Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering) 2023 24(5):404-418
www.jzus.zju.edu.cn; www.springer.com/journal/11582
E-mail: jzus_a@zju.edu.cn

Research Article

https://doi.org/10.1631/jzus.A2200555

Fault diagnosis for gearboxes based on Fourier decomposition method and resonance demodulation

Shuiguang TONG[1,2], Zilong FU[2], Zheming TONG[1,2]✉, Junjie LI[2], Feiyun CONG[1,2]

[1]*State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310027, China*
[2]*School of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China*

Abstract: Condition monitoring and fault diagnosis of gearboxes play an important role in the maintenance of mechanical systems. The vibration signal of gearboxes is characterized by complex spectral structure and strong time variability, which brings challenges to fault feature extraction. To address this issue, a new demodulation technique, based on the Fourier decomposition method and resonance demodulation, is proposed to extract fault-related information. First, the Fourier decomposition method decomposes the vibration signal into Fourier intrinsic band functions (FIBFs) adaptively in the frequency domain. Then, the original signal is segmented into short-time vectors to construct double-row matrices and the maximum singular value ratio method is employed to estimate the resonance frequency. Then, the resonance frequency is used as a criterion to guide the selection of the most relevant FIBF for demodulation analysis. Finally, for the optimal FIBF, envelope demodulation is conducted to identify the fault characteristic frequency. The main contributions are that the proposed method describes how to obtain the resonance frequency effectively and how to select the optimal FIBF after decomposition in order to extract the fault characteristic frequency. Both numerical and experimental studies are conducted to investigate the performance of the proposed method. It is demonstrated that the proposed method can effectively demodulate the fault information from the original signal.

Key words: Fourier decomposition method; Singular value ratio; Resonance frequency; Envelope demodulation; Fault diagnosis

1 Introduction

As one of the most crucial parts of mechanical systems, gearboxes are widely used in the machine tool, automotive, wind turbine, helicopter, and engineering machinery industries (Lei et al., 2014). Gearboxes are crucial to the reliable operation of mechanical systems. Due to the harsh working environment, failures of gearboxes occur occasionally, which may cause shutdown of the machinery and the entire production line and even result in huge economic losses (Lei et al., 2013; Zhang et al., 2020). By analyzing vibration signals from gearboxes, condition monitoring and fault diagnosis can prevent serious accidents and provide warning information in the early stages of failure, ensuring machine safety, reliability, and stability, and minimizing unplanned machine downtime (Wang et al., 2020). Fault detection and diagnosis on gears have been attracting increasing attention and many publications on these studies appear in academic journals and technical reports every year.

It is well known that signal processing lays the foundation for condition monitoring and fault diagnosis (Wang et al., 2019; Shi et al., 2022), and that signal decomposition is an effective method for decomposing signals into several superposed components. In the late 1990s, Huang et al. (1998) put forward empirical mode decomposition (EMD) for decomposing a signal into a series of mono-components, called intrinsic mode functions (IMFs). The IMFs can be separated adaptively in a frequency order from high to low according to the waveform of the raw signal itself. However, EMD has some known shortcomings, including mode mixing between adjacent components, end effects, and overfitting due to cubic spline interpolation (Rato et al., 2008; Feng et al., 2017; Li et al., 2017). To improve the EMD method, enhanced versions such as ensemble empirical mode decomposition

✉ Zheming TONG, tzm@zju.edu.cn
Zheming TONG, https://orcid.org/0000-0003-1129-7439

Received Nov. 23, 2022; Revision accepted Feb. 16, 2023;
Crosschecked Mar. 3, 2023; Online first Apr. 19, 2023

(EEMD) and complete ensemble empirical mode decomposition with adaptive noise, have been proposed to overcome the mode mixing problem and end effects by incorporating white noise into the raw signal and treating the mean of a large ensemble as the true result (Wu and Huang, 2009; Torres et al., 2011). These improved methods take full advantage of the characteristics of white noise to eliminate the background noise contained in the original signal. Furthermore, Yeh et al. (2010) designed a complementary EEMD method by replacing the single noise in each decomposition with a pair of noises which are composed of a positive noise and a negative noise. The new method decreased the number of the ensemble and improved the computational efficiency compared to EEMD.

Recently, Dragomiretskiy and Zosso (2014) designed the variational mode decomposition (VMD) with a complete mathematical theory. This algorithm generalizes the classic Wiener filter to multiple and adaptive bands, and decomposes a complicated multi-component signal into constituent mono-components non-recursively. VMD quickly became another popular signal decomposition method. Researchers took advantage of various optimization algorithms to determine the penalty factor and the number of components adaptively, thus contributing to the development of the theory of VMD. Yi et al. (2016) improved the determination of the penalty factor and number of components by combining VMD with a particle swarm optimization algorithm. To avoid the impact caused by over-decomposition or under-decomposition, Liu et al. (2016) designed a criterion based on detrended fluctuation analysis to determine the number of components in VMD. Zhang et al. (2018) proposed a weighted kurtosis index based on kurtosis and correlation coefficient and took advantage of the grasshopper optimization algorithm to obtain the optimal decomposition parameters. This improved method can determine the parameters in VMD adaptively to achieve satisfactory analysis results. So far, EMD and VMD are still the most representative and widely used signal decomposition methods.

Despite the considerable success of EMD and VMD, the components after decomposition are difficult to interpret in terms of their physical meaning, and it is hard to select the optimal component to conduct fault diagnosis. Drawing inspiration from the EMD algorithms and their filter bank properties, Singh et al. (2017) proposed a novel Fourier decomposition method (FDM) based on the mature Fourier theory to decompose a signal into Fourier intrinsic band functions (FIBFs). To extend the applications of the method, Singh (2018) introduced new formulations of the FDM using the discrete cosine transform and applied it to many real-life non-stationary signals, such as earthquake data, speech signals, and electrocardiogram data. Singhal et al. (2020) combined FDM with a zero-phase filtering to separate baseline wander and power-line interference from the electrocardiogram signal. Inspired by the empirical wavelet transform, Zhou et al. (2022) developed an improved segmentation technique to produce more accurate spectrum segmentation results and proposed the empirical Fourier decomposition method to decompose a signal into mono-components accurately and effectively. In order to shorten convergence time, researchers developed Fourier spectrum bandwidth optimization to reduce the computational time and proposed a variable initialization method to accelerate the convergence process (Deng et al., 2019, 2020). For clarity, the representative signal decomposition methods and their enhanced versions are summarized in Table 1.

We wanted to apply FDM to gearbox fault diagnosis. Different from those low-frequency signals, the vibration signals measured from gearboxes contain many frequency components, and the quantity of FIBFs is large. Hence, selecting a proper FIBF containing rich fault information adaptively is of vital significance. The vibration impacts are thought to excite the structural resonance, and the fault characteristic frequency (FCF) is modulated to the resonance frequency (RF) and its surroundings (Wang, 2001; Wang et al., 2017). Our previous work (Tong et al., 2019, 2020) has also verified the phenomenon. Thus, we will focus on how to determine the RF and pick out the optimal FIBF for fault diagnosis.

As the resonance phenomena are closely related to impulses, extracting the impulse signals is important for the calculation of RF. Since spectral kurtosis (SK) was proposed (Antoni, 2006, 2007; Antoni and Randall, 2006), it has been widely used to locate the impulses and calculate the RF. However, its performance is unsatisfactory when the signal-to-noise ratio is low (Moshrefzadeh and Fasana, 2018; Wang, 2018). Therefore, it is necessary to propose a new method

Table 1 Representative decomposition methods

Method	Reference	Main finding
EMD	Huang et al., 1998	Proposed the original algorithm of EMD to decompose a signal into mono-components adaptively
	Wu and Huang, 2009	Incorporated white noise into the EMD method and took the mean as the final result to overcome the mode mixing problem
	Torres et al., 2011	Added a particular noise at each decomposition stage to reduce noise contained in the decomposition result
	Yeh et al., 2010	Replaced a single noise with a pair and reduced the residual noise in the result
VMD	Dragomiretskiy and Zosso, 2014	Put forward the original algorithm of VMD with complete mathematical theory
	Yi et al., 2016	Used particle warm optimization algorithm to determine the penalty factor and number of components
	Zhang et al., 2018	Combined VMD with grasshopper optimization algorithm to determine the parameters adaptively
	Liu et al., 2016	Took advantage of detrended fluctuation analysis to design a criterion for determination of the number of components
FDM	Singh et al., 2017	Proposed the original algorithm of FDM to decompose a signal according to the definition of mono-component
	Singhal et al., 2020	Combined the FDM with a zero-phase filtering to remove power-line interference and baseline wander
	Zhou et al., 2022	Predefined the number of components and introduced a zero-phase filter bank to improve the accuracy of the decomposition result
	Deng et al., 2020	Proposed a new bandwidth estimation and an improved variable initialization strategy based on spectral energy distribution

for determining the RF. For time-domain signals, singular value decomposition (SVD) performs well in feature extraction and decomposition. Zhao and Ye (2011) put a forward-difference spectrum of singular values and applied the method to successfully separate the modulation signal caused by the fault gear from the complicated turning force signal. Li et al. (2019) studied the distribution of singular values with different frequency components. Zhao and Jia (2017) introduced the periodic modulation intensity index to select the decomposed components for the reconstruction of the denoised signal. Yang et al. (2023) proposed an SVD-based strategy to extract the crack-induced impulses and applied it to detect multiple tooth cracks successfully. Based on SVD, Cong et al. (2013) improved the construction method of short-time vector series and proposed the singular value ratio (SVR) spectrum to identify the local impulses in rolling bearing vibration signals. To locate the RF, we propose the maximum singular value ratios (MSVRs) of different windows and extract the impulse signals to calculate the RF. With the help of precise RF, one can pick out the FIBFs containing rich fault information and analyze the envelope spectra for fault diagnosis.

