IOP Publishing Measurement Science and Technology
Meas. Sci. Technol. **31** (2020) 065103 (12pp) https://doi.org/10.1088/1361-6501/ab7406

A fault information-oriented weighted nuclear norm minimization method and its application to fault feature extraction in a rolling bearing

Baoxiang Wang[1,2], Yuhe Liao[1,2], Yang Lv[1,2] and Xining Zhang[2]

[1] Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, People's Republic of China
[2] Shaanxi Key Laboratory of Mechanical Product Quality Assurance and Diagnostics, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, People's Republic of China

E-mail: yhliao@xjtu.edu.cn

Received 5 December 2019, revised 4 February 2020
Accepted for publication 7 February 2020
Published 2 April 2020

Abstract

How to effectively extract fault features from noisy vibration signals is a key problem to be solved in bearing fault diagnosis. The difficulty here mainly lies in the fact that the components of interest are almost submerged by background noise and external interference. To accurately implement extraction of fault features, nuclear norm minimization (NNM) has been introduced to fault feature extraction due to its powerful denoising capability. However, NNM treats all singular values equally, which may result in components of interest with small singular values being abandoned, especially when a bearing defect is at its incipient stage. Therefore, to enhance the flexibility and effectiveness of NNM in handling practical problems, fault information-oriented weighted nuclear norm minimization (FIWNNM) is proposed for extracting bearing fault features, in which singular values are treated differently by assigning appropriate weights according to the physical meanings. Through the proposed FIWNNM, fault features can be drawn from a noisy signal, even when contaminated by strong noise and external interferences. The analysis results from both simulated and real measured signals are used as the verification of the proposed FIWNNM.

Keywords: fault information-oriented weighted nuclear norm minimization (FIWNNM), fault feature extraction, rolling element bearings, fault diagnosis

(Some figures may appear in colour only in the online journal)

1. Introduction

Rolling element bearings (REBs) are important supporting components and used extensively in rotating machinery, thus their health is an important factor affecting the normal operation of integral rotating machines. In real applications, the operating conditions of bearings is often not ideal and unpredictable factors in the operation process will cause bearings to fail quickly. Therefore, to guarantee the reliable operation of rotating machines, monitoring the condition of rolling bearings has received considerable attention in industrial and academic fields.

With the rapid development of measurement techniques, vibration signal analysis has been one of the most extensively applied technologies for the fault diagnosis of REBs [1–3]. When a localized fault exists on the REB, repetitive transients with a particular period will be generated [4, 5]. The occurrence of repetitive transients in vibration signals is important evidence for rolling bearing faults. However, fault-induced transient features are always obscured by background noise

and other irrelevant interference, especially when the bearing defect is at its incipient stage. Therefore, various methods have been proposed for transient feature enhancement and fault diagnosis in REBs, such as spectral kurtosis (SK) [6, 7], singular value decomposition (SVD) [8, 9], time frequency analysis (TFA) [3, 10], sparse representation (SR) [11, 12] and so on.

In addition to the methods mentioned above, nuclear norm minimization (NNM)-based methods utilizing the low-rank properties exhibited by fault features are increasingly investigated for extraction of fault features. The core idea of NNM-based methods is to recover the inherent low-rank matrix from a noisy matrix. Due to the fact that NNM has been proved to be easily solved and has a great capability of recovering a low-rank matrix [13], it has achieved rapid development in the fields of computer vision and image processing [14–16].

Despite the superior performance of NNM in computer vision and image processing, the use of fault feature extraction based on the low-rank property for bearing fault diagnosis is rare and insufficient. When the bearing has a localized defect, there will be repetitive transients in the vibration signal, and these repetitive transients indicate the low-rank property. Due to its low-rank property, the intrinsic fault features of bearings can be extracted from noisy signal by NNM. Recently, Xin *et al* [17] established a feature extraction model exploiting the low-rank property and applied it to the fault diagnosis of bearings. However, it should be noted that NNM treats all singular values (SVs) equally by shrinking the SVs with the same threshold, while the physical meaning of SVs and their related singular components are totally ignored. In practical application, SVs should be treated differently according to their physical meanings to preserve the data components of interest [18], which illustrates that NNM is not flexible enough to deal with many practical problems. To improve the flexibility and effectiveness of NNM, weighted NNM methods were proposed and successfully applied in the field of image denoising [18–20]. These weighted NNM methods all adopt the weighting strategy with its weights in non-descending order, which can effectively overcome the problem that NNM tends to underestimate nonzero SVs. In addition, this weighting strategy was also applied to fault feature extraction and fault diagnosis of a generator bearing [21], in which the weight is concretely set as the reciprocal of the SV. In the paper, this method is referred to as the SV-based NNM method (SVNNM). Note that the hypothesis of the SVNNM is that the low-rank components of interest are highly related to the large SVs of the matrix, thus a satisfactory fault extraction result can be obtained in the case where the bearing fault features are relatively obvious. However, in practice, bearing fault features are usually weak and submerged by other large-energy components, thus the obtained result of the SVNNM is not satisfactory.

