



# Sparse representation by novel cascaded dictionary for bearing fault diagnosis using bi-damped wavelet

Long Zhang<sup>1</sup> · Lijuan Zhao<sup>1</sup> · Chaobing Wang<sup>1</sup>

Received: 22 May 2022 / Accepted: 26 November 2022 / Published online: 12 December 2022  
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

## Abstract

Vibration-based bearing condition monitoring of rotating machinery is of great importance for improving production efficiency and ensuring operational safety in the manufacturing industry. Sparse representation is able to effectively extract inherent impulse features from fault vibration signals corrupted by noise and harmonic interference, of which the performance is directly determined by dictionaries. In this study, the typical drawbacks of commonly used dictionaries are addressed using a novel cascaded dictionary. Period-assisted bi-damped wavelets with specific shapes are employed as the initial dictionary atoms to achieve overall matches with impulse features. Subsequently, the initial atoms are subjected to the K-singular value decomposition (K-SVD) for a secondary learning to obtain a cascaded dictionary that matches the real impulse features globally and locally. Finally, faulty vibration signals are recovered in segments using the cascaded dictionary and orthogonal matching pursuit (OMP). The results on the signals from the simulations, experiments, and real-world engineering confirm that the proposed cascaded dictionary consistently outperforms three other leading methods. Furthermore, the proposed cascaded dictionary is proved to be suitable for practical engineering diagnosis because of its outstanding anti-noise capabilities and self-adaptability.

**Keywords** Sparse representation · Cascaded dictionary · Bi-damped wavelet · Fault diagnosis · Feature extraction

## 1 Introduction

Rolling bearings play a paramount role in machine tools and are primarily employed to support rotations and carry the radical dynamic loads of spindles [1–3]. As a safety-critical component, rolling bearings usually work under harsh operating conditions, and are therefore prone to various failures, which may lead to production efficiency reduction and economic losses in the manufacturing industry [4]. Accurate fault diagnosis of rolling bearings is thus of great importance for ensuring the high safety and high-efficiency operation of manufacturing equipment. When a localised

bearing fault occurs, the defect passes its mating surface resulting in a short force pulse. As the bearing rotates in a continuous fashion, the transient pulse repeatedly arises, and thus, periodic impulse features take place in the vibration signal. Hence, the bearing fault diagnosis technique based on vibration signal analysis has attracted significant attention in recent years [5–8]. Nevertheless, vibration signals are normally contaminated by noise and harmonic interference due to poor operating conditions.

Numerous effective methods have sprung up for extracting periodic fault features from vibration signals. Commonly applied vibration signal analysis methods include the short-time Fourier transform (STFT) [9, 10], wavelet transform [11, 12], ensemble empirical mode decomposition (EEMD) [13], and tuneable Q-factor wavelet transform (TQWT) [14, 15]. However, these methods suffer from threshold setting and high computational costs, as well as the requirement of abundant human expertise. Recently, sparse representation, emerging as an active and valid signal analysis approach, has garnered considerable attention because of its outstanding performance in feature extraction.

---

✉ Long Zhang  
longzh@126.com  
Lijuan Zhao  
zlj1813871956@126.com  
Chaobing Wang  
chaobing@163.com

<sup>1</sup> Key Laboratory of Conveyance and Equipment of the Ministry of Education of China, East China Jiaotong University, Nanchang 330013, China

Sparse representation has two key components: dictionary construction, and the determination of sparse coefficients. The matching pursuit (MP) [16] and basis pursuit (BP) algorithms [17] are two typical approaches employed to determine sparse coefficients. The advantage of the MP algorithm over the BP algorithm is that the former is computationally simpler and easier to implement [18]. The orthogonal matching pursuit (OMP) algorithm is an improved derivative of the MP algorithm, which has been extensively utilised in the determination of sparse coefficients. For example, Li et al. [19] applied the OMP algorithm and a self-adaptive complete dictionary to obtain a sparse signal. Song et al. [20] proposed a cluster contraction stagewise OMP (CcStOMP) algorithm to detect the fault features of rolling bearings. In fact, the coefficient solving algorithm mainly affects the speed of feature reconstruction, whereas it is the dictionary which truly underlies the accuracy of feature extraction [21]. Hence, designing an appropriate dictionary is essential to fault feature extraction.

Among the existing dictionaries, wavelet parameter dictionaries are extensively explored because of the flexibility and variability of wavelet waveforms. Notably, the similarity between the dictionary atoms and fault impulse directly determines the performance of feature extraction [21]. The unit impulse response, Laplace wavelet, and Morlet wavelet are widely used as mathematical models of wavelet dictionary atoms, where the first two are both single-sided and damped exponential functions. By coincidence, a bearing fault-induced impulse response is generally acknowledged to be single-sided attenuation, which leads a wide application of such two wavelets to bearing fault feature extraction. For instance, Jiang et al. [21] used a unit impulse response function as the mathematical model of atoms to design a dictionary, which was employed in conjunction with the MP algorithm to realise fault feature extraction for bearings. Li et al. [22] applied the Laplace wavelet dictionary and a secondary selection-based OMP (SS-OMP) algorithm to successfully detect the fault features of rolling bearings, whilst Sun et al. [23] also utilised the Laplace wavelet parametric dictionary to extract fault features of rolling bearings. Meanwhile, Qin et al. [24] constructed a new impulsive wavelet as the dictionary atom model for a sparse representation of rolling bearing fault-induced impulses. On the contrary, the distinctive feature of the Morlet wavelet is that its waveform has a shape similar to that of a gear fault-induced impulse response with double-sided attenuation. Hence, Fan et al. [25] constructed a parameter dictionary to diagnose gearboxes based on Morlet wavelets, whereas Wang et al. [26] used the Morlet wavelet as a transient model and identified the model parameters via a correlation filtering algorithm (CFA) to diagnose gear faults within rotating machinery. In the case of compound faults where gear and bearing faults co-exist, Deng et al. [27] introduced the Laplace and Morlet

wavelets as the mathematical models of dictionary atoms to design corresponding dictionaries for diagnosing bearing and gear faults simultaneously. He et al. [28] designed two types of dictionaries for the sparse representation of gear and bearing faults, where the mathematical models of their dictionary atoms were a steady harmonic and unit impulse response function.

The review of the previous literature indicates that the wavelet mathematical model used for sparse representation has a significant effect on the fault feature extraction. However, the commonly used wavelet functions exhibit single-sided attenuation or bilateral symmetric attenuation because of a single damping ratio. In fact, the impulse response components in the real-life faulty bearing vibration signal generally exhibit bilateral asymmetric attenuation because of the influence of such factors as fault mechanism, transmission paths, and sensor characteristics. Therefore, such wavelet functions fail to be used to good effect when serving as mathematical models for dictionary atoms in practical engineering applications. To address the above issues, this work makes use of a bi-damped wavelet as the mathematical model of an atom to construct an initial parameter dictionary, where the bi-damped wavelet is able to take an asymmetric bilateral shape resembling the fault-induced impulse response [29]. As such, this initial dictionary is expected to achieve a global match with the fault-induced impulses in vibration signals.

Most wavelet dictionaries are paved by changing the time-shift parameter of an optimal wavelet basis, so that the waveforms of all the atoms are essentially identical. This is also the case of the bi-damped wavelet parameter dictionary. However, the fault-induced impulses in the vibration signals are not identical in terms of amplitudes and oscillation characteristics. A data-driven learning dictionary can automatically update each column of the dictionary according to the fault-induced impulses in the signal at hand. As a typical dictionary learning model, K-singular value decomposition (K-SVD) has been applied widely in fault diagnosis [30–32]. However, when the noise and harmonic components are intensive, the K-SVD algorithm is prone to learning components that are unrelated to fault-induced impulse features.

To improve the global and local matching of the initial dictionary to the fault-induced impulse, we developed a cascaded dictionary construction method. In this approach, a bi-damped wavelet is first employed as the wavelet mathematical model of the atom, which globally matches the fault-induced impulses. Motivated by the periodicity characteristic of the impulses in the vibration signal, the parameter dictionary atom is constructed with period-assisted bi-damped wavelets. Next, the initial bi-damped wavelet parameter dictionary is substituted into the K-SVD algorithm to adjust the amplitude and oscillation characteristics of each atom to achieve a local match with the fault

impulses. Thus, a cascaded dictionary is obtained that takes into account the global and local features of the vibration signals.

The main contributions of this paper are as follows: (1) The cascaded dictionary is proposed to merge the parametric dictionary and the learning one, integrating the power of individual dictionary and overcomes their own shortcomings simultaneously. (2) Period-assisted bi-damped wavelets are employed as the dictionary atoms, considering not only the shape of the impulse response but also the periodicity in real vibration signals. (3) The proposed method can achieve excellent performance in the fault feature extraction of an actual locomotive wheel-set bearing.

The structure of this paper is arranged as follows. In Section 2, the individual methods utilised in the developed approach are briefly described. The implementation details of the proposed approach are introduced in Section 3. In Section 4, the effectiveness of the proposed method is validated by simulation analysis. In Section 5, the performance and superiority of the proposed method are analysed using experimental and actual engineering signals. Finally, concluding remarks are summarised in Section 6.

## 2 Methods

### 2.1 Sparse representation and OMP algorithm

A bearing fault vibration signal typically consists of periodic transient impulse components and background noise. Thus, the observed fault signal  $y$  can be described as.

$$y = h + e \tag{1}$$

where  $h$  represents the periodic transient impulse components and  $e$  denotes the background noise. In accordance with the theory of sparse decomposition, the signal  $y \in \mathbf{R}^{n \times 1}$  can be expanded with an over-complete dictionary  $D \in \mathbf{R}^{n \times M}$ ,  $D = [d_1, d_2, \dots, d_M]$ . Then, the signal  $y$  can be modelled as.

$$y = Dx + \epsilon \tag{2}$$

where  $x \in \mathbf{R}^{M \times 1}$  denotes the sparse coefficient vector and  $\epsilon$  is the reconstruction error. The purpose of sparse representation is to use as few atoms as possible from the over-complete dictionary  $D$  to represent the vibration signal  $y$  with the minimum  $l_0$  norm  $\|x\|_0$ . The sparse representation is defined as.

$$\min_x \|x\|_0, \text{ s.t. } \|y - Dx\|_2^2 < \epsilon \tag{3}$$

Determining the minimum  $l_0$  norm  $\|x\|_0$  is a non-deterministic polynomial hard problem. The OMP algorithm can transform the  $l_0$ -minimisation problem into an  $l_1$ -minimisation problem, as follows:

$$\min_x \|x\|_1, \text{ s.t. } \|y - Dx\|_2^2 < \epsilon \tag{4}$$

To ensure that the solution is optimal, the OMP algorithm performs a Gram-Schmidt orthogonalisation operation on all selected atoms during each iteration. For clarity, a basic description of the OMP algorithm is presented in Table 1.

### 2.2 Cascaded sparse dictionary

For better extraction of the transient impulse responses from a vibration signal submerged in noise and harmonic components, the construction of a dictionary that allows atoms to follow impulse changes and not easily be affected by noise interference is a subject worthy of attention. This paper proposes a cascaded dictionary construction method that satisfies the aforementioned demands and the accuracy of rolling bearing feature extraction is expected to be improved. Considering the asymmetric bilateral attenuation and periodicity of the real-life impulse responses, a period-assisted bi-damped wavelet is employed as the atom to design an initial parameter dictionary. Then, the K-SVD algorithm is utilised to refine the initial parameter dictionary locally. Finally, a cascaded dictionary can be obtained that matches the real impulse response globally and locally.

