Journal of Physics: Conference Series

Fault feature extraction and recognition of vibration signal based on CWD-SK and HMM

To cite this article: Fan Cai and Tongbo Zhu 2022 *J. Phys.: Conf. Ser.* **2383** 012042

Fault feature extraction and recognition of vibration signal based on CWD-SK and HMM

Fan Cai, Tongbo Zhu

Minnan University of Science and Technology, Quanzhou Fujian 362700, China

[1]442567679@qq.com

Abstract. Aiming at the problem of rolling bearing fault diagnosis caused by interference information in the process of vibration transmission, this paper puts forward the theory and method of vibration signal fault diagnosis based on Choi Williams distribution spectral kurtosis (SK) combined with hidden Markov model (HMM). The fault location method of local mean decomposition (LMD) is used to screen the PF component with obvious fault characteristic information, CWD-SK is calculated and set as the characteristic parameter to realize the feature extraction of the initial fault of the vibration signal. CWD-SK is input into HMM as the feature vector to classify and identify the fault features of the vibration signal. Finally, the effectiveness of the algorithm is verified by the simulation of the actual data. It provides a robust theoretical and methodological basis for the establishment of fault diagnosis and classification recognition technology suitable for initial rolling mechanical vibration signals.

1. Introduction

Vibration fault diagnosis method is an indirect diagnosis method based on vibration measurement signal analysis [1]. The vibration signal reaches the vibration sensor through the complex transmission path of the mechanical system, and the fault signal is inevitably disturbed in the transmission process. The uncertain fault diagnosis information after transmission through the complex path is very easy to cause false diagnosis [2]. The occurrence of false diagnosis will lead to excessive maintenance or false maintenance, and even lead to major safety accidents [3]. Taking the rolling bearing as the research object, this paper studies the vibration characteristics of the outer ring, inner ring and rolling fault of the rolling bearing, and puts forward the fault location and fault diagnosis method of the rolling bearing. Study the interior of rolling bearing and carry out initial fault diagnosis, so as to find the early faults of internal equipment as soon as possible. Once the initial fault is found, notify the relevant staff for emergency repair as early as possible, so as to minimize the loss caused by the fault of rolling bearing [4]. This has far-reaching significance for rolling bearings and other rotating machinery and equipment to maintain a good running state for a long time.

2. Vibration signal preprocessing algorithm based on LMD

2.1. Formatting the title

Under the normal operation of rolling bearing, its vibration signal will appear periodic impact components caused by natural wear [5]. Once the initial failure of the rolling bearing begins, it is difficult to distinguish between the normal periodic impact component and the impact component caused by the failure. Therefore, the difficulty of fault feature extraction of rolling bearing will increase greatly. Then the vibration mechanism and vibration model of rolling bearing are analyzed, and finally the vibration signal is preprocessed based on LMD (local mean decomposition).

For the preparation work before fault diagnosis of rotating machinery such as various rolling bearings, the collected vibration signals will be preprocessed. It is usually processed from the following aspects, such as eliminating data outliers in the signal, eliminating trend items, signal denoising, etc. At present, data preprocessing includes the following mainstream methods: least square method, LMD method, EMD method, etc.

2.2. Basic algorithm principle of LMD

Considering the advantages and disadvantages of various methods, this paper finally uses LMD method to preprocess the rolling bearing data, in order to study the fault location method of the collected vibration signal. The LMD method can distinguish the pure FM signal from the envelope signal in the original signal, which is decomposed as follows:

(1) Set the original signal $x(t)$, calculate the local extreme points n_i of the original signal, and finally calculate the mean value of adjacent extreme points to obtain:

$$m_i = \frac{n_i + n_{i+1}}{2} \tag{1}$$

m_i represents the adjacent mean points, and the smoothing of LMD method adopts the moving average method to obtain the local mean function $m_{11}(t)$.

(2) Calculate the envelope estimation value according to step (1):

$$a_i = \frac{n_i - n_{i+1}}{2} \tag{2}$$

a_i is the adjacent envelope estimation point. The smoothing of LMD method adopts the moving average method to obtain the envelope estimation function.

