

Feature Mode Decomposition: New Decomposition Theory for Rotating Machinery Fault Diagnosis

Yonghao Miao , Boyao Zhang, Chenhui Li, Jing Lin , and Dayi Zhang

Abstract—In this article, a new decomposition theory, feature mode decomposition (FMD), is tailored for the feature extraction of machinery fault. The proposed FMD is essentially for the purpose of decomposing the different modes by the designed adaptive finite-impulse response (FIR) filters. Benefiting from the superiority of correlated Kurtosis, FMD takes the impulsiveness and periodicity of fault signal into consideration simultaneously. First, a designed FIR filter bank by Hanning window initialization is used to provide a direction for the decomposition. The period estimation and updating process are then used to lock the fault information. Finally, the redundant and mixing modes are removed in the process of mode selection. The superiority of the FMD is demonstrated to adaptively and accurately decompose the fault mode as well as robust to other interferences and noise using simulated and experimental data collected from bearing single and compound fault. Moreover, it has been demonstrated that FMD has superiority in feature extraction of machinery fault compared with the most popular variational mode decomposition.

Index Terms—Adaptive filter, correlated Kurtosis (CK), feature extraction, feature mode decomposition (FMD), machinery fault diagnosis.

I. INTRODUCTION

CONDITION monitoring and fault diagnosis of machinery equipment have critical signification to improve their availability, safety, and reliability, thus reducing the operation

Manuscript received 13 October 2021; revised 20 October 2021, 17 December 2021, and 21 January 2022; accepted 11 February 2022. Date of publication 9 March 2022; date of current version 5 October 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 51905017 and Grant 91860205 and in part by the Science and Technology on Reliability and Environmental Engineering Laboratory under Grant WDZC20220502. (Corresponding author: Jing Lin.)

Yonghao Miao is with the School of Reliability and Systems Engineering, Beihang University, Haidian 100191, China, and also with Advanced Manufacturing Center, Ningbo Institute of Technology, Beihang University, Ningbo 315100, China (e-mail: miao.yonghao@buaa.edu.cn).

Boyao Zhang, Chenhui Li, and Jing Lin are with the School of Reliability and Systems Engineering, Beihang University, Haidian 100191, China (e-mail: zhangboyao@buaa.edu.cn; lichenhui@buaa.edu.cn; jinglin@mail.xjtu.edu.cn).

Dayi Zhang is with the School of Energy and Power Engineering, Beihang University, Jiangsu 100191, China (e-mail: dayi@buaa.edu.cn).

The detailed code files of FMD can be found at <https://www.mathworks.com/matlabcentral/fileexchange/?term=profileid:6610099>.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TIE.2022.3156156>.

Digital Object Identifier 10.1109/TIE.2022.3156156

and maintenance costs, and achieving downtime minimization and productivity maximization. The related research has received much attention from both academic and industrial fields [1], [2].

Feature extraction, which uses signal processing techniques to extract the fault-related information from the complex measured signal, is the most important procedure in the machinery fault diagnosis. When a localized fault, defect, or damage occurs in the most prevalent components, such as rolling element bearing or gear, etc., a series of impulses will be periodically generated in the measured signal [3]. Therefore, the periodic impulses in the vibration signal are always considered as the important indicators of machinery fault. Generally, the core of feature extraction is to refine the periodic impulses from the complicated signal by matching and filtering, etc. The most-used methods include spectrum kurtosis [4], sparse filtering [5], deconvolution [6], etc. Yet, with the advancement of modern equipment towards high integration, the structure and composition are increasingly complex resulting in more components and interferences in the measured signal that seriously decreases the signal-to-noise ratio (SNR) of the signal as well as affects the extraction of fault information. Consequently, conventional feature extraction methods are confronting more and more challenges.

Decomposition methods can separate the different components from the original signal into several regular simple modes, that can be easily analyzed in the time and frequency domain [7]. They are considered one of the most effective tools for signal multicomponent analysis. Among them, empirical mode decomposition (EMD), which Huang *et al.* [8] pioneered, is the most prevalent and powerful time-frequency analysis technique [9]. Derived from the local characteristic time scales of a signal, EMD could adaptively decompose the superposed signal into a set of complete and almost orthogonal components, which are called intrinsic mode functions (IMFs). However, some limitations, i.e., mode mixing and boundary effect etc., that restrict its performance in machinery fault diagnosis. To overcome the shortcoming of EMD, the variants and new decomposition theories are successively proposed. As the classic improvement, ensemble EMD (EEMD) was proposed by Wu and Huang [10] assisted with background noise to eliminate the mode mixing problem. Regarding the product function as an amplitude-modulated and frequency-modulated signal and obtaining envelop signals and frequency-modulated signals respectively, Smith [11] designed the concept of local mean

decomposition (LMD). Although plenty of improvements [12], [13] have mitigated the problems derived from these decomposition theories, EEMD and LMD are essentially recursive as well as data-driven methods that never consider the form of fault feature. Motivated by this, Gilles [14] proposed empirical wavelet transform (EWT) whose principle is to build an adaptive wavelet basis to decompose the signal into adaptive subbands. Yet, its superiority largely depends on the segmentation of the Fourier spectrum. Similarly, inspired by the concept of IMF, a new adaptive, non-recursive decomposition, called variational mode decomposition (VMD) was proposed by Dragomiretskiy and Zosso [15] in 2014. They dexterously transform mode decomposition into a variational solution problem. Essentially, VMD is to design a group of Wiener filters and could separate the modes with different center frequencies restricted by minimizing the sum of the estimated bandwidth of each mode. But, the selection of mode number and balance parameter determines the decomposition performance of VMD. The mode number could control the location of the decomposed modes and indirectly affect the center frequency of the Wiener filter. As well, the balance parameter has been verified to determine the bandwidth of the mode component [16], [17]. Although numerous efforts have been devoted to parameter selection, it is still elusive to use VMD without any prior knowledge.

Given these literature reviews above, it is reasonable to conclude that current decomposition methods generally have the following issues in practice.

- 1) Without the complete consideration for the machinery fault feature, i.e., impulsiveness and periodicity of the signal, it is difficult for them to thoroughly separate the different components. For example, the application of VMD in machinery fault diagnosis is based on the characteristic. That is, the fault component is considered the narrow-band component that yet generally contains some interferences aside from the fault component in the measured signal.
- 2) The decomposition performance highly depends on the selection of the filter shape and bandwidth of the filter. For instance, EWT and VMD are to design the wavelet basis and Wiener filter to decompose the superposed signal. If the bandwidth is too big, different monocomponents, as well as the noise, might be blended into a single mode. In contrast, it might result in more redundant modes and some important elaborate information missing [16], [18].

Motivated by these limitations, a new adaptive decomposition theory, called feature mode decomposition (FMD), is proposed in this article. Inspired by the deconvolution principle that the adaptive finite-impulse response (FIR) filter is built by the iteratively updating filter coefficients to make the filtered signal infinitely approach the deconvolution objective function, FMD uses the adaptive FIR filter bank to decompose the signal into the different modes. And correlated Kurtosis (CK) [19] which can evaluate the impulsiveness and periodicity of the signal simultaneously is the best candidate to be the objective function in this article. In the process of FMD, a group of uniformly distributed FIR filters covering the whole frequency band are first initialized to decompose the original signal. Secondly, the filter

bank is iteratively updated using CK as the objective function as well as the fault period is estimated in the process of each filtering. Then, by comparing the correlation coefficient (CC) of every two modes, the one with less CK is abandoned from the two modes with maximum CC. Finally, the decomposition is ended when the terminal condition is achieved. The main contributions of this article are summarized as follows.

- 1) Taking the impulsiveness and periodicity of signal into consideration simultaneously, the decomposition target of FMD orientates the machinery fault as well as is robust to other interferences and noise.
- 2) The decomposed mode is extracted by the adaptive FIR filter. Without the restriction of the filter shape, bandwidth, and center frequency of the filter, the decomposition is more thorough.
- 3) Without the fault period as prior knowledge, FMD can accurately decompose the machinery fault information, especially the challenging compound fault that dominates the different frequency bands.