In this paper, to reliably mine the components of interest from noisy vibration signal, an adaptive weighted NNM method called fault information-oriented weighted nuclear norm minimization (FIWNNM) is proposed. In the proposed FIWNNM, the low-dimensional intrinsic subspace of the bearing fault features is extracted through solving a weighted NNM problem. Specifically, in the optimization problem of the proposed FIWNNM, all SVs are treated differently by assigning appropriate weights according to the amount of fault information contained in the corresponding singular components, therefore the components of interest rather than the components with large SVs are preserved. With the adaptive weighting strategy, the fault features of interest can be extracted from vibration signal even when disturbed by heavy noise and harmonic components. Meanwhile, the correlated kurtosis (CK) is selected as an indicator to measure the fault information of each singular component in this paper. To solve the non-convex optimization problem of the proposed FIWNNM, an iterative algorithm is derived and described in detail. Finally, the simulated and experimental signals are processed by the proposed FIWNNM to validate its effectivenessin fault feature extraction, while traditional NNM and SVNNM are used as comparison methods.

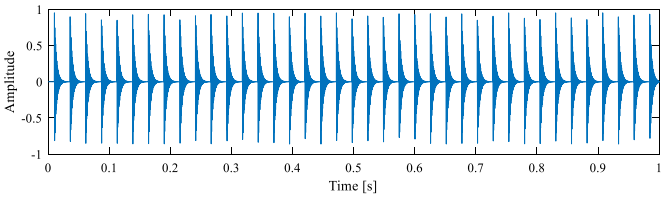

Figure 1. Time domain waveform of repetitive defect impulses.

The main contributions are summarized as follows.

(a) The low-rank property of fault features is explored in detail. Based on this, fault information-oriented weighted nuclear norm minimization (FIWNNM) is proposed for fault feature extraction, in which all SVs are treated differently through an adaptive weighting strategy.
(b) An iterative optimization algorithm is derived to solve the non-convex optimization problem of the proposed FIWNNM.

The paper is organized as follows. Low-rank prior knowledge and the principle of NNM are briefly introduced in section 2. Section 3 proposes FIWNNM for the fault feature extraction and fault diagnosis of rolling bearings. Section 4 presents the rudimentary validation of FIWNNM by utilizing a simulated signal. Subsequently, vibration signals acquired from an experimental rig are used to further verify the performance of the proposed FIWNNM in section 5. In section 6, the conclusions are summarized.

2. Low-rank prior knowledge and nuclear norm minimization

2.1. Low-rank prior knowledge

When a localized defect occurs on a rolling bearing, repetitive impulses with fixed period are produced, as shown in figure 1. However, fault information—namely, fault-induced repetitive impulses—are usually disturbed by other components in the real vibration signal. Concretely, besides the fault

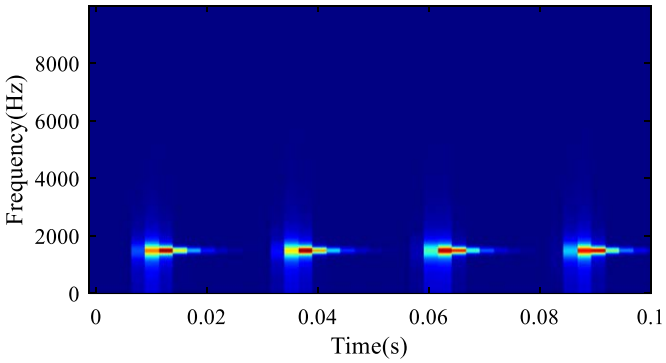

Figure 2. Time–frequency diagram of repetitive impulses.