**Table 1** Description of OMP algorithm

---

<b>Input:</b> signal $y$ , dictionary $D$ , and sparsity threshold $T$
<b>Initialise:</b> residual $r_0 = y$ , iteration number $s = 1$ , initial atom set $D_0 = \emptyset$ , and support index set $\Lambda_0 = \phi$
<b>Procedure:</b>
(1) <b>for</b> $s = 1, 2, \dots, T$ , <b>do</b>
(2) Select the index $\lambda_s$ that solves the optimisation problem: $\lambda_s = \arg_{m=1, \dots, M}^{\max}   \langle r_{s-1}, d_m \rangle  $
(3) Update the index set $\Lambda_s$ : $\Lambda_s = \Lambda_{s-1} \cup \{ \lambda_s \}$ , update initial set of atoms: $D_s = [D_{s-1}, d_{\lambda_s}]$
(4) Search the approximate solution of least squares problem: $\hat{x}_s = \arg_{\hat{x}}^{\min} \ y - D_s \hat{x}\ _2$
(5) Update the residual: $r_s = y - D_s \hat{x}_s$ , iteration time: $s = s + 1$
(6) <b>end for</b>
<b>Output:</b> $\hat{x}_T$

---

### 2.2.1 Bi-damped wavelet

Some researchers consider that the transient impulses generated by a bearing fault exhibit strictly single-sided attenuation characteristics [33, 34], with the impulse model described as Eq. (5):

$$\varphi_{imp}(\omega, \xi, \tau, t) = e^{\frac{-\xi}{\sqrt{1-\xi^2}}\omega(t-\tau)} \sin\omega(t - \tau), t \geq \tau \quad (5)$$

where  $\omega = 2\pi f$  represents the oscillation frequency, which is denoted by  $f$  in the following equations. Meanwhile,  $\tau$  denotes the time shift;  $\xi$  represents the damping ratio, which

is used to control the rate of wavelet attenuation; and  $t$  is the signal duration.

Figure 1 illustrates the commonly used wavelet functions, from which we can see that the Laplace wavelet with single-sided attenuation and the Morlet wavelet with double-sided symmetric attenuation match poorly with the fault-induced double-sided asymmetric attenuation impulses as shown in Fig. 2, which come from the bearing vibration signals provided by the Western Reserve University. Therefore, based on the characteristics of the bearing fault impulses, and considering the weaknesses of commonly used wavelet functions, this study uses a bi-damped wavelet  $g_{imp}(\omega, \xi, \zeta, \tau, t)$  as the atom mathematical model to construct an initial parameter dictionary, which is expressed as Eq. (6):

$$g_{imp}(\omega, \xi, \zeta, \tau, t) = \begin{cases} K_{imp}(t - \tau, 0) \cdot e^{\frac{-\xi}{\sqrt{1-\xi^2}}\omega(t-\tau)} \cos(\omega(t - \tau)) - K_{imp}(\tau - t, 0) \cdot e^{\frac{\zeta}{\sqrt{1-\zeta^2}}\omega(t-\tau)} \cos(\omega(t - \tau)), & t \in [\tau - W_s, \tau + W_s] \\ 0, & \text{else} \end{cases} \quad (6)$$

$$K_{imp}(t, 0) = \begin{cases} 1, & t > 0 \\ 0, & \text{else} \end{cases}$$

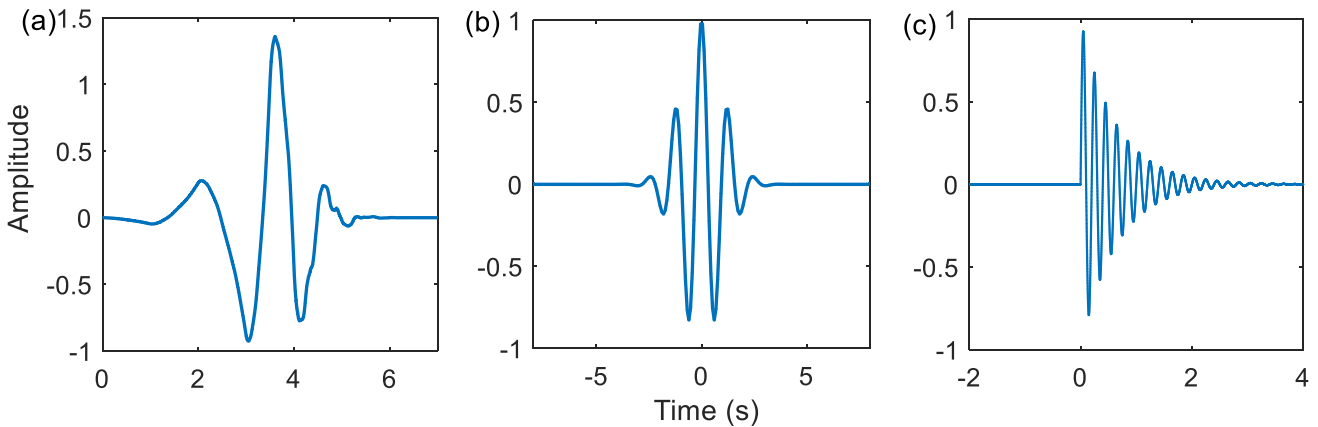


Fig. 1 Commonly used wavelet functions: (a) db4 wavelet; (b) Morlet wavelet; (c) Laplace wavelet

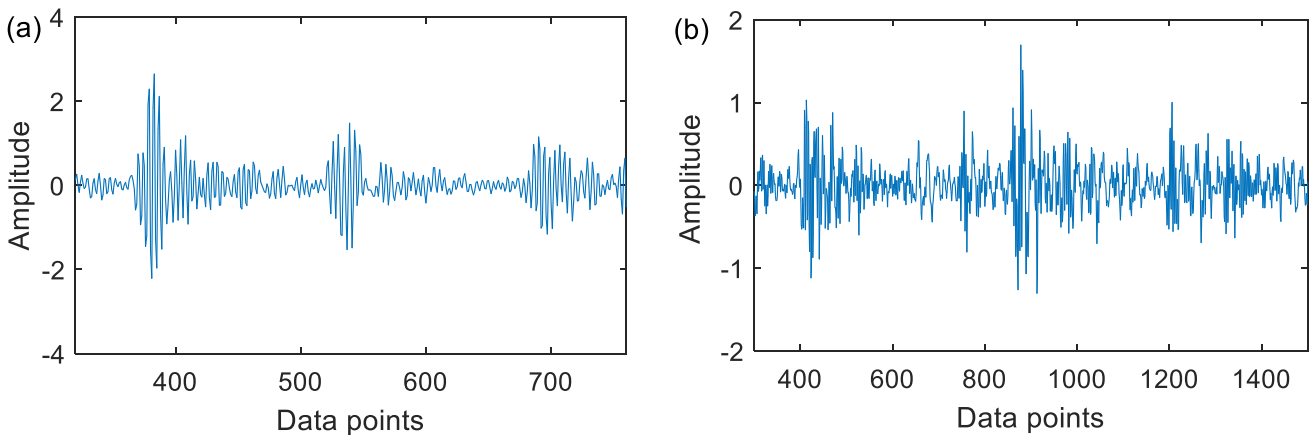
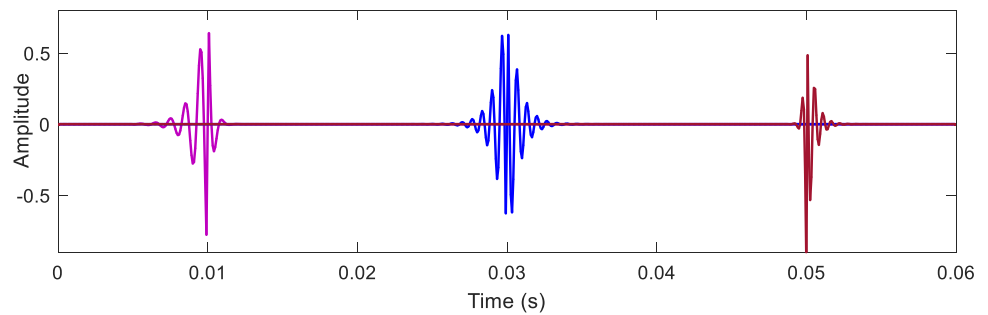


Fig. 2 Bearing fault signals: a outer ring fault signal; b inner ring fault signal

**Fig. 3** Waveforms of bi-damped wavelet with different parameter combinations



where by carefully adjusting the two damping ratios  $\xi$  and  $\zeta$ , we can refine the oscillation attenuations on the left and right sides of the wavelet respectively, so as to match the inherent impulse responses in faulty bearing vibration signals. Figure 3 displays three bi-damped wavelets with the parameter combinations of  $g_{imp}(1000, 0.5, 0.2, 0.01)$ ,  $g_{imp}(1500, 0.15, 0.15, 0.03)$ , and  $g_{imp}(1800, 0.2, 0.5, 0.05)$ . It is seen that a suitable selection of wavelet parameters has the potential to yield a bi-damped wavelet matching perfectly with the fault-induced impulse response.

### 2.2.2 Building bi-damped wavelet dictionary

- (1) Optimise the bi-damped wavelet by CFA with respect to the parameters  $(\omega, \xi, \zeta, \tau)$  ( $\omega = 2\pi f$ ) to approximate fault-induced impulses in the processed signal. First, in accordance with Ref. [34], let  $f \in [0, f_s/2]$ ,  $\tau \in [0, T_c]$ , and both  $\xi$  and  $\zeta \in [0, 1]$ , where  $f_s$  is the sampling frequency, and  $T_c$  is the signal time series used for correlation filtering.

Next, the inner products of the wavelets of various parameter combinations and the signal of interest are calculated by CFA. The inner product measures the correlation degree between the wavelet atom and the signal, which can be defined as Eq. (7).

$$\langle g_{imp}(t), y(t) \rangle = \|g_{imp}(t)\|_2 \|y(t)\|_2 \cos(\theta) \tag{7}$$

As can be seen from Eq. (7), the inner product is affected by the amplitude of the wavelet  $g_{imp}(t)$  and the signal  $y(t)$ , as well as the angle  $\theta$ . Actually, the smaller the angle  $\theta$  is, the more similar of  $g_{imp}(t)$  and  $y(t)$  will be. To eliminate the amplitude effect of the wavelet and signal, a correlation coefficient  $K_r$  can be defined to quantify the magnitude of the angle, i.e., the similarity level of the wavelet to the signal. The correlation coefficient  $K_r$  is described as Eq. (8).

$$K_r = \cos\theta = \sqrt{2} \frac{|\langle g_{imp}(t), y(t) \rangle|}{\|g_{imp}(t)\|_2 \cdot \|y(t)\|_2} \tag{8}$$

Considering that the process of identifying the optimal parameters is computationally expensive, the

CFA is performed in conjunction with the particle swarm optimisation (PSO) algorithm to speed up the optimisation process. The correlation coefficient  $K_r$  is employed as the fitness function in the PSO algorithm, as shown in Eq. (9). Finally, the optimal parameters are selected as those maximising  $K_r$ . In the PSO algorithm, the main parameters are the particle swarm size  $A$ , maximum number of iterations  $M$ , and acceleration factors  $c_1$  and  $c_2$ , prescribed as  $A = 50$ ,  $M = 10$ , and  $c_1 = c_2 = 1.49445$ .

$$fitness = \max_{\{f, \xi, \zeta, \tau\}} (K_r) \tag{9}$$

- (2) Construct the initial bi-damped wavelet parameter dictionary. After the above parameter optimisation, the optimal bi-damped wavelet  $g_{imp}(i)$  is obtained for the subsequent dictionary atom construction. The length of the optimal wavelet is set to  $L_w = \text{round}(f_s/f_i)$  (where  $f_i$  is the fault characteristic frequency of the bearings). Inspired by the periodicity of fault impulses, period-assisted optimal wavelets are employed as the dictionary atoms. The period number of the optimal wavelet in an atom is set to 4 [29]. Thus, the length of an atom is  $L_a = 4 \times L_w$ . Meanwhile, we can observe from Fig. 3 that the parameter  $\tau$  controls only the position of the bi-damped wavelet on the time axis and has no effect on the wavelet waveform. Thus, paving the optimal wavelet with varied time-shift parameter  $\tau$  gives rise to an initial bi-damped wavelet parameter dictionary.