The rest of this paper is arranged as follows: In Section 2, the theory and principles of FDM are briefly reviewed. In Section 3, the proposed MSVR is introduced for the determination of RF, and an algorithm combining FDM and MSVR is described for the fault diagnosis of the gearbox. In Section 4, the performance of FDM-MSVR is illustrated by numerical simulation. In Section 5, the proposed method is applied to the analysis of gear vibration signals measured from a test rig. In Section 6, conclusions are summarized and some discussion on future work is presented.

2 Discrete Fourier decomposition method

Usually, the signals analyzed on computers are in discrete form, so FDM is introduced in discrete form with the help of the discrete Fourier transform. Assuming that $\boldsymbol{x}[n]$ is a discrete signal, it can be expressed as

$$\boldsymbol{x}[n]=\sum_{k=0}^{N-1}\boldsymbol{X}[k]\exp\left(\mathrm{j}\frac{2\pi n}{N}k\right),\quad 0\leqslant k\leqslant N-1, \tag{1}$$

where N is the length of the signal, $\boldsymbol{X}[k]$ is the discrete Fourier transform of signal $\boldsymbol{x}[n]$ and its formulation is as follows:

$$\boldsymbol{X}[k]=\frac{1}{N}\sum_{n=0}^{N-1}\boldsymbol{x}[n]\exp\left(-\mathrm{j}\frac{2\pi k}{N}n\right),\ 0\leqslant n\leqslant N-1. \quad (2)$$

Without loss of generality, N is assumed to be an even number. In this case, $\boldsymbol{X}[0]$ and $\boldsymbol{X}[N/2]$ can be simplified, and both of them are real values.

$$\boldsymbol{X}[0]=\frac{1}{N}\sum_{n=0}^{N-1}\boldsymbol{x}[n],\ \boldsymbol{X}\left[\frac{N}{2}\right]=\frac{1}{N}\sum_{n=0}^{N-1}\boldsymbol{x}[n]\cdot(-1)^n. \quad (3)$$

Further, Eq. (1) can be expanded as

$$\boldsymbol{x}[n]=\boldsymbol{X}[0]+\sum_{k=1}^{\frac{N}{2}-1}\boldsymbol{X}[k]\exp\left(\mathrm{j}\frac{2\pi n}{N}k\right)+\boldsymbol{X}\left[\frac{N}{2}\right]\cdot(-1)^n+\sum_{k=\frac{N}{2}+1}^{N-1}\boldsymbol{X}[k]\exp\left(\mathrm{j}\frac{2\pi n}{N}k\right). \quad (4)$$

As $\boldsymbol{x}[n]$ is a real-value signal, the second and fourth parts must be complex conjugate pairs, which are denoted by $\boldsymbol{z}[n]$ and $\boldsymbol{z}^*[n]$, respectively. Then, $\boldsymbol{x}[n]$ can be rewritten in a more concise form:

$$\boldsymbol{x}[n]=\boldsymbol{X}[0]+2\mathrm{Re}[\boldsymbol{z}(n)]+\boldsymbol{X}\left[\frac{N}{2}\right]\cdot(-1)^n,\boldsymbol{z}[n]=\sum_{k=1}^{\frac{N}{2}-1}\boldsymbol{X}[k]\exp\left(\mathrm{j}\frac{2\pi n}{N}k\right), \quad (5)$$

where $\mathrm{Re}(\boldsymbol{z}[n])$ denotes the real part of an analytic signal $\boldsymbol{z}[n]$.

$\boldsymbol{z}[n]$ can be regarded as the sum of P analytic signals, which are analytic forms of FIBFs. It is obvious that such a decomposition always exists. The expression of $\boldsymbol{z}[n]$ is

$$\sum_{k=1}^{\frac{N}{2}-1}\boldsymbol{X}[k]\exp\left(\mathrm{j}\frac{2\pi n}{N}k\right)=\sum_{m=1}^{P}\boldsymbol{A}_m[n]\exp(\mathrm{j}\cdot\boldsymbol{\varphi}_m[n]), \quad (6)$$

where $\boldsymbol{A}_m[n]$ and $\boldsymbol{\varphi}_m[n]$ are the instantaneous amplitude and instantaneous phase of the mth analytic signal, respectively.

The mth analytic signal $\boldsymbol{z}_m$ can be represented by

$$\boldsymbol{z}_m=\boldsymbol{A}_m[n]\exp(\mathrm{j}\cdot\boldsymbol{\varphi}_m[n])=\sum_{k=N_{m-1}+1}^{N_m}\boldsymbol{X}[k]\exp\left(\mathrm{j}\frac{2\pi n}{N}k\right), \quad (7)$$

where N_m is the end index of component $\boldsymbol{z}_m$ and $N_0=0<\cdots<N_{m-1}<N_m<\cdots<N_P<N/2-1$. In this way, the analytic signal can be reconstructed as

$$\boldsymbol{z}[n]\doteq\begin{bmatrix}\boldsymbol{z}_1\\ \vdots\\ \boldsymbol{z}_m\\ \vdots\\ \boldsymbol{z}_P\end{bmatrix}=\begin{bmatrix}\sum_{k=1}^{N_1}\boldsymbol{X}[k]\exp\left(\mathrm{j}\frac{2\pi n}{N}k\right)\\ \vdots\\ \sum_{k=N_{m-1}+1}^{N_m}\boldsymbol{X}[k]\exp\left(\mathrm{j}\frac{2\pi n}{N}k\right)\\ \vdots\\ \sum_{k=N_{P-1}+1}^{N_P}\boldsymbol{X}[k]\exp\left(\mathrm{j}\frac{2\pi n}{N}k\right)\end{bmatrix}, \quad (8)$$

where the symbol $\doteq$ means that the left can be represented by the right. In order to obtain FIBFs, we can take the real parts of analytic FIBFs.

$$\boldsymbol{x}[n]\doteq\begin{bmatrix}\boldsymbol{x}_1\\ \vdots\\ \boldsymbol{x}_m\\ \vdots\\ \boldsymbol{x}_P\end{bmatrix}=\begin{bmatrix}\mathrm{Re}[\boldsymbol{z}_1]\\ \vdots\\ \mathrm{Re}[\boldsymbol{z}_m]\\ \vdots\\ \mathrm{Re}[\boldsymbol{z}_P]\end{bmatrix}. \quad (9)$$

For the FIBFs, they should satisfy the following conditions:

(1) Each FIBF $\boldsymbol{x}_m$ is a zero-mean function, i.e., $\sum\boldsymbol{x}_m=0$.

(2) The FIBFs are orthogonal functions, i.e., $\boldsymbol{x}_p\cdot\boldsymbol{x}_q=0$ ($p\neq q$).

(3) The instantaneous frequency (IF) $\boldsymbol{f}_m[n]$ and instantaneous amplitude $\boldsymbol{A}_m[n]$ of every FIBF $\boldsymbol{x}_m$ should be non-negative for each n, i.e., $\boldsymbol{f}_m[n]=(\boldsymbol{\varphi}[n]-\boldsymbol{\varphi}[n-1])/(2\pi)\geqslant 0$, $\forall n$ and $\boldsymbol{A}_m[n]=|\boldsymbol{z}_m[n]|\geqslant 0$, $\forall n$.

Theoretically, the FIBFs are mono-component signals, as IF is defined for the signal that has a single frequency or a narrow range of frequencies. The physical meaning of IF constraints is that IF $\boldsymbol{f}_m[n]$ must be positive or zero and that phase $\boldsymbol{\varphi}[n]$ is an increasing or non-decreasing function of n. To minimize the number of components, the end index N_m of the mth analytic FIBF should be as large as possible, provided that the FIBF $\boldsymbol{x}_m$ satisfies all the conditions.

