Improved Spectral Amplitude Modulation Based on Sparse Feature Adaptive Convolution for Variable Speed Fault Diagnosis of Bearing

Jiawei Lin[1], Changkun Han[1,*], Wei Lu[1,2], Liuyang Song[1], Peng Chen[4] and Huaqing Wang[1,3*]

1 College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029, China

2 Institute of Engineering Technology, Sinopec Catalyst Company Limited, Beijing 100029, China

3 Beijing Key Laboratory of Health Monitoring and Self Recovery for High-end Mechanical Equipment, Beijing University of Chemical Technology, Beijing 100029, China

4 Graduate School of Environmental Science and Technology, Mie University, Tsu 514-8507, Japan

*Huaqing Wang: hqwang@mail.buct.edu.cn

*Changkun Han: hanck2023@163.com

Received Month X, XXXX | Accepted Month X, XXXX | Posted Online Month X, XXXX

Difficulty in extracting nonlinear sparse impulse features due to variable speed conditions and redundant noise interference leads to challenges in diagnosing variable speed faults. Therefore, an improved spectral amplitude modulation based on sparse feature adaptive convolution (ISAM-SFAC) is proposed to enhance the fault features under variable speed condition. First, an optimal bi-damped wavelet construction method is proposed to learn signal impulse features, which selects the optimal bi-damped wavelet parameters with correlation criterion and particle swarm optimization (PSO). Second, a convolutional basis pursuit denoising model based on optimal bi-damped wavelet is proposed for resolving sparse impulses. A model regularization parameter selection method based on weighted fault characteristic amplitude ratio (WFCAR) assistance is proposed. Then, an improved spectral amplitude modulation method based on kurtosis threshold is proposed to further enhance the fault information of sparse signal. Finally, the type of variable speed faults is determined by order spectrum analysis. Various experimental results, such as spectral amplitude modulation and Morlet wavelet matching, verify the effectiveness and advantages of the ISAM-SFAC method.

Keywords: bearing fault diagnosis; variable speed; sparse representation; spectral amplitude modulation; feature enhancement

1 Introduction

Rolling bearing is the crucial component to support the rotating machinery, which is essential for the reliable operation and efficient working of the equipment [1],[2]. Equipment is usually operated under harsh and complex working conditions in industrial scenarios [3]. Time domain fault information is difficult to recognize due to the interference of background noise components, which brings major challenges to equipment condition monitoring and fault diagnosis [4]. Therefore, it is crucial for equipment fault diagnosis to mine fault information from complex variable speed signals [5],[6].

Scholars have proposed many signal processing methods to analyze the fault information of variable speed signals. Zhao et al [7] used generalized modulation to extract particular harmonics of the bearing rotational frequency and achieved order tracking without tachometer information. Wang et al [8] combined the principles of generalized demodulation and resonant demodulation to solve the problem of transient component extraction under variable speed conditions. Moshrefzadeh et al [9] proposed the spectral amplitude modulation (SAM) method for adaptive filtering to separate the fault features of bearing variable speed signals. Zhang et al [10] and Wu et al [11] combined the improved symplectic geometry mode decomposition method and nonlinear chirp modal decomposition method with order analysis to extract the fault features of variable speed signals, respectively. However, the above methods rely on prior knowledge for parameter settings or are less robust when interfered with by strong background noise to perform optimally.

The robust signal feature mining capability of sparse representation has led to a wide application in fault diagnosis [12],[13]. Qiu et al [14] sparse optimized singular value vectors of noise time-frequency maps to efficiently extract fault features. Cao et al [15] improved the categorical probabilistic model and sparse Bayesian learning framework to extract sparse impulse signals with high accuracy and robustness. Liu et al [16] integrated a strengthen-operate-subtract boosting strategy and dictionary learning into a multiscale transform to adaptively and efficiently extract fault impulses. Ma et al [17] developed a physically-inspired sparse model combined with a generalized sparse cyclic spectrum to sense compound faults in ultra-low-speed operating turbines efficiently. Li et al [18] proposed the period-assisted correlation kurtosis of the envelope spectrum as a wavelet sparse weighting strategy, which enhances the capture of information on bearing fault features. Yao et al [19] proposed a multiband weighted generalized minimax-concave induced sparse representation to diagnose the initial faults of bearings effectively.

However, the above methods are primarily applied in stationary fault diagnosis. Under variable speed conditions, the fault information modes within the collected signals are mixed, and the periodic transient pulse forms are more complex [20],[21]. Some scholars have attempted to use sparse representation methods for variable speed fault diagnosis. Shi et al [22] constructed a Laplace wavelet complete dictionary and reconstructed the variable speed signal using stage-wise orthogonal matching pursuit. Hou et al [23] used sparse time-frequency representations to reveal potential fault feature frequencies and effectively improved the signal time-

frequency resolution. All these methods have achieved some effect. However, there are limitations in matching fault pulses using mathematical models with single-sided attenuation or bilateral symmetric attenuation wavelets [24],[25]. The parameters of the sparse coding often need to be selected using posterior knowledge. All these issues create additional difficulties when applying sparse representation to variable speed fault diagnosis.

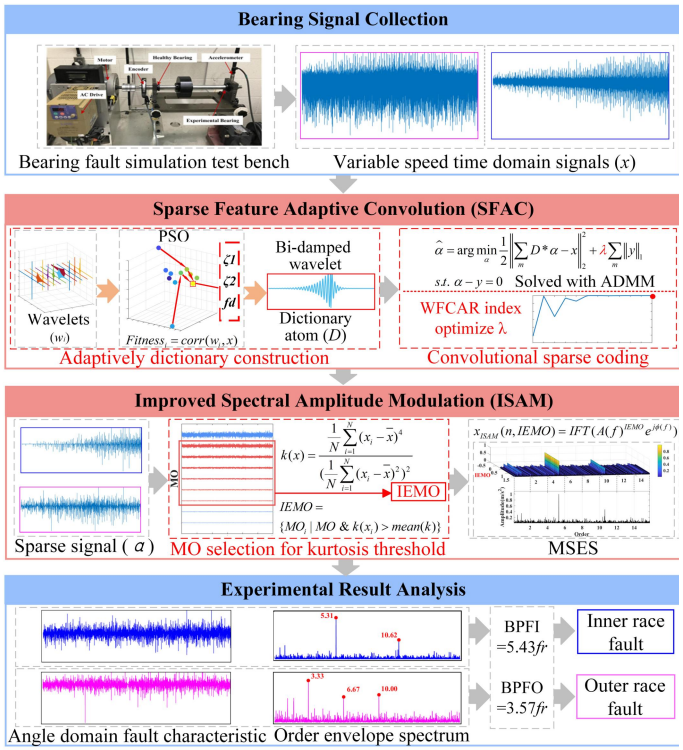


Fig. 1. Schematic diagram of ISAM-SFAC.

Therefore, an improved spectral amplitude modulation (ISAM) based on sparse feature adaptive convolution (SFAC) is proposed for fault diagnosis of bearings with variable speed. First, an adaptive bi-damped wavelet matching algorithm is proposed. The particle swarm optimization (PSO) algorithm and correlation analysis are combined to adaptively select the optimal bi-damped wavelet in the parameter space. Second, a convolutional sparse coding model based on optimal bi-damped wavelet is used to reconstruct sparse impulse information in the time domain. The regularization parameter of the model is adaptively chosen by the weighted fault characteristic amplitude ratio (WFCAR) index. Finally, an improved SAM based on kurtosis thresholding is used to analyze the sparse reconstructed signal in terms of fault order. The main fault information in the signal is identified. The schematic diagram of the method is shown in Fig. 1. The main contributions are as follows:

1) An efficient convolutional sparse model is proposed. A signal-matched optimal bi-damped wavelet is used for convolutional basis pursuit denoising to extract fault impulse features.

2) A parameter selection method for SFAC is proposed. PSO constructs the optimal bi-damped wavelet, and the WFCAR index is constructed to select the regularization parameter adaptively.

3) An improved method for SAM is proposed. The kurtosis threshold adaptively preserves the magnitude order of SAM to construct an order envelope spectrum with fault impulse features.

The remaining chapters of this paper are arranged as follows. Section 2 introduces the theories of sparse representation and SAM. Section 3 introduces the steps for implementing the proposed method in this paper. Section 4 and Section 5 demonstrate the effectiveness of the proposed method on the variable speed simulation signal, inner race, and outer race experimental signals, respectively, and verify the robustness and applicability of the method through several comparison tests. The study is summarized in Section 6.