The rest of this article is organized as follows. The proposed FMD is elaborated through four parts in Section II. To verify the effectiveness of FMD, two simulation cases with bearing single fault and compound fault are applied in Section III. In Section IV, the real data from wind turbine bearing single fault and wheel bearing compound fault is further used for verification. The results and comparison of the proposed FMD and most popular decomposition method, VMD are discussed in Section V. Finally, the conclusion is drawn in Section VI.

II. PROPOSED METHOD

In this article, an FMD method is proposed for the adaptive filtering mode and machinery fault feature-oriented decomposition target. The proposed method will be presented in four parts in the following subsections.

A. Adaptive FIR Filter Bank

Different from EMD and LMD, the proposed FMD is designed to be a non-recursive decomposition method. By initializing an FIR filter bank and updating filter coefficients, the different modes can be adaptively chosen simultaneously. The filtering process is inspired by the deconvolution theory, that is the adaptive FIR filter is finally designed by the iteratively updating filter coefficients to make the filtered signal infinitely approach the deconvolution objective function. However, in the conventional deconvolution methods [19]–[21], the significance of filter initialization is often belittled. To verify the importance of filter initialization, a numerical simulation modeled the bearing fault is used. The simulated signal is generated according to the vibration model in (1). The synthetic simulated signal contains four parts. The first part is the periodic impulses caused by the bearing outer race fault. To simulate the random slip of the roller in bearing, the parameter v is set as 1%–2% of the fault period. The second part is the random impulses that may be derived from external knocks on the bearing housing, or electromagnetic interferences mixed into the measured signal. The third part is the harmonic interference from the gear meshing

TABLE I
PARAMETERS OF SIMULATED SIGNAL

	Impulses			Harmonics/Noise		
	Amplitude A_i or B_j	Period or number T_a or M_1	Resonant frequency f_n (Hz)	Amplitude C_1/C_2	Frequency f_1/f_2 (Hz)	Initial phase φ_1/φ_2 (rad)
Periodic impulses	1	1/29 s	2000	0.025/0.025	7/14	$1/6\pi / -1/3\pi$
Random impulses	<i>Random</i>	1	5100	SNR (dB)		-13

or shaft rotating. The last part is the background noise

$$x(t) = \underbrace{\sum_{i=1}^{M_0} A_i s_i(t - iT_a - v_i)}_{\text{Part 1}} + \underbrace{\sum_{j=1}^{M_1} B_j s_j(t - jT_d)}_{\text{Part 2}} + \underbrace{\sum_{k=1}^{M_2} C_k \sin(2\pi f_k t + \varphi_k)}_{\text{Part 3}} + n(t) \quad (1)$$

where A , B , and C are used to adjust the amplitudes of the different signal components. M is used to control the number of impulses and the harmonics. T_a and T_d are the time interval of the adjacent impulses. f and φ define the frequency and phase, respectively. The sampling rate is set as 20 kHz and the time is 1s. The impulse component is widely accepted as the impulse response of an underdamped single-degree-of-freedom system as

$$s(t) = e^{-\eta t} \cos(2\pi f_n t + \theta) \quad (2)$$

where f_n , η , and θ denote the frequency, coefficient of resonance damping, and the phase, respectively.

CK as the objective function is used in FMD to update the filter. Miao *et al.* [22] proposed an improved maximum CK deconvolution (IMCKD) that can obtain the optimal FIR filter coefficients without fault period as the prior knowledge. Here, IMCKD is used to process simulated signal with the parameters in Table I to test the filtering performance of the different initialization ways of the FIR filter. Simple initialization that is the most-used way in current methods is $\mathbf{f}_0 = [01 0 \dots 0]$. The proposed initialization is to use the Hanning window by the cut-off frequency. To facilitate interpretation, the resonant information of the fault is assumed to be known and the cut-off frequency is set as [1800 2200]. The result is shown in Fig. 1 when other parameters are the same. It can be found that IMCKD rapidly locks the fault period using Hanning window initialization after two iterations. Although the simple initialization still reaches the fault period after 16 iterations, it is obviously time-consuming and requires a higher computational resource. Furthermore, the traditional deconvolution methods just optimize an FIR filter once. In this regard, it provides more possibility for the decomposition theory based on the adaptive FIR filter bank. To achieve the Hanning window initialization, the frequency band of the raw signal firstly is divided into K segments uniformly. Then, the upper and lower cut-off frequencies of these segments f_u and f_l are assigned as

$$\begin{cases} f_l = k \cdot f_s / 2K \\ f_u = (k+1) \cdot f_s / 2K \end{cases} \quad k = 0, 1, 2, \dots, K-1 \quad (3)$$

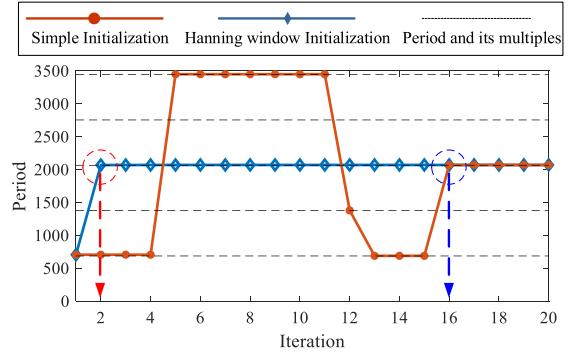


Fig. 1. Filtering performance of the different initialization ways of FIR filter.

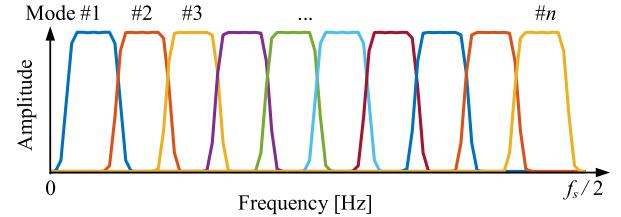


Fig. 2. FIR filter bank by Hanning window initialization.

where f_s is the sampling frequency of the raw signal. Therefore, K FIR filters with cut-off frequencies $[f_l, f_u]$ with the change of k from 0 to $K-1$ are generated by using the function *firl* in MATLAB. The filter length is L and the “window” is selected as “Hanning.” As shown in Fig. 2, a group of uniformly distributed FIR filters covering the whole frequency band could constitute an FIR filter bank by the proposed initialization. The symbol “#” denotes the serial number in this article. $\#n$ shows the location of the n th filter in Fig. 2. Through filter updating and mode selection, the fault modes can be decomposed adaptively.

B. Filter Updating and Period Estimation

The FIR filter bank by the proposed initialization just provides a coarse denoising function that may result in the condition that noise and interference may be contained in the candidate modes. The same fault component is divided into different modes. Therefore, a more elaborate filtering operation is highly desirable. The success of FMD lies in the selection of the objective function since it is the orientation of the decomposition. Generally, there are four signal components in the machinery fault signal, which is described in the fault mode in (1). They include the periodic impulses (signal #1), the random impulse interference (signal #2), the periodic harmonics (signal #3), and

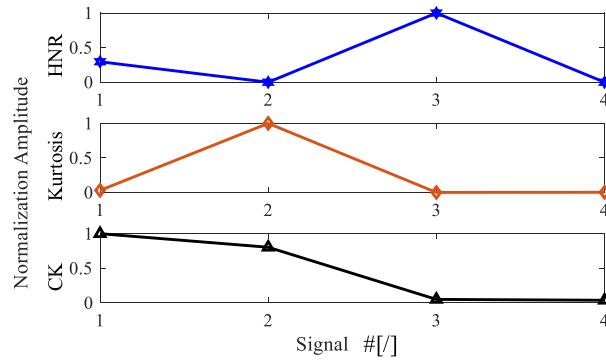


Fig. 3. HNR, Kurtosis, and CK values of four signal components.

other background noise (signal #4). To measure their characteristics, three indexes, including harmonic-to-noise ratio (HNR) [23], kurtosis [24], and CK [19], are used to evaluate the impulsiveness and periodicity. Four signal components are generated using the parameters in Table I. These indexes of four signals are computed and the result is normalized shown in Fig. 3. HNR and kurtosis are widely used to evaluate the periodicity and impulsiveness of the signal, respectively. From the figure, it can be clearly observed that the fault component (signal #1) does not have the most periodicity as well as the most impulsiveness. Yet, it has the maximum CK value that takes the periodicity and impulsiveness of the signal into consideration simultaneously. Therefore, FMD with CK as the objective function would prefer to extract the fault component rather than other components.

