing KPCA and wear fault identification of planetary gear

**Yan He[1,2], Linzheng Ye[1,2], Xijing Zhu[1,2] and Zongyan Wang[1]**

[1]School of Mechanical Engineering, North University of China, Taiyuan 030051, China, [2]Shanxi Key Laboratory of Advanced Manufacturing Technology, North University of China, Taiyuan 030051, China

**Abstract** The feature extraction problem of coupled vibration signals with multiple fault modes of planetary gear has not been solved effectively. At present, kernel principal component analysis (KPCA) is usually used for nonlinear feature extraction, but the blind setting of kernel function parameters greatly affects the performance of KPCA algorithm. For the optimization of kernel parameters, it is necessary to study theoretical modeling to improve KPCA performance. In this paper, employing a Fisher criterion (FC) discriminant function in pattern recognition, the optimization mathematical model of the kernel parameter was presented and the improved particle swarm optimization algorithm (PSO) was applied to search for the optimum value, and the performance of the Kernel principal component analysis for nonlinear problems was improved. The optimized KPCA was applied for feature extraction of different wear fault modes of a planetary gear, and the feature dimensions were reduced from 27 to 10. The feature parameters with 92.9 % contribution rates were retained and sample sets were formed to feed the support vector machine (SVM) for final classification and identification. The intelligently optimized KPCA based on the PSO-FC has improved the structural distribution of data in the feature space and showed a good scale clustering effect in planetary gear wear state recognition. The accuracy of the SVM classification was improved by 17.5 %.

## 1. Introduction

A planetary gearbox is a composite gear transmission system. Due to its poor working conditions, the gear teeth often produce fatigue cracks, wear, and pits, and different degrees of failure may occur in multiple gears [1]. At present, some investigations on the fault identification and diagnosis of the planetary gearbox have been done from the aspects of modeling, signal processing, and intelligent diagnosis [2, 3]. Considering the time-varying effect of the vibration transfer path, Lei et al. [4] established a corresponding dynamic model and obtained the dynamic response of a planetary gear train in normal, crack, and spalling conditions, and then analyzed their spectrum characteristics to achieve the goal of health monitoring. Wu et al. [5] proposed local oscillatory characteristic decomposition, by which the fault characteristics of gear cracks were analyzed and the fault characteristics were efficiently and accurately extracted. Liu et al. [6] established a composite vibration signal model to realize the fault diagnosis of an incipient tooth crack on the sun gear in a planetary gear, and finally verified the simulation results through a test rig experiment. Cheng et al. [7] used entropy feature fusion of the ensemble empirical mode decomposition (EMD) for planetary gear fault diagnosis. Chen and Feng [8] presented an iterative generalized time-frequency redistribution method and successfully extracted the fault symptoms of time-varying gears under unsteady operating conditions to acquire good diagnostic results for the planetary gearbox. Lei et al. [9] proposed a method based on the multi-kernel relevance vector machine (MRVM), which was applied to multilevel planetary gearboxes for health status identification and diagnosis. Khazaee et al. [10] combined a fast Fourier transform and a least-square support vector machine for condition monitoring and

classification of a planetary gearbox, which improved diagnostic accuracy. Liu et al. [11] proposed a hybrid intelligent model based on the redundant second-generation wavelet package transform (RSGWPT), the kernel principal component analysis (KPCA), and the twin support vector machine (SVM) to realize multi-fault detection of rotating machinery, and the experiment results demonstrated its effectivity. Although some achievements have been made in the previous research, multiple components such as synchromesh excite similar vibration in different phases during the planetary gearbox's transmission, which are coupled with each other and results in the neutralization or weakening of the vibration signals from the fault components. So, typical nonlinear characteristics are presented. Therefore, it is still an urgent task to extract effective and sensitive features from the coupled vibration signals of the multi-fault mode for fault diagnosis of a planetary gearbox.

In recent years, the kernel learning method has made some progress in the research fields of feature extraction, pattern recognition, data mining, image and signal processing. To a certain extent, it solves the nonlinear problem in the actual system and improves the accuracy of pattern recognition and prediction. KPCA is a method based on "kernel learning", which is widely used for nonlinear feature extraction at present. It uses a kernel function instead of the inner product of nonlinear mapping to solve nonlinear problems. Liu et al. [12] used hybrid kernel feature selection and kernel Fisher discriminant analysis to carry out fault level diagnosis for a planetary gearbox. Vo and Durlofsky [13] adjusted the KPCA to achieve the efficient parameterization of complex geological models. However, the kernel learning method has encountered a bottleneck problem that affects the performance of the algorithm, that is, the selection of kernel function and its parameter. Existing studies showed that the kernel function and its parameters directly affect the data distribution structure in nonlinear feature space. So, the optimization of kernel parameters is the most direct method to improve the performance of the kernel learning algorithm. Currently, it is mainly determined by a large number of experiments or adopted by cross-checking methods, and there are few systematic studies from the perspective of modeling. Therefore, this is a new research hotspot to optimize and improve the performance of KPCA by using new intelligent methods.

The particle swarm optimization algorithm (PSO), a typical swarm intelligence method, has been rapidly developed and applied to solve optimization problems [14, 15] due to inherent parallelism and distributed processing characteristics in recent years. In this paper, the improved PSO was compounded with the Fisher criterion (FC) discrimination method in pattern recognition to solve the optimization of the kernel parameters by modeling optimization. A simulated failure experiment of a planetary gearbox was carried out, and the optimized KPCA was applied to the nonlinear feature extraction. Then, SVM was used for wear fault state recognition of a planetary gear to solve the problem where the fault boundary is fuzzy and identifying and classifying of the fault are difficult.

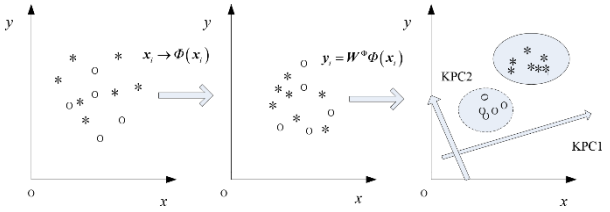



Fig. 1. Geometric sketch of the KPCA.

## 2. Theory

### 2.1 The principle of KPCA

Firstly, nonlinear mapping is used to map the input vectors of the original space to the feature space, and then the corresponding linear transformation is carried out in the feature space. In essence, this method of kernel transformation realizes the nonlinear mapping of the data space, feature space, and category space, and thus greatly enhances the capability for nonlinear data processing [16]. A geometric sketch of the KPCA is shown in Fig. 1.

$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N, \boldsymbol{x}_i \in \boldsymbol{X}$, are set as data sample points in the original space $\boldsymbol{X}$ ($\boldsymbol{X} \subseteq \boldsymbol{R}^n$). When the KPCA is carried out, the input space $\boldsymbol{X}$ is firstly nonlinearly mapped to the feature space $\boldsymbol{F}$, that is, $\Phi$: $\boldsymbol{X} \to \boldsymbol{F}$, $\boldsymbol{x}_i \to \Phi(\boldsymbol{x}_i)$, $i$ = 1, 2, …, $N$, $\boldsymbol{F} = \{\Phi(\boldsymbol{x}) | \boldsymbol{x} \in X\}$, $\boldsymbol{F} \subseteq \boldsymbol{R}^m$.