2 Basic theory

2.1 Sparse representation

Signal sparse representation is a linear combination of as few atoms as possible in the constructed dictionary. This method can reduce the redundant noise component interference and obtain the main features to approximate the original signal. The constrained problem of signal sparse reconstruction is transformed into an unconstrained optimization problem as in Equation (1) based on the Lagrange multiplier method.

$$\alpha = \arg\min_{\alpha} \|x - D\alpha\|_2^2 + \lambda R(\alpha) \quad (1)$$

where x denotes the original signal, D denotes the dictionary matrix, α denotes the sparse coefficients, λ denotes the regularization parameter and $R(\cdot)$ denotes the sparsity-inducing function, e.g., L_0 norm or L_1 norm. The efficiency of sparsity is usually expressed in terms of the sparsity of the coefficients α.

The sparse representation model described above yields sparse results for the primary components of the signal by employing a sparse dictionary. The constraint of convolutional sparse coding is like those of the L_1 sparse model, which represents the signal as a set of convolutional sparse mappings to extract the principal features within the signal and suppress noise components. How to design a set of efficient dictionary filters is the key to improve the performance of the convolutional sparse coding algorithm.

2.2 Spectral amplitude modulation

Filtering can increase the signal to noise ratio of the signal. The complex cepstrum can assist the filtering to process the signal and its mathematical model is shown in Equation (2).

$$x_c(n) = IFT\{(\log A(f)) + j\phi(f)\} \quad (2)$$

where $IFT\{\cdot\}$ denotes a Fourier inverse transform, $A(f)$ denotes the signal amplitude, $\Phi(f)$ denotes the signal phase. Its phase can completely or partially reconstruct a signal to preserve important features [26]. The cepstrum pre-whitening method zeroes out the real cepstrum of the signal. The phase information of the original signal with fault-related features is preserved in the pure phase reconstruction shown in Equation (3) [27].

$$x_{cpw}(n) = IFT\{A(f)^0 e^{j\phi(f)}\} \quad (3)$$

However, when the real part of the complex inverted spectrum is completely preserved or eliminated, it is difficult to represent the complete characterization of the signal in the frequency domain. Thus, the spectral amplitude modulation (SAM) method shown in Equation (4) results from a further refinement of the cepstrum pre-whitening, where MO denotes the magnitude orders of the modified components. SAM constructs multiple signal components to represent the spectrum's overall behavior. The fault features of the signal can be identified by observing the maximum square envelope spectrum.

$$x_{SAM}(n, MO) = IFT\{A(f)^{MO} e^{j\phi(f)}\} \quad (4)$$

As a nonlinear filtering method, SAM enables simple, fast, and adaptive extraction of fault features without priori knowledge of the signal.

3 Improved spectral amplitude modulation with convolutional sparsity

A variable speed fault diagnosis method is proposed. This method can extract variable speed fault features by sparse feature adaptive convolution (SFAC) and improved spectral amplitude modulation (ISAM). SFAC effectively enhances the signal's time domain fault impulse and reduces the noise component's interference, and ISAM further separates the fault features.

3.1 Sparse feature adaptive convolution

The shift-invariant convolutional basis pursuit denoising is chosen as the model for sparse coding. As shown in Equation (5), the model replaces the multiplication operation of the dictionary and sparse coefficients in the sparse representation model with convolution operation. The impulse filter convolves the signal to extract the fault features [28],[29]. The bi-damped wavelet constructed as a convolutional filter dictionary reduces the computational effort of the solution and effectively captures the global faulty pulse features of the signal.

$$\begin{aligned} &a = \arg\min_{a} \frac{1}{2}\left\|\sum_{m} D*a - x\right\|_2^2 + \lambda\sum_{m}\|y\|_1 \\ &s.t.\ a - y = 0 \end{aligned} \tag{5}$$

where D denotes the dictionary, α denotes the sparse coefficients, x denotes the original signal, * denotes the convolution operation, and y is the introduced auxiliary variable Equation. The model often uses the alternating direction multiplier method (ADMM) as in Equation (6), Equation (7), and Equation (8) to iteratively solve the operator efficiently and obtain the sparse signal containing the fault features.

$$a^{(j+1)} = \arg\min_{a} \frac{1}{2}\|D*a - x\|_2^2 + \frac{\rho}{2}\left\|a - y^{(j)} + u^{(j)}\right\|_2^2 \tag{6}$$

$$y^{(j+1)} = \arg\min_{y} \lambda\|y\|_1 + \frac{\rho}{2}\left\|a^{(j+1)} - y + u^{(j)}\right\|_2^2 \tag{7}$$

$$u^{(j+1)} = u^{(j)} + a^{(j+1)} - y^{(j+1)} \tag{8}$$

The solution of subproblem Equation (6) requires the introduction of the auxiliary variable $z = y^{(j)} - u^{(j)}$ and the conversion of the optimization problem from the time domain to the frequency domain, transforming the convolution problem to a multiplication problem, and solving the resulting linear system Equation (9). D, α, x, and z in the Fourier transform domain are denoted as $\hat{D}$, $\hat{a}$, $\hat{x}$, and $\hat{z}$.

$$(\hat{D}^H\hat{D} + \rho I)\hat{a} = \hat{D}^H\hat{x} + \rho\hat{z} \tag{9}$$

The subproblem Equation (7) is solved using the soft thresholding method denoted as

$$y^{(j+1)} = \mathrm{Soft}_{\lambda/\rho}(a^{(j+1)} + u^{(j)}) \tag{10}$$

where Soft(·) is the soft threshold function, expressed as Equation (11), Sgn (·) denotes the sign function, $\odot$ denotes element-wise multiplication.

$$\mathrm{Soft}_{\gamma}(a) = Sgn(a)\odot\max(0, |a| - \gamma) \tag{11}$$

The model converges by setting an iterative error threshold. The convolution coefficients obtained from the final solution are sparse, and the reconstructed signal has more obvious impulse features than the original signal.

The setting of the regularization parameter λ in the convolutional sparse coding model affects the sparsity of the processed signal in the time domain, and the effect of fault features representation in the order domain. Therefore, this paper proposes a maximized weighted fault characteristic amplitude ratio (WFCAR) index, which combines the golden ratio trichotomy to efficiently search for the optimal regularization parameter to improve the identifiability of the fault order of the processed signal.

$$WFCAR = \frac{\sum_{i=1}^{H=floor(\frac{O_s}{O_c})} \omega_i \times A_{i\times O_c}}{\sum_{j=1}^{n} A_j} \quad (12)$$

where O_s denotes the identifiable order within the order spectrum, O_c denotes the fault order, *floor* $(\cdot)$ denotes rounding down, $A_i \times O_c$ denotes the amplitude of the i th-fold fault order harmonic, A_j denotes the amplitude of the j th data point, and H denotes the largest identifiable order. The weighting coefficients ω_i shown in Equation (13) are added to the indices for different multiples of harmonics, respectively.

$$\omega_i = \frac{H-i+1}{H(H+1)/2} \quad (13)$$

The algorithm is considered to have converged and regularization parameters adaptively selected when the index stabilized during iteration, i.e., when the index error is less than 0.01 in two iterations.

The commonly used overcomplete dictionaries have significant complexity in solving coefficients due to their large dimensions. Therefore, signal analysis computational efficiency can be improved by using convolutional sparse coding of the reduced dimensional dictionary. The bi-damped wavelet can tune its bilaterally attenuated waveform. This property enables the construction of rich wavelet set structures, making them more adaptable and capable of representing different types of fault impulse signals. An adaptive bi-damped wavelet parameters optimization method is proposed to construct an efficient dictionary with a strong correlation with variable speed fault information. An optimal bi-damped wavelet with the fault impulse feature is constructed by optimally matching the signal correlation.