(3) Subtract the local mean function $m_{11}(t)$ from the original signal $x(t)$, and the specific steps are as follows:

$$h_{11}(t) = x(t) - m_{11}(t) \tag{3}$$

(4) Divide $h_{11}(t)$ by envelope estimation function $a_{11}(t)$ to obtain:

$$s_{11}(t) = h(t) / a_{11}(t) \tag{4}$$

Repeat the above steps of equation (2-4) n times until $s_{1n}(t)$ is an FM signal, that is, the envelope estimation a of $a_{1(n+1)}(t) = 1$. The envelope signal $a_1(t)$ and the pure FM signal $s_{1n}(t)$ are multiplied to obtain the first pf component of the original signal, that is:

$$PF_1(t) = a_1(t) s_{1n}(t) \tag{5}$$

Among them, $PF_1(t)$ component mainly represents the high-frequency component in its original signal. As the first separated amplitude modulation signal, the related envelope signal is called instantaneous amplitude modulation. The instantaneous frequency $f_1(t)$ can be calculated from the pure FM signal $s_{1n}(t)$ as:

$$f_1(t) = \frac{1}{2\pi}\frac{d[\arccos(s_{1n}(t))]}{dt} \tag{6}$$

(5) The $PF_1(t)$ component is removed from the original signal $x(t)$ as the first PF component. $x(t)$ separates $PF_1(t)$ to get $u(t)$, from which the original data can be obtained. Repeat the above process n times, and the sign of the end of the final repeated process is that $u(t)$ is a monotonic function. The original signal $x(t)$ can be reconstructed by all PF components and $u(t)$, that is:

$$x(t) = \sum_{p=1}^{k} PF_p(t) + u_k(t) \tag{7}$$

The LMD will be decomposed from high frequency to low frequency, and each pf component will be in a corresponding frequency range. For the gear fault vibration signal in the gearbox, there may be many signals at different frequencies after LMD. Some frequency signals are not generated by the fault part or interference signals. Therefore, only some pf components of certain specific frequencies need to be selected. On the other hand, too minor faults have little impact on the equipment, and the PF_n decomposed component is very small. Therefore, the PF_n component that basically does not reflect the fault information with very small energy decomposed by LMD is generally ignored as an interference signal.

3. Spectral kurtosis model based on Choi Williams distribution

For the analysis and research of non-stationary signals, time-frequency tools have been widely used for a long time. At present, the focus of research and attention is how to improve time-frequency aggregation and how to suppress cross terms. From Wigner Ville's time-frequency distribution [9], similar to Cohen's time-frequency distribution, kernel function is introduced to improve this characteristic. It shows that the time-frequency characteristics not only have the characteristics of time-frequency set, but also fully suppress the interference of cross terms of multi-component signals due to the introduction of kernel function. Therefore, it has been widely studied in the field of time-frequency distribution.

CWD-SK has excellent time-frequency performance, suppresses the interference of its cross term, and is insensitive to the selection of window function. When the initial fault occurs, it can accurately reflect the transient variable components in the signal. The kernel function [10] is introduced to suppress the ambiguity far from the origin. The function is:

$$\phi(\tau, \nu) = \mathrm{e}^{[-\alpha(\tau\nu)^2]} \tag{8}$$

The inverse Fourier transform of the exponential kernel function of equation (8) is as follows:

$$\varphi(t, v) = \int_{-\infty}^{+\infty} \phi(\tau, v)\mathrm{e}^{-jvt}\mathrm{d}v = \frac{1}{\sqrt{4\pi\alpha\tau^2}}\mathrm{e}^{\frac{t^2}{4\alpha\tau^2}} \tag{9}$$

For signal $x(t)$, its CWD can be defined as follows:

$$C_x(t, f) = \int_{-\infty}^{+\infty} \sqrt{\frac{\sigma}{\sqrt{4\pi\tau^2}}}\mathrm{e}^{-\frac{\sigma t^2}{4\tau^2}} x(t + f + \frac{\tau}{2}) x^*(t + f - \frac{\tau}{2})\mathrm{e}^{-j2\pi ft}\mathrm{d}\tau \tag{10}$$

Wherein, in formula (10) is τ time shift parameter; $x*$ is the convolution of x; f is frequency; t is the time; σ is the scale factor (usually constant). As can be seen from the above formula, the kernel function $\phi(0,\nu) = \phi(\tau,0) = 1$. When it is not 0, $\phi(\tau, \nu) < 1$, σ is directly proportional to the self term resolution. When σ increases, the self term resolution increases. When σ decreases, the inhibition of its cross term will increase. Therefore, the selection of σ should weigh the self term resolution and the suppression of cross terms. For the selection of kernel function, not only should it be able to effectively suppress different frequencies and cross interference terms, but also the time-frequency resolution should not be too low, otherwise it will affect the results.