3 Fault diagnosis based on FDM and resonance demodulation

3.1 Resonance frequency curve based on MSVR

In order to analyze the detailed information and locate the impulse in the original signal, a short-time vector series is obtained through intercepting the segmentation with length N_w (even number) along the time sequence. After segmentation of the original signal, the short-time vector series can be expressed as $[\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \cdots, \boldsymbol{x}_m^T]^T$. The number of vectors is calculated by

$$m = \text{floor}\left[\frac{N - N_w}{p} + 1\right], \tag{10}$$

where N_w is the length of each segmentation, p is the moving step length, and floor [·] denotes the largest integer not greater than ·. In this paper, we take p=1 for an accurate determination of RF. Each vector can be transformed into a double-row matrix $\boldsymbol{D}_m$ as illustrated in Fig. 1. According to SVD theory, $\boldsymbol{D}_m$ has only two singular values σ_1 and σ_2 ($\sigma_1 > \sigma_2$), which reflect the relationship between the two rows (Zhao and Ye, 2009, 2011; Cong et al., 2013). However, the singular values are closely related to the energy of the corresponding vectors (Zhao and Jia, 2017; Yang et al., 2023). The larger the length N_w of the short-time vectors, the larger the singular values obtained. Hence, to eliminate the effect of length N_w on the magnitude of singular values, SVR is introduced to represent the correlation between these two singular values, and the equation is

$$R = \frac{\sigma_1}{\sigma_2},\ \sigma_1 > \sigma_2\ . \tag{11}$$

As each short-time vector corresponds to an SVR, the sequence of SVRs is given by **SR**=[R_1, R_2, ⋯, R_m]. The maximum value R_k indicates that the front segment and back segment of the short-time vector $\boldsymbol{x}_k$ have the least similarity, which means that one is a noise signal and the other is a vibration signal. When a failure occurs in the gearbox (for example, a tooth crack), impulses will periodically appear in the vibration signal. For a proper N_w, there are two ideal vectors for intercepting the segmentation along the time sequence as illustrated in Fig. 2, and they are equivalent to each other in theory. Faults, in general, will excite gear resonance contained in impulses, and FCF is modulated to the RF and its surroundings.

The length of the short-time vector N_w is the key parameter for analysis. On the one hand, if N_w is too small, it is probable that the nonlinear relationship between two rows in the constructed matrix will be misidentified as a linear one, resulting in false peaks in SVRs. On the other hand, if N_w is too large, the vectors contain too much noise, and the capability of local identification will decrease rapidly. According to our previous work (Cong et al., 2013), two conditions must be satisfied to determine a range for the selection of N_w: one is that the vector should contain at least a quasi-periodic oscillation and a noise with the same length, and the other is that the length of the vector should be smaller than the spacing between two successive impulses. These two conditions give lower and upper limits for the selection of N_w. The amount of data in a quasi-period oscillation can be calculated by F_s/f_n, where F_s is the sampling frequency and f_n is the RF. The spacing between two successive impulses can be calculated by F_s/f_c, where f_c is the largest FCF to ensure the smallest spacing. To sum up, the length of short-time vectors N_w can be determined by the following equations:

$$\begin{cases} (N_w)_{\min} \leqslant N_w \leqslant (N_w)_{\max}, \\ (N_w)_{\min} = 2 \cdot \dfrac{F_s}{f_n}, \\ (N_w)_{\max} = \dfrac{F_s}{f_c}, \end{cases} \tag{12}$$

herein N_w is an even number.

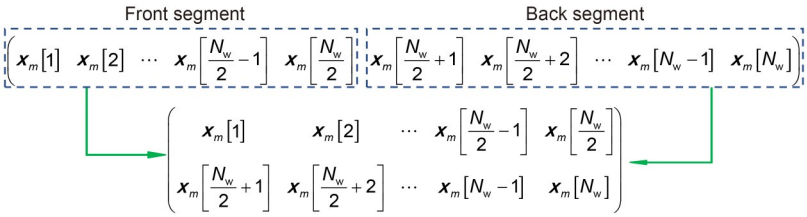

Fig. 1 Construction of double-row matrix

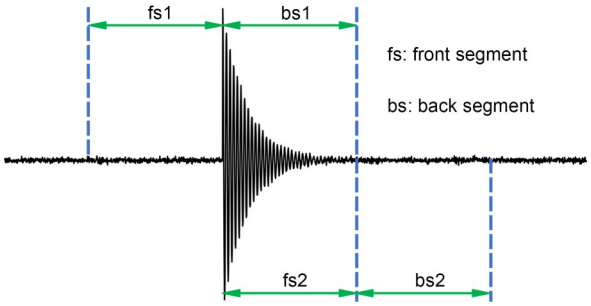

Fig. 2 Two ideal vectors with proper N_w

For a specific length N_w, a series of vectors and SVRs will be obtained. The vector with maximum SVR is regarded as the ideal vector. Then the first half (for example, fs2 in Fig. 2) or the second half (for example, bs1 in Fig. 2) of the ideal vector is taken out to calculate its amplitude spectrum, and the frequency with the maximum amplitude is regarded as the RF. Similarly, we can calculate an RF for each length N_w. To be clear, it is necessary to plot the RF curve with different N_w, and the relationship between RF and N_w is shown in Fig. 3. It is obvious that when N_w is smaller, the calculated RF is unstable. In reverse, when N_w is larger, the calculated RF oscillates around the true value. It can be seen from Fig. 3 that in the late stage, RF oscillates around its true value of 2000 Hz. It is reasonable to regard the average of the last three oscillations as the true value of RF.

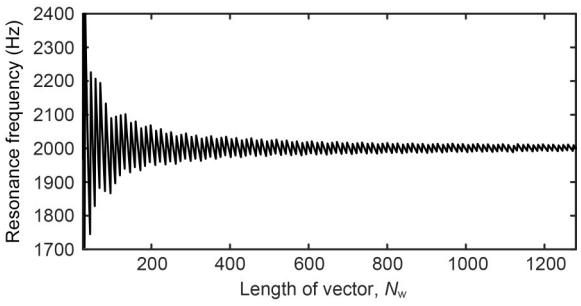

Fig. 3 Resonance frequency curve

3.2 Procedure of the joint FDM and MSVR

For these practical vibration signals measured from rolling bearings and gearboxes, amplitude modulation and frequency modulation effects resulting from faults, assembly errors, and random speed fluctuations make frequencies very complex. FDM can decompose these complex signals into a series of band-limited FIBFs and guarantee that there is no mode mixing problem between different FIBFs. Although FDM decomposes these signals into simpler mono-components, the number of components is so large that we cannot analyze the results directly, and they need to be further processed. Selecting representative components containing rich fault information is of great importance and essential for fault diagnosis. According to previous studies (Wang et al., 2017; Tong et al., 2020), we believe that FCF is modulated to RF and its harmonics when there are impulses in the vibration signal. Hence, we propose the MSVR method to calculate the RF. Then we can select the FIBF whose mean frequency is closest to the RF and utilize demodulation through the envelope spectrum which extracts FCF from sidebands around the RF. The whole procedure of the demodulation method based on FDM-MSVR is summarized in Fig. 4.

4 Numerical investigation

4.1 Simulated signal

The synthetic signal contains three parts (Wang, 2001; Man et al., 2012; Yang et al., 2022): the gear meshing part $\boldsymbol{x}_m(t)$, the gear fault part $\boldsymbol{x}_c(t)$ and a white Gaussian noise $\boldsymbol{n}(t)$ at a signal-to-noise ratio of −3 dB. Consequently, the synthetic signal is expressed as

$$\boldsymbol{x}(t)=\boldsymbol{x}_m(t)+\boldsymbol{x}_c(t)+\boldsymbol{n}(t). \tag{13}$$

The gear meshing part can be presented as

$$\begin{aligned}\boldsymbol{x}_m(t)=&\left[1+A_m\cos(2\pi f_1 t)\right]\cdot\cos\left[2\pi f_m t+B_m\sin(2\pi f_1 t)\right]+\\&\left[1+A_m\cos(2\pi f_2 t)\right]\cdot\cos\left[2\pi f_m t+B_m\sin(2\pi f_2 t)\right]+\\&\sum_{n=1}^{f_m}D_m e^{-\beta(t-nT_m)}\sin\left[2\pi f_{n_1}(t-nT_m)\right].\end{aligned} \tag{14}$$