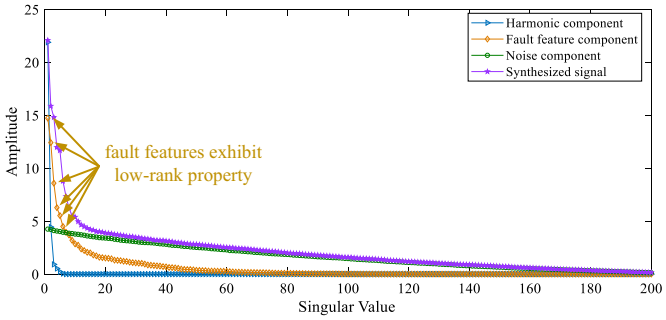

Figure 3. Singular value distribution of the simulated signal.

features induced by bearing fault, harmonic components produced by other components in rotating machines and background noise are also usually found in vibration signals, which undoubtedly increases the signal analysis complexity.

Figure 2 demonstrates the time–frequency diagrams of fault feature components obtained by short-time Fourier transform (STFT). It can be found the fault information in the time–frequency domain is mainly concentrated around the resonance frequency and the STFT coefficients between different frequencies within the resonance frequency band exhibit high similarity. The above phenomenon illustrates the low-rank property of fault features in the time–frequency diagram. To give an intuitive representation of the low-rank property of fault features, the SV distributions of the STFT coefficient matrix of the fault features are illustrated in figure 3. It can be found that most of the energy of the fault feature component is concentrated in a few large singular subspaces, which means there is no significant information loss if those small singular components are ignored [21]. This result further intuitively provides an explanation of the low-rank property of fault features. Moreover, the SV distributions of the STFT coefficients of the harmonic components and noise are also given in figure 3, from which it can be found that harmonic components also exhibit a low-rank property while the Gaussian white noise does not.

2.2. Nuclear norm minimization

The nuclear norm minimization problem is a fundamental problem constructed based on the low-rank property to approximate a low-rank matrix X from its noisy observation Y [22–25]. The corresponding optimization problem is defined as

$$\widehat{X} = \text{prox}_{\lambda\|\bullet\|_*}(Y) = \arg\min_X \|Y-X\|_F^2 + \lambda\|X\|_* \quad (1)$$

where Y is the noisy matrix and $\hat{X}$ is the approximated low-rank matrix, λ is the regularization parameter. Optimization problem equation (1) is convex and its global minimum can be directly obtained utilizing the soft-thresholding function. To be concrete, the solution $\hat{X}$ can be expressed as

$$\hat{X} = US_\lambda(\Sigma)V^T \quad (2)$$

where $Y = U\Sigma V^T$ is the SVD of matrix Y and $S_\lambda(\Sigma)_{ii} = \max(\Sigma_{ii} - \lambda, 0)$ is the soft-thresholding function [26] applied to the diagonal matrix Σ. Therefore, this method of solving optimization problem equation (1) is also known as the SV thresholding method [27].

It can be found from equation (2) that NNM equally shrinks the SVs with the same threshold and the estimation of the low-rank matrix from its noisy observation is realized by retaining a few larger singular components. Due to the low-rank property of bearing fault features, NNM is very suitable for extracting potential fault features from noisy vibration signal when fault features have relatively large amplitude. However, the desirable result may be degenerated when harmonic components that also exhibit the low-rank property are prominent in the vibration signal. To overcome this limitation of NNM in bearing fault feature extraction, fault information oriented weighted nuclear norm minimization (FIWNNM) is proposed and its detailed description is given in section 3.

3. Using the proposed FIWNNM to extractbearing fault features

In this section, the optimization problem of the proposed FIWNNM is described in detail so that the proposed FIWNNM can be used to mine the fault features from noisy vibration signal submerged by harmonic components and other interference components. Moreover, the optimization algorithm is derived to solve the established optimization problem.

3.1. The optimization problem of the proposed FIWNNM

The bearing vibration signal is a one-dimensional discrete sequence, therefore it should be firstly constructed into a two-dimensional matrix before performing low-rank matrix estimation. There are two generic methods used to construct a two-dimensional matrix namely time–frequency transform [28] and Hankel matrix conversion [29]. In this paper, STFT operator A^T is utilized to convert the one-dimensional raw vibration signal y in time domain into two-dimension time–frequency coefficient Y, and this process can be mathematically expressed as $Y = A^T y$. For detailed information about parameter selection of STFT, one can refer to [28].