Therefore, according to the definition of cwd-sk, its k -order mean value is:

$$k_x(f)=\frac{\mathrm{C}_{4x}(f)}{\hat{\mathrm{S}}_{2x}^2(f)}=\frac{\hat{\mathrm{S}}_{4x}^2(f)-2\hat{\mathrm{S}}_{2x}^2(f)}{\hat{\mathrm{S}}_{2x}^2(f)}=\frac{\hat{\mathrm{S}}_{4x}^2(f)}{\hat{\mathrm{S}}_{2x}^2(f)}-2,(f\neq 0) \tag{11}$$

In formula (11), $\hat{\mathrm{S}}_{2x}(f)$ and $\hat{\mathrm{S}}_{4x}(f)$ are the second-order and fourth-order spectral cumulants of $x(t)$, and the spectral kurtosis reflects the degree of deviation of the signal from the Gaussian distribution to a certain extent. In the case of Gaussian distribution, the value of spectral kurtosis is zero; The spectral kurtosis of sinusoidal signal in frequency is -1. It is not difficult to see that deviation from Gaussian distribution is inversely proportional to spectral kurtosis. That is, the smaller the deviation, the greater the spectral kurtosis, and vice versa. This paper uses the above properties of spectral kurtosis to distinguish whether there is a fault in the initial state according to the change of the set value of spectral kurtosis.

4. Fault feature classification model and analysis based on HMM

4.1. Basic principles of HMM

In early 1979, Professor Baum's team proposed hidden Markov model (HMM). As a statistical analysis model, HMM is essentially a Markov process with unknown parameters. Because HMM has good randomness, it has great advantages in the process of model establishment. As a statistical model with incomplete observations, HMM has strong adaptability and wide application range. HMM is essentially a double random process, and a hidden variable sequence corresponds to the change process of system state one by one. The state of the system is described by another random process. Therefore, HMM has a double random process, which can not only reflect the randomness of the object, but also reflect the structural framework of the object. Therefore, the corresponding direct observation results can be used to study the problem a priori, so as to improve the accuracy.

HMM can usually be divided into continuous HMM and discrete HMM according to the statistical properties of variables. Taking the fault diagnosis of bearing as an example, the continuous observation variable is set as the feature extraction factor of vibration signal. The forward backward variable algorithm is used to calculate the probability of CHMM model and re evaluate the model parameters. Then, the vibration signals collected by the rolling bearing are analyzed in time domain and frequency domain respectively, and the observation variables - normal operation vibration signals (health state) and three different degrees of fault vibration signals are set. The observed characteristic variables are used as the basic data of training, and then trained into five CHMM models respectively. Finally, another part of the characteristic factors are input and used as test variables, and five kinds of CHMM models under different states are obtained by using Viterbi algorithm. When the log likelihood probability value of the observation sequence reaches the maximum, it is judged that the relevant state is the operation state of the observation sequence.

4.2. HMM fault characteristic model and analysis

According to the basic theory and algorithm of HMM, its essence is a classification statistical model based on time series. Aiming at the non-stationary dynamic analysis process with difficult feature extraction and poor repeatability of initial fault vibration signal. In the process of setting the experiment, the HMM selected in this paper has the characteristics of fast training speed, avoiding the difficulty of establishing the objective function and high classification and recognition rate. Figure 1 is the diagnosis flow diagram of hidden Markov model.

The main purpose of this process is to establish the corresponding HMM model, obtain the serial number of unknown fault state, then input each model in turn, and calculate and compare the possibility. Finally, the unknown signal fault type is obtained from the output probability of the maximum model.

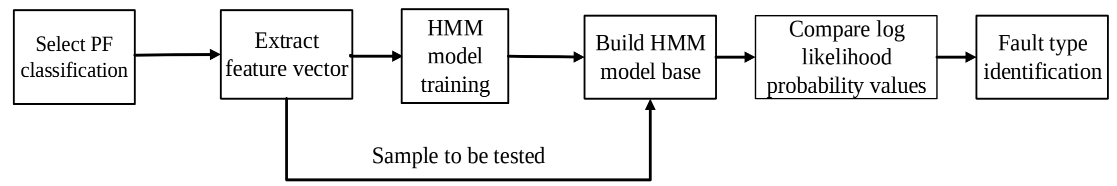

Figure 1. Diagnosis process based on Hidden Markov model (HMM).