For the gear fault part, it mainly reflects the amplitude modulation-frequency modulation (AM-FM) phenomenon and impulses caused by the gear fault. The model can be expressed as

$$\begin{aligned}\boldsymbol{x}_c(t)=&\left[1+A_c\cos(2\pi f_c t)\right]\cdot\\&\cos\left[2\pi f_m t+B_c\sin(2\pi f_c t)\right]+\\&\sum_{i=1}^{f_c}D_c e^{-\beta(t-iT_c)}\sin\left[2\pi f_{n_2}(t-iT_c)\right].\end{aligned} \tag{15}$$

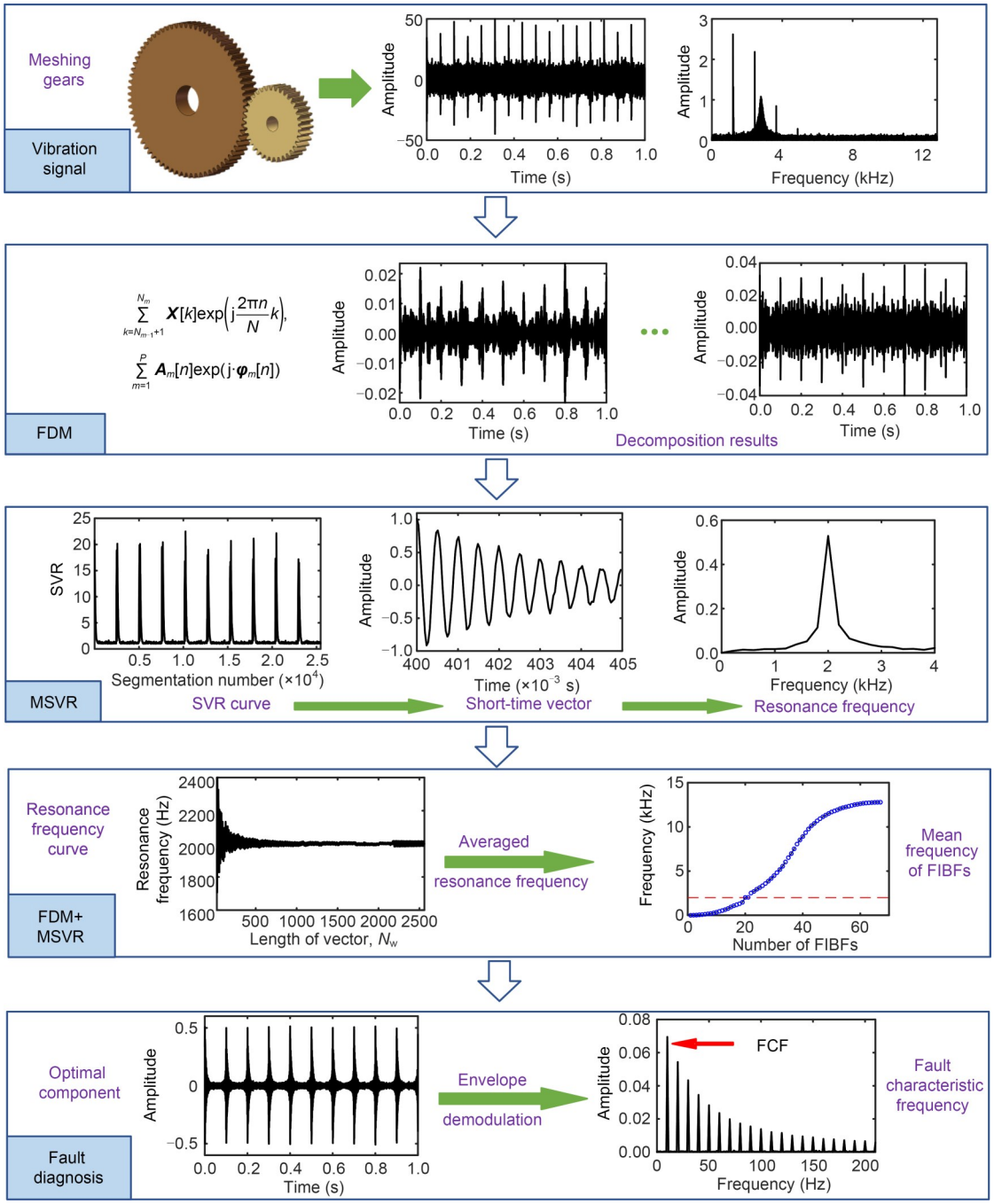

Fig. 4 Demodulation method based on FDM-MSVR

The meaning and values of all the parameters are listed in Table 2.

The waveform and spectrum of the simulated synthetic signal are shown in Fig. 5. Although the frequency spacing between peaks in the envelope spectrum can indicate FCF, the harmonics of FCF are also dominant frequencies, and their amplitudes are almost as high as that of FCF. Considering the fact that the amplitudes and decaying rates of impulses are various and that the envelope spectrum reveals the dominant modulation frequency in the overall frequency band, the relationship between FCF and its harmonics is inconsistent with the known law that the amplitude of harmonics decreases as the order increases. The envelope spectrum demodulation is incomplete, and the distribution of harmonic amplitudes may transfer

Table 2 Definitions of the parameters

Parameter	Value
Magnitude of the AM due to gear meshing, A_m	0.5
Magnitude of the FM due to gear meshing, B_m	0.8
Amplitude of the impulses due to gear meshing, D_m	4
Period of gear meshing (T_m=1/f_m), T_m (s)	0.0008333
Meshing frequency, f_m (Hz)	1200
Rotational frequency of input shaft, f_1 (Hz)	33.3
Rotational frequency of output shaft, f_2 (Hz)	16
Resonance frequency caused by gear meshing, f_{n_1} (Hz)	2800
Structural resonance damping coefficient, β	600
Magnitude of the AM due to gear fault, A_c	0.5
Magnitude of the FM due to gear fault, B_c	0.8
Amplitude of the impulses due to gear fault, D_c	40
Period of impulses caused by gear fault, T_c (T_c=1/f_c) (s)	0.0625
Fault characteristic frequency, f_c (Hz)	16
Resonance frequency caused by gear fault, f_{n_2} (Hz)	2800

incorrect information, resulting in failure in the gearbox fault diagnosis.

4.2 Signal decomposition and envelope spectrum analysis

Compared to EMD, EEMD has a better effect in terms of removing mode mixing and reducing noise. In this section, the ratio of the standard deviation of the added noise to that of the original signal is set to 0.2, and the ensemble number is set to 50. In total, 14 IMFs are obtained through EEMD. The representative component, IMF2, is shown in Fig. 6. Although there is less residual noise left in the component after decomposition, the distribution and relationship between harmonics in the envelope spectrum are no better than that of the original signal. VMD is another widely used decomposition method. There are five components in the result, among which IMF3 is the representative component and is illustrated in Fig. 7. Compared to IMF2 in the EEMD method, IMF3 in the VMD method contains less noise and has a higher amplitude. However, the structure of the envelope spectrum in VMD is similar to that in EEMD.

Similar to EEMD and VMD, FDM is also a decomposition method. After decomposition by FDM, there are 69 FIBFs, and each FIBF is an independent component. To perform envelope demodulation, it is necessary to select the key component with the help of RF by MSVR. First of all, the range of vector length N_w needs to be determined, i.e., the calculation of the lower limit $(N_w)_{min}$ and the upper limit $(N_w)_{max}$ by Eq. (12). The FCF and the RF are estimated at $\dot{f}_c$=34 and $\dot{f}_n$=3000 Hz, respectively. As a result, the range of the vector length is $10 \leqslant N_w \leqslant 750$. It is worth mentioning that the initial estimated value of RF $\dot{f}_n$ has little influence on the result of RF because it determines the lower limit of N_w, and that the FCF $\dot{f}_c$ should be estimated as large as possible to ensure that there is only one impulse at most included in each vector.