Then the space transformation is once again carried out on the data $\Phi(\boldsymbol{x}_i)$ in the feature space $\boldsymbol{F}$, so the linear mapping feature after the space transformation matrix $\boldsymbol{W}^{\Phi}$ can be written as:

$$\boldsymbol{y}_i = \boldsymbol{W}^{\Phi}\Phi(\boldsymbol{x}_i) \tag{1}$$

where $\boldsymbol{y}_i (i = 1, 2, \ldots, N)$ is the eigenvector, and the $N$ eigenvectors are standardized first and then linearly combined into several vectors, which must be orthogonal to each other. When the data $\Phi(\boldsymbol{x}_i)$ in $\boldsymbol{F}$ are projected on these vectors, the first vector should fully reflect the maximum difference of the data $\Phi(\boldsymbol{x}_i)$. The degree of difference reflected by other vectors decreases in turn. These vectors are defined as the kernel principal component (KPC). Thus, a vector is calculated to describe the maximum change direction of feature space data, which is regarded as the first KPC (KPC1). Then, the second KPC (KPC2) is calculated and used to describe the maximum change direction of the remaining data.

There are many forms of kernel function. In this paper, the most commonly used radial basis kernel function is applied and analyzed. Its specific form is as follows:

$$k(x,z) = \exp\left(-\frac{\|x-z\|^2}{h}\right) \tag{2}$$

where $\boldsymbol{z}$ is the kernel function center, $h$ is the scale parameter of the kernel function, and its value affects the data distribution scale of KPCA after nonlinear projection.

### 2.2 Establishment of the scale parameter optimization model of the kernel function based on the Fisher criterion

The basic design idea of the Fisher criterion discrimination is to project the N-class m-dimensional data as far as possible in one direction (a straight line), so that the class and other classes can be separated as much as possible.

For a data matrix $\boldsymbol{x}$, the first class $\boldsymbol{w}_1$ has $N_1$ samples, the second class $\boldsymbol{w}_2$ has $N_2$ samples and $N = N_1+N_2$.

Through nonlinear mapping, the vector set $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, …, $\boldsymbol{x}_N$ of input space $\boldsymbol{X}$ is mapped to the vector set $\Phi(\boldsymbol{x}_1)$, $\Phi(\boldsymbol{x}_2)$, …, $\Phi(\boldsymbol{x}_N)$ of feature space $\boldsymbol{F}$. The mean vectors of two classes in the feature space are as follows:

$$\mu_i = \frac{1}{N_i}\sum_{x_p \in w_i} \Phi\left(x_p\right) \quad (i = 1,2) . \tag{3}$$

In the feature space $\boldsymbol{F}$, the within-class discrete matrix $\boldsymbol{S}_i$ in each sample class and the total within-class discrete matrix $\boldsymbol{S}_w$, are:

$$\boldsymbol{S}_1 = \sum_{x_p \in w_i} [\Phi\left(\boldsymbol{x}_p\right) - \boldsymbol{\mu}_i][(\Phi\left(\boldsymbol{x}_p\right) - \boldsymbol{\mu}_i)^T \qquad i = 1,2 \tag{4}$$

$$\boldsymbol{S}_w = \boldsymbol{S}_1 + \boldsymbol{S}_2 = \sum_{j=1}^{2}\sum_{p \in N} [\Phi\left(\boldsymbol{x}_p\right) - \boldsymbol{\mu}_j)][(\Phi\left(\boldsymbol{x}_p\right) - \boldsymbol{\mu}_j]^T . \tag{5}$$

The between-class discrete matrix $\boldsymbol{S}_b$ of the samples is:

$$\boldsymbol{S}_b = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T . \tag{6}$$

A projection direction is set: $\boldsymbol{\eta} \in \boldsymbol{F}$, so the projections of both the mean vectors in this direction are:

$$\overline{\boldsymbol{\mu}}_i = \boldsymbol{\eta}^T \boldsymbol{\mu}_i = \frac{1}{N_i}\sum_{x_p \in w_i} \boldsymbol{\eta}^T \boldsymbol{\Phi}\left(\boldsymbol{x}\right) \quad (i = 1,2) . \tag{7}$$

The mean difference of any projected samples can be obtained by adjusting $\boldsymbol{\eta}$. To separate the two types of samples from the projection area, the larger mean difference and the smaller gap between the samples of the same class are desired. This is the Fisher criterion (FC). These two conditions need to be represented by the between-class discrete matrix $\boldsymbol{S}_b$, and the within-class discrete matrix $\boldsymbol{S}_w$, respectively.

In the feature space after projection, the sum of squares of the total within-class divergence errors of the two classes is:

$$\begin{aligned} \overline{\boldsymbol{S}}_w &= \sum_{j=1}^{2}\sum_{p \in N} [\boldsymbol{\eta}^T \Phi\left(\boldsymbol{x}_p\right) - \boldsymbol{\eta}^T \overline{\boldsymbol{\mu}}_j]^2 \\ &= \boldsymbol{\eta}^T \left[\sum_{j=1}^{2}\sum_{p \in N} [\Phi\left(\boldsymbol{x}_p\right) - \boldsymbol{\mu}_j][\Phi\left(\boldsymbol{x}_p\right) - \boldsymbol{\mu}_j)^T\right] \boldsymbol{\eta} . \\ &= \boldsymbol{\eta}^T \boldsymbol{S}_w \boldsymbol{\eta} \end{aligned} \tag{8}$$

The sum of the squared errors of the dispersion between the sample classes is:

$$\overline{\boldsymbol{S}}_b = (\overline{\boldsymbol{\mu}}_1 - \overline{\boldsymbol{\mu}}_2)^2 = (\boldsymbol{\eta}^T \boldsymbol{\mu}_1 - \boldsymbol{\eta}^T \boldsymbol{\mu}_2)^2 = \boldsymbol{\eta}^T \boldsymbol{S}_b \boldsymbol{\eta} . \tag{9}$$

According to the above optimal criterion, the FC discriminant function is:

$$J(\boldsymbol{\eta}) = \frac{\overline{\boldsymbol{S}}_w}{\overline{\boldsymbol{S}}_b} = \frac{\boldsymbol{\eta}^T \boldsymbol{S}_w \boldsymbol{\eta}}{\boldsymbol{\eta}^T \boldsymbol{S}_b \boldsymbol{\eta}} . \tag{10}$$

The optimal projection direction is obtained from the above formula

$$\boldsymbol{\eta}^* = (\boldsymbol{S}_b)^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) . \tag{11}$$