Because the eigenvalues of vibration signal include time domain characteristics, frequency domain characteristics, spectrum demodulation spectrum, various entropy, kurtosis and so on. How to reasonably select the eigenvalue parameters instead of using all eigenvalues to build HMM model will lead to low recognition rate. Usually, only a few groups of characteristic vector data with typical high sensitivity need to be selected as training data, and the recognition rate of vibration signal is greatly improved by the trained hidden Markov model. After training, the model can be used to classify and identify the vibration signals in different modes.

5. Experiment and numerical analysis

5.1. Raw data collection

This paper uses the measured vibration data of gearbox rotor rolling bearing to verify the effectiveness of the vibration signal fault feature extraction and diagnosis method based on cwd-sk and HMM. Taking the accuracy and the number of training samples as the evaluation criteria, the diagnosis results of fault diagnosis methods are compared and studied. Based on the rolling bearing fault signals at different measuring points, the rolling bearing fault is identified, and the influence of different fault signal transmission paths on the fault diagnosis results is analyzed.

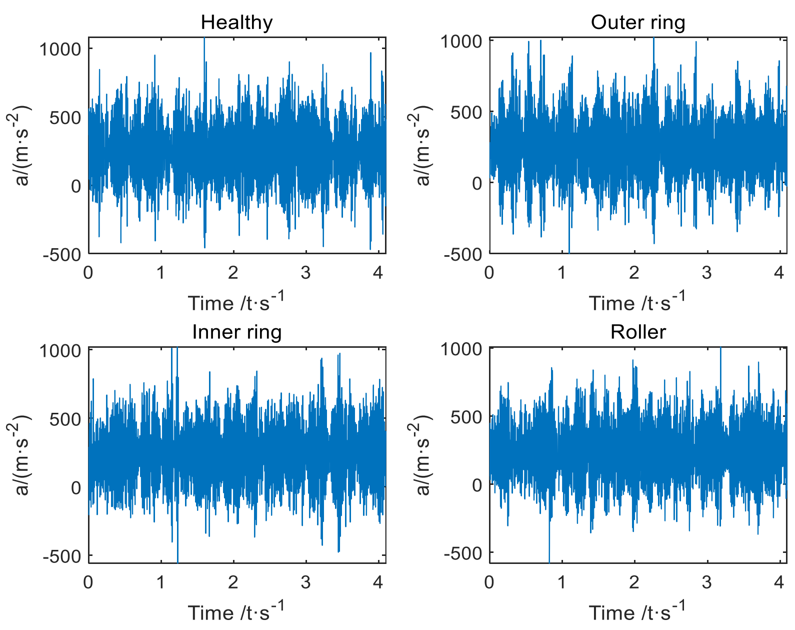

Figure 2. Vibration time domain signal of rolling bearing under fault mode

The vibration data of rolling bearing in four different states without load are collected as the test data of rolling bearing diagnosis method, which are: healthy bearing, outer ring fault bearing, inner ring fault bearing and roller fault bearing. The single sampling time is 4S, and the output shaft speed is set to 800rpm. The time domain waveform is shown in Figure 2. In order to ensure the integrity of samples of various state types, state vibration signals are selected under different loads. 60 groups of signals under each operating state are sampled respectively, the sampling frequency is 2.56khz, and the sampling time of each group is 4S.

In order to better describe the characteristics of different signals, the signals in the above four states of rolling bearing are subjected to fast Fourier transform to obtain the corresponding spectrum diagram, and figure 3 is obtained.The following is the time domain diagram and amplitude frequency diagram of rolling bearing in four states.

It can be seen that the time domain diagram is in a disordered state, extremely complex and basically random distribution. It is almost impossible to judge the operation state of the gearbox according to this. Therefore, in order to extract the initial fault and subsequent processing of fault classification, first find the decomposed non fault signal, which requires local mean decomposition LMD.

5.2. Simulation experiment of vibration signal preprocessing based on LMD

Based on the component, the pure fault signal and interference signal are obtained by subtracting the component from the original signal. Due to the slight interference, it is not obvious and will usually be submerged by strong noise. The energy decomposed by LMD is very small, which basically does not reflect the fault information. Therefore, the interference in the fault signal can be ignored, and the final result is the fault signal under the fault state of rolling bearing.