### 2.2.3 K-SVD algorithm

At present, K-SVD is the most widely used dictionary learning algorithm which can effectively reduce the sparsity of the coefficient matrix. That is to say, the bi-damped wavelet dictionary can represent signal features more effectively, if the K-SVD is further utilised. The K-SVD can be defined as an objective optimisation problem:

$$\min_{D, X} \|Y - DX\|_F^2 \text{ s.t. } \|x_i\|_0 \leq T_0, \text{ for } i = 1, 2, \dots, n \tag{10}$$

where  $Y$  denotes the training sample,  $x_i$  is the column vector of the coefficient matrix  $X$ , and  $T_0$  represents the sparsity constraint factor. The K-SVD algorithm first fixes the dictionary  $D$  and then updates the sparse coefficients  $X$  using the OMP

algorithm. Each column of the dictionary is then sequentially updated during the above optimisation process with reference to the signal features to be extracted. The mathematical formula for updating the dictionary is as follows:

$$\|Y - DX\|_2^2 = \|Y - \sum_{j=1}^P d_j x_T^j\|_2^2 = \|(Y - \sum_{j \neq l} d_j x_T^j) - d_l x_T^l\|_2^2 = \|E_l - d_l x_T^l\|_2^2 \tag{11}$$

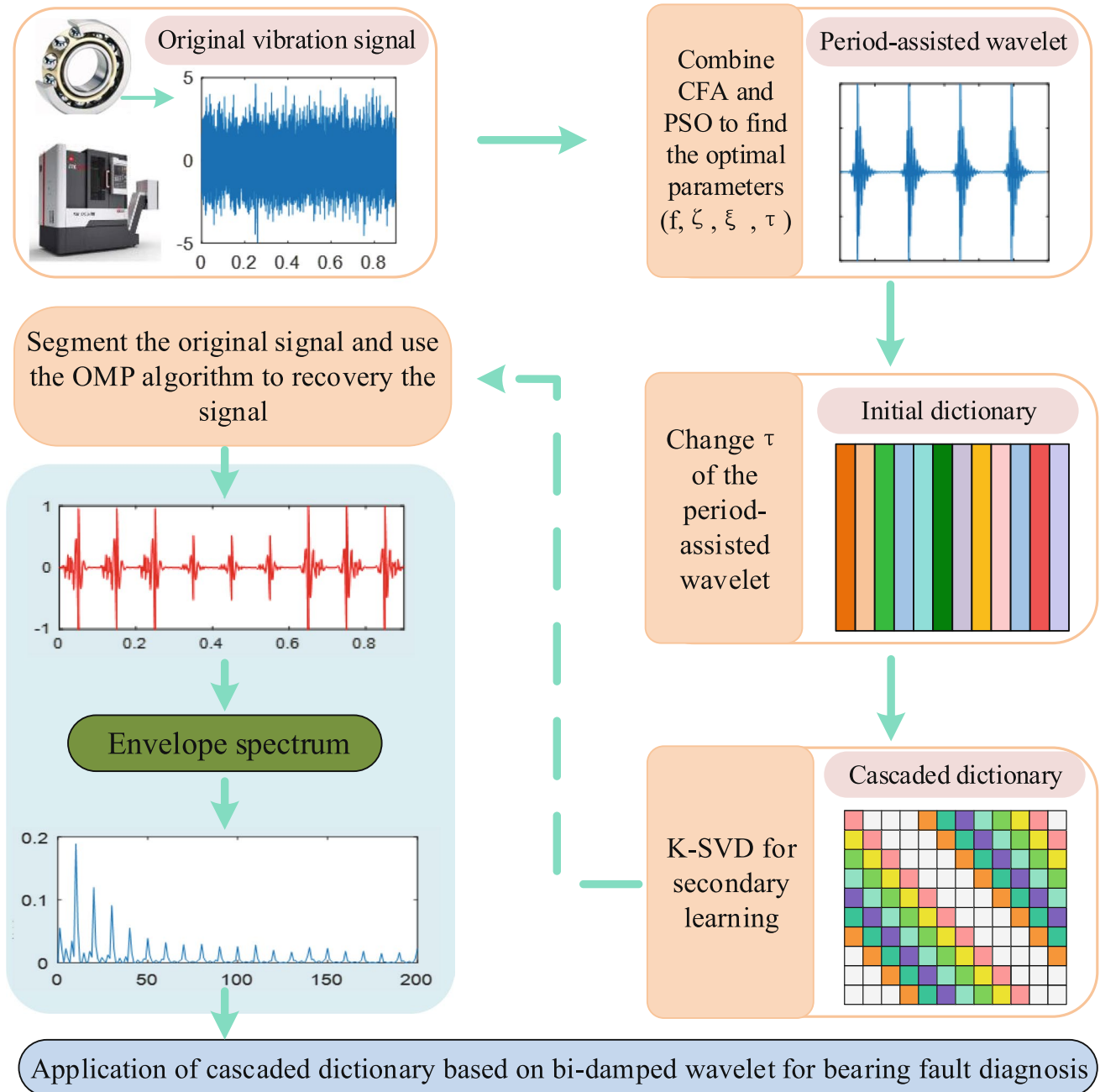
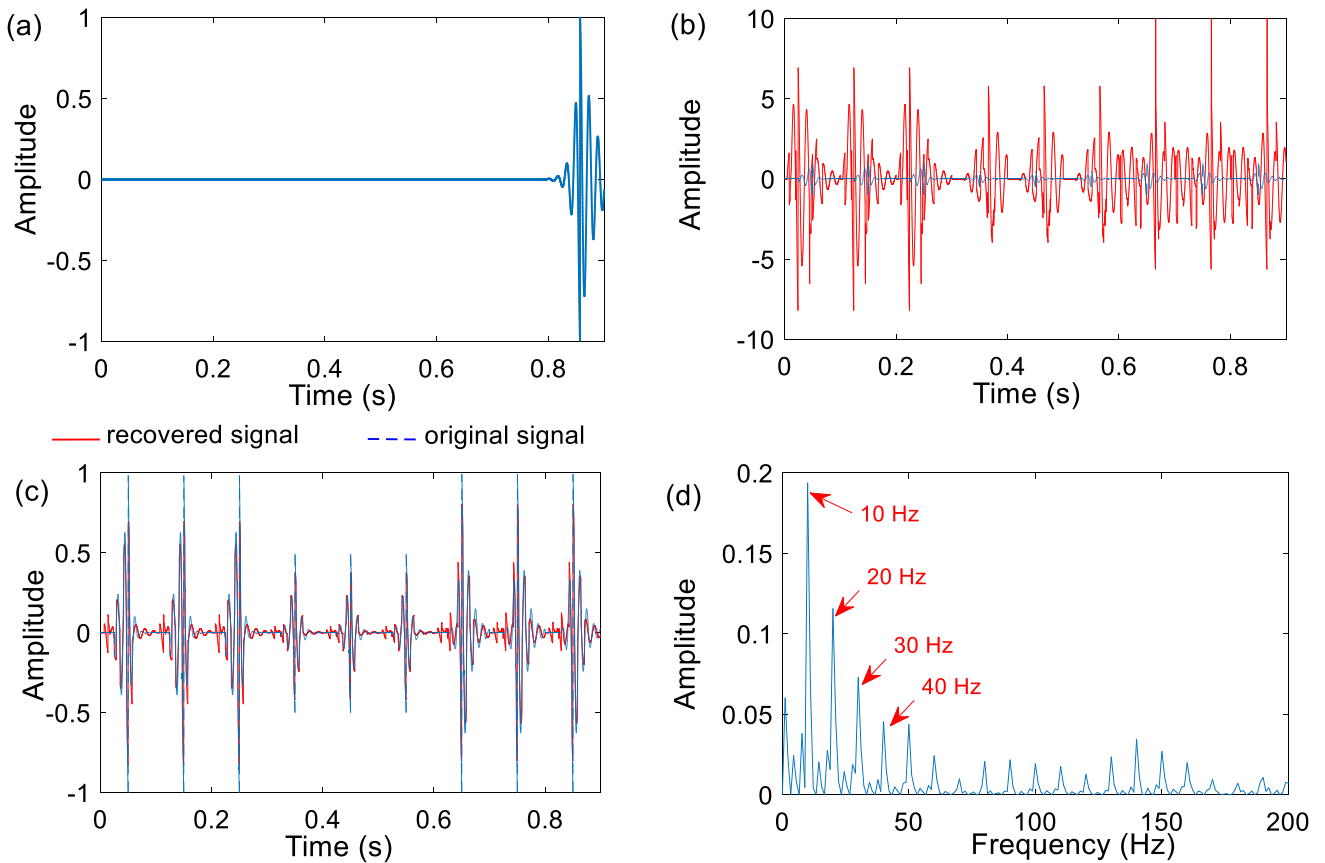
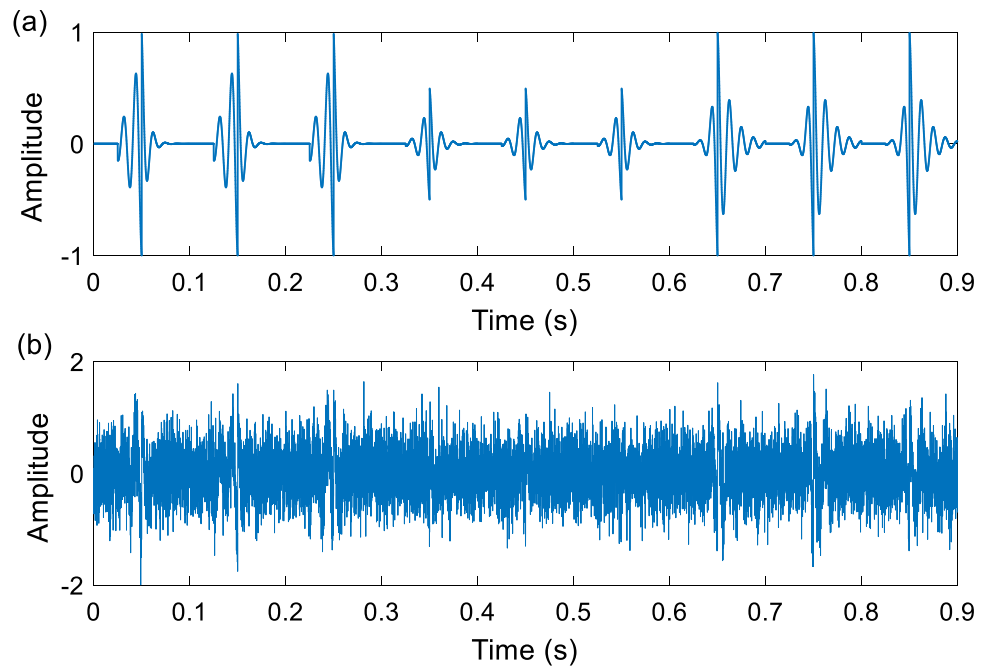


Fig. 4 Flowchart of the proposed approach for fault diagnosis



**Fig. 5** Simulation signal: **a** pure signal; **b** noisy signal



**Fig. 6** The representation results of the simulated signal by using the proposed method: **a** the optimal wavelet atom; **b** the result of wavelet dictionary; **c** the result of cascaded dictionary; **d** the envelope spectrum

where  $E_l$  denotes the residual matrix. Then, the residual matrix is decomposed into various singular values via the singular value decomposition (SVD) method.

$$E_l = U\Delta V^T = \sum_{j=1}^m E_{lj} \quad (12)$$

where  $E_{lj} = o_j u_j v_j^T$  denotes the  $j$ -th singular component. Meanwhile,  $o_j$  denotes the  $j$ -th singular value and  $u_j$  represents the  $j$ -th column vector of the singular matrix  $U$ . Besides,  $v_j$  is the  $j$ -th column vector of the singular matrix  $V$ , and  $\Delta = \text{diag}(o_1, o_2, \dots, o_n)$  is the diagonal matrix. According to the principle of K-SVD,  $d_l$  is updated by the columns of  $U$ , and the solution to  $x^l$  is  $o_l v_l^T$ . All the atoms  $d_l$  ( $l = 1, 2, \dots, L$ ) in  $D$  are updated sequentially.