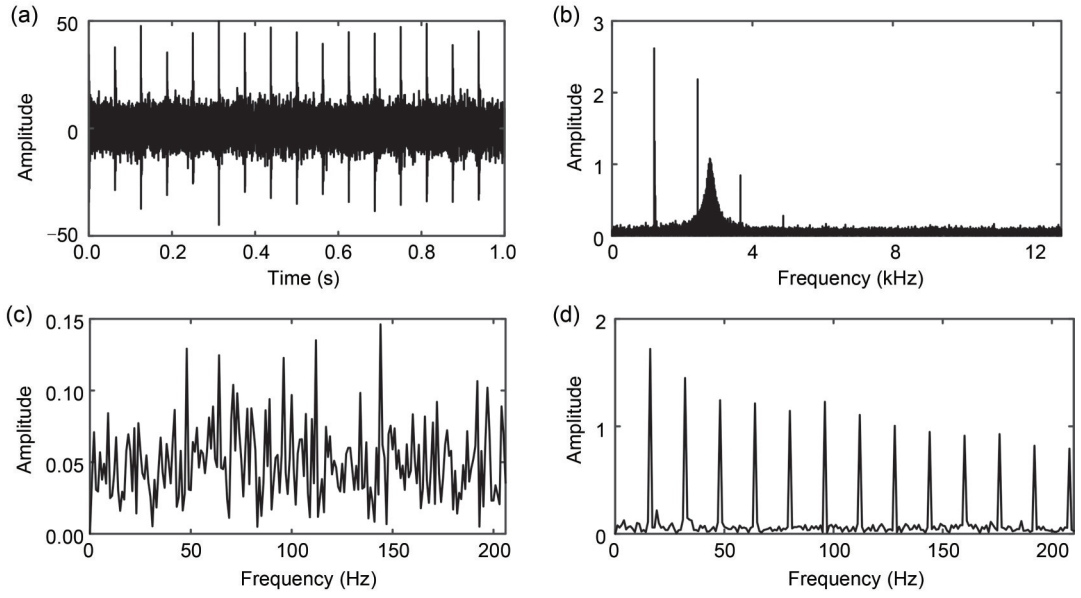

Fig. 5 Waveform and spectrum of original signal: (a) original signal; (b) amplitude spectrum; (c) spectrum in [0, 200] Hz; (d) envelope spectrum

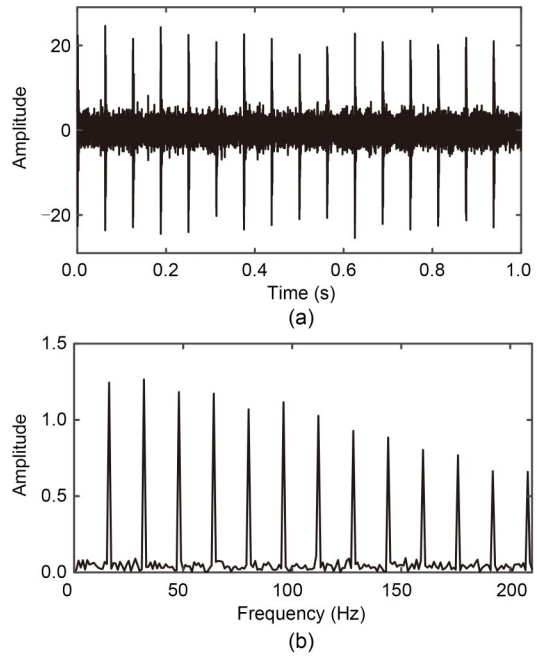

Fig. 6 Representative IMF (IMF2) in EEMD: (a) waveform; (b) amplitude spectrum

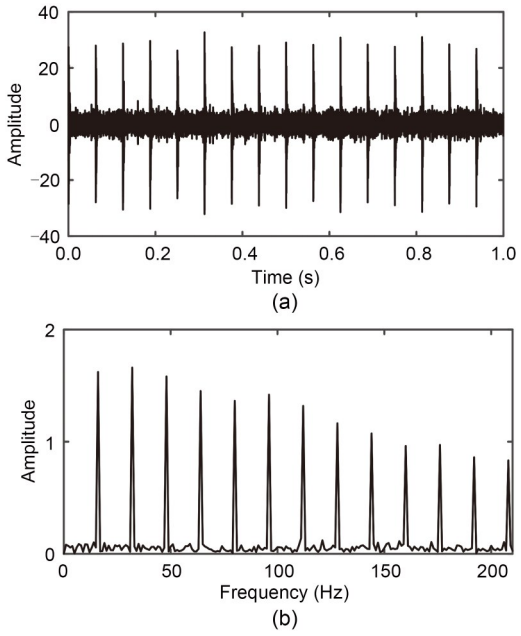

Fig. 7 Representative IMF (IMF3) in VMD: (a) waveform; (b) amplitude spectrum

The RF curve is plotted in Fig. 8a. In the initial stage, the estimated RF oscillates so violently that we cannot take the estimated value as an approximation of the true value. Fortunately, in the later stages, the estimated RF reaches convergence. Therefore, it is reasonable to calculate the average of the estimated values in the last three quasi-periodic oscillations and regard it as the true RF. Consequently, the final estimated RF is 2791 Hz. For comparison, the Kurtogram by SK is also presented in Fig. 8b, and it indicates that the RF is 2933 Hz and the bandwidth (B_w) is 533 Hz. Hence, the RF generated by MSVR is closer to the true value of 2800 Hz. To select the optimal FIBFs, the mean frequencies of FIBFs and the estimated RFs are plotted on the same graph in Fig. 9. We can see that the 21st and 22nd FIBFs are the optimal components for MSVR and SK, respectively.

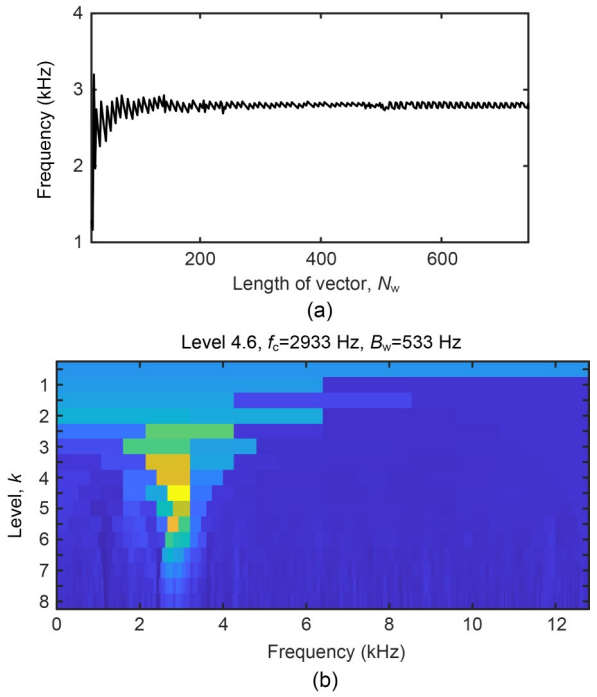

Fig. 8 Resonance frequency and mean frequency curve: (a) resonance frequency curve; (b) Kurtogram

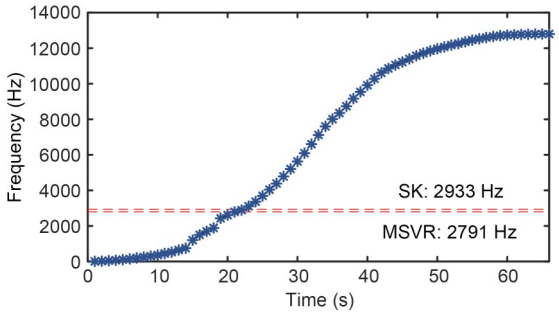

Fig. 9 Mean frequencies of FIBFs

The envelope spectra of optimal FIBFs are shown in Fig. 10, from which one can see that the frequency of 16 Hz at the largest peak is the FCF. It is obvious

that the envelope spectra of optimal FIBFs perform better than those of EEMD and VMD. Not only does the largest peak in the envelope spectrum have a higher amplitude, but the relationship between harmonics is consistent with known laws, i.e., the amplitude of harmonics decreases rapidly as the harmonics order increases. Different from the envelope spectrum in EEMD and VMD, there is only one dominant peak in the envelope spectrum of optimal FIBF in the proposed method. It is apparent in Fig. 10 that the first peak or dominant peak at 16 Hz is the FCF. In addition, the amplitude of the envelope spectrum of the 21st FIBF is larger than that of the 22nd FIBF, indicating that the FIBF selected by MSVR is a little better than that selected by SK.

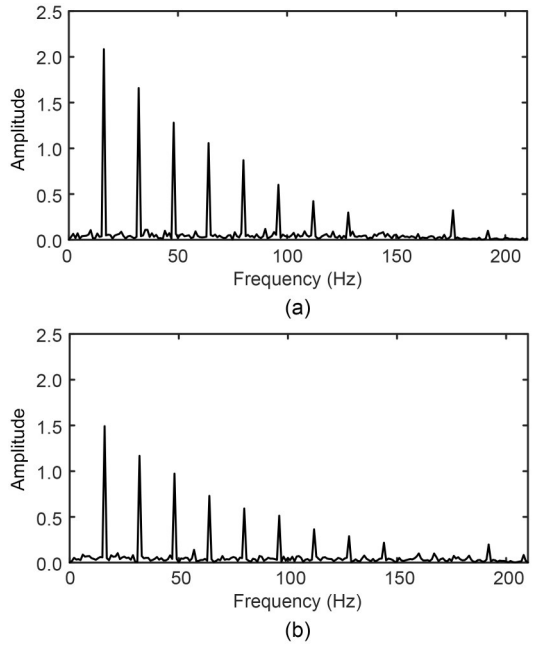

Fig. 10 Optimal FIBFs: (a) envelope spectrum of the 21st FIBF; (b) envelope spectrum of the 22nd FIBF

5 Experimental validation

In this section, we conduct the experiment using a gearbox with a cracked gear and collect the vibration signals from an accelerometer installed on the gearbox casing. To validate the performance of the proposed FDM-MSVR method, EEMD and VMD are also adopted to analyze the experimental signal.