The projection of $\boldsymbol{\Phi}(\boldsymbol{x})$ on $\boldsymbol{\eta}^*$ is

$$\boldsymbol{y} = \boldsymbol{\eta}^{*T} \boldsymbol{\Phi}\left(\boldsymbol{x}\right) . \tag{12}$$

All the above calculations are carried out in the feature space. Because of the high dimension of the feature space, it is impossible to operate directly. Considering that it can be expressed linearly by $\Phi(\boldsymbol{x}_1)$, $\Phi(\boldsymbol{x}_2)$, …, $\Phi(\boldsymbol{x}_N)$, i.e.,

$$\boldsymbol{\eta} = \sum_{i=1}^{N} \boldsymbol{\alpha}_i \Phi\left(\boldsymbol{x}_i\right) . \tag{13}$$

Combined with Eqs. (3), (7) and (13):

$$\begin{aligned} &\boldsymbol{\eta}^T \boldsymbol{\mu}_i = \frac{1}{N_i}\sum_{j=1}^{N}\sum_{p=1}^{N_i} \boldsymbol{\alpha}_j k(\boldsymbol{x}_j, \boldsymbol{x}_p) = \boldsymbol{\alpha}^T \boldsymbol{M}_i \\ &(i = 1,2; j = 1,2,\cdots,N; \boldsymbol{x}_p \in \boldsymbol{w}_i) \end{aligned} \tag{14}$$

where $\boldsymbol{M}_i$ is defined as an $N$×1 matrix,

$$(\boldsymbol{M}_i)_j = \frac{1}{N_i}\sum_{p=1}^{N_i} k(\boldsymbol{x}_j, \boldsymbol{x}_p)(i = 1,2; j = 1,2,\cdots,N; \boldsymbol{x}_p \in \boldsymbol{w}_i)$$

Gaussian radial basis function is taken as kernel function, that is,

$$k(\boldsymbol{x}_j, \boldsymbol{x}_p) = \exp(-\frac{\left\|\boldsymbol{x}_j - \boldsymbol{x}_p\right\|^2}{h}) . \tag{15}$$

Then,

$$\begin{aligned} &\boldsymbol{\eta}^T \boldsymbol{S}_b \boldsymbol{\eta} = \boldsymbol{\eta}^T (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\eta} \\ &= \boldsymbol{\alpha}^T (\boldsymbol{M}_1 - \boldsymbol{M}_2)(\boldsymbol{M}_1 - \boldsymbol{M}_2)^T \boldsymbol{\alpha} = \boldsymbol{\alpha}^T \boldsymbol{M} \boldsymbol{\alpha} \end{aligned} \tag{16}$$

where $\boldsymbol{M} = (\boldsymbol{M}_1 - \boldsymbol{M}_2)(\boldsymbol{M}_1 - \boldsymbol{M}_2)^T$

$$\begin{aligned}\boldsymbol{\eta}^T \boldsymbol{S}_w \boldsymbol{\eta} &= \boldsymbol{\eta}^T (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\eta} \\ &= \boldsymbol{\eta}^T \sum_{j=1}^{2} \sum_{p \in N} (\boldsymbol{x}_p - \boldsymbol{\mu}_j)(\boldsymbol{x}_p - \boldsymbol{\mu}_j)^T \boldsymbol{\eta} = \boldsymbol{\alpha}^T \boldsymbol{H} \boldsymbol{\alpha}\end{aligned} \tag{17}$$

where $\boldsymbol{H} = \sum_{i=1,2} \boldsymbol{K}_i (\boldsymbol{I} - \boldsymbol{L}_i) \boldsymbol{K}_i^T$ .

$\boldsymbol{K}_i$ is the kernel matrix of $i$th class that satisfies with $(\boldsymbol{K}_i)_{r,s} = k(x_r, x_s)$, $\boldsymbol{I}$ is a matrix of $N_i \times N_i$ size, and all elements are $1/N_i$. Then Eq. (10) is equivalent to the following formula:

$$J(\boldsymbol{\alpha}) = \frac{\boldsymbol{\alpha}^T \boldsymbol{H} \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \boldsymbol{M} \boldsymbol{\alpha}} . \tag{18}$$

It can be seen that the essence of $\boldsymbol{\alpha}$ is the eigenvector corresponding to the maximum eigenvalue of the matrix $\boldsymbol{H}^{-1}\boldsymbol{M}$, which can be obtained directly as follows:

$$\boldsymbol{\alpha} = \boldsymbol{H}^{-1} (\boldsymbol{M}_1 - \boldsymbol{M}_2) . \tag{19}$$

Therefore, the best projection vector $\boldsymbol{\eta}^*$ is transformed into the solution $\boldsymbol{\alpha}$, which needs to be solved by the kernel function. Thus, the projection of features $\Phi(\boldsymbol{x}_1)$, $\Phi(\boldsymbol{x}_2)$, …, $\Phi(\boldsymbol{x}_N)$ on $\boldsymbol{\eta}$ is transformed into the projection of $k(\cdot, x)$ on $\boldsymbol{\alpha}$.

Here, the optimization objective is transformed into the projection vector $\boldsymbol{\eta}$ to obtain the minimum value of $J(\boldsymbol{\eta}^*)$. Since the space transformation is realized by the radial basis kernel function, the problem is transformed to find the optimal parameter $h^*$, so that $J(h^*)$ is minimized. $J(h)$ is used as the fitness function of the PSO optimization in intelligent optimization.

For multiple classes of problems, the FC discriminant method can be promoted. The set of samples contains $q$-class problems, which can be converted into $q(q\text{-}1)/2$ two-class problems, namely, the $q(q\text{-}1)/2$ discriminant functions are required.

## 3. Algorithm

### 3.1 Improved PSO algorithm

As a global optimization algorithm, the proposed PSO is based on the behavior rules of birds and the similarity of optimization problems. At present, there are some problems in the application of the PSO, such as being caught in a local optimal when the convergence speed of the PSO is accelerated. Therefore, researchers mixed it with other algorithms and increased the depth of local exploration and the breadth of the global search.

The standard PSO algorithm is expressed by Eqs. (20) and (21):

$$v_{id}(t+1) = \omega v_{id}(t) + c_1 r_1 (p_{id} - x_{id}(t)) + c_2 r_2 (p_{gd} - x_{id}(t)) \tag{20}$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \tag{21}$$

where $d$ is the dimension of the particle, $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id})$ and $\mathbf{v}_i = (v_{i1}, v_{i2}, \ldots, v_{id})$ denote the current position and velocity of the particle, respectively. $v_{id}$ is within the range $[-v_{max}, v_{max}]$, where $v_{max}$ is the maximum particle velocity. $p_{id}$ and $p_{gd}$ are the optimal location of particle $i$ and the optimal location of the population, respectively. $c_1$ and $c_2$ are cognitive factor and social factor, respectively, which are constants. $r_1$ and $r_2$ are random numbers within [0, 1]. $\omega$ is the inertial weight. Currently, linear decreasing weight is used more, as shown in Eq. (22).

$$\omega = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{T_{max}} \times t \tag{22}$$

where $\omega_{max}$ is the maximum value of $\omega$ and typically set within [0.9 1.4], $\omega_{min}$ is the minimum and set as 0.4, $t$ is the current evolution generation, and $T_{max}$ is the maximum evolution generation.