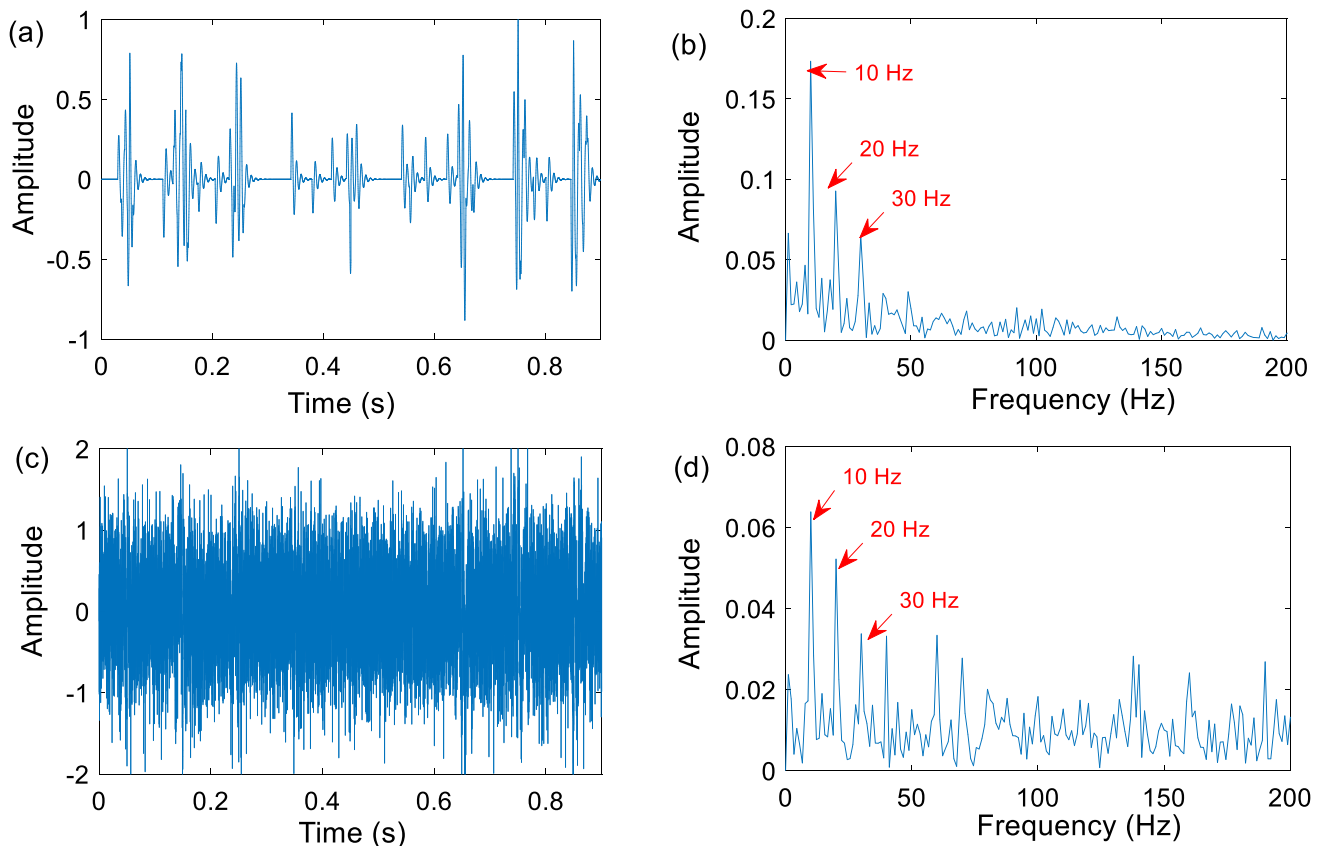
The K-SVD has been recognised as an effective tool to detect inherent fault features. However, the updating principle of the K-SVD algorithm is based on the maximum singular value. When the noise and harmonic components are intense, it will learn some features unrelated to fault impulses [29]. Therefore, through substituting the bi-damped wavelet parameter dictionary into the K-SVD dictionary, the influence of unrelated features can be avoided, and the

dictionary atoms can be learned in the way of a secondary learning to obtain a cascaded dictionary that further suits the local features of the fault vibration signals. Consequently, the cascaded dictionary is promising in the sense that it can achieve better feature extraction in bearing fault diagnosis.

### 3 Proposed procedures for feature extraction

Based on the theory of sparse representation, this paper presents a novel cascaded dictionary construction method using a bi-damped wavelet and the K-SVD algorithm. The steps of the proposed approach are provided in detail as follows, with the corresponding flowchart presented in Fig. 4.

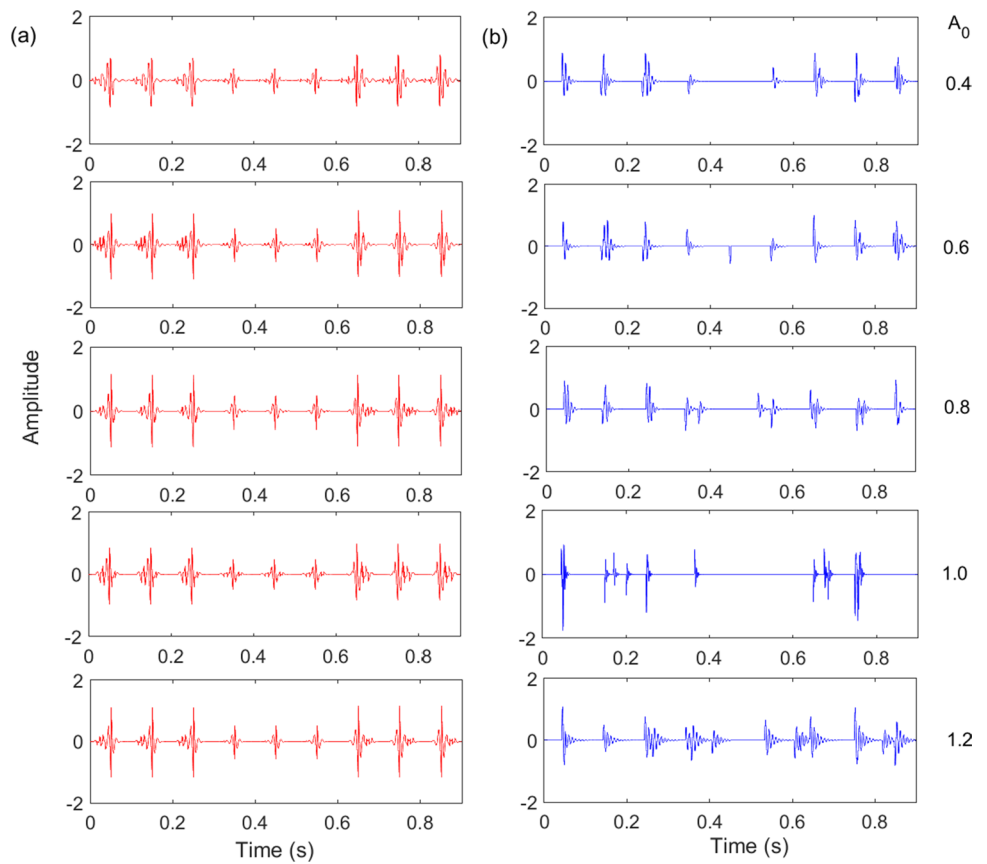
Step 1: Input the original fault vibration signal of a rolling bearing. First, a cascaded dictionary is constructed based on the method mentioned in Section 2.2, which can reduce noise and harmonic interference and learn the characteristics of fault impulses.



**Fig. 7** a, b The processing results of the Laplace wavelet dictionary; c, d the processing results of the K-SVD algorithm



**Fig. 8** Recovered signals at various noise intensities (i.e., the noise amplitudes  $A_0=0.4, 0.6, \dots, 1.2$ ): **a** cascaded dictionary; **b** Laplace dictionary



Step 2: Determine the sparse coefficients. The fault vibration signal is divided into several segments with the lengths equal to those of the atoms in the cascaded dictionary. The OMP algorithm is then applied to obtain a sparse coefficient matrix  $\hat{x} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_i)$ .

Step 3: Detect the fault feature frequency. The fault vibration signals are recovered using the sparse coefficient matrix and the cascaded dictionary. The recovered signal is then demodulated by the Hilbert transform to get the envelope spectra. Finally, bearing fault can be diagnosed by identifying the fault characteristic frequency in the envelope spectra.

### 4 Simulation verification and result analysis

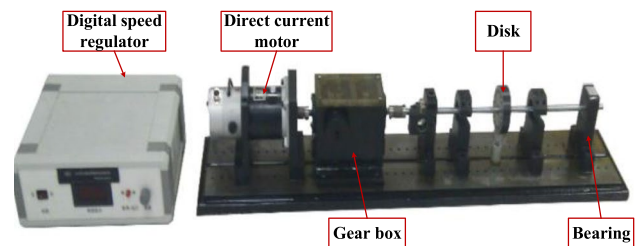
To validate the effectiveness and adaptiveness of the proposed approach to feature extraction, a quasi-periodic signal  $y(t)$  is simulated, which is composed of three kinds of impulse responses  $g_{imp}(t)$  and a random noise  $n(t)$ . The simulated signal  $y(t)$  is formulated as follows:

$$y(t) = \sum_i A_i g_{imp}(t - iT_r) + A_0 n(t) \tag{13}$$

where the cyclic period  $T_r = 0.1s$ , and  $A_i$  and  $A_0$  represent the amplitude of the impulse response and noise, respectively. The parameters for the three types of impulse responses are predefined as  $g_{imp}(80, 0.35, 0.15, 0.05)$ ,  $g_{imp}(80, 0.25, 0.25, 0.35)$ , and  $g_{imp}(80, 0.15, 0.35, 0.65)$ ,

**Table 2** Comparative analysis of reconstructed signals at different SNRs

Noise level ( $A_0$ )	Polluted signal (SNR, dB)	Cascaded dictionary (RMSE)	Laplace dictionary (RMSE)
0.4	-6.532	0.059	0.139
0.6	-10.085	0.050	0.143
0.8	-12.457	0.062	0.158
1.0	-14.545	0.073	0.179
1.2	-16.153	0.064	0.183