The mathematical model of the bi-damped wavelet is shown in Equation (14), where $\omega=2\pi f$ and τ denote the angular frequency and time shift, respectively, the two damping ratios ζ_1 and ζ_2 adjust the attenuation of the oscillations on the left and the right side of the wavelet, respectively. Its asymmetric bilateral exponential attenuation waveform can better match the implied fault impulse response inside the signal. The two attenuation damping ratios and the natural frequency in the parameters are crucial parameters affecting the constructed wavelet morphology.

$$\begin{aligned} g_1 &= Sgn(t-\tau,0)\times e^{\omega(t-\tau)\frac{-\zeta 1}{\sqrt{1-\zeta 1^2}}}\times\cos(\omega(t-\tau)) \\ g_2 &= Sgn(\tau-t,0)\times e^{\omega(\tau-t)\frac{-\zeta 2}{\sqrt{1-\zeta 2^2}}}\times\cos(\omega(t-\tau)) \\ g &= g_2 - g_1 \end{aligned} \quad (14)$$

Unlike wavelet transform methods that use fixed wavelet base, SFAC constructs optimal wavelet base adapted to the features of signal fault impulses through correlation analysis. The method dynamically and adaptively optimizes the time and frequency domain features of the wavelet for different signal types by introducing two independent

damping factors and intrinsic frequencies, making it more flexible and efficient. A parameter space containing two attenuation damping ratios and the natural frequency is constructed, in which different combinations of parameters can be selected to construct bi-damped wavelets with different morphologies. The degree of similarity between wavelet and signal internal impulses can be evaluated quantitatively using the correlation coefficients shown in Equation (15), where w denotes the wavelet vector, x denotes the one-dimensional signal vector, and <·> denotes the inner product of the vectors, $\|\cdot\|_2$ denotes the L_2 norm of the vector.

$$C = \frac{\left|\langle w(t), x(t)\rangle\right|}{\|w(t)\|_2 \|x(t)\|_2} \tag{15}$$

The schematic diagram of using correlation analysis to assist the particle swarm algorithm in selecting wavelet parameters is shown in Fig. 2. The wavelet parameter intervals are taken as follows: the natural frequency fd is taken as [500:3000], the attenuation damping coefficient ζ_1 and enhance damping coefficient ζ_2 are taken as [0:1]. The correlation coefficient, as the fitness for PSO iterations, will converge after the iterations to obtain the optimal wavelet parameters. The optimal parameters are used to construct a bi-damped wavelet, which is used as a convolutional filter dictionary to sparsely code the original signal to enhance fault features.

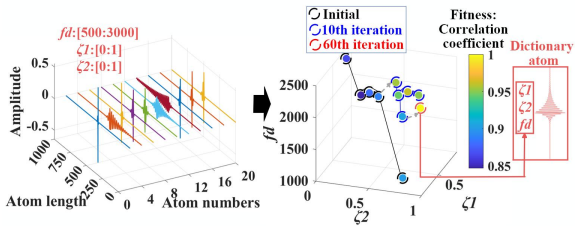


Fig. 2. Schematic of particle swarm construction of optimal bi-damped wavelet.

3.2 Improved spectral amplitude modulation

The range of custom amplitude magnitude order (MO) selection in spectral amplitude modulation (SAM) significantly impulses identifying the final maximum squared envelope spectrum obtained. When $MO > 1$, the effect of frequency components with larger amplitude within the signal is amplified, and the low-energy components are hidden by the high-energy components, which can distinguish the impulse resonance frequency from the noise-related frequency. When $0 \leqslant MO \leqslant 1$, high-energy and low-energy components are preserved, essential in detecting defective signals when fault components are hidden in the noise background. When $MO < 0$, the primary frequency amplitude is filtered out, and information related to the low-energy defects hidden by the high energy of the other defect components is extracted. The MO selection range is usually set from -0.5 to 1.5 to make the processing as general as possible.

A reasonable magnitude order interval ensures that the subsequent plotting of the maximum squared envelope spectrum retains complete signal information. However, the modified component constructed in the low energy level interval may contain many interfering noises and abnormal amplitudes generated by uncorrelated components. Fault components in the high energy interval cannot be effectively extracted when drowned by strong noise. Therefore, there is an urgent need for an adaptive magnitude order selection approach to screen the magnitude order of SAM further to enhance the fault features of the signal.

Kurtosis is a statistic that describes the sharpness of the distribution pattern of

signal data and reflects the impulse features of the signal [30],[31]. Therefore, the kurtosis coefficient was chosen as a reference indicator that the signal contains impulse components.

A kurtosis-based magnitude order selection method is proposed to make modified components constructed from selected magnitude orders have stronger fault impulse features. The magnitude orders of the enhanced impulse features are further filtered by calculating the kurtosis coefficients of all modified components in the interval. The impulse-enhanced magnitude order (IEMO) components with higher kurtosis coefficients than the mean of all kurtosis coefficients will be kept. The other magnitude order components will be removed. The magnitude order is chosen as shown in Equation (16), in which μ_k denotes the mean value of kurtosis, k denotes the kurtosis coefficient.

$$\begin{aligned} &IEMO = \{MO_i \mid MO_i \in MO \ and \ k(x_i) > \mu_k\} \\ &\mu_k = \sum_{i=1}^{num} k(x_i) \end{aligned} \tag{16}$$

The modified components of Equation (17) are constructed separately for each selected IEMO, and the analytic signal (A) of the enhancement component, as shown in Equation (18), is calculated using the Hilbert transform. The squared envelope spectrums of the enhancement component in Equation (19) can be superimposed to obtain the maximum squared envelope spectrum to get the fault information.

$$x_{ISAM}(n, IEMO) = IFT(\mathrm{A}(f)^{IEMO} e^{j\phi(f)}) \tag{17}$$

$$\begin{aligned} &\mathrm{A}\{x_{ISAM}(n, IEMO)\} \\ &= x_{ISAM}(n, IEMO) + jH\{x_{ISAM}(n, IEMO)\} \end{aligned} \tag{18}$$

$$\begin{aligned} &SES\{x_{ISAM}(n, IEMO)\} \\ &= \left| FT\{|\mathrm{A}\{x_{ISAM}(n, IEMO)\}|^2\} \right| \end{aligned} \tag{19}$$

The component selection and the construction of the maximum order envelope spectrum of ISAM are shown in Fig. 3.

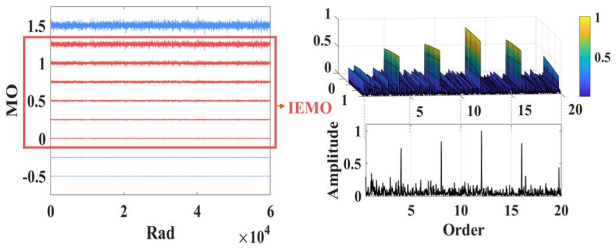


Fig. 3. Schematic of components selection and maximum order envelope spectrum for ISAM.

The lower orders of the order envelope spectrum and their corresponding amplitudes are removed to minimize the effect of the direct current component of the signal on the order spectrum. The appearance of abnormally large values of the maximum square envelope spectrum at low orders can be avoided.

4 Simulation verification

4.1 Simulation signal modeling

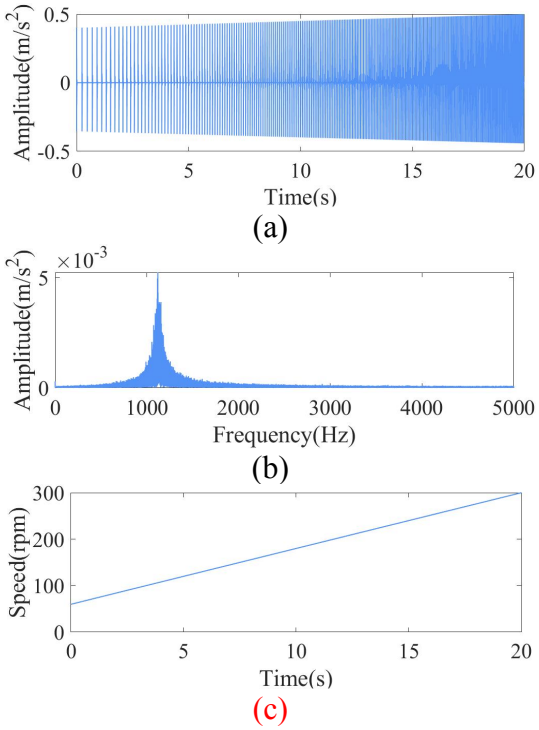
To verify the effectiveness of the ISAM-SFAC method in diagnosing the variable speed signal, a rolling bearing fault simulation signal under variable speed is constructed [32],[33]. The simulated signal contains the angular domain periodic impulse and noise components. The mathematical expression is shown in Equation (20).

$$x(t)=\sum_{i=1}^{K} A_i e^{-d(t-\sum_{j=1}^{i} T_j)} \cos(2\pi f_0(t-\sum_{j=1}^{i} T_j))+n_0(t) \tag{20}$$

In this Equation, K denotes the number of fault impulses, A_i denotes the impulse amplitude, d denotes the attenuation coefficient, T_j denotes the time interval between the occurrence of the j th fault impulse and the j-1 th fault impulse, f_0 denotes the resonance frequency, n_0 is the noise component. The signal crucial parameters are set as shown in Table II.