5.1 Experimental setting

Fig. 11 shows the experimental equipment. The test rig consists of a motor, a torque-speed transducer, a gearbox, a magnetic powder brake, and three couplings, as illustrated in Fig. 11a. In the experiment, we created a crack in one tooth root of the gear to mimic gearbox damage, and a picture of the damaged gear is

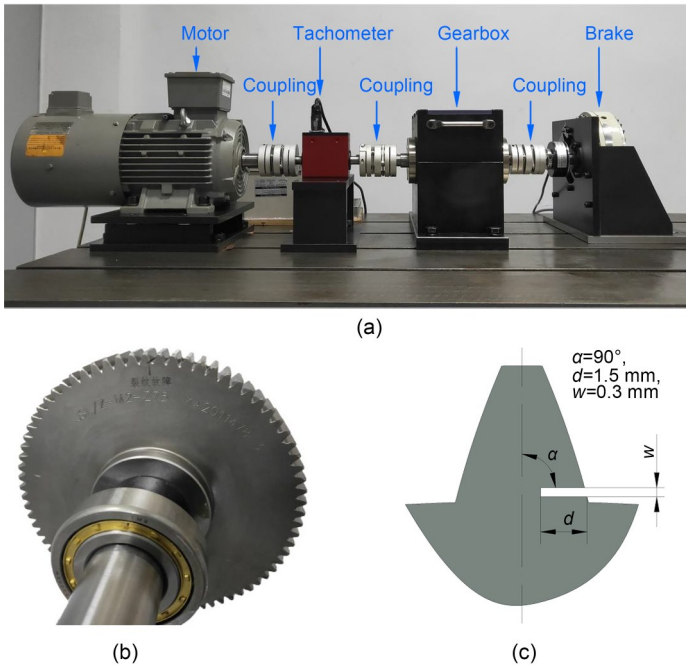

Fig. 11 Experimental equipment: (a) gearbox test rig; (b) cracked gear; (c) parameters of the tooth crack

shown in Fig. 11b. The parameters of the tooth crack are illustrated in Fig. 11c, where α is the angle, d is the depth of the crack, and w is the width of the crack; the length of the crack is equal to the width of the tooth. An accelerometer is mounted on the top of the gearbox casing to collect the vibration signals, and the sampling frequency is 25600 Hz. In the experiment, the speed of the motor is set to 2500 r/min, and the parameters of the gearbox and characteristic frequencies are listed in Table 3.

Table 3 Gearbox parameters and characteristic frequencies

Modulus	Number of pinion teeth	Number of gear teeth
2	36	75
FCF of pinion	FCF of gear	Meshing frequency
41.67 Hz	20 Hz	1500 Hz

5.2 Signal analysis

Fig. 12 presents the waveform and spectrum of the experimental signal. It can be observed that the experimental signal is more complicated than the simulated signal. To find the fault characteristic frequencies, the spectrum in the low-frequency range is enlarged, as shown in Fig. 12c. It is surprising that the FCF of the pinion, 41.67 Hz, is the dominant frequency while the FCF of the gear, 20 Hz, does not arise, which is the opposite of what we expected. This could be due to assembly errors and the rotating frequency of the input shaft being higher than that of the output shaft. More importantly, the gear's FCF is modulated to the RF band. Further, envelope demodulation is conducted to analyze the signal, and the envelope spectrum is shown in Fig. 12d. From the figure, one can see that the largest peak occurs at the frequency of the fourth harmonic, 80 Hz, rather than at the fundamental frequency of 20 Hz, which shows that the demodulation by envelope analysis is not thorough and may give wrong information.

In order to show the performance of the proposed method, EEMD and VMD are also applied to the experimental signal for comparison. In EEMD, the ensemble number and the standard deviation ratio are set to 50 and 0.2, respectively. There are 13 IMFs in total, and the first two are highly informative and occupy most of the energy of the original signal. Consequently, the envelope spectra of the first two components are shown in Fig. 13. IMF1 indicates that 80 Hz is the dominant frequency, while IMF2 shows that both 20 Hz and 80 Hz are the characteristic frequencies. Similarly, the decomposition results by VMD are presented in Fig. 14. IMF2, IMF3, and IMF4 are the major components. IMF3 is excluded for simplicity because its envelope spectrum has a structure that is nearly identical to that of IMF2. In terms of the envelope spectrum, IMF2 shows that 80 Hz stands out,

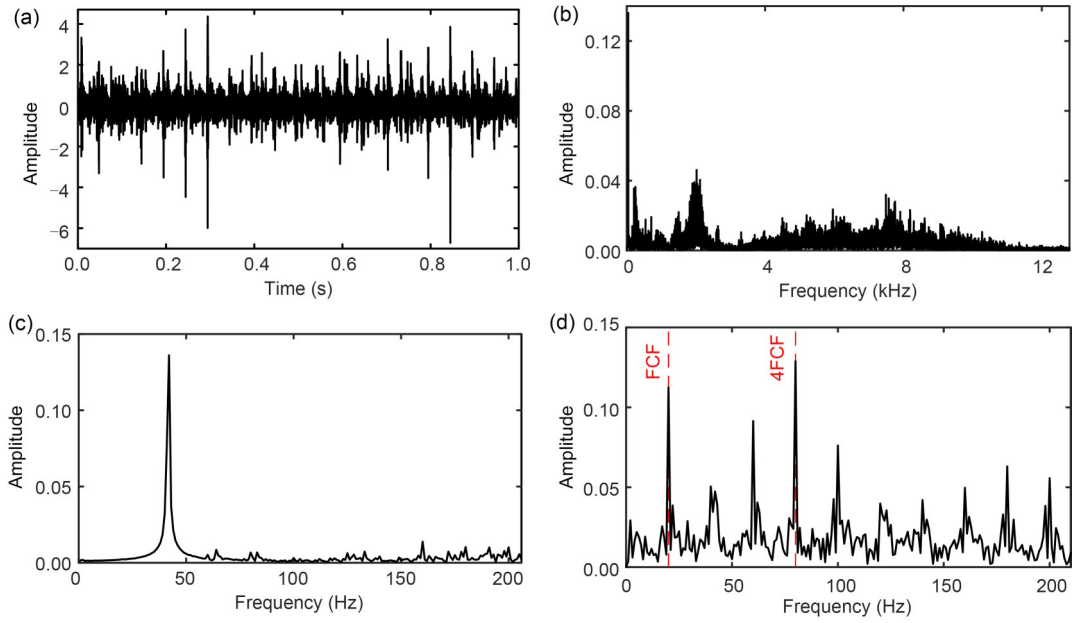

Fig. 12 Experimental signal and its spectrum analysis: (a) original signal; (b) amplitude spectrum; (c) spectrum in [0, 200] Hz; (d) envelope spectrum

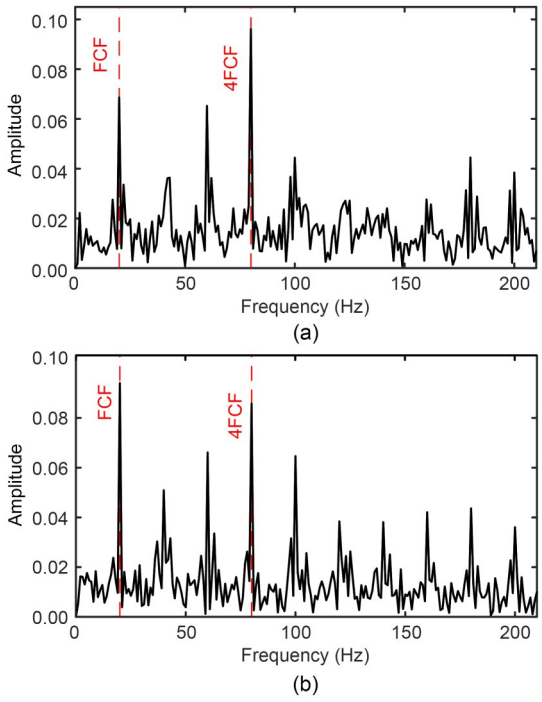

Fig. 13 Decomposed signals by EEMD: (a) envelope spectrum of IMF1; (b) envelope spectrum of IMF2