**Fig. 9** COINV-1618 test bench

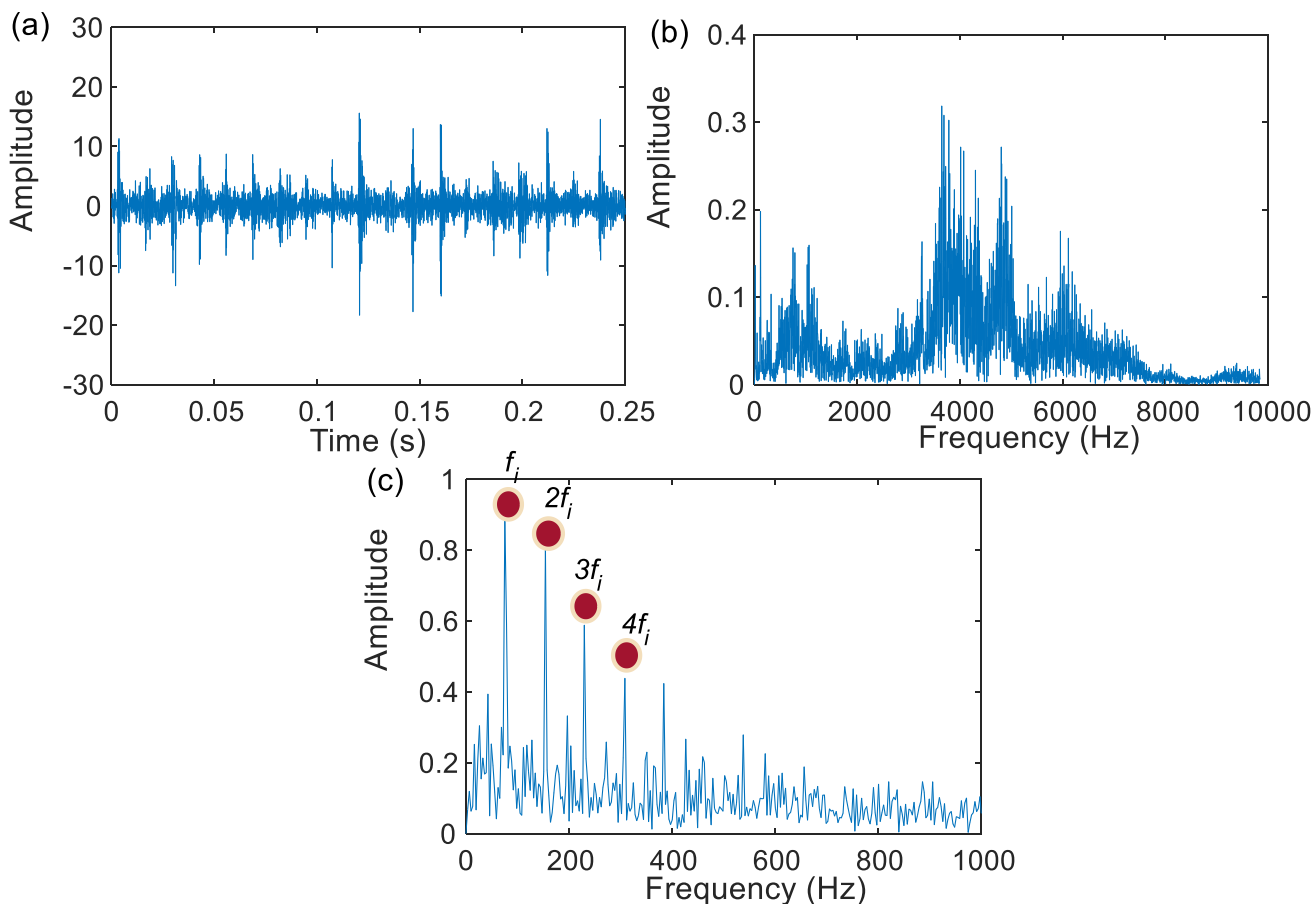


Fig. 10 Inner race fault signal: **a** time-domain waveform; **b** spectra; **c** envelope spectra

and their amplitudes are set as 1, 0.5, and 1, respectively. The sampling frequency is 12 kHz. Figure 5a and b show the waveforms of pure signal and polluted signal with noise level  $A_0 = 0.4$ , respectively.

According to the procedure shown in Fig. 4, the first step is to search for the optimal bi-damped wavelet atom using the CFA-PSO method by maximising the correlation coefficient  $K_r$  (Eq. 8). For the polluted signal in Fig. 5b, the

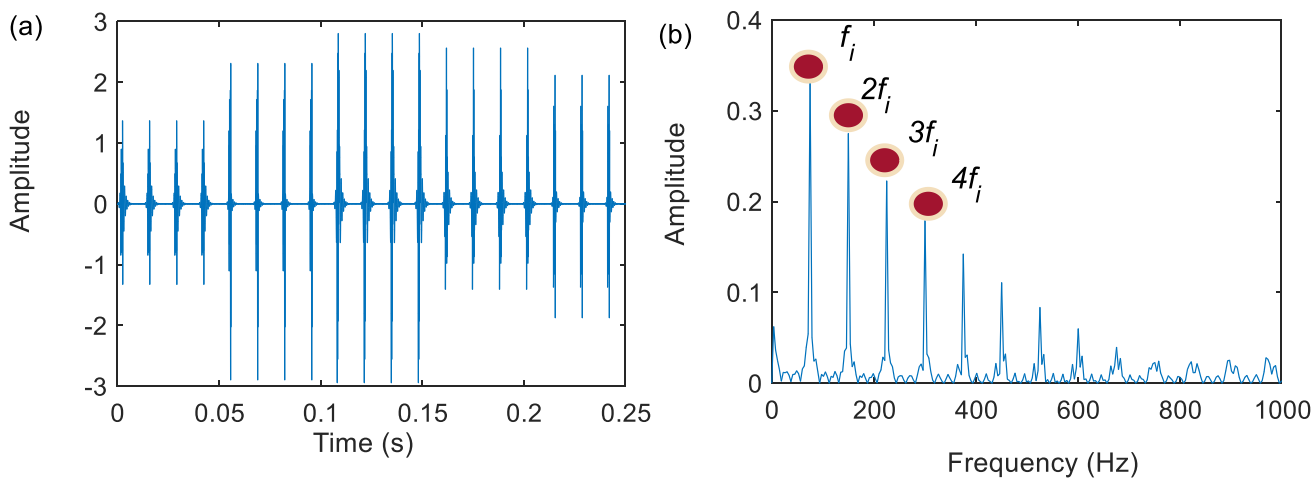


Fig. 11 Analysis results of cascaded dictionary: **a** recovered signal; **b** envelope spectra

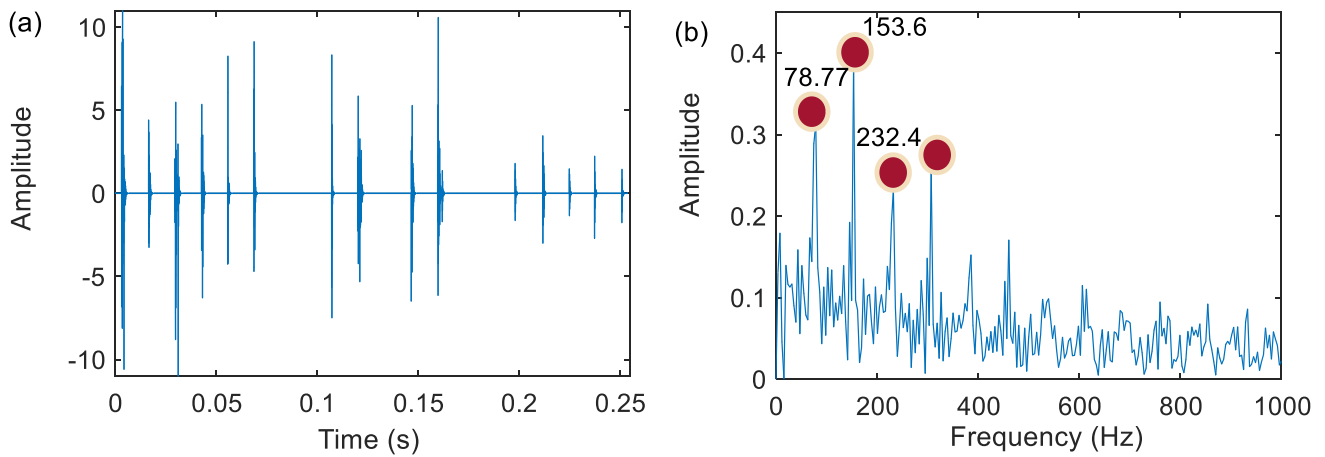


Fig. 12 Analysis results of Laplace dictionary: **a** recovered signal; **b** envelope spectra

identified optimal parameters of the bi-damped wavelet are  $\tilde{f} = 62$ ,  $\tilde{\xi} = 0.1058$ ,  $\tilde{\zeta} = 0.2419$ , and  $\tilde{\tau} = 0.8568$ , with the corresponding atom shown in Fig. 6a. From a global perspective, the optimal wavelet atom has a waveform similar to the impulse responses in the simulated signal. An initial bi-damped wavelet dictionary is then paved by changing the time-shift parameter  $\tau$  to form a period-assisted optimal wavelet. Figure 6b shows the signal reconstructed by the initial wavelet dictionary and original pure signal. It can be observed that there is a large difference between the impulse responses in the two signals in terms of both amplitude and decay characteristics, but the impulse locations are correctly identified. Next, the initial bi-damped wavelet dictionary is substituted into the K-SVD algorithm to perform a secondary learning to form a cascaded dictionary, where the initial wavelet atoms are locally adjusted to match globally and locally with the fault features of interest. Finally,

with the help of the OMP algorithm, Fig. 6c and d present the recovered signal resultant from the cascaded dictionary and its envelope spectrum. As shown in Fig. 6c, almost all impulse responses are perfectly recovered. More specifically, the locations, amplitudes, and decay characteristics of the impulse responses in the recovered signal are extremely identical to those of the pure signal. It is noteworthy that the reconstructed signal is almost completely free of noise. In the envelope spectrum (Fig. 6d), the spectrum peaks indicative of fault frequency and its harmonics are prominent.

For comparison, a Laplace wavelet parameter dictionary is also applied to analyse the polluted signal. The Laplace wavelet is a complex wavelet being single-sided and damped exponential as shown in Fig. 1c. It is reported that the Laplace wavelet perform well on bearing fault feature extraction due to the fact that its waveform resembles the impulses induced by bearing faults. The Laplace wavelet is defined as.