To improve the performance of the PSO algorithm, a large cognitive factor and a small social factor are set in the early evolutionary algorithm, so the individual cognitive accounts for a larger proportion, which can traverse the whole search space without quickly clustering to the local optimum. In the later stage of evolution, the social part dominates, and the smaller cognitive factor and larger social factor make the particles approach the global optimal region gradually. For this purpose, the improved PSO is a linear adaptive change for factors $c_1$ and $c_2$ within the evolution process, and the mathematical expressions are as follows [17]:

$$c_1 = R_1 + \frac{R_2}{T_{max}} \times t \tag{23}$$

$$c_2 = R_3 - \frac{R_4}{T_{max}} \times t \tag{24}$$

where $R_1$, $R_2$, $R_3$, and $R_4$ are the initial values. Through the simulation of test function, $0 < R_1 + R_2 \le 2$, $R_3 - R_4 \ge 1$, the algorithm can overcome the disadvantages of local search and accelerate the convergence.

### 3.2 The procedure for kernel function scale parameters optimized by the PSO-FC

The FC discriminant function $J(h)$ is taken as the fitness of the improved PSO algorithm, and then the kernel function scale parameter $h$ is optimized. The detailed process is shown in Fig. 2.

## 4. Experiment

### 4.1 Planetary gearbox testbed

As shown in Fig. 3, the planetary gearbox fault diagnosis testbed consists of a control cabinet, a variable speed drive motor, a helical gearbox, a planetary gearbox, a magnetic powder brake, and a vibration isolation pad. The planetary gearbox includes an inner gear ring, a sun gear, and three planetary gears. Table 1 illustrates the equipment's technical parameters.

Table 1. The equipment's technical parameters.

| Component | Name | Parameters |
|---|---|---|
| Helical gear box | Gear | Big gear: modulus 2, tooth number 77<br>Pinion: modules 2, tooth number 55 |
| | Bearing | Deep groove rolling ball bearing 6206 |
| Planetary gear | Gear | Inner gear ring: modulus 2, tooth number 72 |
| | | Planetary gears: modulus 2, tooth number 27, quantity 3 |
| | | Sun gear: modules 2, tooth number 18 |
| | Bearing | Deep groove rolling ball bearing: planetary gear 6202, planetary rack 6206, sun gear 6205 |
| Brake | - | The loading form is magnetic, loading torque is 0-100 N·m |
| Motor | Converter motor | 2.2 kW, rotational speed 1500 RPM, rated speed 1410 RPM |

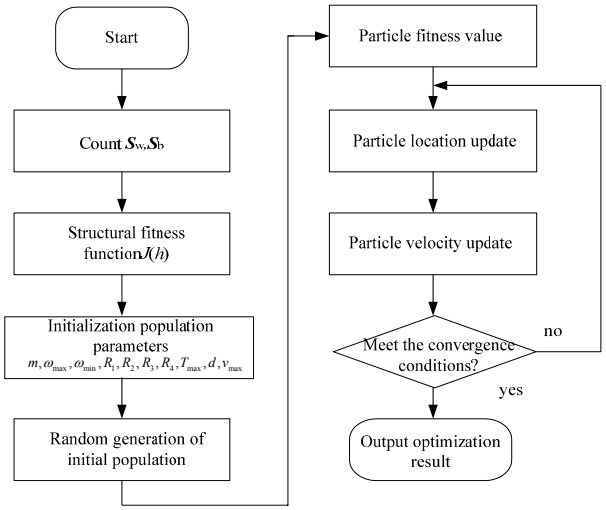



Fig. 2. Steps for optimizing kernel parameter with the PSO.

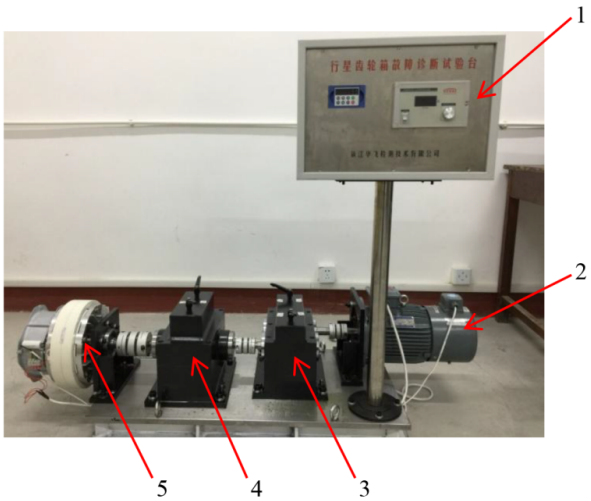



Fig. 3. Fault diagnosis testbed of the planetary gearbox (1. Control cabinet, 2. Motor, 3. Helical gearbox, 4. Planetary gearbox, 5. Magnetic powder brake).

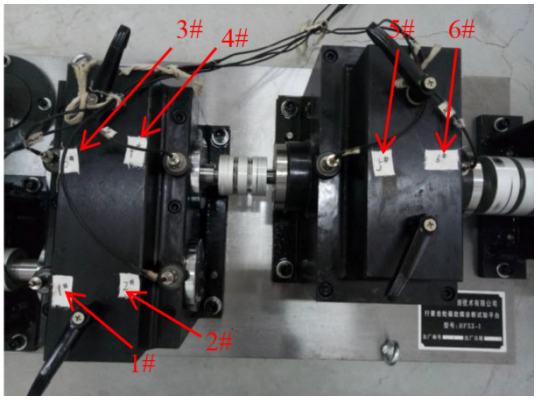



Fig. 4. The layout of the six measuring points on the planetary gearbox.

(a) One-tooth wear (b) Two-tooth wear (c) Three-tooth wear

Fig. 5. The wear fault modes of the planetary gear.

## 4.2 Planetary gearbox signal acquisition

The vibration signal was measured by six unidirectional piezoelectric acceleration sensors (CA-YD-186G) in this experiment, and Fig. 4 indicates the layout of the measuring points. Among them, four measuring points (1 to 4) were arranged on the helical gear-box, and the others (5 to 6) were on the planetary gearbox. In the experiment, the different degrees of wear damage of planetary gear were simulated, and four kinds of vibration state signals of the planetary gearbox were acquired in the experiment, including the normal state (model A), one-tooth wear (model B), two-tooth wear (model C) and three-tooth wear (model D). Fig. 5 shows the wear of planetary gear teeth. Figs. 5(a)-(c) show one-tooth wear, two-tooth wear, and three-tooth wear of planetary gear, respectively. Among them, Fig. 5(a) simulates the slight damage and shedding of the single tooth tip. Fig. 5(b) simulates moderate wear, that is, when the planetary gear is under heavy load or poor lubrication, the adjacent tooth surfaces are damaged and the tooth thickness is reduced. Fig. 5(c) simulates severe wear, that is, three consecutive tooth surfaces are damaged, and tooth profile deforms and tooth side clearance increases. The sampling frequency of the signal was 10.24 kHz and the sampling points were 4096.