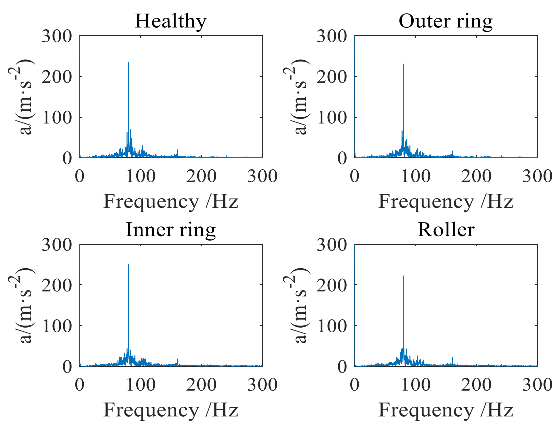

Figure 3. Spectrum characteristics of vibration signals of rolling bearing under four kinds of fault modes

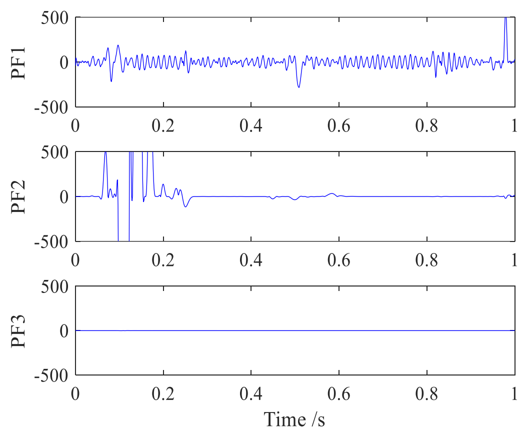

Figure 4. Vibration signal LMD in roller state.

The components are basically concentrated in the first two, accounting for the vast majority of the total energy of the signal, leaving little for the later components, so the later components can be ignored in the analysis. Based on the component, the pure fault signal and interference signal are obtained by subtracting the component from the original signal. Due to the slight interference, it is not obvious and will usually be submerged by strong noise. The energy decomposed by LMD is very small, which basically does not reflect the fault information. Therefore, the interference in the fault signal can be ignored, and the final result is the fault signal under the fault state of rolling bearing.

5.3. Initial fault feature extraction based on CWD-SK

The styles listed in Table 2 automatically add extra spacing before and/or after paragraphs: SPIE title, SPIE authors-affiliations, SPIE section heading, SPIE subsection heading, and SPIE body text. The 1.1

Heading 2 style automatically goes into the body text style after one paragraph return.

The Spectral Kurtosis Algorithm Based on Choi Williams distribution analyzes and processes the vibration signal. The test data of rolling bearing vibration in four different states as the rolling bearing diagnosis method are: healthy bearing, outer ring fault bearing, inner ring fault bearing and roller fault bearing. The average value of spectral kurtosis of Choi Williams distribution in these four states is:

Table 1. Comparison of CWD-SK for fault recognition.

Type	Healthy bearing	Outer ring fault bearing	Inner ring fault bearing	Roller bearing failure
Average value based on CWD-SK	2.596	3.165	3.214	3.454

According to table 1, it can be observed that according to the data calculated based on CWD-SK, when the rolling bearing is in a healthy state, the CWD-SK value is always less than 3, and the average value of CWD-SK under normal conditions is 2.596. When the rolling bearing has a slight initial fault, the CWD-SK value is always greater than 3, and its average value is 3.165. According to the literature , the fault characteristics are consistent when the spectral kurtosis value is greater than 3. Therefore, it shows that CWD-SK can well identify the initial fault of gear, the algorithm is relatively simple, and there is no fuzzy existence of cross interference term based on CWD-SK, which is consistent with the actual situation.Because the four types of fault features with different degrees of discrimination are not obvious, further fault classifier HMM model is needed to classify and identify fault types.

5.4. Fault feature type identification

The most likely path through the sequence is observed by Viterbi algorithm. During HMM modeling, cwd-sk eigenvalues are used as the observation sequence. During HMM modeling, cwd-sk eigenvalues are used as the observation sequence. The initial probability distribution vector, initial state transition matrix and observation matrix are obtained from the random function for normalization. The test matrix model is 10*4. Taking the input cwd-sk eigenvalue as the training sample, the number of samples is 20 groups (5 groups for each of 4 types of States).