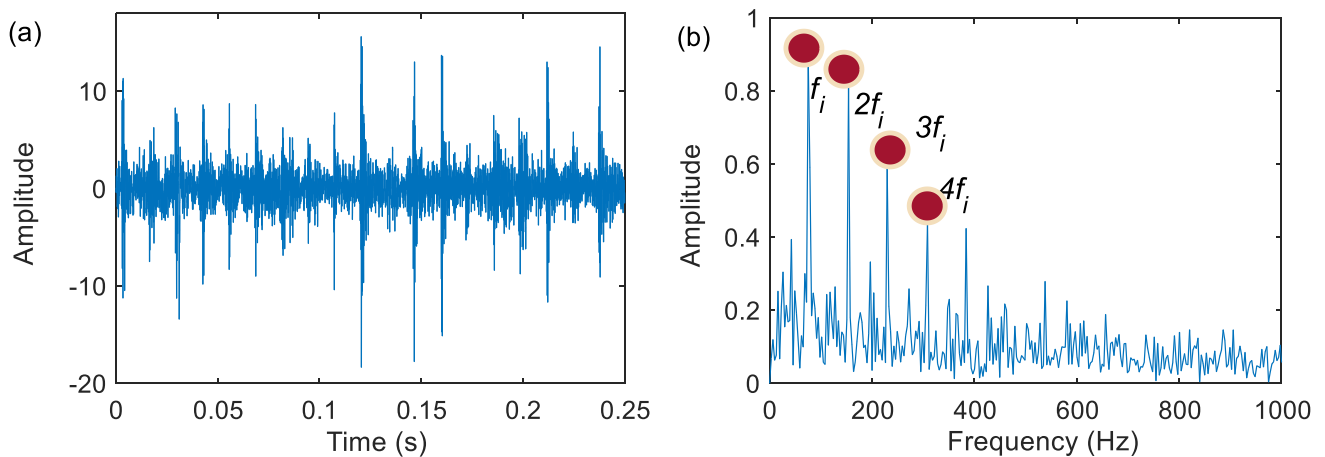


Fig. 13 Analysis results of K-SVD dictionary: **a** recovered signal; **b** envelope spectra

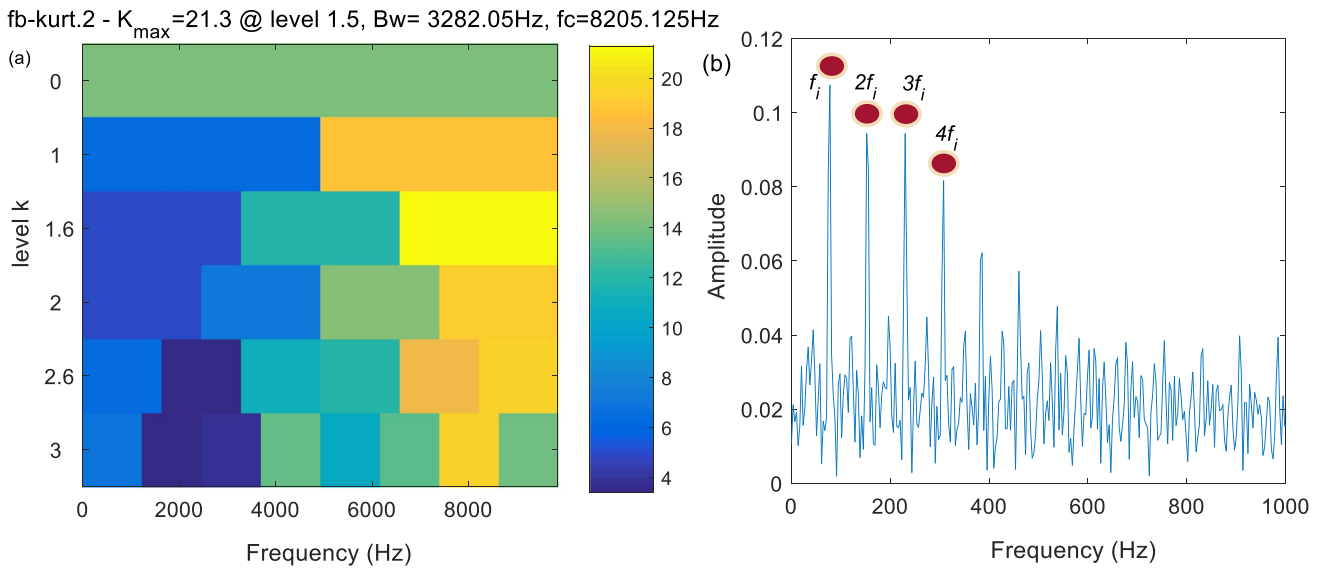


Fig. 14 Analysis results of FK method: **a** FK; **b** envelope spectra

$$\psi(t) = \begin{cases} \frac{-\xi}{\sqrt{1-\xi^2}} 2\pi f(t-\tau) & e^{-j2\pi f(t-\tau)}, t \in [\tau, \tau + W_s] \\ 0, & else \end{cases} \quad (14)$$

The Laplace wavelet is determined by three parameters, i.e.,  $(f, \xi, \tau)$ . To test the validity of the Laplace wavelet as a mathematical model for dictionary atoms, the CFA is also employed to choose its optimal parameters. With the help of the OMP algorithm, Fig. 7a and b show the recovered signal and its envelope spectrum obtained by using the Laplace wavelet dictionary with optimal parameters  $\tilde{f} = 100$ ,  $\tilde{\xi} = 0.2000$ , and  $\tilde{\tau} = 0.0600$ . We can see from Fig. 7a that the impulse period cannot be easily observed from the time-domain signal and the shape of the impulses differentiates greatly from the pure signal, although relatively obvious spectral peaks appear at the fault frequencies (see Fig. 7b).

To further demonstrate the superiority of the cascaded dictionary with a secondary learning, the K-SVD algorithm is employed individually to handle the polluted simulation signal. Figure 7c and d illustrate the results of the K-SVD algorithm. In Fig. 7c, the periodic impulses are severely masked by noise. The feature extraction by the single use of the K-SVD algorithm is absolutely inferior to the proposed approach (Fig. 6d), which is also manifested in its envelope spectrum shown in Fig. 7d. In fact, the K-SVD algorithm along may be unsuited for extracting features directly as it is extremely susceptible to noise and harmonic interferences. Based on the analysis above, it can be inferred that the quality of the recovered signal using the cascaded dictionary is distinctly superior to those obtained by the single use of the Laplace dictionary or K-SVD algorithm.

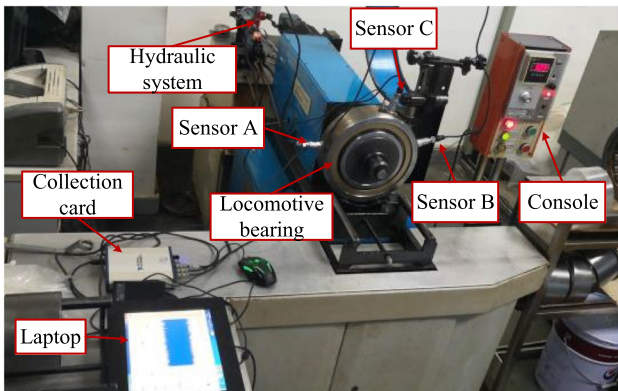


Fig. 15 JL-501 locomotive bearing dynamic testing platform

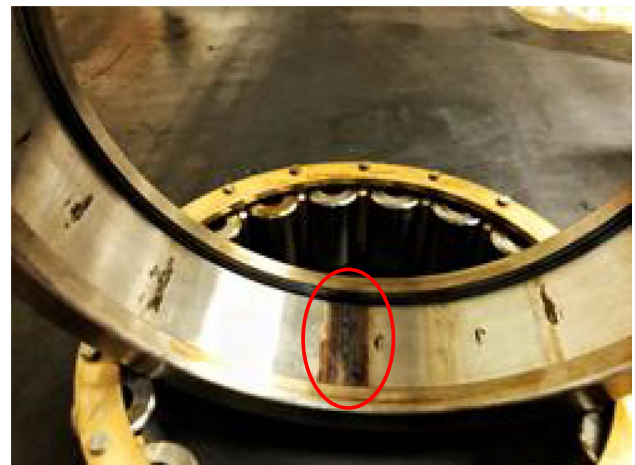


Fig. 16 Wheel-set bearing with outer race fault

**Table 3** Parameters of NJ2232WB bearing

Bearing type	NJ2232WB
Motor speed (rpm)	506
Sampling frequency (Hz)	20 k
Thickness (mm)	110
Pitch diameter (mm)	225
Contact angle $\alpha$ ( $^\circ$ )	0
Outer race diameter (mm)	160
Inner race diameter (mm)	290
Roller number	17

To validate the anti-noise performance of the proposed approach, we increase the amplitude  $A_o$  of the random noise from 0.4 to 1.2 with a step of 0.2, resulting in five simulated signals. The reconstructed signals using the cascaded dictionary are shown in Fig. 8a. In addition, the root-mean-square error (RMSE) is employed to evaluate the quality of the recovered signals with the results listed in Table 2.

In the forgoing comparative study (Fig. 7), the performance of the Laplace wavelet dictionary is obviously superior to that of the K-SVD algorithm. Therefore, the reconstruction results of the Laplace wavelet dictionary other than the K-SVD in various signal-to-noise ratio (SNR) scenarios are shown in Fig. 8b, and the RMSE of the recovered signals are presented in Table 2. It is evident in Fig. 8a that the periodic impulse responses are clearly revealed with a good noise interference cancellation. In contrast, the shape and periodicity of the impulse responses are poorly captured by the Laplace wavelet dictionary (Fig. 8b). From Table 2, we can observe that the RMSEs of the cascaded dictionary are significantly lower than those of the Laplace dictionary. Therefore, the proposed approach has excellent anti-noise performance, and the feature extraction effect of the cascaded dictionary is significantly superior to that of the Laplace dictionary.

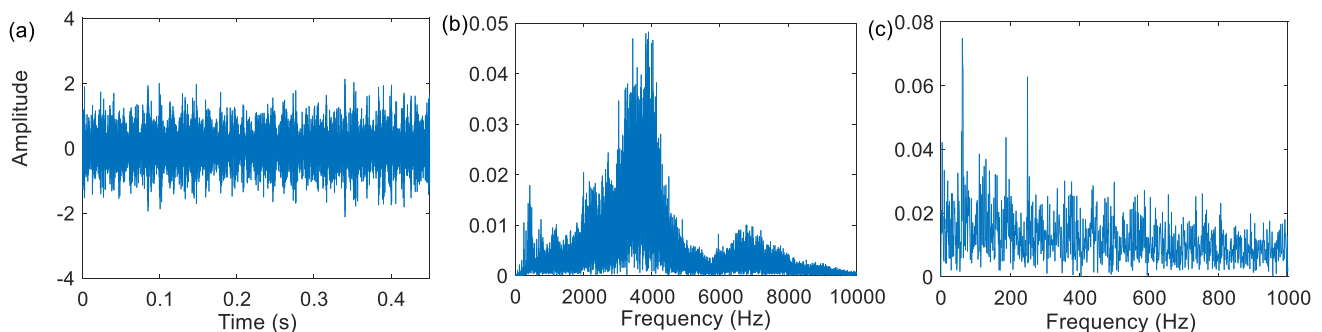
## 5 Application validation

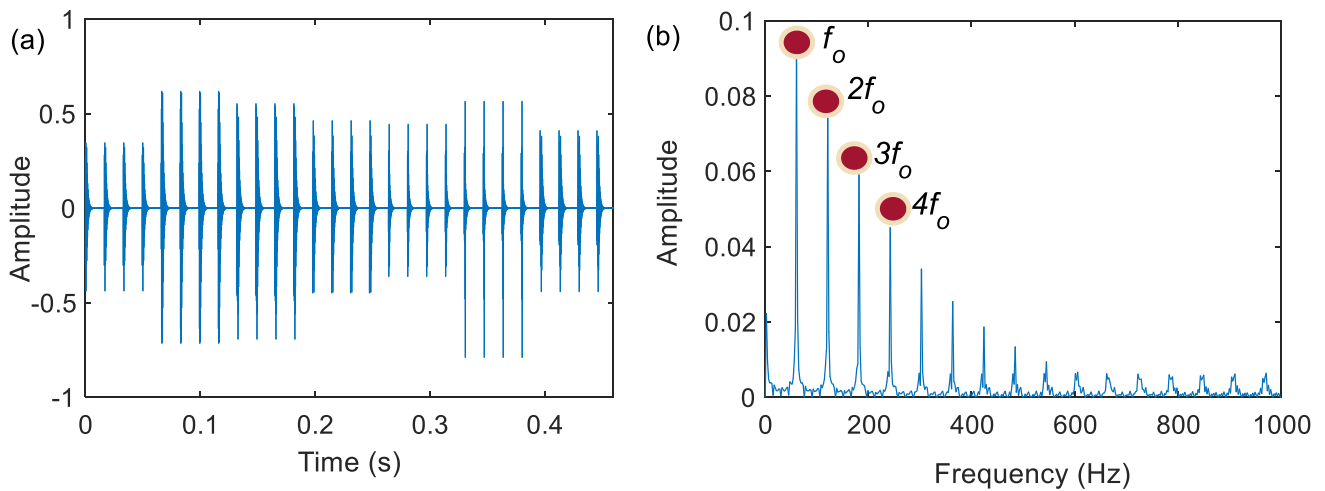
### 5.1 Case 1: inner race fault from laboratory

The experimental signal was obtained from the fault simulation test bench (COINV-1618) shown in Fig. 9, which is provided by the China Orient Institute of Noise & Vibration. Different faulty bearings can be installed on this bench to conduct vibration experiments. The platform includes a base plate, motor, gearbox, rotor, rolling bearings, and acceleration sensor. The bearing type used in the test was 6200Z with eight rolling elements. There was a fracture in the inner raceway of the bearing. The acceleration signals were sampled at the frequency of 19,692.3 Hz as the bearing rotates at 1000 rpm. Correspondingly, the inner race fault characteristic frequency ( $f_i$ ) is 75 Hz. The signal and its spectrum of the inner race fault are shown in Fig. 10a and b, respectively, although  $f_i$  and its harmonics can be identified from the envelope spectra (Fig. 10c), where substantial noise components remain.