As described above, NNM treats all SVs equally without consideration of signal itself, which may result in undesirable components being contained in the estimated low-rank matrix $\hat{X}$, especially in complex vibration signals. Therefore, to tackle the inherent limitation of NNM and improve its flexibility in analyzing vibration signal, FIWNNM that introduces an adaptive weighting strategy to NNM is proposed, which can effectively extract fault features even in the presence of harmonic interferences. The optimization problem of FIWNNM is defined as

$$\hat{X} = \arg\min_{X} \|Y - X\|_F^2 + \|X\|_{w,*} \tag{3}$$

where w is the weight vector, $\|X\|_{w,*}$ is the weighted nuclear norm of matrix X and can be explicitly expressed as

$$\|X\|_{w,*} = \sum\nolimits_i w_i \sigma_i(X) \tag{4}$$

where w_i represents the weight of the ith SV $\sigma_i(X)$ and $X, Y \in \mathbb{C}^{M_1 \times M_2}$ are the STFT coefficients. In this paper, we define weight vector w as $w = [w_1, w_2, \ldots, w_n]$ with $n = \min(M_1, M_2)$. The purpose of applying the weighted nuclear norm is to more effectively restore the low-rank matrix $\hat{X}$ composed of fault feature components while eliminating other interference. To assign suitable weights that are beneficial to the extraction of bearing fault features, an adaptive weighting strategy based on fault information is proposed in this paper. In the proposed weighting strategy, to effectively preserve the singular components with the abundant fault features, a small weight is applied to the corresponding SV. Conversely, SVs related to the singular components with almost no fault information are assigned a large weight, which means the SVs are severely shrunk and interference components are reduced.

In the paper, the amount of fault information contained in a singular component is measured by the correlated kurtosis (CK) indicator, which is defined as

$$CK(M,T) = \frac{\sum_{i=1}^{N} \left(\Pi_{k=0}^{M} s_{i-kT} \right)^2}{\left(\sum_{i=1}^{N} s_i^2 \right)^{M+1}} \tag{5}$$

where s is the signal to be measured with length of N, T represents the point number of interested period and M denotes the shift number.

Compared with indicators that only measure impulsiveness or periodicity, including kurtosis [30], Gini [31] and harmonic-to-noise ratio [32], CK has powerful capability of simultaneously emphasizing impulsiveness and periodicity [33], and is conducive to measuring the amount of fault information contained in signal s. There are two parameters in the calculation formula of CK, i.e. T and M. Parameter T can be easily calculated based on formula $T = f_s \times P$, where f_s and P represent sampling frequency and fault period, respectively. As regards shift number M, higher M can much emphasize the periodicity of fault impulses. However, higher M requires the more accurate estimation of fault period and increases computational complexity [33]. Therefore, M is recommended to be 2–5 and $M = 2$ is adopted in this paper.

As described above, to effectively mine the fault features from noisy signal, the SVs associated with fault features are assigned small weights and the SVs related to interference components are assigned large weights. In the paper, the weight w_i is inversely proportional to the value of CK, and the calculation formula of the weight w_i is explicitly expressed as

$$w_i = \lambda \frac{1}{CK_i} \tag{6}$$

where CK_i represents the CK of ith singular component and λ is the regularization parameter.

Regarding the regularization parameter λ, it has a great impact on the final result of FIWNNM and must be appropriately selected according to the signal itself. Since the purpose of FIWNNM is to extract fault features from noisy signal and CK is an excellent criterion to indicate the amount of fault information, the parameter λ that corresponds to the maximum CK of extracted fault features is regarded as the optimal value. After obtaining the optimal λ, the desired fault features can be efficiently extracted and the potential bearing fault can be detected.

3.2. Optimization algorithm of the proposed FIWNNM

In the proposed weighting strategy of FIWNNM, the weight w_i is determined based on the fault information of the singular component rather than SV, which leads to the issue that the weights in the weight vector w are in an arbitrary order. Correspondingly, optimization problem equation (3) is nonconvex and is difficult to be directly solved like NNM. In this section, an iterative optimization algorithm is derived and the detailed procedure is as follows.