The original signal with length N is marked $x(n)$. The FMD theory then becomes the solution of a constrained problem and can be presented in (4).

$$\begin{aligned} \arg \max_{\{f_k(l)\}} & \left\{ CK_M(\mathbf{u}_k) = \sum_{n=1}^N \left(\prod_{m=0}^M u_k(n - mT_s) \right)^2 \right. \\ & \left. \left/ \left(\sum_{n=1}^N u_k(n)^2 \right)^{M+1} \right\} \right. \\ \text{s.t. } & u_k(n) = \sum_{l=1}^L f_k(l)x(n - l + 1) \end{aligned} \quad (4)$$

where $u_k(n)$ is the k th decomposed mode. f_k is the k th FIR filter and its length is L , where T_s is the input period which is measured using the sample number. M is the order of shift.

An iterative eigenvalue decomposition algorithm [25] is proposed to solve the constrained problem in (4). The decomposition mode can be represented in the matrix form

$$\mathbf{u}_k = \mathbf{X}\mathbf{f}_k \quad (5)$$

where

$$\mathbf{u}_k = \begin{bmatrix} u_k[1] \\ \vdots \\ u_k[N-L+1] \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x(1) & \cdots & x(L) \\ \vdots & \ddots & \vdots \\ x(N-L+1) & \cdots & x(N) \end{bmatrix}, \quad \mathbf{f}_k = \begin{bmatrix} f_k(1) \\ \vdots \\ f_k(L) \end{bmatrix}.$$

Then, the CK of the decomposition mode can be defined as

$$CK_M(\mathbf{u}_k) = \frac{\mathbf{u}_k^H \mathbf{W}_M \mathbf{u}_k}{\mathbf{u}_k^H \mathbf{u}_k} \quad (6)$$

where unnumbered eqs shown at the bottom of this page.

The superscripts H is the conjugate transpose operation. \mathbf{W}_M as the intermediate variable is used to control the weighted correlation matrix.

Substituting (5) in (6), the following expression can be obtained:

$$CK_M(\mathbf{u}_k) = \frac{\mathbf{f}_k^H \mathbf{X}^H \mathbf{W}_M \mathbf{X} \mathbf{f}_k}{\mathbf{f}_k^H \mathbf{X}^H \mathbf{X} \mathbf{f}_k} = \frac{\mathbf{f}_k^H \mathbf{R}_{XWX} \mathbf{f}_k}{\mathbf{f}_k^H \mathbf{R}_{XX} \mathbf{f}_k} \quad (7)$$

where \mathbf{R}_{XWX} and \mathbf{R}_{XX} are the weighted correlation matrix and the correlation matrix, respectively. Mathematically, the maximization of (7) with respect to filter coefficients is equivalent to the eigenvector associated with the maximum eigenvalue λ of the following generalized eigenvalue problem:

$$\mathbf{R}_{XWX} \mathbf{f}_k = \mathbf{R}_{XX} \mathbf{f}_k \lambda. \quad (8)$$

Therefore, with the iterative process, the k th filter coefficients will be updated by the solution of (8) to continually approach the setting objective, which is the filtered signal with maximum CK.

Intuitively, the accurate input period T_s plays a decisive role in the update of the filter coefficients. Yet, the fault period is hard to be estimated accurately in industrial practice due to the unknown condition and the fluctuant speed. IMCKD provides a solution to estimate the fault period from the measured signal. The scheme is based on the autocorrelation theory. That is, the autocorrelation spectrum of the signal would generate the local maximum value at the period location. If $R_x(\tau)$ denotes the

$$\mathbf{W}_M = \begin{bmatrix} \left(\prod_{m=0}^M u_k[1 - mT_s] \right)^2 & 0 & \cdots & 0 \\ 0 & \left(\prod_{m=0}^M u_k[2 - mT_s] \right)^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\prod_{m=0}^M u_k[N - L + 1 - mT_s] \right)^2 \end{bmatrix} \frac{1}{\sum_{n=1}^{N-L+1} u_k[n]^{M-1}}$$

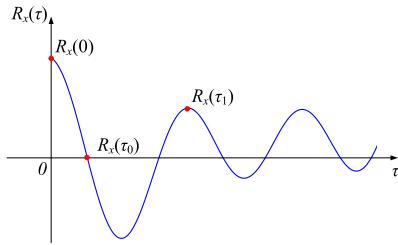


Fig. 4. Autocorrelation spectrum of the signal.

autocorrelation function of the signal $x(n)$, the expression of $R_x(\tau)$ about the lag τ can be defined as

$$R_x(\tau) = \int_{n=1}^N x(n)x(n+\tau)dn. \quad (9)$$

Fig. 4 shows the schematic diagram of the autocorrelation spectrum of a hypothetical signal. Since the point τ_0 in $R_x(\tau_0) = 0$ is called the zero-crossing point, the estimated period of each filtered signal is chosen as the point where the autocorrelation spectrum reaches the local maximum value $R_x(\tau_1)$ after the zero-crossing point, that is $T_s = \tau_1$.

It is worth mentioning here, with the update of the FIR filter, the estimated period would be updated and more and more accurate that has been verified in Fig. 1. And from the process of the estimated period in Fig. 1, it can be found that the order of shift $M = 1$ in CK has met the requirement.

C. Mode Selection

Obviously, it is undesirable to update each FIR filter initialized by the Hanning window from the start to the end since it would result in lots of repetitive and redundant modes being decomposed as well as a high computational resource. Substantially, the FIR filter bank by Hanning window initialization provides a fuzzy direction for the decomposition. Further, the period estimation and updating process can accurately lock the fault information. Motivated by this, the mode selection is proposed here.

Theoretically, the final modes will contain the specific components in the raw signal. FMD only chooses the modes with maximum CK value. Therefore, lots of modes might contain the same components if all initialized filters are updated from the start to the end in FMD. To eliminate the mode mixing or redundant modes, the modes with the biggest CC are firstly locked since a higher CC means the two modes contain more same components [26], [27]. Meanwhile, to reserve the mode with more fault information, the one with less CK is abandoned from the two modes with maximum CC. The CC of two modes \mathbf{u}_p and \mathbf{u}_q is defined as [28]

$$CC_{pq} = \frac{\sum_{n=1}^N (u_p(n) - \bar{u}_p)(u_q(n) - \bar{u}_q)}{\sqrt{\sum_{n=1}^N (u_p(n) - \bar{u}_p)^2} \sqrt{\sum_{n=1}^N (u_q(n) - \bar{u}_q)^2}} \quad (10)$$

where \bar{u}_p and \bar{u}_q are the mean value of modes \mathbf{u}_p and \mathbf{u}_q , respectively.

D. Feature Mode Decomposition

Based on the description earlier, the procedure of FMD can be summarized as follows.

Step 1: Load the original signal and input the parameters, i.e., the mode number n and the filter length L .

Step 2: Initialize the FIR filter bank by Hanning window using K filters which are recommended to set as 5–10, and start iteration $i = 1$.

Step 3: Obtain the filtered signals (i.e., the decomposed modes) by $\mathbf{u}_k^i = \mathbf{x} * \mathbf{f}_k^i$ where $k = 1, 2, \dots, K$, and $*$ is the convolution operation.

Step 4: Update the filter coefficients using raw signal \mathbf{x} , decomposed mode \mathbf{u}_k^i , and the estimated period T_k^i that is chosen as the point where the autocorrelation spectrum of \mathbf{u}_k^i reaches the local maximum value R_k^i after the zero-crossing point. Accomplish one iteration and set $i = i+1$.

Step 5: Judge whether the iteration number reaches the pre-iteration number. If not, return to step 3 otherwise enter step 6.

Step 6: Compute CC of every two modes to construct a $K \times K$ matrix $\mathbf{CC}_{(K \times K)}$. Lock the two modes with the biggest CC value CC_{max} and compute their CK using the estimated period. Then, the one with less CK is abandoned from the two modes and set $K = K-1$.