TABLE II. Simulation signal parameters

Parameters	Numerical value
Attenuation coefficient d	500
Resonance frequency f_0	2000
Amplitude A_i	-0.005t+0.4
Sampling frequency fs	10000
Sampling points N	200000



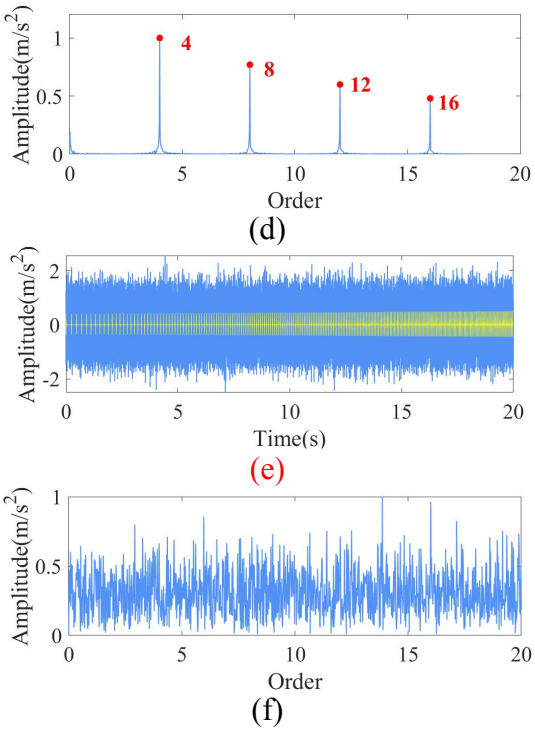


Fig. 4. Simulation signal. (a) Time domain waveform; (b) Frequency spectrum; (c) Speed curve; (d) Order envelope spectrum; (e) Time domain waveform with noise; (f) Order envelope spectrum with noise.

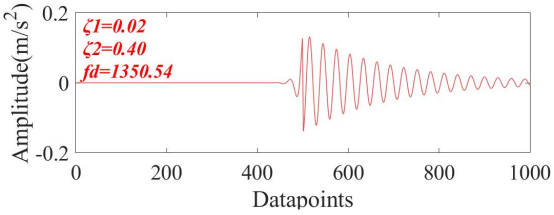
The type of fault in the variable speed signal can be recognized by observing the order spectrum. In this simulated signal, the fault order of the signal is 4. The rotational speed of the simulated reference axis is set to ω(t)=12t+60, and the rotational speed curve is shown in Fig. 4(c).

Fig. 4(a) plots the time domain impulse components in the fault simulation signal for variable speed. The signal amplitude exhibits an increasing trend with increasing speed. As shown in Fig. 4(b), a trend of energy concentration can be observed in the frequency spectrum of the impulse signal. The angle domain of the simulated impulse signal is resampled, and its order envelope spectrum, shown in Fig. 4(d), is plotted. The fault order 4 can be identified from the order

envelope spectrum. Gaussian white noise with a signal to noise ratio (SNR) of 5 dB is added to the signal. Fig. 4(e) shows that the strong noise drowns the time domain fault impulses, the time domain noise amplitude is about 3-4 times the amplitude of the fault impulse. The fault order is unrecognizable within the order envelope spectrum of Fig. 4(f).

4.2 Simulation signal analysis

The simulated signal is analyzed using the ISAM-SFAC method. In Fig. 5(a), the optimal bi-damped wavelet is obtained using the adaptive bi-damped wavelet matching algorithm, similar to the impulse pattern constructed by the simulation. The optimal regularization parameter is iteratively derived based on the weighted fault characteristic amplitude ratio (WFCAR) index, with its trend shown in Fig. 5(b). After 12 rounds of iterations, the difference in the WFCAR index between two iterations is less than the set threshold of 0.01, indicating that the corresponding regularization parameter λ converges to an optimal value of 0.14. The sparse processed time-domain signal in Fig. 5(c) shows effectively weakened noise components. In Fig. 5(d), the preset fault order 4 and its harmonics are identifiable, verifying the ISAM-SFAC method's feature enhancement and noise suppression capabilities.



(a)

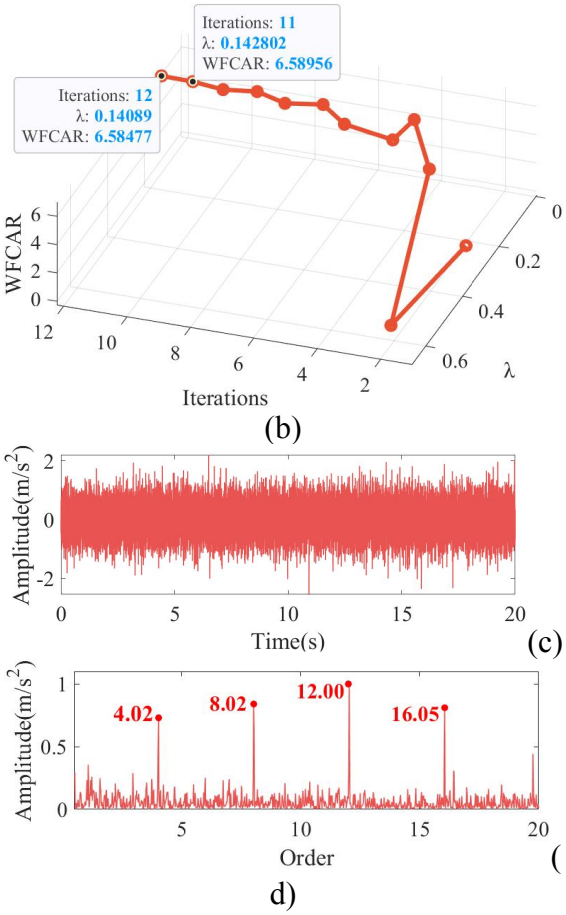


Fig. 5. Simulation signal after ISAM-SFAC. (a) Optimal bi-damped wavelet waveform; (b) Iterative trend of the regularization parameter; (c) Time domain waveform; (d) Order envelope spectrum.

TABLE III. Comparison methods

Serial number	Methods	Type of methods
(1)	SAM	Nonlinear filtering
(2)	Morlet Wavelet Matching	Another wavelet atom
(3)	GME-TV	Sparse assisted filtering

4.3 Comparison methods analysis

The comparison methods shown in Table III process the simulation signal separately to validate the advantages of the proposed ISAM-SFAC method in feature enhancement and robustness under comparable conditions.

Part of the proposed method is the improvement of Method (1). Spectral amplitude modulation (SAM) is simple to implement and effectively separates fault components. The constructed noise-added simulated signal is analyzed by SAM. Fig. 6(a) contains a large amount of redundant noise components that have not been effectively suppressed, significantly affecting the analysis results. As shown in Fig. 6(b), the maximum order envelope spectrum after SAM processing cannot recognize the fault order, indicating that SAM has limited ability to resist noise interference and is not effective under harsh conditions.

In Method (2), the bi-damped wavelet used in the proposed method is replaced with the Morlet wavelet to validate further the effect of the wavelet base mathematical model choice on the experimental results. The Morlet waveform obtained by matching the simulated signal is shown in Fig. 7(a). The optimal regularization parameter λ=0.77 is obtained after 12 iterations, as in Fig. 7(b). The time domain and order envelope spectrum of the signal processed using Method (2) are shown in Figs. 7(c) and (d), respectively, where the time domain noise component is suppressed in Fig. 7(c), and the fault order and harmonics can be recognized in the order envelope spectrum in Fig. 7(d). However, the recognition effect is not as good as the ISAM-SFAC method.

Table IV shows the weighted fault characteristic amplitude ratio indices calculated for the proposed method and two comparison methods. The ISAM-SFAC method has the highest value of the index, which proves that the ISAM-SFAC method has a prominent feature extraction advantage.