## 4.3 Vibration signal analysis

The measured signals were processed and analyzed. When the planet gear is running normally, there is always a meshing frequency due to the periodic engagement of the teeth. At the same time, the rotating vibration signal of gear shaft can form a double frequency component, and thus two side bands around the meshing frequency and frequency doubling composition are formed on the power spectral density curve. According to

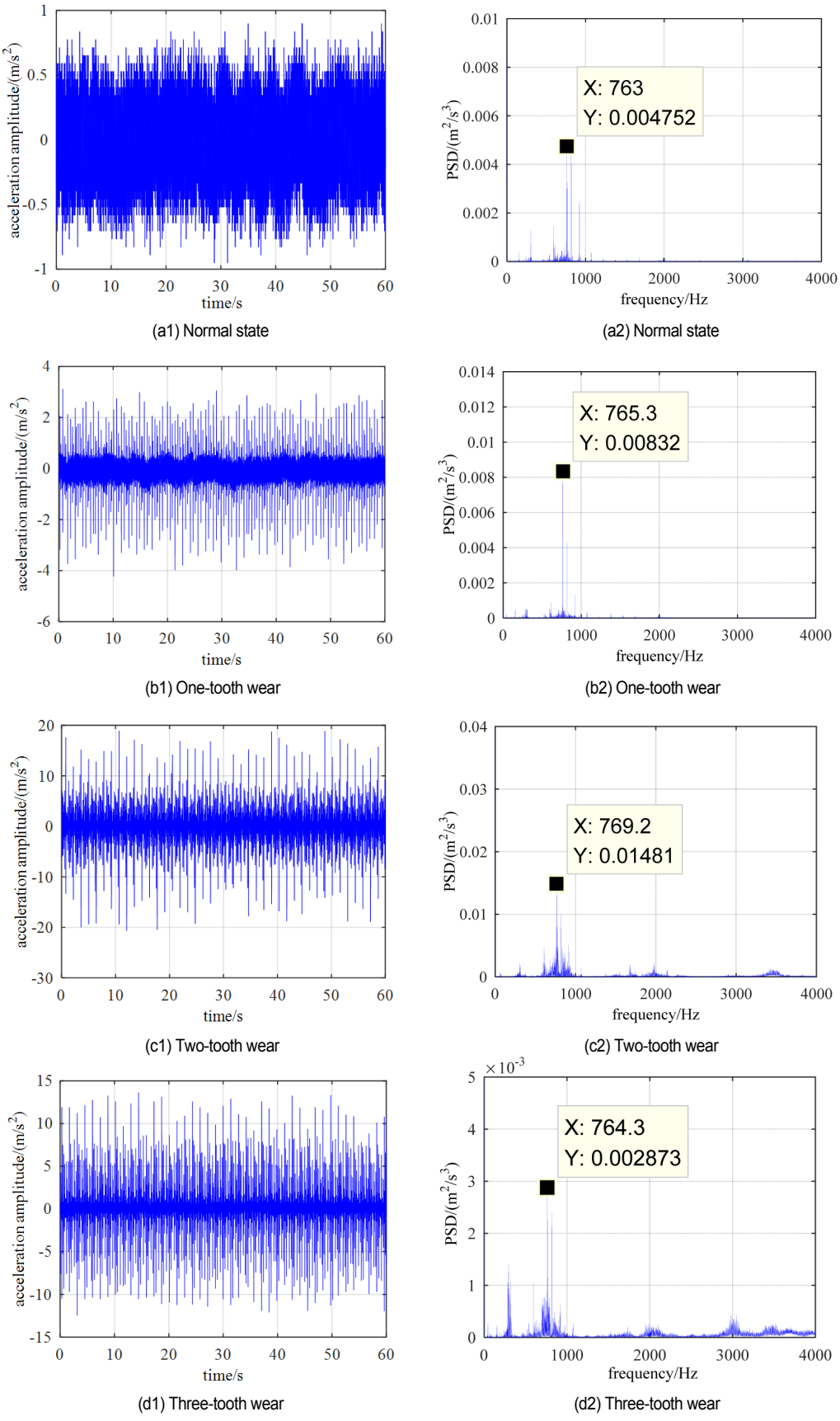



Fig. 6. The time-domain curves and power spectral density curves of measuring point 5.

the parameters in Table 1, the theoretical meshing frequency of planet gear was calculated as 720 Hz. The measuring point 5 on the planetary gearbox was taken as the research object. Figs. 6(a)-(d) show the time-domain curves (marked a1, b1, c1, and d1) and the power spectral density (PSD) curves (marked a2, b2, c2, and d2) of the four modes, respectively. As seen from Fig. 6, the meshing frequency of vibration signal is 763 Hz under the normal state, while it is 765.3 Hz, 769 Hz, and 764.3 Hz, respectively, under the other three kinds of wear states, which are close to the theoretical values, but the side band and amplitude under the three fault modes (seen in Figs. 6(b1), (c1) and (d1)) have larger changes compared with the normal state.

It is found that the amplitude of the vibration signal is small under normal working conditions, and the maximum PSD is $4.7523\times10^{-3}$ $m^2/s^3$ and the signal is relatively stable. When the planetary gear is worn, the stability of the vibration signal becomes poor and the amplitude begins to increase. When a tooth is worn, the maximum PSD is twice that of the normal state, reaching $8.32\times10^{-3}$ $m^2/s^3$. The PSD of the two-tooth wear is $1.481\times10^{-2}$ $m^2/s^3$, reaching 3.11 times the normal state. However, when the planetary gear operates with three-tooth wear, the maximum PSD is only $2.837\times10^{-3}$ $m^2/s^3$, which is smaller than normal and much smaller than the two-tooth wear. The reason is that the amplitude modulation and frequency modulation will occur when the planet gear wears out, so that the amplitude of gear meshing frequency will become larger or smaller, and the surrounding edge frequency band will become narrower or wider. As can be seen from Figs. 6(b2) and (c2), the PSD amplitudes of one-tooth wear and two-tooth wear increase to different degrees, and the edge frequency bands are narrowed to different degrees compared with the normal state. When three-tooth wear occurs, its wear is serious and makes the load fluctuations and rotational speed of planet gear uneven, so amplitude modulation and frequency modulation are presented, and the side frequency components of the curve are the results of two modulations. That is, after the side frequency components are superimposed with different phases, some side frequency amplitudes increase and some decrease, which destroys the symmetry of the frequency band under the normal state and reduces the overall PSD. Therefore, the signal energy is neutralized and weakened due to coupling and modulation. It is difficult to distinguish the fault modes of B, C, and D from the vibration signal and judge the fault degree.