Step 7: Judge whether mode number K reaches the specified n . If not, return to Step 3 otherwise enter step 8.

Step 8: Obtain the reserved modes as the final decomposed modes.

To clearly depict the algorithm flow of FMD, a flow diagram is plotted in Fig. 5.

III. SIMULATION ANALYSIS

In this section, two simulated cases with different bearing faults, whose simulated model is built in Section II-A, are utilized to verify the feasibility of the proposed FMD. The most prevalent and powerful decomposition in the field of machinery fault diagnosis at present, VMD, is used as the comparison reference in this article.

A. Bearing Single Fault

To highlight the advantage of FMD that the filter design is adaptive, the resonant frequency band is discretized in 2000–5000 Hz and SNR is set as –10 dB. Other parameters are the same as the case in Section II-A. Fig. 6 shows the time waveform and the frequency spectrum. It is hard to find the periodic impulses from the raw signal. The discrete frequency band of the fault component is highlighted in Fig. 6(b).

In this case, the mode number $n = 1$ and the filter length $L = 40$ are set in advance. Fig. 7 provides the decomposition result. The obvious period impulses in the time waveform as well as the clear fault characteristic frequency and its multiples indicate

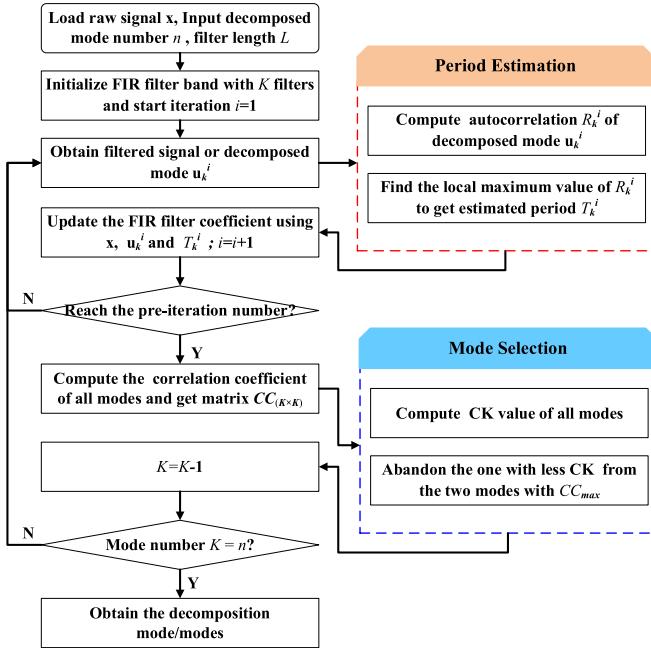


Fig. 5. Flow diagram of the proposed FMD.

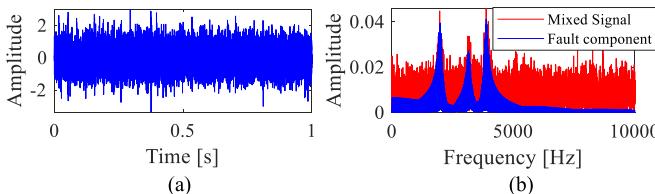


Fig. 6. (a) Raw signal and (b) its frequency spectrum in simulation A.

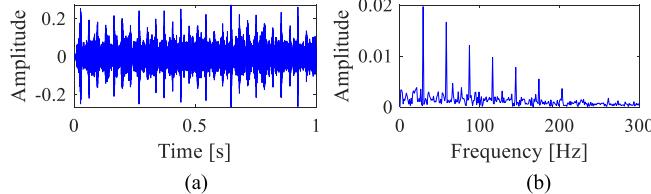


Fig. 7. (a) Decomposed mode of FMD and (b) its envelope spectrum in simulation A.

the outer race fault. Fig. 8 depicts the iteration process of FMD. After five preiterations, the mode selection starts. And one mode is removed after every two iterations. Finally, mode #3 that is defined by the Hanning window initialization is reserved.

Fig. 9 shows the result of VMD. After some trials, the fault information can be extracted when the mode number is set as 2. Mode #1 shows the existence of the outer race fault, but mode #2 is occupied by a single random impulse. According to the different SNRs of the decomposed modes of FMD and VMD, it can be inferred that there is some discrepancy in the modes. Fig. 10 provides information on the frequency spectra of the two modes. The decomposed mode of FMD covers the discrete frequency band and the wave crests are highly consistent

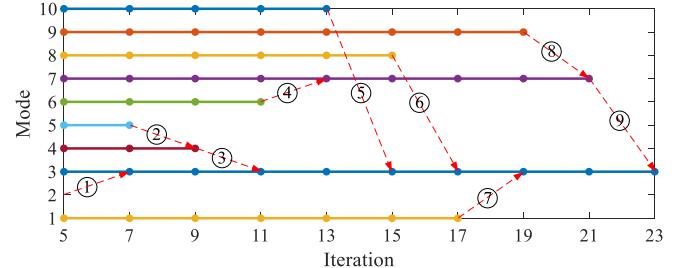


Fig. 8. Iteration process of FMD in simulation A.

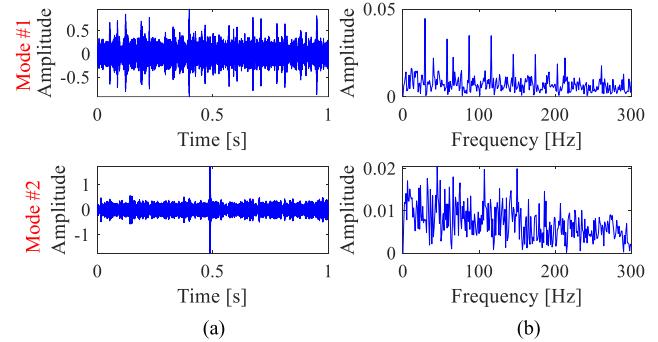


Fig. 9. (a) Decomposed modes of VMD and (b) their envelope spectra in simulation A.

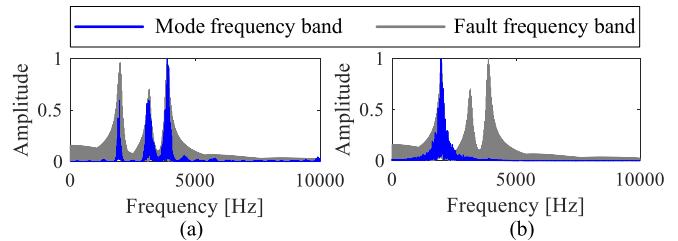


Fig. 10. Frequency spectra of (a) decomposed mode of FMD and (b) decomposed mode #1 of VMD in simulation A.

with the fault frequency band from Fig. 10(a). However, mode #1 of VMD just covers the frequency band around 2000 Hz from Fig. 10(b). To compare their performance, CK values of different frequency bands filtered by Hanning window, including decomposed mode of FMD and VMD are computed and shown in Fig. 11. It can be found the proposed FMD has the best filtering performance.

B. Bearing Compound Fault

Bearing compound fault diagnosis is always considered a challenging task compared with the single fault diagnosis [29]. To show the superiority of the proposed FMD, a simulation with bearing inner and outer race faults is used in this section. The resonant frequencies of the inner and outer race fault are set as 3605 and 2000 Hz, respectively. Their fault periods are 1/28s and 1/22s, respectively. SNR is set as -10 dB. Other parameters are the same as the case in Section II-A.

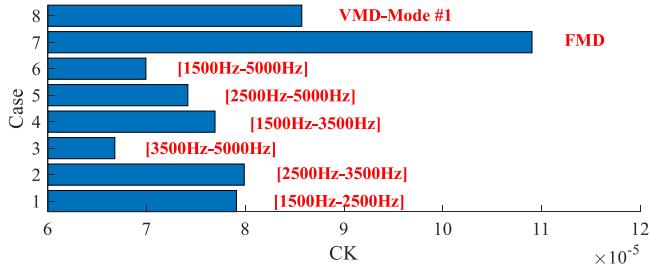


Fig. 11. CK values of different filtered frequency bands in simulation A.