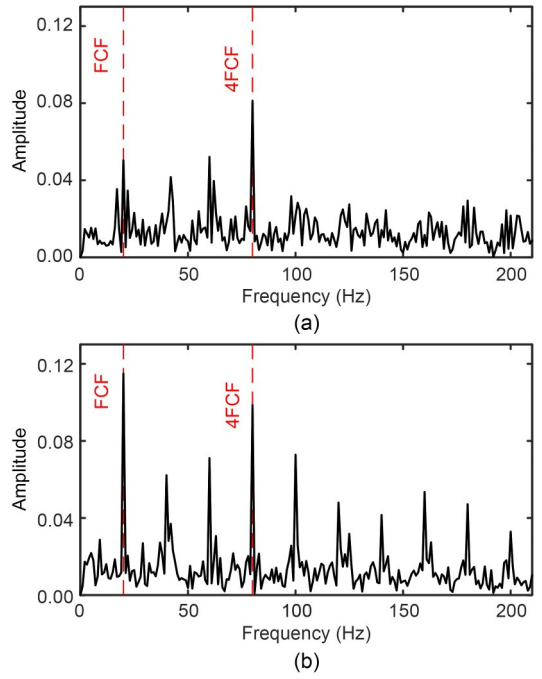

Fig. 14 Decomposed signals by VMD: (a) envelope spectrum of IMF2; (b) envelope spectrum of IMF4

while IMF4 indicates that both 20 Hz and 80 Hz are prominent frequencies. It seems that 80 Hz is the FCF. However, it is inconsistent with any of the theoretical characteristic frequencies.

To analyze the experiment signal more effectively, we conduct FDM and calculate the mean frequency of each FIBF. In total, there are 64 FIBFs after decomposition. To pick out the optimal component containing rich fault information, it is of vital importance to obtain the RF. The upper limit and lower limit of vector length N_w can be calculated by Eq. (12), and the corresponding range is $26 \leqslant N_w \leqslant 608$. Finally, the RF curve is plotted in Fig. 15a. As mentioned above, the mean value (2021 Hz) of the last three quasi-periodic oscillations is taken as the final estimated RF. For comparison, the Kurtogram of SK is shown in Fig. 15b. It is obvious that the center frequency is 4000 Hz at level 4.6. To make it clear, the mean frequencies and the RF are plotted on the same graph shown in Fig. 16. It is apparent that the 17th and 24th FIBFs are the optimal

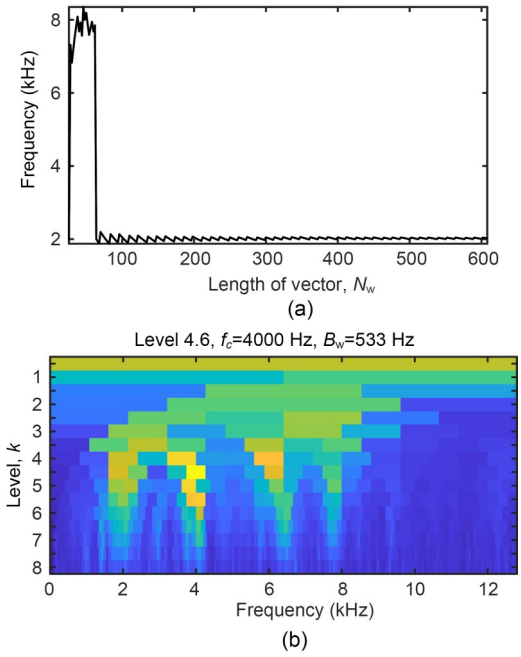

Fig. 15 Resonance frequency by MSVR and spectral kurtosis: (a) resonance frequency curve; (b) Kurtogram

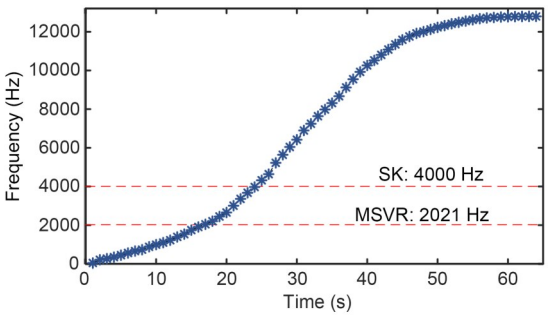

Fig. 16 Mean frequencies of FIBFs

components, according to MSVR and SK, respectively. Fig. 17 shows the envelope spectra of the optimal FIBFs, from which we can see that the frequency of 20 Hz at the largest peak is the FCF. It coincides with the FCF of the pinion, so we can conclude that the pinion is damaged. In Fig. 17a, there is one major peak, which is exactly the FCF, and three minor peaks. However, the first four harmonics of FCF are all major peaks in Fig. 17b, although the first one is a little higher than the others. There is no doubt that the envelope spectrum of the 17th FIBF performs better than that of the 24th FIBF in terms of fault diagnosis. Hence, the proposed MSVR is superior to SK in selecting the optimal component.

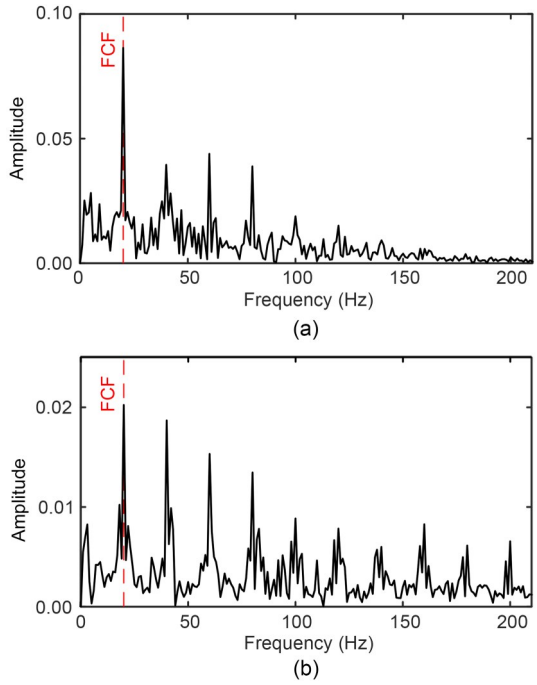

Fig. 17 Optimal FIBFs: (a) envelope spectrum of the 17th FIBF; (b) envelope spectrum of the 24th FIBF

In summary, the proposed FDM-MSVR method performs better than EEMD and VMD. On the one hand, the selection of the optimal component in EEMD and VMD relies on expert knowledge and experience, while the FDM-MSVR decomposes a signal according to its nature, and the optimal FIBF is determined by RF. The optimal FIBF selected by MSVR also performs better than that of SK. On the other hand, the algorithm of FDM-MSVR can select the optimal FIBF automatically; hence, it is well suited for embedding in a condition monitoring system.

6 Conclusions

In this paper, a novel demodulation method named FDM-MSVR is proposed for the determination of the optimal demodulation component. A given signal can be decomposed adaptively into a series of FIBFs by FDM, and the IF of each FIBF is non-negative due to the superposition principle. Characteristic frequencies associated with gear faults (especially tooth cracks and missing teeth) are modulated by the RF band and its surroundings. MSVR is used to estimate the RF in order to select the best FIBF. The joint FDM-MSVR method can detect the optimal FIBF precisely and demodulate the FCF effectively. The performance of FDM-MSVR is compared with the widely used envelope spectra, EEMD and VMD, through a simulated signal, and the results show that FDM-MSVR can accurately locate the optimal component containing rich fault information. The FCF can be successfully identified by the envelope demodulation analysis of the optimal component. In addition, the result by SK is also presented to demonstrate the superiority of the proposed MSVR method. The experimental signal analysis demonstrates its effectiveness and superiority over other decomposition methods. In the future, we are going to investigate how to extend the proposed method to the diagnosis of gearboxes in non-stationary working conditions.

Acknowledgments

This work is supported by the National Key R&D Program of China (No. 2019YFB2004604), the National Natural Science Foundation of China (No. 52075477), and the Key R&D Program of Zhejiang Province (No. 2021C01139), China.

Author contributions

Shuiguang TONG and Zilong FU designed the research and wrote the first draft of the manuscript. Feiyun CONG and Junjie LI processed the data and helped to organize the manuscript. Zheming TONG revised the final version and provided the funding support.

Conflict of interest

Shuiguang TONG, Zilong FU, Zheming TONG, Junjie LI, and Feiyun CONG declare that they have no conflict of interest.