Next, the proposed approach is employed to enhance the fault feature information. The identified optimal parameters of the bi-damped wavelet are  $\tilde{f} = 1140$ ,  $\tilde{\xi} = 0.1688$ ,  $\tilde{\zeta} = 0.5095$ , and  $\tilde{\tau} = 0.1271$ . The recovered signal from the cascaded dictionary sparse representation and its envelope spectrum are shown in Fig. 11a and b respectively, where the fault-induced impulses as well as the characteristic frequency and its harmonics are clearly evident.

To illustrate the advantage of the proposed approach in fault feature enhancement, three other signal reconstruction methods involving the Laplace dictionary, K-SVD algorithm, and fast kurtogram (FK) method were applied to analyse the experimental signal. The optimal parameters of the Laplace wavelet are identified as  $\tilde{f} = 1968$ ,  $\tilde{\xi} = 0.02$ , and  $\tilde{\tau} = 0.01$ . Figure 12a and b present the results using the Laplace dictionary. We can observe from Fig. 12a that noise is essentially reduced, but some impulse responses are absent and the characteristic frequencies deviate considerably in Fig. 12b. The

**Fig. 17** Signal of locomotive bearing: **a** waveform; **b** spectra; **c** envelope spectra



**Fig. 18** Locomotive bearing analysis results of cascaded dictionary: **a** recovered signal; **b** envelope spectra

results of the K-SVD are shown in Fig. 13. From the envelope spectra in Fig. 13b, the fault characteristic frequencies can be observed but submerged in rather heavy noise.

Meanwhile, the results of the FK method are shown in Fig. 14. The fault characteristic frequency  $f_i$  and its harmonics up to  $4f_i$  can be observed but with certain noise interference in Fig. 14b. Based on the results, the FK method demonstrates no superiority to the proposed approach. The investigations on the experimental signal prove that the cascaded dictionary can effectively enhance fault-induced impulses and thus has great potential to improve the extraction accuracy of fault features.

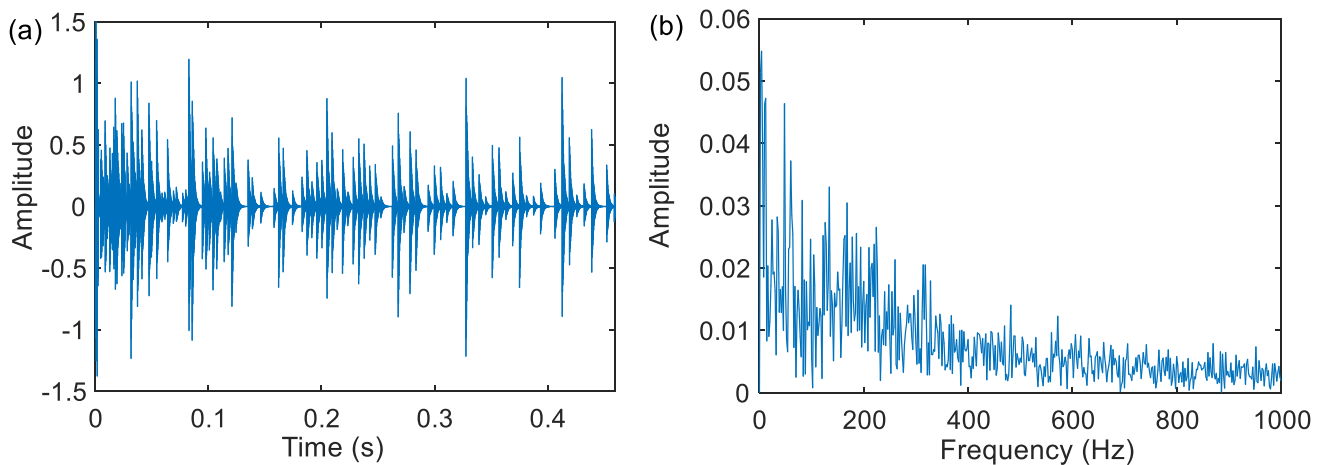
## 5.2 Case 2: outer race fault from engineering

The fault impulses of the experimental signal presented above are easily observable because of an artificial defect

and a relatively low external interference. However, in real engineering applications, the working environment is much more complicated than that in a laboratory, resulting in more noise components in vibration signals.

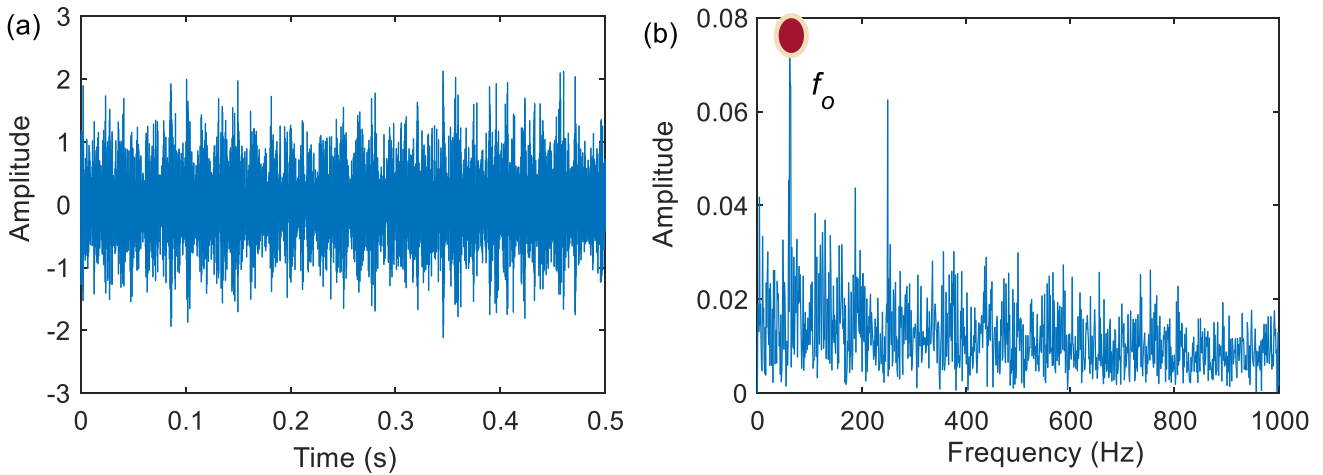
To verify the practicality of the proposed approach, the signal from a wheel-set bearing with a real outer ring fault dismantled from a DF4 diesel locomotive was studied. The bearing testing platform is shown in Fig. 15 with the tested bearing presented in Fig. 16. Meanwhile, Table 3 lists the main parameters of the testing platform and the bearing. Figure 17a implies that the fault impulse responses in the real-life signal are masked by severe noise. Figure 17b shows the spectra of the signal. The envelope spectra in Fig. 17c indicate that the fault characteristic frequency is completely overwhelmed by noise.

Next, the proposed approach is applied to the real-life outer race fault signal. The identified bi-damped wavelet parameters are  $\tilde{f} = 3333$ ,  $\tilde{\xi} = 0.0473$ ,  $\tilde{\zeta} = 0.0334$ , and



**Fig. 19** Locomotive bearing analysis results of Laplace dictionary: **a** recovered signal; **b** envelope spectra





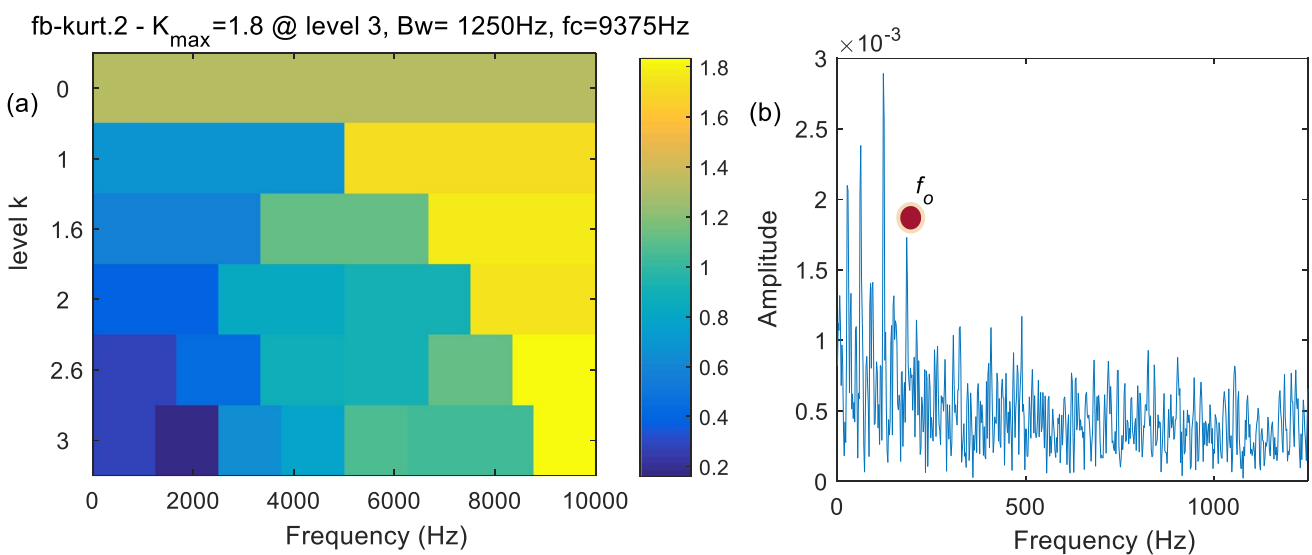
**Fig. 20** Locomotive bearing analysis results of K-SVD dictionary: **a** recovered signal; **b** envelope spectra

$\tilde{\tau} = 0.1045$ . Figure 18a and b show the analysis results using the cascaded dictionary. It can be inferred from Fig. 18a that the fault impulse responses are clearly recovered in time domain without any interference components. The fault characteristic frequency and its harmonics appear distinctly in the envelope spectra shown in Fig. 18b. Undoubtedly, this reliable evidence demonstrates the practicality of the proposed approach in engineering applications.