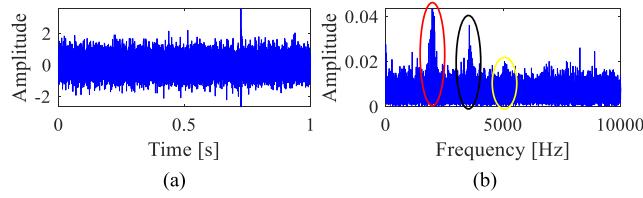


Fig. 12. (a) Raw synthesized signal and (b) its frequency spectrum in simulation B.

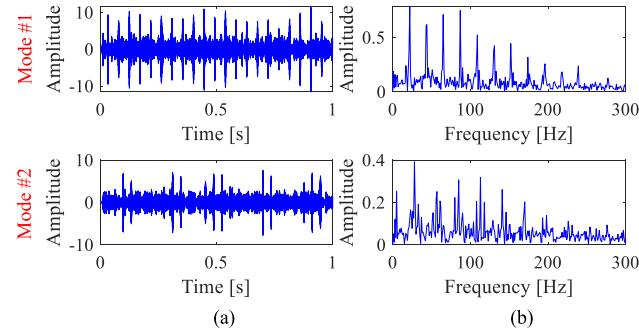


Fig. 13. (a) Decomposed modes of FMD and (b) its envelope spectra in simulation B.

Fig. 12 shows the time domain and frequency spectrum of the synthesized signal. The single random impulse is obvious and it cannot be found any fault information from Fig. 12(a). The red, black, and yellow ellipses respectively mark the locations of outer race fault, inner race fault, and random impulse in the frequency domain.

In this case, the mode number $n = 2$ and the filter length $L = 40$ of FMD are set in advance. The results of decomposition are presented in Fig. 13. Two different fault information is clearly decomposed into the two modes. Both the outer and inner race faults in mode #1 and mode #2 have high SNRs.

For comparison, VMD is used in the same signal and the mode number is set as 2. After some trials through changing the balance parameter, it is still challenging to extract two faults simultaneously. Fig. 14 shows the results of VMD when the balance parameter is 5000. The outer race fault is decomposed into mode #1. But, mode #2 is the random impulse. Furthermore, the frequency spectra of the decomposed modes of the two

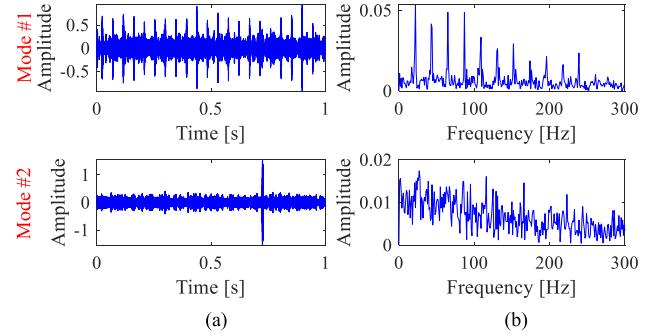


Fig. 14. (a) Decomposed modes of VMD and (b) its envelope spectra in simulation B.

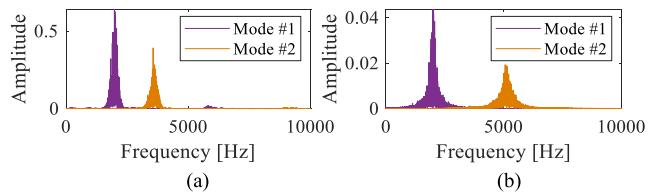


Fig. 15. Frequency spectra of (a) decomposed modes of FMD and (b) VMD in simulation B.

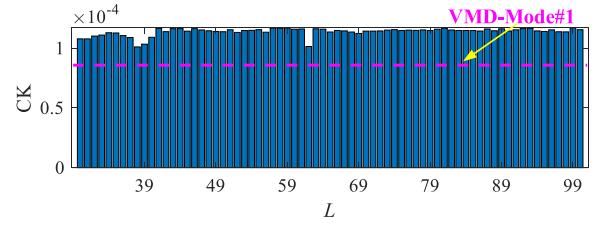


Fig. 16. CK values of the filtered signal of FMD with the change of the filter length L in simulation A.

methods are plotted in Fig. 15. It can be observed that there is no mode mixing phenomenon in the two decomposition methods.

C. Performance Analysis and Parameter Selection Rule

The parameters L , K , and n , as the input of FMD, play an important role in the decomposition performance. This section would make a discussion about the parameter selection rule through the performance analysis based on the simulation.

L is the filter length that is the important parameter of the FIR filter. A short filter length is easy to lead to a coarsely filtering result. A too-long one may result in distortion and further increase the computational burden [30]. In other methods, such as deconvolution methods, a small difference, for example, $L = 45$ and $L = 46$, may generate entirely different results [30]. Fortunately, in the proposed FMD, the selection of the filter length has little effect on the filtering performance when it fluctuates at a suitable range. To verify the performance of FMD, a group of simulations based on simulation A is used to test. Fig. 16 shows CK values of these filtered signals with different filter lengths from 30 to 100 and other conditions are kept unchanged. It can be observed that their CK values that

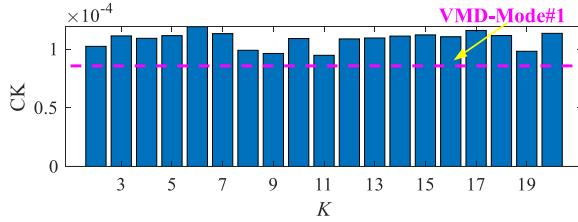


Fig. 17 CK values of the filtered signal of FMD with the change of the parameter K in simulation A.

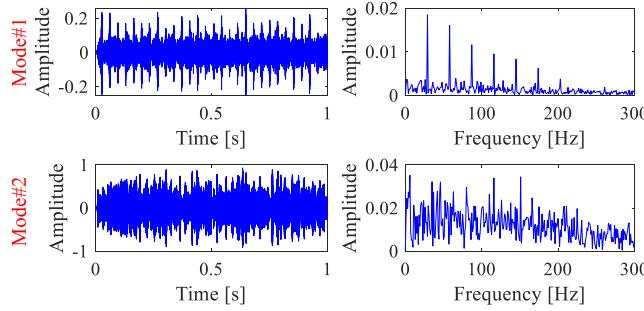


Fig. 18. Decomposition result of FMD with $n = 2$ in simulation A.

have little fluctuation are higher than that of mode #1 of VMD in simulation A. It means that all of them have the higher SNR that verifies the selection of the filter length has little effect for the filtering performance. In general, [30, 100] is the suitable range for the filter length of FMD.

The parameter K can be used to determine the segment number of the frequency band that is divided for the initialization of the FIR filter. First, a bigger K brings a heavier computational burden. Second, $K \geq n$ is another important restrictive condition to ensure that the n modes can be decomposed. Finally, to some extent, the selection of parameter K can affect the decomposition performance. Therefore, a tradeoff between the decomposition performance and computational burden may be made for the optimal selection of the parameter K . Similarly, to evaluate the influence of parameter K , a group of simulations based on simulation A is used to test the performance of FMD when other conditions are kept unchanged. The result is shown in Fig. 17. It can be observed that CK values with different parameter K from 2 to 20 are higher than that of mode #1 of VMD in simulation A even though they have a tiny fluctuation. It means that FMD is robust to the change of the parameter K in the range of [2, 20]. In this article, the parameter K is recommended to select in the range of [5, 10].