References

Antoni J, 2006. The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mechanical Systems and Signal Processing*, 20(2):282-307.

https://doi.org/10.1016/j.ymssp.2004.09.001

Antoni J, 2007. Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing*, 21(1):108-124. https://doi.org/10.1016/j.ymssp.2005.12.002

Antoni J, Randall RB, 2006. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*, 20(2):308-331. https://doi.org/10.1016/j.ymssp.2004.09.002

Cong FY, Chen J, Dong GM, et al., 2013. Short-time matrix series based singular value decomposition for rolling bearing fault diagnosis. *Mechanical Systems and Signal Processing*, 34(1-2):218-230. https://doi.org/10.1016/j.ymssp.2012.06.005

Deng MQ, Deng AD, Zhu J, et al., 2019. Adaptive bandwidth Fourier decomposition method for multi-component signal processing. *IEEE Access*, 7:109776-109791. https://doi.org/10.1109/access.2019.2933897

Deng MQ, Deng AD, Zhu J, et al., 2020. Bandwidth Fourier decomposition and its application in incipient fault identification of rolling bearings. *Measurement Science and Technology*, 31(1):015012. https://doi.org/10.1088/1361-6501/ab4069

Dragomiretskiy K, Zosso D, 2014. Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3): 531-544. https://doi.org/10.1109/tsp.2013.2288675

Feng ZP, Zhang D, Zuo MJ, 2017. Adaptive mode decomposition methods and their applications in signal analysis for machinery fault diagnosis: a review with examples. *IEEE Access*, 5:24301-24331. https://doi.org/10.1109/access.2017.2766232

Huang NE, Shen Z, Long SR, et al., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 454(1971):903-995. https://doi.org/10.1098/rspa.1998.0193

Lei YG, Lin J, He ZJ, et al., 2013. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*, 35(1-2):108-126. https://doi.org/10.1016/j.ymssp.2012.09.015

Lei YG, Lin J, Zuo MJ, et al., 2014. Condition monitoring and fault diagnosis of planetary gearboxes: a review. *Measurement*, 48:292-305. https://doi.org/10.1016/j.measurement.2013.11.012

Li H, Li Z, Mo W, 2017. A time varying filter approach for empirical mode decomposition. *Signal Processing*, 138: 146-158. https://doi.org/10.1016/j.sigpro.2017.03.019

Li Z, Li WG, Zhao XZ, 2019. Feature frequency extraction based on singular value decomposition and its application on rotor faults diagnosis. *Journal of Vibration and Control*, 25(6):1246-1262. https://doi.org/10.1177/1077546318818690

Liu YY, Yang GL, Li M, et al., 2016. Variational mode decomposition denoising combined the detrended fluctuation analysis. *Signal Processing*, 125:349-364. https://doi.org/10.1016/j.sigpro.2016.02.011

Man ZH, Wang WY, Khoo S, et al., 2012. Optimal sinusoidal modelling of gear mesh vibration signals for gear diagnosis and prognosis. *Mechanical Systems and Signal Processing*, 33:256-274. https://doi.org/10.1016/j.ymssp.2012.07.004

Moshrefzadeh A, Fasana A, 2018. The autogram: an effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis. *Mechanical Systems and Signal Processing*, 105:294-318. https://doi.org/10.1016/j.ymssp.2017.12.009

Rato RT, Ortigueira MD, Batista AG, 2008. On the HHT, its problems, and some solutions. *Mechanical Systems and Signal Processing*, 22(6):1374-1394. https://doi.org/10.1016/j.ymssp.2007.11.028

Shi JC, Ren Y, Tang HS, et al., 2022. Hydraulic directional valve fault diagnosis using a weighted adaptive fusion of multi-dimensional features of a multi-sensor. *Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)*, 23(4):257-271. https://doi.org/10.1631/jzus.A2100394

Singh P, 2018. Novel Fourier quadrature transforms and analytic signal representations for nonlinear and non-stationary time-series analysis. *Royal Society Open Science*, 5(11): 181131. https://doi.org/10.1098/rsos.181131

Singh P, Joshi SD, Patney RK, et al., 2017. The Fourier decomposition method for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2199):20160871. https://doi.org/10.1098/rspa.2016.0871

Singhal A, Singh P, Fatimah B, et al., 2020. An efficient removal of power-line interference and baseline wander from ECG signals by employing Fourier decomposition technique. *Biomedical Signal Processing and Control*, 57:101741. https://doi.org/10.1016/j.bspc.2019.101741

Tong SG, Huang YY, Jiang YQ, et al., 2019. The identification of gearbox vibration using the meshing impacts based demodulation technique. *Journal of Sound and Vibration*, 461:114879. https://doi.org/10.1016/j.jsv.2019.114879

Tong SG, Huang YY, Tong ZM, et al., 2020. A novel short-frequency slip fault energy distribution-based demodulation technique for gear diagnosis and prognosis. *International Journal of Advanced Robotic Systems*, 17(2). https://doi.org/10.1177/1729881420915032

Torres ME, Colominas MA, Schlotthauer G, et al., 2011. A complete ensemble empirical mode decomposition with adaptive noise. IEEE International Conference on Acoustics, Speech and Signal Processing, p.4144-4147. https://doi.org/10.1109/ICASSP.2011.5947265

Wang D, 2018. Some further thoughts about spectral kurtosis, spectral *L2*/*L1* norm, spectral smoothness index and spectral Gini index for characterizing repetitive transients. *Mechanical Systems and Signal Processing*, 108:360-368. https://doi.org/10.1016/j.ymssp.2018.02.034

Wang J, Du GF, Zhu ZK, et al., 2020. Fault diagnosis of rotating machines based on the EMD manifold. *Mechanical Systems and Signal Processing*, 135:106443. https://doi.org/10.1016/j.ymssp.2019.106443

Wang TY, Chu FL, Han QK, et al., 2017. Compound faults detection in gearbox via meshing resonance and spectral kurtosis methods. *Journal of Sound and Vibration*, 392: 367-381. https://doi.org/10.1016/j.jsv.2016.12.041

Wang TY, Han QK, Chu FL, et al., 2019. Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: a review. *Mechanical Systems and Signal Processing*, 126:662-685. https://doi.org/10.1016/j.ymssp.2019.02.051

Wang WY, 2001. Early detection of gear tooth cracking using the resonance demodulation technique. *Mechanical Systems and Signal Processing*, 15(5):887-903. https://doi.org/10.1006/mssp.2001.1416

Wu ZH, Huang NE, 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(1):1-41. https://doi.org/10.1142/s1793536909000047

Yang XK, Zuo MJ, Tian ZG, 2022. Development of crack induced impulse-based condition indicators for early tooth crack severity assessment. *Mechanical Systems and Signal Processing*, 165:108327. https://doi.org/10.1016/j.ymssp.2021.108327

Yang XK, Wei DD, Zuo MJ, et al., 2023. Analysis of vibration signals and detection for multiple tooth cracks in spur gearboxes. *Mechanical Systems and Signal Processing*, 185:109780. https://doi.org/10.1016/j.ymssp.2022.109780

Yeh JR, Shieh JS, Huang NE, 2010. Complementary ensemble empirical mode decomposition: a novel noise enhanced data analysis method. *Advances in Adaptive Data Analysis*, 2(2):135-156. https://doi.org/10.1142/s1793536910000422

Yi CC, Lv Y, Dang Z, 2016. A fault diagnosis scheme for rolling bearing based on particle swarm optimization in variational mode decomposition. *Shock and Vibration*, 2016: 9372691. https://doi.org/10.1155/2016/9372691

Zhang D, Liu YY, Feng ZP, 2020. Demodulation analysis based on Fourier decomposition method and its application for gearbox fault diagnosis. International Conference on Sensing, Diagnostics, Prognostics, and Control, p.329-334. https://doi.org/10.1109/sdpc49476.2020.9353119

Zhang X, Miao Q, Zhang H, et al., 2018. A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery. *Mechanical Systems and Signal Processing*, 108:58-72. https://doi.org/10.1016/j.ymssp.2017.11.029

Zhao M, Jia XD, 2017. A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery. *Mechanical Systems and Signal Processing*, 94:129-147. https://doi.org/10.1016/j.ymssp.2017.02.036

Zhao XZ, Ye BY, 2009. Similarity of signal processing effect between hankel matrix-based SVD and wavelet transform and its mechanism analysis. *Mechanical Systems and Signal Processing*, 23(4):1062-1075. https://doi.org/10.1016/j.ymssp.2008.09.009

Zhao XZ, Ye BY, 2011. Selection of effective singular values using difference spectrum and its application to fault diagnosis of headstock. *Mechanical Systems and Signal Processing*, 25(5):1617-1631. https://doi.org/10.1016/j.ymssp.2011.01.003

Zhou W, Feng ZR, Xu YF, et al., 2022. Empirical Fourier decomposition: an accurate signal decomposition method for nonlinear and non-stationary time series analysis. *Mechanical Systems and Signal Processing*, 163:108155. https://doi.org/10.1016/j.ymssp.2021.108155