## 5. Results and discussions

### 5.1 Establishment of the feature parameter set of the planetary gearbox

In the fault diagnosis, the acceleration signals collected can be used for signal processing and statistical analysis in the time domain and frequency domain. There are 27 features extracted. They include 21 time-domain features, such as mean value, mean square value, maximum value, minimum value, variance, root-mean-square value, root amplitude, absolute average amplitude, skewness, kurtosis, peak, six-order moments and so on, and six frequency-domain features, such as fequency domain variance, correlation factor, power spectrum barycenter index, mean square spectrum, harmonic factor, origin moment of spectrum.

Table 2. Parameters of the improved PSO.

| Parameters | $\omega_{max}$ | $\omega_{min}$ | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $d$ | $m$ | $T_{max}$ | $v_{max}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Values | 1.2 | 0.4 | 1 | 0.5 | 6 | 2 | 5 | 20 | 50 | 1.0 |

Feature extraction has a great influence on the accuracy and timeliness of fault diagnosis. Time-domain features can directly reflect the change of characteristic parameters with time. But they can not obtain the information of signal frequency change. Frequency-domain features mainly reflect the energy change of vibration signal with global frequency. In fact, a characteristic parameter is only sensitive to certain types of faults. Therefore, 27 feature attributes were selected in this paper, which could integrate some sensitive features in the time and frequency domain. According to these features, a fast diagnosis of planetary gearbox fault is realized.

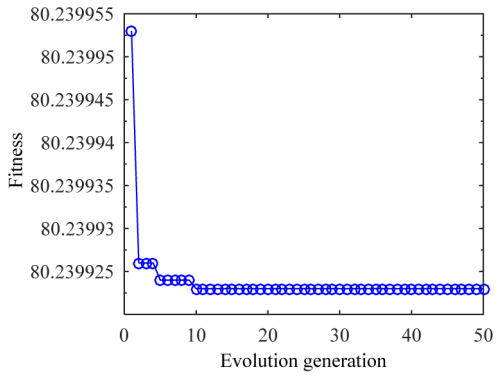
In this paper, the data of measuring point 5 on the shaft bearings of a planetary gearbox were selected as the samples for observation and analysis. 60 groups of training samples and 60 groups of test samples were extracted from the four models, respectively, and the kernel parameter of the KPCA was optimized and the sensitive features were extracted after a standardized treatment.

### 5.2 Optimization of the scale parameters of the kernel function based on the improved PSO

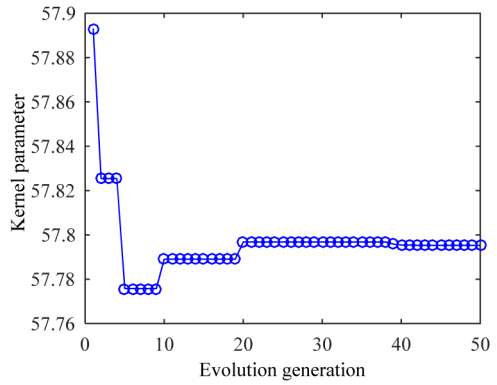
For the data sets of the four models, the scale parameter of the kernel function was optimized by the improved PSO. Table 2 shows the parameters of the improved PSO. Figs. 7(a) and (b) indicate the particle evolution process and the kernel parameter evolution process, respectively. It can be seen that the scale parameter $h$ can stabilize at an optimal value of 57.7953 within 50 iterations, and the minimum value of fitness $J(h^*)$ is 80.2399. Therefore, it may achieve the goal of maximizing classes spacing and minimizing intra class spacing by parameter optimization. The optimal value $h^*$ is the scientific basis for KPCA feature extraction.

### 5.3 Feature extraction by KPCA based on PSO-FC optimization

Based on the test data of the normal state and three different wear failure modes of the planetary gear in the experiment, KPCA was used for feature extraction after the kernel parameters were optimized by PSO-FC [18, 19]. Before optimization of the kernel parameter, Fig. 8(a) is the projection diagram for $h$ = 10, and Fig. 8(b) indicates the accumulative contribution rate of the kernel principal component. It is found that A can be sepa-

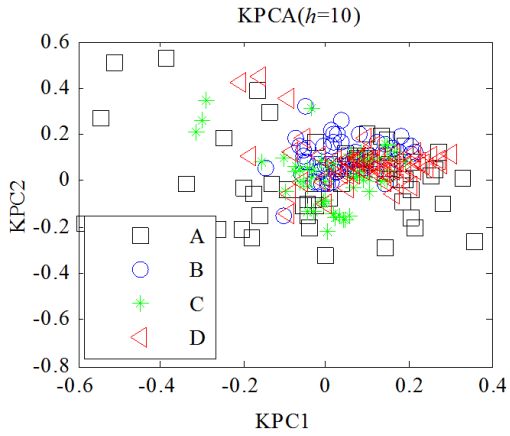



(a) Particle evolutionary process

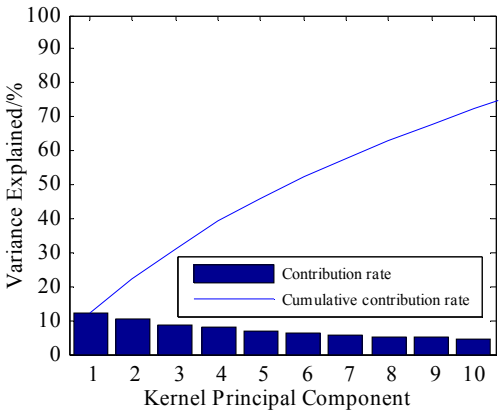



(b) Kernel parameter evolutionary process

Fig. 7. Evolution of the kernel parameter.




(a) KPCA projection diagram




(b) Contribution rate of the kernel principal component

Fig. 8. KPCA analysis results for different fault models before optimization of kernel parameter h (A: Normal, B: One-tooth wear, C: Two-tooth wear, D: Three-tooth wear).

rated from B and C, but there are many cross points among B, C, and D, which are mixed together and cannot be recognized. The degree of wear and the category cannot be distinguished at all. For comparison, the KPCA projection diagram of the four states after the optimization is exhibited in Fig. 9(a), and the accumulative contribution rate of the kernel principal component by KPCA is shown in Fig. 9(b). It is found that the data for the normal state are widely distributed but can be clearly distinguished from the other wear faults. Several sample points are in a transitional stage with individual crossover behavior from model B to model C. In general, B, C, and D are obviously different from each other, and the KPCA projection features have smaller intra class spacing. Hence, the projections of the KPCA in four states have an obvious scale segmentation effect after optimization of the kernel parameter.

Table 3 illustrates the results of the kernel principal component extraction feature. As seen from Table 3 and Fig. 9(b), the cumulative contribution rate of the first 10 kernel principal components can reach 92.9 %. That is, when the feature attributes are compressed from 27 to 10 dimensions, 92.9 % of the original features can be retained, which avoids the interference of redundant features on fault identification. The ten features can be used to form feature vectors for fault identification and classification, including the root-mean-square value, the absolute average amplitude, the variance, the root square amplitude, the waveform index, the peak index, the impulsion index, the tolerance index, the kurtosis index, and the power spectrum barycenter index. However, Fig. 8(b) indicates the cumulative contribution rate of the first ten kernel principal components cannot reach 80 %. This means the feature attributes cannot be completely retained when the feature dimensions are reduced.