The parameter n determines the number of the decomposition mode. It has been highlighted and verified that FMD that uses CK as the objective function would prefer to extract the fault component rather than other components. The mode selection strategy proposed in Section II-C further eliminates the mode mixing. A too-big parameter n may generate the redundant mode. Simulation A and simulation B are used to verify. In simulation A, parameter n is changed as 2, the decomposition result shown in Fig. 18(a). Mode #1 is the fault information that is the same as the result with $n = 1$ shown in Fig. 7. Mode #2 is the redundant mode that is full of noise. Similarly, the parameter n

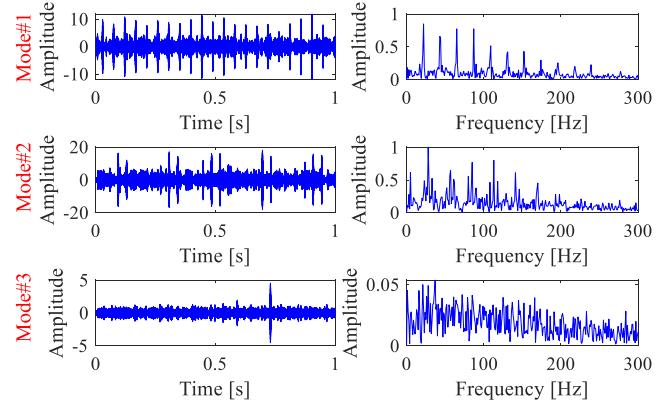


Fig. 19. Decomposition result of FMD with $n = 3$ in simulation B.

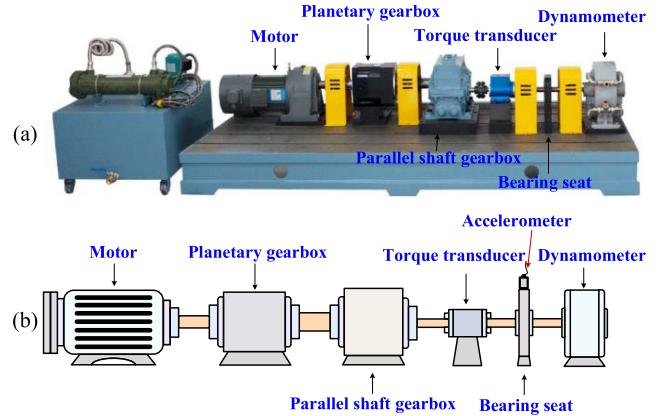


Fig. 20. (a) Real experimental setup and (b) schematic diagram of wind turbine experiment bench.

is changed as 3 in simulation B, the redundant mode is generated in Fig. 19. There is no doubt that a too-small parameter n would lose the potential fault information. In this article, the parameter n of FMD is empirically selected from 3, 2, 1. Actually, just as VMD, it is difficult to provide a priori estimation of the suitable parameter n in FMD that needs to be further investigated.

IV. EXPERIMENTAL STUDY

Further, the real data with bearing fault is used to verify the effectiveness and generalizability of the proposed FMD. Similarly, VMD is considered the reference for the comparison.

A. Wind Turbine Bearing Single Fault

In this section, the data was measured from the wind turbine bearing with the inner race fault. Fig. 20 presents the real figure and schematic diagram of the experiment bench. There are two gearboxes, including a planetary gearbox and a parallel shaft gearbox, being used to power transmission. Motor and dynamometer are used to provide power and resistance, respectively. Torque transducer can measure the speed information for the motor control feedback. The bearing seat is used to support the shaft. The accelerometer is mounted on the bearing seat to collect the vibration information with a 20 k Hz sampling rate. A wear fault is planted on the inner race of the bearing in this

TABLE II
FAULT CHARACTER FREQUENCY AND ROTATING FREQUENCY

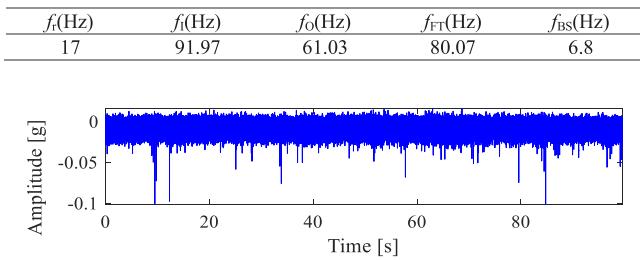


Fig. 21. Measured vibration signal.

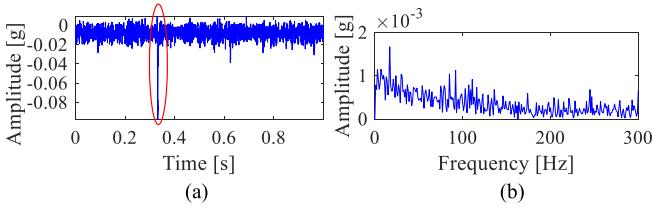


Fig. 22. Analyzed vibration signal (a) time domain and (b) its frequency spectrum.

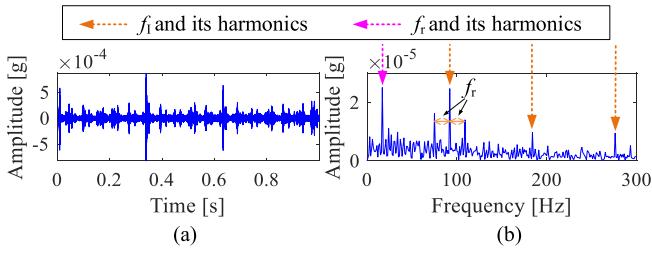


Fig. 23. (a) Decomposed mode of FMD and (b) its envelope spectrum in experiment A.

case. According to the design parameter of bearing and speed information, the fault character frequency of different parts of the bearing can be calculated in Table II. In this article, f_1 , f_0 , f_{FT} , f_{BS} denote the fault character frequency of inner race, outer race, roller, and fundamental train, respectively. f_r denotes the rotating frequency.

Fig. 21 presents a measured vibration signal with 99.57 s. As can be seen in the waveform, lots of impulses without any rule dominate the landscape due to the heavy work condition. Therefore, the measured signal is considerably complex.

It is not necessary to analyze the whole data with 99.57 s. The decomposition methods can achieve much in terms of fault diagnosis with a little measured signature. Therefore, the data from 12–13 s is cut for the analysis. Fig. 22 depicts the time waveform and its envelope frequency spectrum. A single impulse with high amplitude is very prominent in Fig. 22(a). Generally, most signal processing methods are hard to be free from the interference. First, FMD is used to process the signal. The mode number $n = 1$ and the filter length $L = 40$ are set in advance. Fig. 23 shows the result of FMD. The method not only is proved to be immune to the random impulse, but has excellent visual inspection ability for the fault feature. The clear f_1 and its

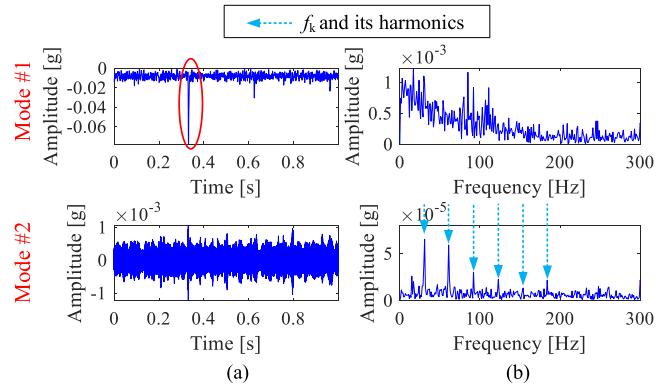


Fig. 24. (a) Decomposed modes of VMD and (b) its envelope spectra in experiment A.

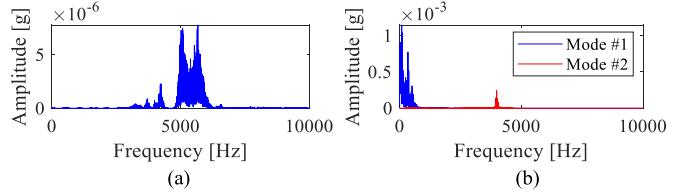


Fig. 25. Frequency spectra of (a) decomposed modes of FMD and (b) VMD in experiment A.

multiples as well as the side frequencies indicate the fault on the inner race of the bearing.

Second, VMD is used to test the same signal. It is difficult to adjust the suitable mode number and balance parameter for the feature extraction of VMD. The mode with the single impulse is preferentially extracted. Even so, the fault information cannot be detected from the other modes. To show the effectiveness, the result of VMD with two modes is presented in Fig. 24. Mode #1 is the random impulse. From the envelope spectrum, mode #2 is the gear meshing information of the gearboxes with frequency $f_k = 31$ Hz. Furthermore, Fig. 25 provides the frequency spectra of decomposed modes of FMD and VMD, from which both the two modes of VMD miss the fault frequency band. Therefore, in this case, VMD can be considered to fail in the fault diagnosis.