Theorem 1. *$\forall Y \in R^{m \times n}, Y = U\Sigma V^T$ is the SVD of matrix Y [18]. Then the solution $\hat{X}$ of optimization problem equation (3) with non-negative weight vector w can be written as*

$$\hat{X} = U\hat{B}V^T \tag{7}$$

where $\hat{B}$ is the solution of the following optimization problem:

$$\hat{B} = \arg\min_B \|\Sigma - B\|_F^2 + \|B\|_{w,*} \tag{8}$$

To get the solution of equation (8), we first denote $B = P\Lambda Q^T$ as the SVD of B. The problem in equation (8) can be transformed into the following optimization problem:

$$\begin{aligned} &(\widehat{P}, \widehat{\Lambda}, \widehat{Q}) = \arg\min_{P,\Lambda,Q} \|P\Lambda Q^T - \Sigma\|_F^2 + \|P\Lambda Q^T\|_{w,*} \\ &\text{s.t. } P^T P = I, Q^T Q = I \end{aligned} \tag{9}$$

where I is the identity matrix. The solution of problem equation (9) can be obtained by iteratively performing the following two steps composed of sorting the diagonal elements and shrinking the SVs:

$$\begin{cases} \left(P_{(k+1)}^T, \Phi, Q_{k+1}^T \right) = SVD\left(\Lambda_{(k+1)} \right) \\ \Lambda_{(k+1)} = P_{(k+1)}^T S_w(\Sigma) Q_{(k+1)}. \end{cases} \tag{10}$$

Algorithm 1: Iterative optimization algorithm for solving the optimization problem equation (3) in the proposed FIWNNM.

Require: convert one-dimensionalraw vibration signal y into two-dimensional time –frequency coefficients Y using STFT operator A^T, set maximum iteration number K.
1. **Input**: Y
2. **Initialization**: X, P, Q, Λ
3. **for** $k \in [1, K]$
4. Calculate weight w_i of each singular component of matrix X through the expression $w_i = \lambda \frac{1}{CK_i}$ and further construct weight vector $w = [w_1, w_2, \ldots, w_n]$.
5. $\left(P^T, \Phi, Q^T\right) = SVD\left(\Lambda\right)$
6. $\Lambda = P^T S_w\left(\Sigma\right) Q$
7. $X = UP^T S_w\left(\Sigma\right) QV^T$
8. **end**
9. **Output**: low-rank matrix X and fault features $x = AX$.

where $S_w(\Sigma)_{ii} = \max\left(\Sigma_{ii} - w_i, 0\right)$ *is the soft-thresholding function applied to the diagonal matrix* Σ.

By combining the above steps in equations (7)–(10), the final solution of optimization problem equation (3) in the proposed FIWNNM can be obtained by

$$\hat{X} = U\hat{P}^T S_w\left(\Sigma\right)\hat{Q}V^T. \tag{11}$$

After the optimal solution $\hat{X}$ *is obtained, the estimated fault features* x *can be obtained by* $x = A\hat{X}$, *where* A *represents the inverse STFT operator. According to the above description, the detailed procedures for solving optimization problem equation (3) in the proposed method FIWNNM is summarized in Algorithm 1.*

4. Simulation analysis

In this section, the performance of the proposed FIWNNM is preliminarily verified using simulated signals. Meanwhile, two other methods including NNM and spectral kurtosis (SK) are selected for comparison to further demonstrate the superiority of the proposed FIWNNM.

To establish a realistic simulated signal, four frequent signal components are contained in this simulation analysis. Concretely, the expression of simulated signal y is as follows:

$$y(t) = \sum_i D_i S_d(t - \mathrm{i}T_d - \tau_i) + \sum_n A_n \cos\left(2\pi f_n t + \alpha_n\right) + \sum_i R_i S_i(t - T_i) + n(t) \tag{12}$$

$$S(t) = \mathrm{e}^{-\alpha_r t}\sin\left(2\pi f_r t\right). \tag{13}$$

In equation (12), the first item represents the repetitive fault features caused by a localized bearing defect, where D_i is amplitude of the ith impulse, T_d represents the time interval between two transient impulses, τ_i represents the random slip and is usually set to 1%–2% of the fault period, α_r and f_r are the damping ratio and the resonance frequency of the impulse response, respectively. The second item represents the harmonic component caused by the shaft or rotor, where A_n and f_n are the amplitude and the frequency of nth harmonic component, α_n stands for initial phase. The third item represents the random shocks caused by random impulses or electromagnetic interference during data acquisition, where the amplitude and the corresponding occurrence time of the ith random impulse are represented by R_i and T_i. Moreover, the last item $n(t)$ is Gaussian white noise.