### 5.4 Identification and classification of wear faults of the planetary gear based on SVM

Support vector machine (SVM) is a pattern recognition method based on kernel learning and widely used in fault diagnosis due to its advantages in dealing with small samples and uncertain problems [20]. The basic principle is to map the low dimensional data into the high dimensional space through nonlinear transformation, and to find the classification hyperplane with the maximum boundary to minimize the structural risk.

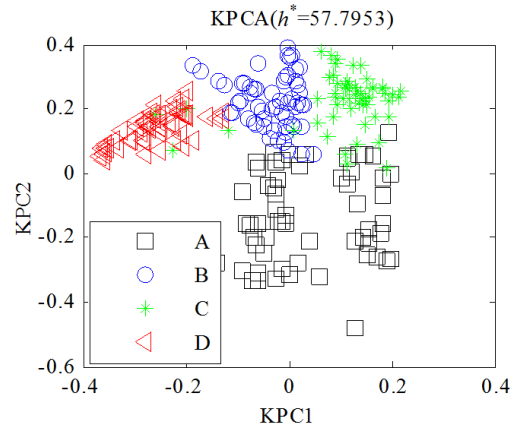
The specific steps are as follows [21, 22].

(1) Get training samples and test samples

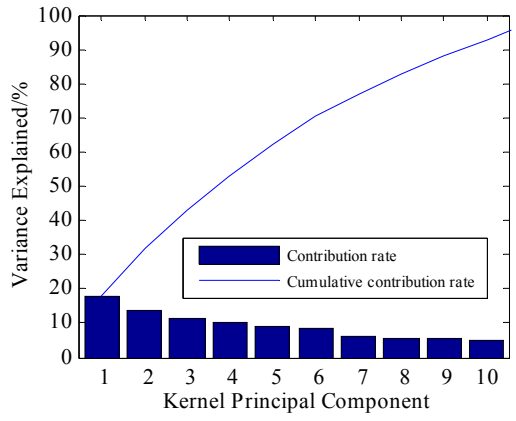
The original signal obtained by data acquisition was proc-

Table 3. KPCA analysis results.

| Number | Eigenvalue | Contribution rate | Cumulative contribution rate |
|---|---|---|---|
| 1 | 3.8157 | 17.8277 | 17.8277 |
| 2 | 2.9557 | 13.8095 | 31.6372 |
| 3 | 2.4352 | 11.3776 | 43.0138 |
| 4 | 2.1791 | 10.1811 | 53.1949 |
| 5 | 1.9396 | 9.0620 | 62.2569 |
| 6 | 1.8068 | 8.4418 | 70.6987 |
| 7 | 1.3144 | 6.1413 | 76.8400 |
| 8 | 1.2040 | 5.6255 | 82.4655 |
| 9 | 1.1484 | 5.3656 | 87.8311 |
| 10 | 1.0827 | 5.0588 | 92.8899 |




(a) KPCA projection diagram




(b) Contribution rate of kernel principal component

Fig. 9. KPCA analysis results for different fault models after optimization of kernel parameter *h* (A: Normal, B: One-tooth wear, C: Two-tooth wear, D: Three-tooth wear).

essed by time-domain and frequency-domain analysis to obtain corresponding characteristic parameters. After the KPCA optimization, ten features were selected. Fifty groups of optimized features were extracted from four states, respectively. Among them, 30 groups were used as the training samples and another 20 groups were used as the test samples. So a total of 120 groups of training samples and 80 groups of test samples were provided for fault identification and diagnosis by SVM.

Table 4. The identification results of four states of the planetary gear by SVM.

| Parameter set | Training samples | Test samples | Misjudgment numbers | Accuracy % |
|---|---|---|---|---|
| 10d feature | 120 | 80 | 2 | 97.5 |
| 27d feature | 120 | 80 | 16 | 80 |

(2) Constructing multi classification model

SVM belongs to two classifiers. For the complex system of planetary gear, there are many fault types. Therefore, it is necessary to construct a multi classifier to diagnose and analyze its faults. The basic idea is constructing $q(q\text{-}1)/2$ classifiers to solve the problem of $q$ classification, and each classifier can classify two kinds of samples. The Libsvm toolbox of SVM model can be used to implement classification, which adopts the one-to-one multi classification SVM method. In this paper, six classifiers were constructed, and RBF kernel function was used, and ten features in the time and frequency domain were taken as input, and four fault classification labels are obtained as output.

(3) Parameter pre-optimization

The radial basis kernel function was chosen as the kernel function of SVM. The kernel parameter of RBF was 40.34, and the penalty factor was 1, which were determined by the cross-validation method.

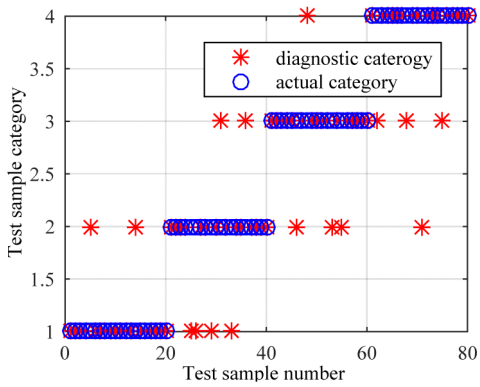
(4) Fault diagnosis

Eighty groups of test samples were put into the trained multi classification model, and the diagnosis results were obtained.

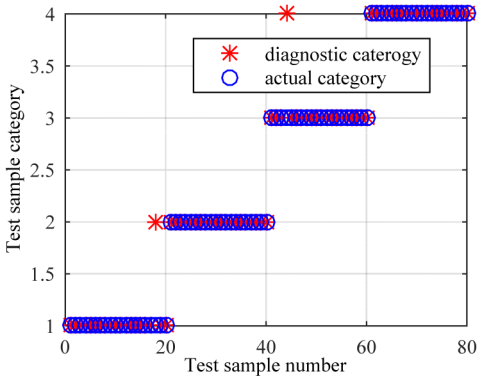
The classification results before and after feature optimization are shown in Figs. 10(a) and (b), where "o" denotes the category of the actual test sample and "*" denotes the category identified by SVM. If "*" and "o" coincide, the discrimination is right, and otherwise, it is wrong. 1, 2, 3, and 4 correspond to the state modes A, B, C, and D, respectively.