B. Wheel Bearing Compound Fault

In this section, the real data from the train wheel bearing with compound fault is applied. Different from experiment A, the fault on the bearing is naturally evolved rather than the artificial. It is found that the naturally evolved defects are usually more difficult to be detected than those seeded faults in the experimental environment [27]. The real experimental setup and schematic diagram of the train wheel safety test bench are plotted in Fig. 26. The working principle of this bench is: motor provides the power to drive the driving wheel. The tested bearing and driving wheel close contact so that the outer race of the bearing rotates with the driving wheel by the friction. An accelerometer is mounted on the end face of the bearing to collect the vibration information with a sampling rate of 76.8 kHz. According to the design parameter of bearing and speed information, the fault character frequency of different parts of the bearing can be

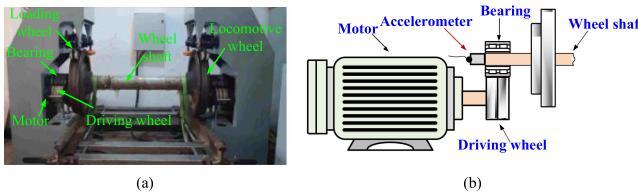


Fig. 26. (a) Real experimental setup and (b) schematic diagram of wheel safety test bench.

TABLE III
FAULT CHARACTER FREQUENCY AND ROTATING FREQUENCY

f_i (Hz)	f_i (Hz)	f_o (Hz)	f_{iT} (Hz)	f_{BS} (Hz)
5.1	50.133	38.562	16.503	2.213

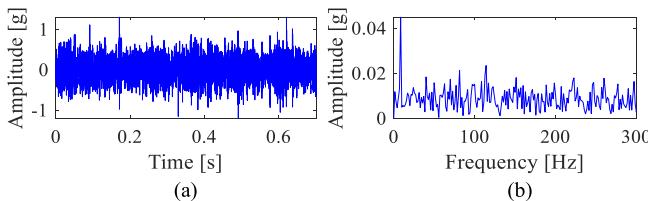


Fig. 27. Raw vibration signal (a) time domain and (b) its frequency spectrum in experiment B.

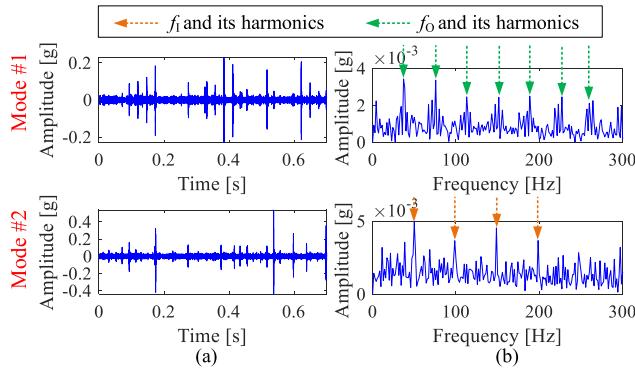


Fig. 28. (a) Decomposed modes of FMD and (b) its envelope spectra in experiment B.

calculated in Table III. It is worth mentioning here the tested bearing is equipped with the rotating outer race rather than the rotating inner race that is different from the common bearing.

Due to the high sampling rate, the data with 0.7 s is applied in this case. Fig. 27 depicts the original signal. It is hard to find any fault information from the time domain from Fig. 27(a). Additionally, from the envelope spectrum in Fig. 27(b), a spectrum line with 8.571 Hz does not have any relationship with the concerned frequencies. Therefore, according to the simple analysis, we can consider the bearing is healthy that will be very dangerous for the train.

Fortunately, the fault can be detected using the proposed method. The mode number $n = 2$ and the filter length $L = 40$ of FMD are set in advance. The decomposed result is presented in Fig. 28. The outer race fault and inner race fault are decomposed into mode #1 and mode #2, respectively. The high SNR in the

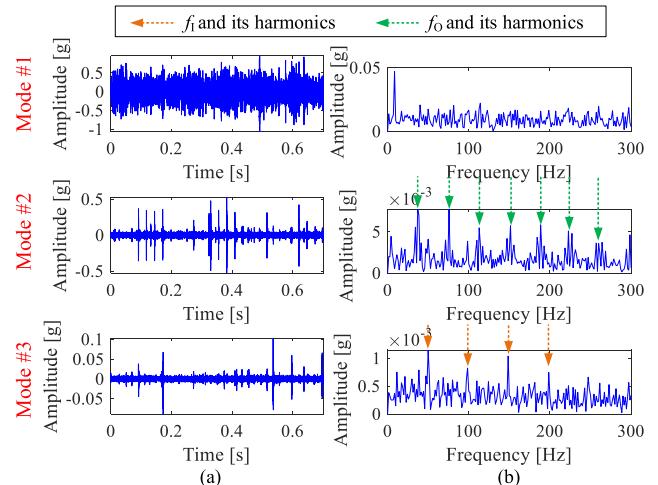


Fig. 29. (a) Decomposed modes of VMD and (b) its envelope spectra in experiment B.

time domain, obvious fault character frequency, and multiples in envelope spectra show the capacity of the proposed FMD.

Fig. 29 plots the result of VMD applied in the same signal. Through several repeated attempts, only the mode number is set as 3, could all fault information be extracted. The result is obtained when the balance parameter is 2000. The outer race fault and inner fault are respectively decomposed into mode #2 and mode #3. Mode #1, which is full of interference and noise, is considered as the redundant mode although the bearing compound fault is identified. It inevitably decreases decomposition performance.

V. DISCUSSION

VMD as the existing state-of-the-art and most popular decomposition method has been widely applied in the field of machinery fault diagnosis. It has been verified that VMD has solved the problem of mode mixing that has troubled most decomposition methods. However, VMD aims at separating the modes with different center frequencies restricted by minimizing the sum of the estimated bandwidth of each mode rather than utilizing the fault feature as the decomposition objective. Therefore, the interference may be preferentially decomposed using VMD since the narrow-band components may contain some interference aside from the fault component in the measured signal. Motivated by this, FMD is tailored to extract machinery fault features, i.e., impulsiveness and periodicity. Benefiting from filter updating and period estimation, FMD can decompose the fault information accurately even facing the challenging compound fault. The results from simulations A and B, as well as experiments A and B, verify the conclusion. Additionally, compared with VMD, the filter of FMD is more adaptive without the restrictions about filter shape and bandwidth of the filter. It means that the more abundant fault information can be extracted. The superiority of FMD can be verified from simulation A and experiment A. More importantly, FMD is much robust against the filter length whose selection brings much trouble for other

methods, such as deconvolution methods. The verification is elaborated in the performance analysis in Section III-C.

The computational efficiency of FMD and VMD is tested. Since their decomposition mode is 2 in simulation B, under this scenario we count the calculation time of the two decomposition methods in computing 20000 time series: VMD 6.17 s and FMD 5.37 s at MATLAB (The MathWorks, Inc., Natick, MA, USA) using Core (TM) I7-8550U @ 1.80 GHz, 16GB RAM.

VI. CONCLUSION

Decomposition methods were considered one of the most effective tools for signal multicomponent analysis. However, traditional decomposition methods encounter different challenges in machinery fault diagnosis. For example, without the complete consideration of the machinery fault feature, i.e., impulsiveness and periodicity of signal, it was difficult for them to accurately decompose the fault information. The performance of decomposition methods based on filtering highly depends on the filter shape and bandwidth of the filter. Motivated by this, a new adaptive decomposition theory, FMD, was proposed in this article. Using the adaptive FIR filter, the component separation was more thorough. Moreover, benefitting from the superiority of CK, FMD takes the impulsiveness and periodicity of signal into consideration simultaneously without the fault period as the prior knowledge. These superiorities were verified by the numerical and experimental data from the bearing fault. The results reveal that the proposed FMD was a more suitable alternative for the machinery signal decomposition analysis than the most popular VMD.