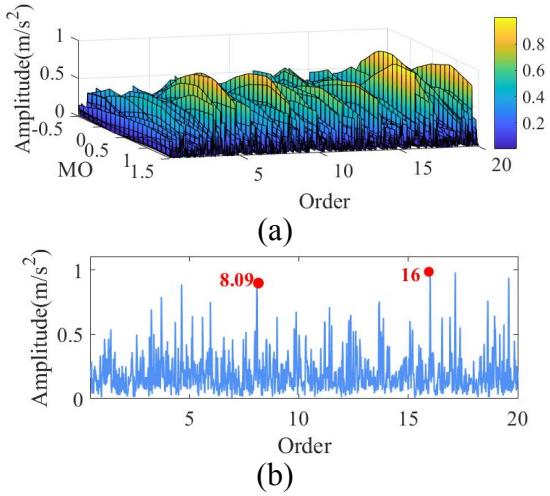
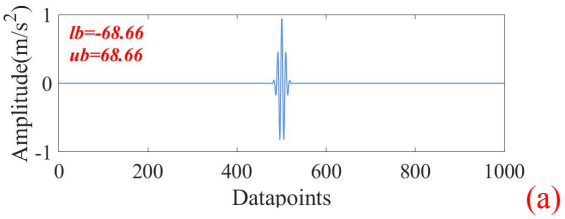


Fig. 6. Simulation signal after SAM. (a) 3D plot; (b) Order envelope spectrum.

TABLE IV. WFCAR values for different methods processing simulation signal

Methods	WFCAR
Original signal	1.13
Method (1)	1.52
Method (2)	4.11
ISAM-SFAC	**6.58**



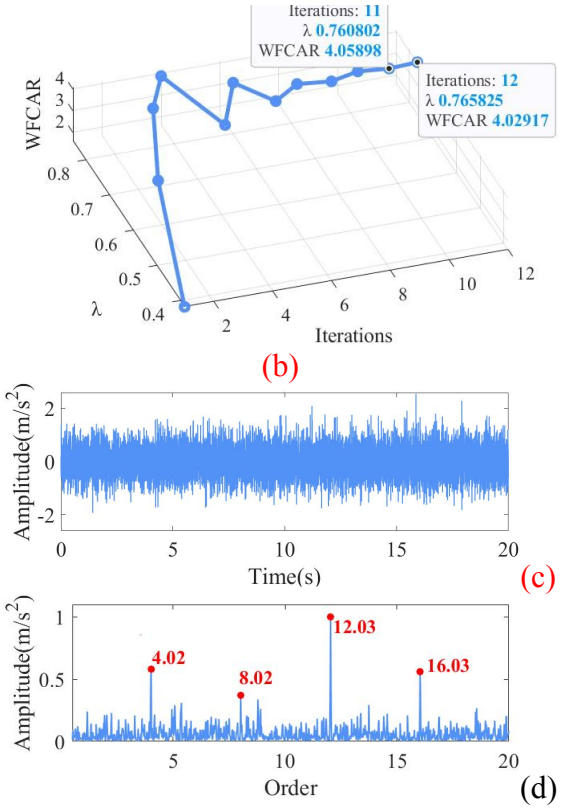


Fig. 7. Simulation signal with noise after Morlet wavelet matching. (a) Optimal Morlet wavelet waveform; (b) Iterative trend of the regularization parameter; (c) Time domain waveform; (d) Order envelope spectrum.

4.4 Anti-noise analysis

To verify the anti-noise performance of the ISAM-SFAC method, the simulated signals under different noise levels are processed with SAM and the ISAM-SFAC method, respectively. Gaussian white noises with SNRs of 10dB, 5dB, and 2dB are set to join the simulated fault impulses, and three sets of comparison experiments are carried out, respectively.

The results of processing the signal with a 10 dB SNR using the two methods are shown in Figs. 8(a). The fault orders and harmonics in the order envelope spectrum of the SAM are interfered with by much noise, whereas the results of processing using the ISAM-SFAC method still clearly identify all the fault harmonics in the order interval. The advantages of the ISAM-SFAC method can be observed from the comparison of the indices in Table V.

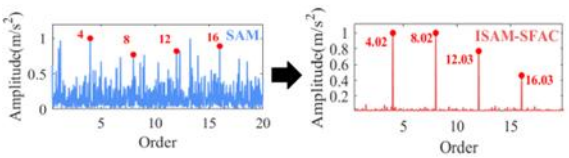
The signal with a SNR of 5 dB is unrecognizable from the fault order after processing with SAM. In contrast, the fault order and harmonics remain stable and high in amplitude in the experimental result of the ISAM-SFAC method, as in Fig. 8(b).

The processing results at a SNR of 2 dB are shown in Fig. 8(c). No fault-related information can be observed in the order envelope spectrum obtained by processing the signal using SAM. However, the fault order and its harmonic become recognizable in the order spectrum after processing by the ISAM-SFAC method.

The results of the three experiments illustrate that the ISAM-SFAC method always ensures better noise suppression and fault feature enhancement with strong robustness in the case of gradual noise enhancement.

TABLE V. WFCAR values for different noise levels

Methods	10dB	5dB	2dB
SAM	2.04	1.52	
ISAM-SFAC	**20.85**	**6.25**	



(a)

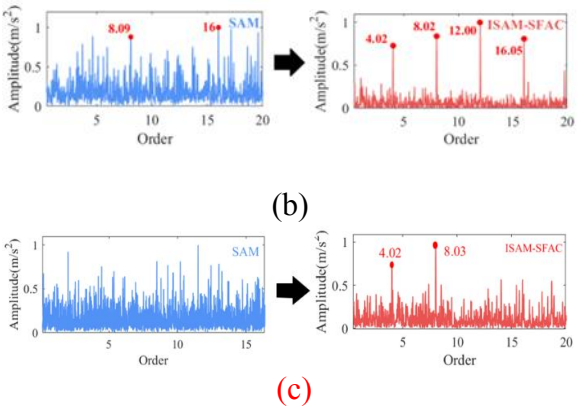


Fig. 8. Simulation signals with different SNR after SAM and ISAM-SFAC. (a) SNR=10dB; (b) SNR=5dB; (c) SNR=2dB.

4.5 Regularization parameter analysis

A weighted fault characteristic amplitude ratio (WFCAR) index is used in convolutional sparse coding to select the best regularization parameter. Therefore, regularization parameter experiments are conducted to verify the validity of the selected parameter index.

Fig. 9(a) shows the iterative graph of the index obtained. It can be observed that the index is taken to the lowest point at the 2nd iteration, which corresponds to the regularization parameter λ=0.62. At the 5th iteration, the index value is taken to the mid-point, corresponding to the regularization parameter λ=0.09. The index converged at λ=0.14 after the 12th iteration.

The result of using the lowest point of WFCAR corresponding to λ=0.62 is shown in Fig. 9(b). The redundant components interfere with the order envelope spectrum, which makes it impossible to recognize the fault order and harmonics. In Fig. 9(c) of the experimental results for the WFCAR midpoint corresponding to λ=0.0902, the feature fault order and harmonics can be observed but are contaminated by many noise components. The fault order and its harmonics can be identified in the experimental results corresponding to selecting the optimal parameter λ=0.14 in Fig. 9(d).

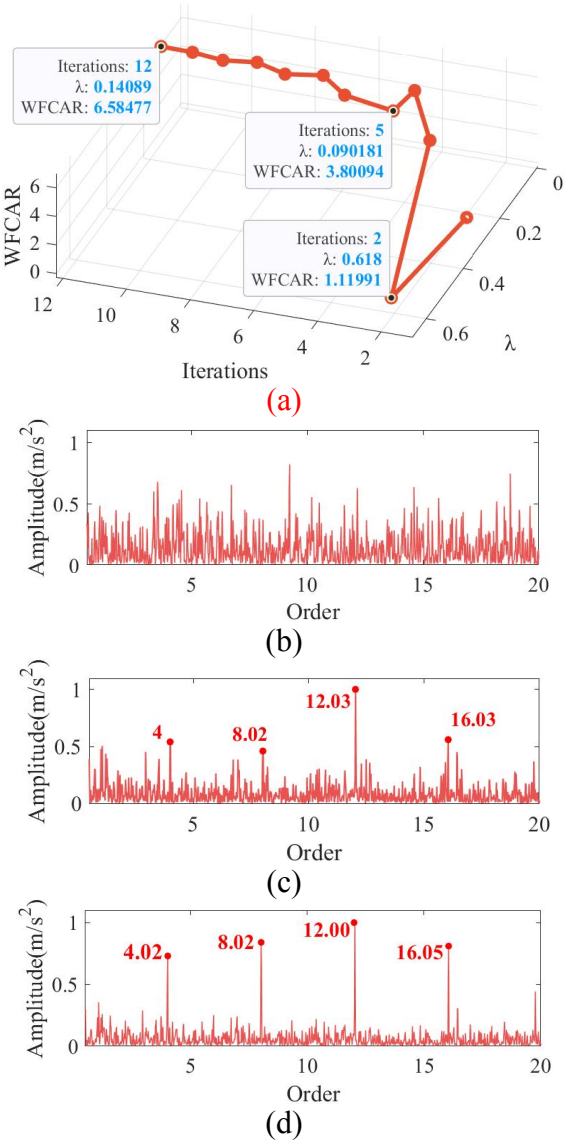


Fig. 9. Simulation signal after ISAM-SFAC with different regularization parameters. (a) Iterative trend of the regularization parameter; (b) λ=0.62; (c) λ=0.09; (d) λ=0.14.