The sampling frequency is 20 000 Hz and time length of the simulated signal is 1 s. In this section, the vibration signal with an outer race fault is simulated, and figure 4 displays the waveform diagrams of four signal components. Meanwhile, the parameters of simulated signal are shown in table 1.

In the first case, the harmonic components are not included in the simulated signal and the synthetic signal is presented in figure 5(a). Then the synthetic signal is process by NNM and the proposed method FIWNNM, and the results are displayed in figures 5(b) and (c), respectively. Note that for the STFT, the window size $R = 64$ with 50% window overlapping and the length of Fourier transform $L = 256$ are used in simulation analysis. Both NNM and FIWNNM successfully extract the fault feature components from the noisy signal, which illustrates the ability of these two methods in low-rank matrix approximation and fault feature extraction. Meanwhile, FIWNNM can better preserve the magnitude of fault-related fault features than NNM, which is attributed to the adaptive weighting strategy introduced by FIWNNM. In addition, root mean square error (RMSE) between extracted and simulated fault features is selected to quantitatively assess the performance of the FIWNNM and NNM methods. As displayed in figure 5, the RMSE value of FIWNNM and NNM are 0.1285 and 0.1711, respectively. These results further demonstrate that FIWNNM has better fault feature extraction ability than traditional NNM.

In the second case, the harmonic components are included in the simulated signal, which is closer to the real industrial environment. The robustness of NNM and FIWNNM to harmonic components is tested in this case. Figure 6(a) shows the synthetic signal, in which fault features are completely submerged by noise and interference. In addition, as displayed in figure 6(b), the frequency related to the harmonic component is very significant, while ball pass frequency outer race (BPFO) and its harmonics are disturbed by other interference

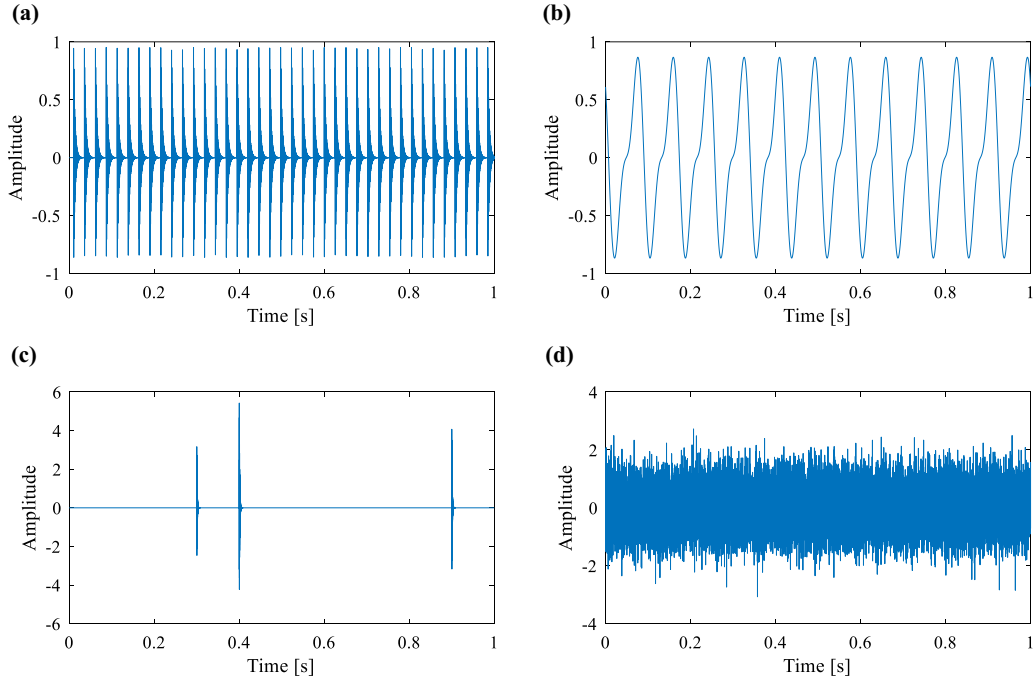

Figure 4. Four components of the simulated signal. (a) Defect impulses, (b) harmonic components, (c) random shocks, (d) Gaussian white noise.