For comparison, the Laplace dictionary, K-SVD, and FK methods are applied to the same signal. The parameters of the Laplace wavelet are ascertained to be  $\tilde{f} = 4100$ ,  $\tilde{\xi} = 0.025$ , and  $\tilde{\tau} = 0.33$ . The results of the Laplace dictionary are shown in Fig. 19. No periodicity of fault impulses is observed in the recovered signal of the Laplace dictionary (Fig. 19a). Accordingly, there is no distinctive fault

characteristic frequency in the envelope spectra (Fig. 19b). In addition, the results from the K-SVD are given in Fig. 20. Limited by the atom updating principle, K-SVD is extremely susceptible to interference components. Thus, in the envelope spectra, a great deal of irrelevant noise interference is retained. Meanwhile, the results obtained using the FK method, which is known to be an effective tool for feature extraction, are shown in Fig. 21. As shown in Fig. 21b, although the fault characteristic frequency  $f_o$  is presented in the envelope spectra, it is heavily disturbed by noise, and thus, this method is weak to extract fault features from a vibration signal in case of low SNRs.

Therefore, the cascaded dictionary outperforms the other three leading methods in terms of fault feature extraction, and therefore provides an effective tool for bearing fault detection.



**Fig. 21** Locomotive bearing analysis results of FK method: **a** FK; **b** envelope spectra

## 6 Conclusions

Based on the properties of the wavelet parameter dictionary and learning dictionary, we designed a cascaded dictionary, which performs better in sparse representation than the Laplace dictionary, K-SVD algorithm, and FK method. In addition, inspired by the characteristics of the fault-induced impulses in real-life vibration signals, period-assisted bi-damped wavelets are adopted as the atoms for the initial parameter dictionary, which is then substituted into the K-SVD for a secondary learning, so as to achieve a much better agreement with the fault features and an effective noise reduction. The main conclusions can be summarised as follows.

- (1) The anti-noise performance of the cascaded dictionary was verified by simulation signals at different SNRs. Investigations on experimental signals confirm that the cascaded dictionary can effectively enhance impulse features, thus improving the accuracy of bearing fault diagnosis. The engineering application results indicate that the cascaded dictionary has outstanding performance in handling signals with heavy background noise, which poses challenging to most existing sparse representation dictionaries.
- (2) The effectiveness of the proposed approach was validated on a simulated signal, an experimental signal, and an actual engineering signal. Meanwhile, the superiority of the proposed approach was verified by comparing it with three other leading methods. In addition, the proposed approach is quite suited for practical engineering applications.

Form all the results in this paper, we can observe that there are large fluctuations in the amplitudes of the recovered signals. Investigation into this issue is to be conducted in the future.

**Author contribution** Long Zhang: methodology, conceptualisation, resources, writing—review and editing, supervision; Lijuan Zhao: methodology, writing—original draft, software; Chaobing Wang: visualisation, analysis.

**Funding** This work is supported by the National Science Foundation of China (no. 51665013), the Natural Science Foundation of Jiangxi Province (no. 20212BAB204007), and Jiangxi Province Graduate Student Innovation Project (YC2021-S422).

**Data availability** The data used to support the finding of this study are available from the corresponding author upon request.

**Code availability** The algorithm described in this paper is still being studied by the research group, so the code has not been publicly disclosed.

## Declarations

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** The authors agree that the content of this article can be published in the journal.

**Conflict of interest** The authors declare no competing interests.

## References

1. Liang M, Cao P, Tang J (2021) Rolling bearing fault diagnosis based on feature fusion with parallel convolutional neural network. *Int J Adv Manuf Technol* 112(3):819–831. <https://doi.org/10.1007/s00170-020-06401-8>
2. Youcef Khodja A, Guersi N, Saadi MN, Boutasseta, N (2020) Rolling element bearing fault diagnosis for rotating machinery using vibration spectrum imaging and convolutional neural networks. *Int J Adv Manuf Technol* 106(5):1737–1751. <https://doi.org/10.1007/s00170-019-04726-7>
3. Ma H, Li S, Lu J, Gong S, Yu T (2022) Impulsive wavelet based probability sparse coding model for bearing fault diagnosis. *Measurement* 194:110969. <https://doi.org/10.1016/j.measurement.2022.110969>
4. Chen G, Yan C, Meng J, Wang H, Wu L (2021) Improved VMD-FRFT based on initial center frequency for early fault diagnosis of rolling element bearing. *Meas Sci Technol* 32(11):115024. <https://doi.org/10.1088/1361-6501/ac1613>
5. Lu Y, Xie R, Liang S (2019) Bearing fault diagnosis with non-linear adaptive dictionary learning. *Int J Adv Manuf Technol* 102(9):4227–4239. <https://doi.org/10.1007/s00170-019-03455-1>
6. Song Y, Liu J, Chu N, Wu P, Wu D (2019) A novel demodulation method for rotating machinery based on time-frequency analysis and principal component analysis. *J Sound Vib* 442:645–656. <https://doi.org/10.1016/j.sv.2018.11.024>
7. Li H, Liu T, Wu X, Chen Q (2020) An optimized VMD method and its applications in bearing fault diagnosis. *Measurement* 166:108185. <https://doi.org/10.1016/j.measurement.2020.108185>
8. Wang H, Du W (2021) Early weak fault diagnosis of rolling element bearing based on resonance sparse decomposition and multi-objective information frequency band selection method. *J Vib Control*. <https://doi.org/10.1177/10775463211020205>
9. Chui C, Jiang Q, Li L, Lu J (2021) Analysis of an adaptive short-time Fourier transform-based multicomponent signal separation method derived from linear chirp local approximation. *J Comput Appl Math* 396:113607. <https://doi.org/10.1016/j.cam.2021.113607>
10. Liu D, Cheng W, Wen W (2020) Rolling bearing fault diagnosis via STFT and improved instantaneous frequency estimation method. *Procedia Manuf* 49:166–172. <https://doi.org/10.1016/j.promfg.2020.07.014>
11. Chen B, Shen B, Chen F, Tian H, Xiao W, Zhang F, Zhao C (2019) Fault diagnosis method based on integration of RSSD and wavelet transform to rolling bearing. *Measurement* 131:400–411. <https://doi.org/10.1016/j.measurement.2018.07.043>
12. Zhang K, Ma C, Xu Y, Chen P, Du J (2021) Feature extraction method based on adaptive and concise empirical wavelet transform and its applications in bearing fault diagnosis. *Measurement* 172:108976. <https://doi.org/10.1016/j.measurement.2021.108976>

13. Gu J, Peng Y (2021) An improved complementary ensemble empirical mode decomposition method and its application in rolling bearing fault diagnosis. *Digital Signal Process* 113:103050. <https://doi.org/10.1016/j.dsp.2021.-103050>
14. Dogan S, Tuncer T (2021) A novel statistical decimal pattern-based surface electromyogram signal classification method using tunable q-factor wavelet transform. *Soft Comput* 25(2):1085–1098. <https://doi.org/10.1007/s00500-020-05205-y>
15. Zhang S, Liu Z, He S, Wang J, Chen L (2022) Improved double TQWT sparse representation using the MQGA algorithm and new norm for aviation bearing compound fault detection. *Eng Appl Artif Intell* 110:104741. <https://doi.org/10.1016/j.engappai.2022.104741>
16. Mallat SG, Zhang Z (1993) Matching pursuits with time-frequency dictionaries. *IEEE Trans Signal Process* 41(12):3397–3415. <https://doi.org/10.1109/78.258082>
17. Qin Y, Mao Y, Tang B (2013) Vibration signal component separation by iteratively using basis pursuit and its application in mechanical fault detection. *J Sound Vib* 332(20):5217–5235. <https://doi.org/10.1016/j.jsv.2013.04.021>
18. Wang S, Selesnick IW, Cai G, Ding B, Chen X (2019) Synthesis versus analysis priors via generalized minimax-concave penalty for sparsity-assisted machinery fault diagnosis. *Mech Syst Signal Process* 127:202–233. <https://doi.org/10.1016/j.ymssp.2019.02.053>
19. Li J, Wang H, Song L (2021) A novel sparse feature extraction method based on sparse signal via dual-channel self-adaptive TQWT. *Chin J Aeronaut* 34(7):157–169. <https://doi.org/10.1016/j.cja.2020.06.013>
20. Song L, Yan R (2019) Bearing fault diagnosis based on cluster-contraction stage-wise orthogonal-matching-pursuit. *Measurement* 140:240–253. <https://doi.org/10.1016/j.measurement.-2019.03.061>
21. Jiang F, Ding K, He G, Du C (2021) Sparse dictionary design based on edited cepstrum and its application in rolling bearing fault diagnosis. *J Sound Vib* 490:115704. <https://doi.org/10.1016/j.jsv.2020.115704>
22. Li Y, Zheng F, Xiong Q, Zhang W (2021) A secondary selection-based orthogonal matching pursuit method for rolling element bearing diagnosis. *Measurement* 176:109199. <https://doi.org/10.1016/j.measurement.2021.109199>
23. Sun R, Yang Z, Zhai Z, Chen X (2019) Sparse representation based on parametric impulsive dictionary design for bearing fault diagnosis. *Mech Syst Signal Process* 122:737–753. <https://doi.org/10.1016/j.ymssp.2018.12.054>
24. Qin Y (2017) A new family of model-based impulsive wavelets and their sparse representation for rolling bearing fault diagnosis. *IEEE Trans Industr Electron* 65(3):2716–2726. <https://doi.org/10.1109/TIE.2017.2736510>
25. Fan W, Cai G, Zhu Z, Shen C, Huang W, Shang L (2015) Sparse representation of transients in wavelet basis and its application in gearbox fault feature extraction. *Mech Syst Signal Process* 56:230–245. <https://doi.org/10.1016/j.ymssp.2014.10.016>
26. Wang S, Huang W, Zhu Z (2011) Transient modeling and parameter identification based on wavelet and correlation filtering for rotating machine fault diagnosis. *Mech Syst Sig Process* 25(4):1299–1320. <https://doi.org/10.1016/j.ymssp.2010.10.013>
27. Deng F, Qiang Y, Yang S, Hao R, Liu Y (2021) Sparse representation of parametric dictionary based on fault impact matching for wheelset bearing fault diagnosis. *ISA Trans* 110:368–378. <https://doi.org/10.1016/j.isatra.-2020.10.034>
28. He G, Li J, Ding K, Zhang Z (2022) Feature extraction of gear and bearing compound faults based on vibration signal sparse decomposition. *Appl Acoust* 189:108604. <https://doi.org/10.1016/j.apacoust.2021.108604>
29. Li J, Tao J, Ding W, Zhang J, Meng Z (2022) Period-assisted adaptive parameterized wavelet dictionary and its sparse representation for periodic transient features of rolling bearing faults. *Mech Syst Signal Process* 169:108796. <https://doi.org/10.1016/j.ymssp.2021>
30. Qin Y, Zou J, Tang B, Wang Y (2019) Transient feature extraction by the improved orthogonal matching pursuit and K-SVD algorithm with adaptive transient dictionary. *IEEE Trans Industr Inf* 16(1):215–227. <https://doi.org/10.1109/TII.2019.2909305>
31. Yuan H, Wu N, Chen X (2021) Mechanical compound fault analysis method based on shift invariant dictionary learning and improved FastICA algorithm. *Machines* 9(8):144. <https://doi.org/10.3390/machines9080144>
32. Li N, Huang W, Guo W, Gao G (2019) Multiple enhanced sparse decomposition for gearbox compound fault diagnosis. *IEEE Trans Instrum Meas* 69(3):770–781. <https://doi.org/10.1109/TIM.2019.2905043>
33. Cui L, Gong X, Zhang J, Wang H (2016) Double-dictionary matching pursuit for fault extent evaluation of rolling bearing based on the Lempel-Ziv complexity. *J Sound Vib* 385:372–388. <https://doi.org/10.1016/j.jsv.2016.09.-008>
34. He G, Ding K, Lin H (2016) Fault feature extraction of rolling element bearings using sparse representation. *J Sound Vib* 366:514–527. <https://doi.org/10.1016/j.jsv.2015.12.020>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.