4.6 Efficiency analysis

The ISAM-SFAC method consists of three main time-consuming phases, and the running efficiency of the method is mainly affected by the signal length. Five experiments were conducted to record the average running time of the proposed method as shown in Table VI.

All simulations and experiments were conducted on Windows 10 22H2 and MATLAB R2023b, using a personal computer with an Intel Core i7 6-CPU (2.20 GHz) and 8 GB RAM.

In addition to evaluating the efficiency of the proposed method, the PSO algorithm is compared with the grid search method for optimal wavelet parameters selection. Both methods varied the attenuation damping coefficient $\zeta 1$ and enhancement damping coefficient $\zeta 2$ within the range [0:1]. For the grid search method, the intrinsic frequency f_d is fixed at 1500 Hz to simplify the optimization process.

Table VI. WFCAR values for different methods processing inner race signal

Step number	Steps	Time consumption (s)
1	PSO adaptive matching of the optimal wavelet	125.35
2	The WFCAR index selects the optimal λ	30.26
3	ISAM based on sparse feature convolution	2.19
Total time consumption		157.80

Under the above conditions, the grid search method takes 223.57 s to select the optimal wavelet parameters for the simulation signal. The PSO algorithm exhibited higher efficiency compared to the grid search method, while also optimizing more parameters. Therefore, the PSO algorithm is considered more suitable for the optimization of bi-damped wavelet parameters in this study.

5 Mechanical fault verification

5.1 Test bench introduction

The validation dataset for the model in this paper is publicly available data on bearing vibration under variable operating conditions from the University of Ottawa, Canada [34]. This dataset was collected from the SpectraQuest Mechanical Failure Simulation test bench shown in Fig. 10. This bearing signal is sampled at fs=200000Hz with a sampling time of 10 seconds. The bearing fault can be judged by identifying the fault order, which is defined by the structural parameters of the bearing. The theoretical fault order of the inner race is 5.43, and the theoretical fault order of the outer race is 3.57.

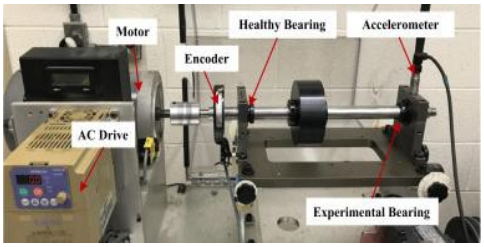


Fig. 10. Fault test bench of rolling bearing.

5.2 Inner race signal experiment

The time-domain waveform of the inner race increasing-speed signal collected by the test bench is shown in Fig. 11(a), exhibiting

noticeable impulse features characteristic of a fault. Significant fault orders and their harmonics are clearly identifiable in the envelope order spectrum presented in Fig. 11(b). To evaluate the feature enhancement capability of the model in the presence of noise, Gaussian white noise with an 8 dB SNR is added to this experimental signal. After adding the noise, the signal's fault impulse components become unrecognizable, as shown in Fig. 11(c), while the fault order information in the order spectrum of Fig. 11(d) is overwhelmed by the noise and cannot be discerned.

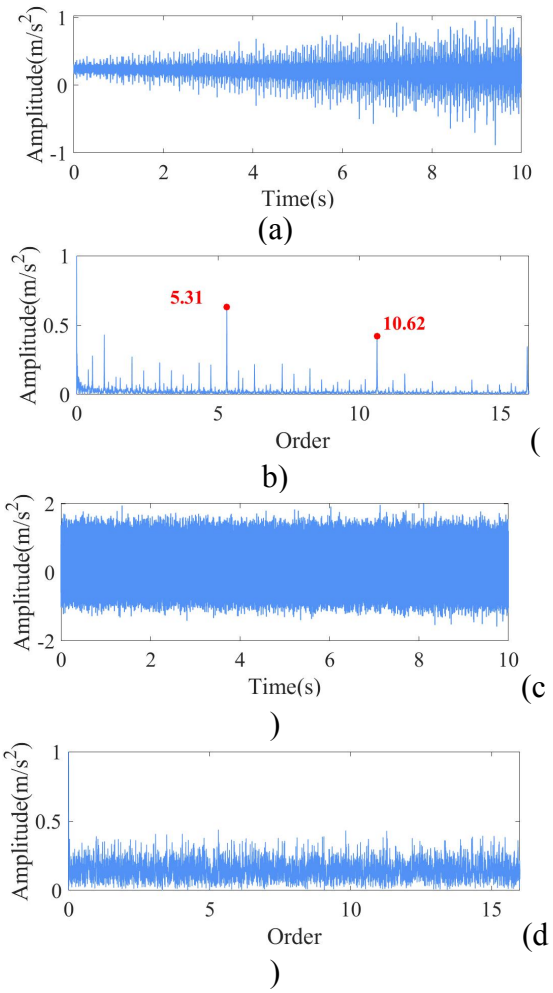


Fig. 11. Measured signal of inner race contains fault under variable speed. (a) Time domain waveform; (b) Order envelope spectrum; (e) Time domain waveform with noise; (f) Order envelope spectrum with noise.

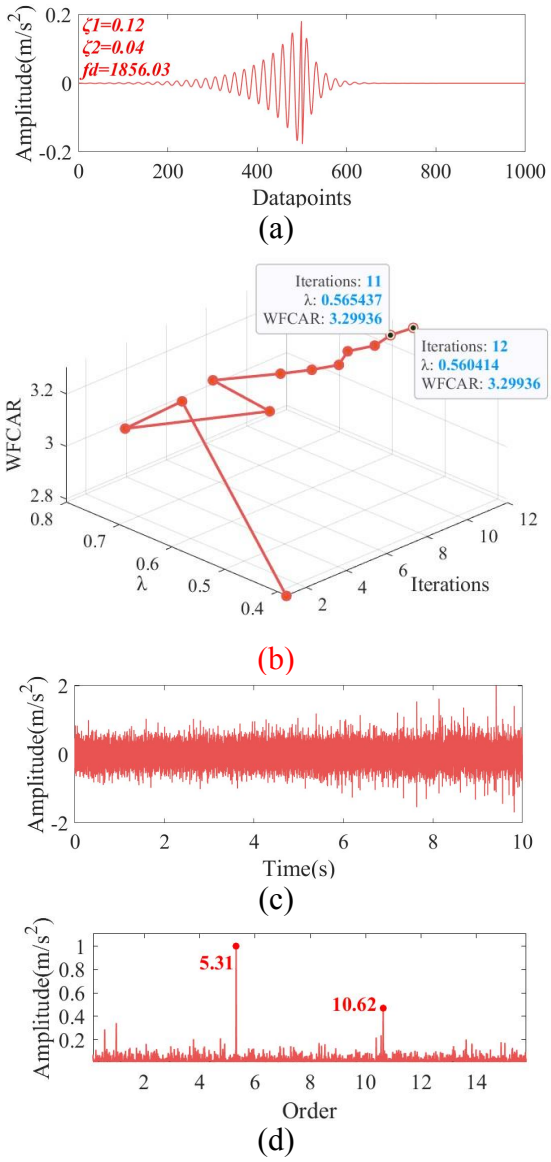


Fig. 12. Inner race signal after ISAM-SFAC. (a) Optimal bi-damped wavelet waveform; (b) Iterative trend of the regularization parameter; (c) Time domain waveform; (d) Order envelope spectrum.

Table VII. WFCAR values for different methods processing inner race signal

Methods	WFCAR
Original signal	1.22

Method (1)	1.61
Method (2)	1.34
ISAM-SFAC	**3.30**

The noise-added signal is then processed using the ISAM-SFAC method. The wavelet waveform obtained through correlation matching is depicted in Fig. 12(a), where the bilaterally attenuated asymmetrical waveform demonstrates its ability to effectively match the fault-induced pulses of the signal. The regularization parameter is optimized to $\lambda = 0.56$, as shown in Fig. 12(b), resulting in a noticeable reduction of noise components in the time-domain waveform, as seen in Fig. 12(c). Furthermore, the envelope order spectrum in Fig. 12(d) successfully reveals the fault order and its harmonics, confirming the model's robustness and effectiveness in extracting fault features from noisy environments.