Despite the advantages, improvements in the proposed FMD still need to be investigated in further research. First, the input parameters, such as the mode number n and the segment number K , have a comparatively noticeable effect on the performance and efficiency of FMD. Under the increasingly complex condition, the research about the optimal selection of these parameters would be significant, which might indicate the future research direction. Second, much of the invalid computation was wasted on the redundant iteration in the current FMD. How to decrease the redundant iteration has an important influence on the efficiency and this should be investigated in further studies.

REFERENCES

- [1] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.
- [2] R. B. Randall, *Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications*, New York, NY, USA: Wiley, 2011.
- [3] X. Wang, Y. Zi, and Z. He, "Multiwavelet denoising with improved neighboring coefficients for application on rolling bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 25, no. 1, pp. 285–304, 2011.
- [4] V. C. Leite *et al.*, "Detection of localized bearing faults in induction machines by spectral kurtosis and envelope analysis of stator current," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1855–1865, Mar. 2015.
- [5] K. Zheng, T. Li, Z. Su, and B. Zhang, "Sparse elitist group lasso denoising in frequency domain for bearing fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 17, no. 7, pp. 4681–4691, Jul. 2021.
- [6] Y. Miao, J. Wang, B. Zhang, and H. Li, "Practical framework of Gini index in the application of machinery fault feature extraction," *Mech. Syst. Signal Process.*, vol. 165, 2022, Art. no. 108333.
- [7] J. Lian, Z. Liu, H. Wang, and X. Dong, "Adaptive variational mode decomposition method for signal processing based on mode characteristic," *Mech. Syst. Signal Process.*, vol. 107, pp. 53–77, 2018.
- [8] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. London A, Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.
- [9] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 35, no. 1–2, pp. 108–126, 2013.
- [10] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adaptive Data Anal.*, vol. 1, no. 01, pp. 1–41, 2009.
- [11] J. S. Smith, "The local mean decomposition and its application to EEG perception data," *J. Roy. Soc. Interface*, vol. 2, no. 5, pp. 443–454, 2005.
- [12] Y. Li, M. Xu, X. Liang, and W. Huang, "Application of bandwidth EMD and adaptive multi-scale Morphology analysis for incipient fault diagnosis of rolling bearings," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6506–6517, Aug. 2017.
- [13] Z. Liu, M. J. Zuo, Y. Jin, D. Pan, and Y. Qin, "Improved local mean decomposition for modulation information mining and its application to machinery fault diagnosis," *J. Sound Vib.*, vol. 397, pp. 266–281, 2017.
- [14] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, Aug. 2013.
- [15] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
- [16] J. Li, X. Yao, H. Wang, and J. Zhang, "Periodic impulses extraction based on improved adaptive VMD and sparse code shrinkage denoising and its application in rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 126, pp. 568–589, 2019.
- [17] X. Jiang, C. Shen, J. Shi, and Z. Zhu, "Initial center frequency-guided VMD for fault diagnosis of rotating machines," *J. Sound Vib.*, vol. 435, pp. 36–55, 2018.
- [18] Y. Huang, J. Lin, Z. Liu, and W. Wu, "A modified scale-space guiding variational mode decomposition for high-speed railway bearing fault diagnosis," *J. Sound Vib.*, vol. 444, pp. 216–234, 2019.
- [19] G. L. McDonald, Q. Zhao, and M. J. Zuo, "Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection," *Mech. Syst. Signal Process.*, vol. 33, pp. 237–255, 2012.
- [20] Y. Miao, M. Zhao, K. Liang, and J. Lin, "Application of an improved MCKDA for fault detection of wind turbine gear based on encoder signal," *Renew. Energy*, vol. 151, pp. 192–203, 2020.
- [21] C. Peeters, J. Antoni, and J. Helsen, "Blind filters based on envelope spectrum sparsity indicators for bearing and gear vibration-based condition monitoring," *Mech. Syst. Signal Process.*, vol. 138, 2020, Art. no. 106556.
- [22] Y. Miao, M. Zhao, J. Lin, and Y. Lei, "Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 92, pp. 173–195, 2017.
- [23] X. Xu, M. Zhao, J. Lin, and Y. Lei, "Envelope harmonic-to-noise ratio for periodic impulses detection and its application to bearing diagnosis," *Measurement*, vol. 91, pp. 385–397, 2016.
- [24] J. Antoni and R. Randall, "The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines," *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 308–331, 2006.
- [25] J. Y. Lee and A. K. Nandi, "Extraction of impacting signals using blind deconvolution," *J. Sound Vib.*, vol. 232, no. 5, pp. 945–962, 2000.
- [26] X. Zhao, P. Wu, and X. Yin, "A quadratic penalty item optimal variational mode decomposition method based on single-objective Salp swarm algorithm," *Mech. Syst. Signal Process.*, vol. 138, 2020, Art. no. 106567.
- [27] M. Zhao, J. Lin, Y. Miao, and X. Xu, "Detection and recovery of fault impulses via improved harmonic product spectrum and its application in defect size estimation of train bearings," *Measurement*, vol. 91, pp. 421–439, 2016.
- [28] P. Kundu, A. K. Darpe, and M. S. Kulkarni, "A correlation coefficient based vibration indicator for detecting natural pitting progression in spur gears," *Mech. Syst. Signal Process.*, vol. 129, pp. 741–763, 2019.
- [29] Y. Miao, M. Zhao, and J. Lin, "Identification of mechanical compound-fault based on the improved parameter-adaptive variational mode decomposition," *ISA Trans.*, vol. 84, pp. 82–95, 2019.
- [30] Y. Miao *et al.*, "A review on the application of blind deconvolution in machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 163, 2022, Art. no. 108202.



Yonghao Miao received the B.S. degree in mechanical engineering and automation from the Wuhan University of Technology, Wuhan, China, in 2013, and the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2018.

He was a Visiting Ph.D. Student with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada, from 2017 to 2018. He is currently an Associate Professor with the School of Reliability and Systems Engineering, Beihang University, Beijing, China. His research interests include mechanical fault diagnosis and prognosis, deep learning, and remaining useful life prediction.



Boyao Zhang received the B.S. degree in engineering mechanics from Tongji University, Shanghai, China, in 2018. He is currently working toward the Ph.D. degree in systems engineering with the School of Reliability and Systems Engineering, Beihang University, Beijing, China.

His current research interests include vibration analysis, cyclo-stationary signal processing for machinery condition monitoring and fault diagnosis.



Chenhui Li received the B.S. degree in air vehicle quality and reliability in 2021 from the Beihang University, Beijing, China, where he is currently working toward the master's degree with the School of Reliability and Systems Engineering.

His research interests include mechanical fault diagnosis and remaining useful life prediction.



Jing Lin received the B.S., M.S., and Ph.D. degrees in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 1993, 1996, and 1999, respectively.

From July 2001 to August 2003, he was a Postdoctoral Researcher and a Research Associate with the University of Alberta, Edmonton, AB, Canada, and the University of Wisconsin-Milwaukee, Milwaukee, WI, USA. From September 2003 to December 2008, he was a Research Scientist with the Institute of Acoustics,

Chinese Academy of Sciences under the sponsorship of the Hundred Talents Program of Chinese Academy of Sciences. From January 2009 to July 2018, he was a Professor with the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China. He is currently the Dean of the School of Reliability and Systems Engineering, Beihang University, Beijing, China. He is also the Changjiang Distinguished Professor with the Ministry of Education of China, Beijing, China. His current research field includes machinery condition monitoring, fault diagnosis and prognosis, vibration analysis, and nonstationary signal processing.

Dr. Lin was the recipient of the Distinguished Young Scholar Funding from the National Natural Science Fund in 2011 and won the State Natural Science Award in 2013.



Dayi Zhang received the B.S. and Ph.D. degrees in energy and power engineering from Beihang University, Beijing, China, in 2001 and 2009, respectively.

From March 2009 to December 2011, he was a Postdoctoral Researcher with the School of Energy and Power Engineering, Beihang University, and he became a permanent Staff Member with the School of Energy and Power Engineering in December 2011, and then promoted as an Associate Professor in August 2014. He has undertaken a variety of studies at both dynamics theory and engineering techniques, resulting in more than 100 peer-reviewed publications. His current research field includes rotor dynamics and vibration control for rotation machines.