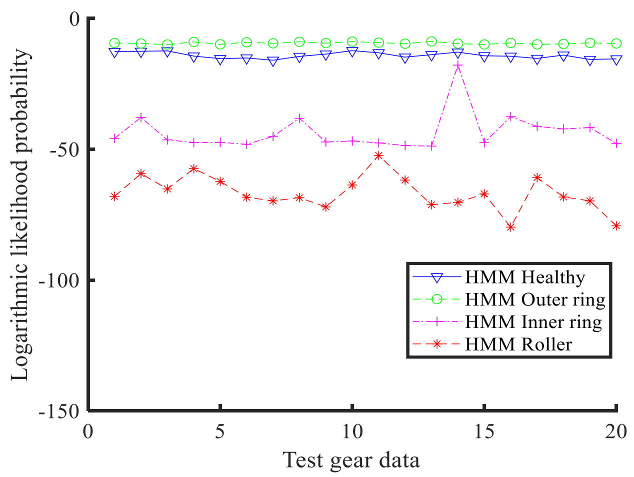

Figure 5. HMM training curves.

Firstly, the training sample vector is scalarized and Lloyd algorithm is used. Then enter the training stage and use Baum Welch algorithm. The number of iterations and logarithm natural estimates show a positive correlation trend until they completely converge to the end of training. After training, the corresponding HMM recognition model is obtained. According to Fig. 5, when the number of iterations is 20, all States converge, and the convergence speed is very fast. Among them, the log likelihood probability estimate represents the similarity of various types of HMM.

After the trained HMM models of each fault are obtained, a classifier will be established based on the three types of fault states of rolling bearings. It can be seen that the three kinds of fault states of rolling bearing converge after 20 iterations. For the trained HMM model, 20 groups of cwd-sk eigenvectors of the other models) are used as test samples to input three kinds of states for recognition.

Table 2. Comparison of CWD-SK for fault recognition.

Recognition mode	Outer ring fault bearing	Inner ring fault bearing	Roller bearing failure	Distinguish
CWD-SK HMM	5	4	5	95%
CWD-SK BP	4	4	4	90%

CWD-SK is used to extract the initial fault and some feature parameters. HMM model is used to train and classify the feature vector of gearbox. The results show that the overall recognition accuracy of cwd-sk model is higher than that of SK as the input of feature vector.

6. Conclusion

Aiming at the problem that the uncertain factors in the process of vibration signal transmission are easy to cause the wrong diagnosis of rolling bearing, this paper carries out the research work on the collection of rolling bearing vibration signal, the selection of data preprocessing LMD method, fault modeling, feature extraction, rolling bearing fault diagnosis method and so on. The proposed spectral kurtosis based on Choi Williams distribution can not only eliminate the interference of cross interference term and save the difficulty of selecting window function, but also improve the computational efficiency of identifying initial fault characteristics. HMM is applied to train the classification model of the extracted rolling bearing fault feature vector cwd-sk, and the four different states of rolling bearing are successfully distinguished and recognized. Finally, the support vector machine classification methods are compared, the rationality of selecting feature vector and the accuracy of HMM selection have played a good effect on rolling bearing fault diagnosis and classification and recognition.

Acknowledgments

This work is sponsored by Quanzhou science and technology project-Research on mechanism and application technology of spherical self rotating polishing (2020c033R).Project fund: 2022 school level scientific research project of MNUST (Project No.: 22kjx007)

References

[1] Ming L Y, et al.(2016)A multidimensional hybrid intelligent method for gear fault diagnosis[J]. Expert Systems with Applications,37(2): 1419-1430.
[2] Kaewkongka T, Au Y H J, Rakowski R. (2001)Continuous Wavelet Transform and Neutral Network for Condition Monitoring of Roto-dynamic Machinery[M]. IEEE Instrumentation and Measurement Technology Conference. Budapest, Hungary, 2(21):21-23.
[3] Cheng J, Yu D, Yang Y. (2004)A Method for Gear Fault Diagnosis Based on the Empirical Mode Decomposition[J]. International journal of plant engineering and management, 4(4):230-235.
[4] Guo Y, Tan K K.and Huang S, et al. (2015)Noise removal in Vold-Kalman order tracking based on independent component analysis[J]. Proceeings of IEEE International Conference on Automation and Logistics, Qingdao, China,1-3 Sept. 2(20):574-579.
[5] Guo Y, Tan K.(2018)High efficient crossing-order decoupling in Vold-Kalman filtering order tracking based on independent component analysis[J]. Mechanical Systems and Signal Processing,24(6): 1756-1766.