Fig. 13(a) and (b) show the order envelope spectrum obtained by processing using the comparison methods (1) and (2), respectively. Both spectra are difficult to recognize the fault order. According to Table VII, both comparison methods are lower than the ISAM-SFAC method in terms of the weighted fault characteristic amplitude ratio (WFCAR) index, which verifies the advantages of the ISAM-SFAC method.

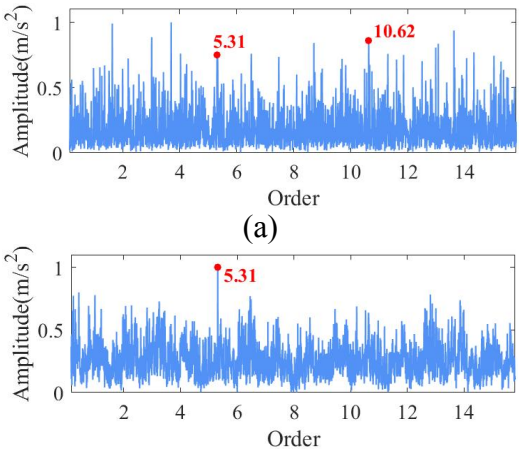


Fig. 13. Inner race signal after comparison methods. (a) Method (1) order envelope spectrum; (b) Method (2) order envelope spectrum.

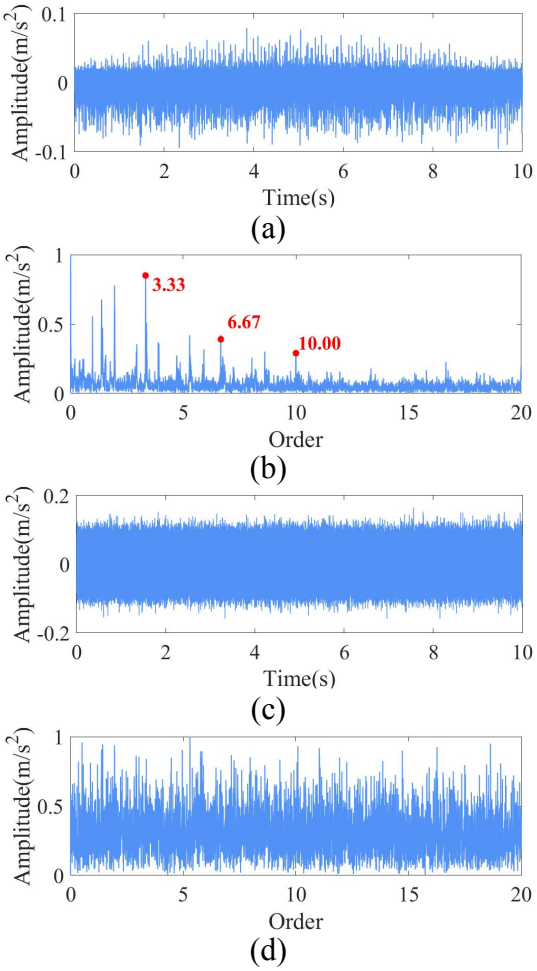


Fig. 14. Measured signal of outer race contains fault under variable speed. (a) Time domain waveform; (b) Order envelope spectrum; (c) Time domain waveform with noise; (d) Order envelope spectrum with noise.

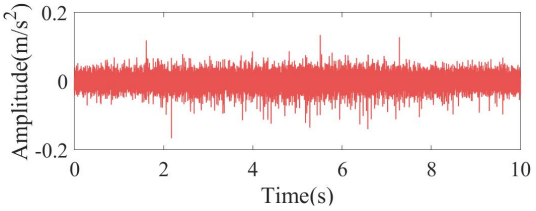
5.3 Outer race signal experiment

The outer race signal with a trend of increasing then decreasing speed is processed using the ISAM-SFAC method.

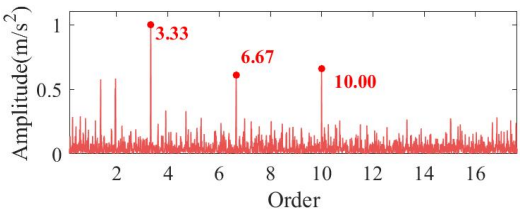
As shown in Fig. 14(a), pulse features and the amplitude trend are evident in the time domain of the original signal. The fault order is identifiable in the order envelope spectrum in Fig. 14(b). A 30 dB SNR noise is added to simulate ambient contamination of the signal. In Fig. 14(c), the pulse is swamped, making fault impulses and amplitude trends difficult to observe. The fault order in the envelope spectrum of Fig. 14(d) cannot be recognized.

The ISAM-SFAC method processes the noise-added signal, and the noise component that drowns the signal impulse in the time domain signal of Fig. 15(a) is effectively weakened. The fault order and its harmonics can be identified by observing the order envelope spectrum of Fig. 15(b).

The results of the comparison Methods (1) and (2) are shown in Figs. 16(a) and (b). Method (1) recognizes only double harmonics of the fault order, while Method (2) can only recognize fault order and not the harmonics. Table VIII shows that both comparison methods are less effective than the ISAM-SFAC fault feature enhancement method. Fig. 17 demonstrates the WFCAR index values obtained by processing different signals using comparison methods. It can be observed that ISAM-SFAC has a significant advantage over other methods for different signal types.

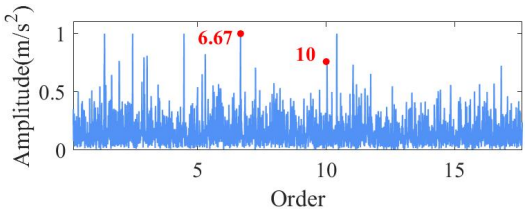


(a)

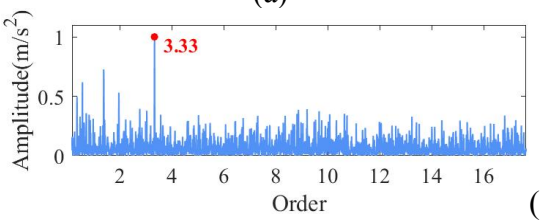


(b)

Fig. 15. Outer race signal after ISAM-SFAC. (a) Time domain waveform; (b) Order envelope spectrum.



(a)



b)

Fig. 16. Outer race signal after comparison methods. (a) Method (1); (b) Method (2).

TABLE VIII. WFCAR values for different methods processing outer race signal

Methods	WFCAR
Original signal	1.20
Method (1)	1.35
Method (2)	2.12
ISAM-SFAC	**3.04**

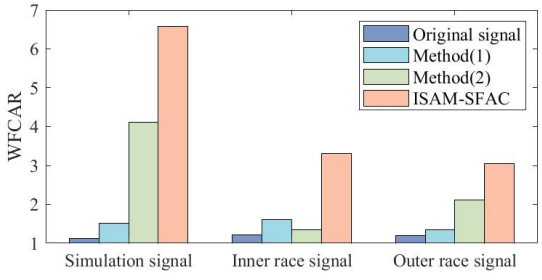


Fig. 17. WFCAR indices processed by different signal comparison methods.

6 Conclusion

In this paper, an improved spectral amplitude modulation based on sparse feature adaptive convolution (ISAM-SFAC) is proposed for variable speed fault diagnosis of bearing. The main conclusions are as follows:

(1) An optimal bi-damped wavelet construction method based on correlation criterion and particle swarm optimization (PSO) is proposed, which is used to analyze the signal impulse information and sparsely learn the fault features.

(2) A convolutional basis pursuit denoising model based on optimal bi-damped wavelet is proposed for sparse reconstruction of signal impulse information. Meanwhile, the model parameter is adaptively selected based on the weighted fault characteristic amplitude ratio (WFCAR).

(3) An improved spectral amplitude modulation method based on kurtosis threshold is proposed to further enhance the fault sparse information. The advantages of the algorithm are verified by simulation signal, fault signal experimental results and comparative analysis.

(4) Research on fault diagnosis for unknown working conditions and full life cycle of speed will be carried out, and the efficiency of fault feature extraction will be further improved in future work.

Acknowledgements

This work is funded by the National Natural Science Foundation of China (Grants No. 52475084 and 52375076) and the Postdoctoral Fellowship Program of CPSF. (Grant No. GZC20230202).