Table 1. The parameters of the simulated signal.

Defect impulses				Rotor and shaft						Random shocks	
D_1	T_d	f_r	α_r	f_1	f_2	A_1	A_2	α_1	α_2	f_r	α_r
1	1/39	1500	300	12	24	0.7	0.3	$\pi/3$	$\pi/6$	6000	1000

lines in the envelope spectrum. Therefore, the traditional NNM and the proposed FIWNNM are applied to the simulated signal to extract fault features. Figure 7 gives the result obtained by NNM. In figure 7, harmonic components are dominant in the waveform and the frequency associated with harmonic components is also clearly detected in the envelope spectrum. However, BPFO and its harmonics are totally invisible, which illustrates that NNM is very sensitive to harmonic components. The reasons for above phenomenon are that the approximation of the low-rank matrix of NNM is implemented by retaining large singular components and the harmonic components with large energy are represented by the first several large singular components. The results obtained by the proposed FIWNNM are described in figure 8. The proposed FIWNNM can successfully extract the periodical fault impulses from the noisy signal, and BPFO and its harmonics are obvious in the envelope spectrum. Based on the above analysis, it can be reasonably concluded that NNM is very sensitive to harmonic components and is not suitable for extracting fault feature extraction from complex signals. On the contrary, FIWNNM has strong robustness to harmonic components and powerful ability in fault feature extraction from the noisy vibration signal.

Moreover, the spectral kurtosis (SK), a widely used fault feature extraction method and a benchmark method for vibration signal analysis, is also applied as comparison method

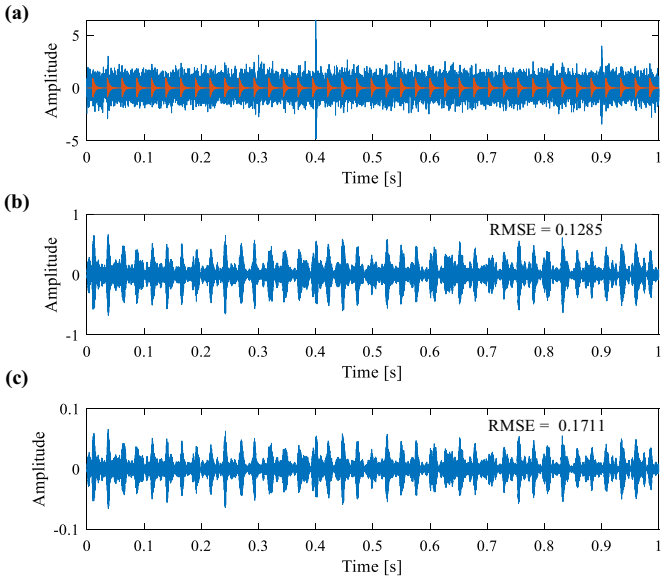

Figure 5. Simulated vibration signal without harmonic components. (a) Raw signal, (b) the result obtained by FIWNNM, (c) the result obtained by NNM.

to further highlight the superiority of the FIWNNM method. Figure 9(a) gives the kurtogram, from which the frequency band with maximum kurtosis is identified whose center frequency and bandwidth are 6042 Hz and 417 Hz. Based on this information, the corresponding band-pass filter is designed and the filtered signal is obtained. Figure 9 displays the filtered signal and its envelope spectrum, from which random shocks rather than repetitive impulses are detected. The result clearly illustrates that SK fails to extract the fault features since SK

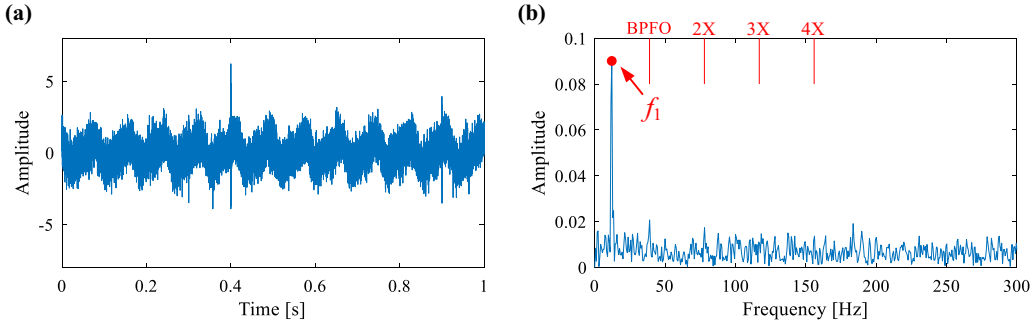

Figure 6. Simulated vibration signal with harmonic components. (a) Raw signal, (b) envelope spectrum.