Among the 80 groups of test samples, only two were misjudged as seen from Fig. 10(b). One was the normal state which was misjudged as one-tooth wear, and another was the actual two-tooth wear fault, which was misjudged as three-tooth wear, and the overall accuracy was 97.5 %. However, there were 27 features before feature optimization, and due to the interference of some redundant features, many misjudgments occurred in the fault diagnosis. Among the 80 groups of the test samples, 16 groups were misjudged, and the accuracy was only 80 %, as seen from Fig. 10(a). Table 4 shows the recognition results of the SVM under both the normal and the wear conditions of the planetary gear, and the fault identification accuracy of SVM is improved by 17.5 %, which is higher than the result of 12.5 % obtained under the same experimental conditions in the Ref. [23]. This shows the superiority of this method.

The method proposed in this paper has been applied on




(a) Before feature optimization




(b) After feature optimization

Fig. 10. SVM classification results before and after feature optimization (1. Normal state, 2. One-tooth wear, 3. Two-tooth wear, 4. Three-tooth wear).

gearbox (JZQ-250 type) [24], in which the simulation failures included the normal working condition of the gearbox, bearing outer ring crack in the intermediate shaft, bearing cage fracture, gear broken tooth, and gear broken tooth combined with bearing outer ring fracture. The analysis results of gearbox fault diagnosis demonstrate its effectiveness.

## 6. Conclusions

Particle swarm optimization and other swarm intelligence algorithms have gradually become the research focus for solving complex optimization problems. In this paper, modeling optimization and an engineering case study are implemented to solve the nonlinear problems of KPCA. The conclusions are as follows:

1) By referring to the Fisher criterion discriminant function in pattern recognition, an intelligent optimization model based on the PSO-FC is established to optimize the kernel parameters, which improves the nonlinear analysis performance of KPCA.

2) KPCA with optimized kernel parameters is applied to feature extraction for different wear failure states of planetary gears, which compresses the feature dimension and retains the feature parameters with 92.9 % contribution rate. The feature extraction problem of the coupled vibration signals with multiple fault modes is effectively solved, and the fault identification accuracy of SVM is improved by 17.5 %. Therefore, the method has a good recognition effect on the nonlinear behavior of fuzzy fault boundaries in mechanical transmission.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 52005455), the Shanxi Province Science Foundation for Youths (Grants Nos. 201901 D211205 and 201901D211201), and the Opening Project of Shanxi Key Laboratory of Advanced Manufacturing Technology (Grants Nos. XJZZ202002).

## References

[1] C. Sun and Y. Wang, Advance in study of fault diagnosis of helicopter planetary gears, *Acta Aeronaut. Astronaut. Sin.*, 37 (7) (2017) 1-13.

[2] Y. Lei et al., Condition monitoring and fault diagnosis of planetary gearboxes: a review, *Measurement*, 48 (1) (2014) 292-305.

[3] X. Liang et al., Dynamic modeling of gearbox faults: a review, *Mech. Syst. Signal Process.,* 98 (2018) 852-876.

[4] Lei et al., A new dynamic model of planetary gear sets and research on fault response characteristics, *J. Mech. Eng.,* 52 (13) (2016) 111-122.

[5] J. Wu et al., Fault feature analysis of cracked gear based on LOD and analytical-FE method, *Mech. Syst. Signal Process.,* 98 (2018) 951-967.

[6] X. Liu et al., Resultant vibration signal model based fault diagnosis of a single stage planetary gear train with an incipient tooth crack on the sun gear, *Renew. Energy*, 122 (2018) 65-79.

[7] G. Cheng et al., Study on planetary gear fault diagnosis based on entropy feature fusion of ensemble empirical mode decomposi-tion, *Measurement*, 91 (2016) 140-154.

[8] X. Chen and Z. Feng, Iterative generalized time-frequency reassignment for planetary gearbox fault diagnosis under nonstationary conditions, *Mech. Syst. Signal Process.*, 80 (2016) 429-444.

[9] Y. Lei et al., Health condition identification of multi-stage planetary gearboxes using a mRVM-based method, *Mech. Syst. Signal Process.,* 60-61 (2015) 289-300.

[10] M. Khazaee et al., An appropriate approach for condition monitoring of planetary gearbox based on fast fourier transform and least-square support vector machine, *Int. J. Multidiscip. Sci. Eng.,* 3 (2012) 22-26.

[11] Z. Liu et al., A hybrid intelligent multi-fault detection method for rotating machinery based on RSGWPT, KPCA and Twin SVM Testing twin SVM, *ISA Trans.,* 66 (2017) 249-261.

[12] Z. Liu et al., Fault level diagnosis for planetary gearboxes using hybrid kernel feature selection and kernel Fisher discriminant analysis, *Int. J. Adv. Manuf. Technol.,* 67 (5-8) (2013) 1217-1230.

[13] H. X. Vo and L. J. Durlofsky, Regularized kernel PCA for the

efficient parameterization of complex geological models, *J. Comput. Phys.,* 322 (2016) 859-881.
[14] R. Liu et al., Artificial intelligence for fault diagnosis of rotating machinery: a review, *Mech. Syst. Signal Process.,* 108 (2018) 33-47.
[15] S. Qing et al., Point cloud simplification algorithm based on particle swarm optimization for online measurement of stored bulk grain, *Int. J. Agric. Biol. Eng.,* 9 (1) (2016) 71-78.
[16] M. Kallas et al., Fault detection and estimation using kernel principal component analysis, *IFAC-PapersOnLine*, 50 (1) (2017) 1025-1030.
[17] X. Wei, *Study on Inteiiigent Fauit Diagnosis of Gearbox Based on Particie Swarm Optimization*, North University of China (2009).
[18] X. Deng and L. Wang, Modified kernel principal component analysis using double-weighted local outlier factor and its application to nonlinear process monitoring, *ISA Trans.*, 72 (2018) 218-228.
[19] D. Kapsoulis et al., Evolutionary multi-objective optimization assisted by metamodels, kernel PCA and multi-criteria decision making techniques with applications in aerodynamics, *Appl. Soft Comput. J.,* 64 (2018) 1-13.
[20] J. Saari et al., Detection and identification of windmill bearing faults using a one-class support vector machine (SVM), *Measurement*, 137 (2019) 287-301.
[21] Z. He et al., Support tensor machine with dynamic penalty factors and its application to the fault diagnosis of rotating machinery with unbalanced data, *Mech. Syst. Signal Process.,* 141 (2020) 106441.
[22] M. Zhang et al., An improved sideband energy ratio for fault diagnosis of planetary gearboxes, *J. Sound Vib.,* 491 (2021) 115712.
[23] H. Yan and W. ZongYan, Fault diagnosis of planetary gearbox based on SFLA-BP model and KPCA feature extraction, *J. Mech. Strength*, 42 (2) (2020) 263-269.
[24] Y. He and Z. Wang, Regularized kernel function parameter of kpca using wpso-fda for feature extraction and fault recognition of gearbox, *J. Vibroengineering*, 20 (1) (2018) 225-239.

**Linzheng Ye** is an Associate Professor of Mechanical Engineering, North University of China, Taiyuan, China. He received his Ph.D. in Mechanical Engineering from North University of China. His research interests include precision and special machining, fault diagnosis and ultrasonics cavitation.