Conflict of interest statement

The authors declare no conflicts of interest.

References

[1] C. Li, K. Luo, L. Yang, S. Li, H. Wang, X. Zhang and Z. Liao, “A Zero-Shot Fault Detection Method for UAV Sensors Based on a Novel CVAE-GAN Model,” *IEEE Sens. J.*, vol. 24, pp. 23239-23254, 2024.

[2] C. Li, S. Li, Y. Feng, K. Gryllias, F. Gu and M. Pecht, “Small data challenges for intelligent prognostics and health management: a review,” *Artif Intell Rev*, vol. 57, p. 214, 2024.

[3] X. Zhang, C. Li, C. Han, S. Li, Y. Feng, H. Wang, Z. Cui and K. Gryllias, “A personalized federated meta-learning method for intelligent and privacy-preserving fault diagnosis,” *Adv Eng Inform*, vol. 62, p. 102781, 2024.

[4] P. Wang, Y. Luo, L. Gong and Y. Zhou, “Rolling Bearing Fault Diagnosis Based on Multiscale Block Convolution Neural Network,” *Meas Sci Technol*, vol. 2508, p. 012019, 2023.

[5] Z. Huang, X. Song, Z. Liao and B. Jia, “Bearing Fault Feature Enhancement and Diagnosis Based on Savitzky–Golay Filtering Gramian Angular Field,” *IEEE Access*, vol. 12, pp. 87991-88005, 2024.

[6] X. Wang, X. Jiang, Q. Song, J. Liu, J. Guo and Z. Zhu, “Spectral structure inducing efficient variational model for

enhancing bearing fault feature" *Signal Process*, vol. 216, p. 109304, 2024.
[7] M. Zhao, J. Lin, X. Xu and Y. Lei, "Tacholess envelope order analysis and its application to fault detection of rolling element bearings with varying speeds," *Sensors*, vol. 13, pp. 10856-10875, 2013.
[8] Y. Wang, G. Xu, A. Luo, L. Liang and K. Jiang, "An online tacholess order tracking technique based on generalized demodulation for rolling bearing fault detection," *J Sound Vib*, vol. 367, pp. 233-249, 2013.
[9] A. Moshrefzadeh, A. Fasana and J. Antoni, "The spectral amplitude modulation: a nonlinear filtering process for diagnosis of rolling element bearings," *Mech Syst Sig Process*, vol. 132, pp. 253–276, 2019.
[10] G. Zhang, Y. Wang, Y. Qin and B. Yang, "An Improved Symplectic Geometry Mode Decomposition Method for Rolling Bearing Fault Diagnosis under Variable Speed Conditions," *2021 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD)*, pp. 1-6, 2021.
[11] Z. Wu, J. Niu, S. Lu, Y. Liu and F. Liu, "Fault Diagnosis of Wind Turbine Bearing Using Variational Nonlinear Chirp Mode Decomposition and Order Analysis," *2020 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD)*, pp. 479-484, 2020.
[12] Y. Sun and J. Yu, "Adaptive Sparse Representation-Based Minimum Entropy Deconvolution for Bearing Fault Detection," *IEEE T Instrum Meas*, vol. 71, p. 3513010, 2022.
[13] L. He, C. Yi, Q. Zhou and J. Lin, "Fast Convolutional Sparse Dictionary Learning Based on LocOMP and Its Application to Bearing Fault Detection," *IEEE T Instrum Meas*, vol. 71, p. 3519012, 2022.
[14] T. Qiu, W. Huang, Y. Liao, C. Ding and J. Wang, "Sparse Optimization Model Based on Sparse Matrix and Singular Value Vector for Fault Diagnosis of Rolling Bearings," *IEEE T Instrum Meas*, vol. 73, p. 3516309, 2024.
[15] Z. Cao, J. Dai, W. Xu and C. Chang, "Sparse Bayesian Learning Approach for Compound Bearing Fault Diagnosis," *IEEE T Ind Inform*, vol. 20, pp. 1562-1574, 2024.
[16] Z. Liu, G. Cai, H. Wei, Y. Hu and S. Wang, "Adaptive Multiscale Boosting Dictionary Learning for Bearing Fault Diagnosis," *IEEE T Instrum Meas*, vol. 73, p. 3513316, 2024.
[17] Z. Ma, M. Zhao, X. Dai and H. Bi, "Compound fault diagnosis of wind turbine bearing under ultra-low speed operations using generalized sparse spectral coherence," *Mech Syst Signal Pr*, vol. 208, p.111027, 2024.
[18] B. Li, C. Li and J. Liu, "Incipient detection of bearing fault using impulse feature enhanced weighted sparse representation," *Tribol Int*, vol. 184, p. 108467, 2023.
[19] R. Yao, H. Jiang, C. Yang, H. Zhu and K. Zhu, "Multiband weights-induced periodic sparse representation for bearing incipient fault diagnosis," *ISA T*, vol. 136, pp. 483-502, 2023.
[20] Q. Li, "New Approach for Bearing Fault Diagnosis Based on Fractional Spatio-Temporal Sparse Low Rank Matrix Under Multichannel Time-Varying Speed Condition," *IEEE T Ind Inform*, vol. 71, p. 3502012, 2022.
[21] G. Yu, X. Huang, T. Lin and H. Dong, "A nonlinear time–frequency tool for machinery fault diagnosis under varying

speed condition," *Mech Syst Signal Pr*, pp. 186, p. 109849, 2023.
[22] J. Shi, N. Wu and X. Jiang, "Transient morphology analysis and sparse representation for bearing fault diagnosis under variable speed condition," *Int. J. Mechatronics and Manufacturing Systems*, vol. 11, pp.17-35, 2018.
[23] F. Hou, I. Selesnick, J. Chen and G. Dong, "Fault diagnosis for rolling bearings under unknown time-varying speed conditions with sparse representation," *J Sound Vib*, vol. 494, p.115854, 2021.
[24] J. Li, J. Tao, W. Ding, J. Zhang and Z. Meng, "Period-assisted adaptive parameterized wavelet dictionary and its sparse representation for periodic transient features of rolling bearing faults," *Mech Syst Signal Pr*, vol. 169, p. 108796, 2022.
[25] L. Zhang, L. Zhao and C. Wang, "Sparse representation by novel cascaded dictionary for bearing fault diagnosis using bi damped wavelet," *Int J Adv Manuf Tech*, vol. 124, pp. 2365–2381, 2023.
[26] A. Oppenheim and J. Lim, "The importance of phase in signals," *Proc. IEEE*, vol. 69, pp.529–541, 1981.
[27] P. Borghesani, P. Pennacchi, R. B. Randall, N. Sawalhi and R. Ricci, "Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions," *Mech Syst Signal Pr*, vol. 36, pp. 370–384, 2013.
[28] F. Veshki and S. Vorobyov, "Efficient ADMM-Based Algorithms for Convolutional Sparse Coding," IEEE Signal Proc Let, vol. 29, pp. 389-393, 2022.
[29] C. Han, W. Lu, H. Wang, L. Song and L. Cui, "Multistate fault diagnosis strategy for bearings based on an improved convolutional sparse coding with priori periodic filter group," *Mech Syst Signal Pr*, vol. 188, p. 109995, 2023.
[30] Y. Li, J. Wang, D. Feng, M. Jiang, C. Peng, X. Geng and F. Zhang, "Bearing fault diagnosis method based on maximum noise ratio kurtosis product deconvolution with noise conditions," *Measurement*, vol. 221 p. 113542, 2023.
[31] R. Bai, Z. Meng, Q. Xu and F. Fan, "Fractional Fourier and time domain recurrence plot fusion combining convolutional neural network for bearing fault diagnosis under variable working conditions," *Reliab Eng Syst Safe*, vol. 232, p. 109076, 2023.
[32] Z. Wang, J. Yang, Y. Guo, T. Gong and Z. Shan, "Positive role of bifurcation on stochastic resonance and its application in fault diagnosis under time-varying rotational speed," J Sound Vib, vol. 537, p. 117210, 2022.
[33] Z. Wang, J. Yang and Y. Guo, "Unknown fault feature extraction of rolling bearings under variable speed conditions based on statistical complexity measures," *Mech Syst Signal Pr*, vol. 172, p. 108964, 2022.
[34] H. Huang and N. Baddour, "Bearing vibration data collected under time-varying rotational speed conditions," *Data in Brief*, vol. 21, pp. 1745-1749, 2018.