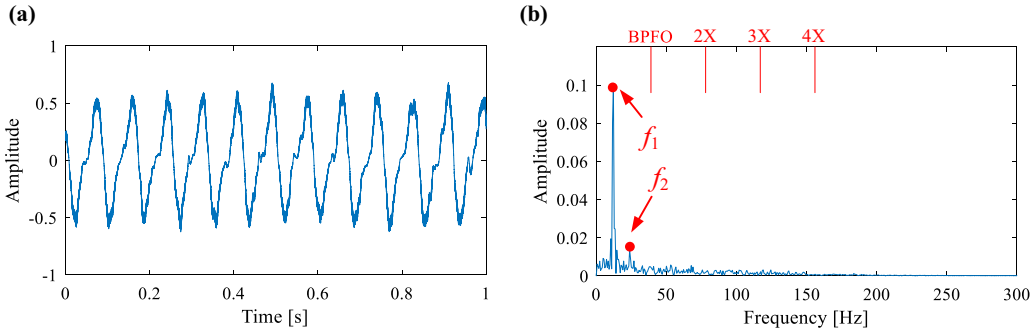

Figure 7. The results obtained by the NNM. (a) Time domain waveform, (b) envelope spectrum.

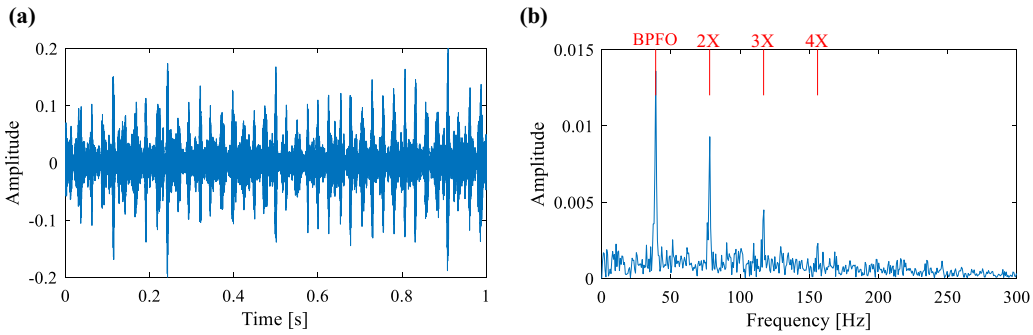

Figure 8. The results obtained by the FIWNNM. (a) Time domain waveform, (b) envelope spectrum.

is sensitive to a single impact and mistakenly locates the frequency band. Therefore, through the above comparative analysis, the effectiveness of the proposed FIWNNM is initially proved using simulated signals.

5. Experiment validation

In this section, experimental research is conducted to further verify the fault feature extraction ability of the proposed FIWNNM. The actual installation diagram and schematic view of the locomotive bearing test rig are demonstrated in figure 10. Meanwhile, it should be pointed out that the bearing outer race is driven by a hydraulic motor, while the bearing inner race is fixed on this test rig. Therefore, there is an amplitude modulation phenomenon for bearing outer race fault and sidebands are expected to be found in the envelope spectrum. Conversely, there is no amplitude modulation phenomenon for the bearing inner race fault. In the experiments, the vibration signals are collected by an accelerometer with sampling frequency 76 800 Hz and time length is 1 s. Meanwhile, rotating speed signal is also obtained by a tachometer. The structural parameters of the tested bearings are listed in table 2. With rotating speed signal and the structural parameters, the theoretical fault characteristic frequencies (FCFs) can be calculated.

5.1. Experiment 1: the bearing with defect on inner race

In experiment 1, the signal acquired from locomotive bearing with inner race defect is analyzed. The traditional NNM, SVNNM and the proposed FIWNNM are all applied to the acquired signal. According to the bearing structural parameters and the rotating speed information, the fault characteristic frequencies are calculated and listed in table 3. It

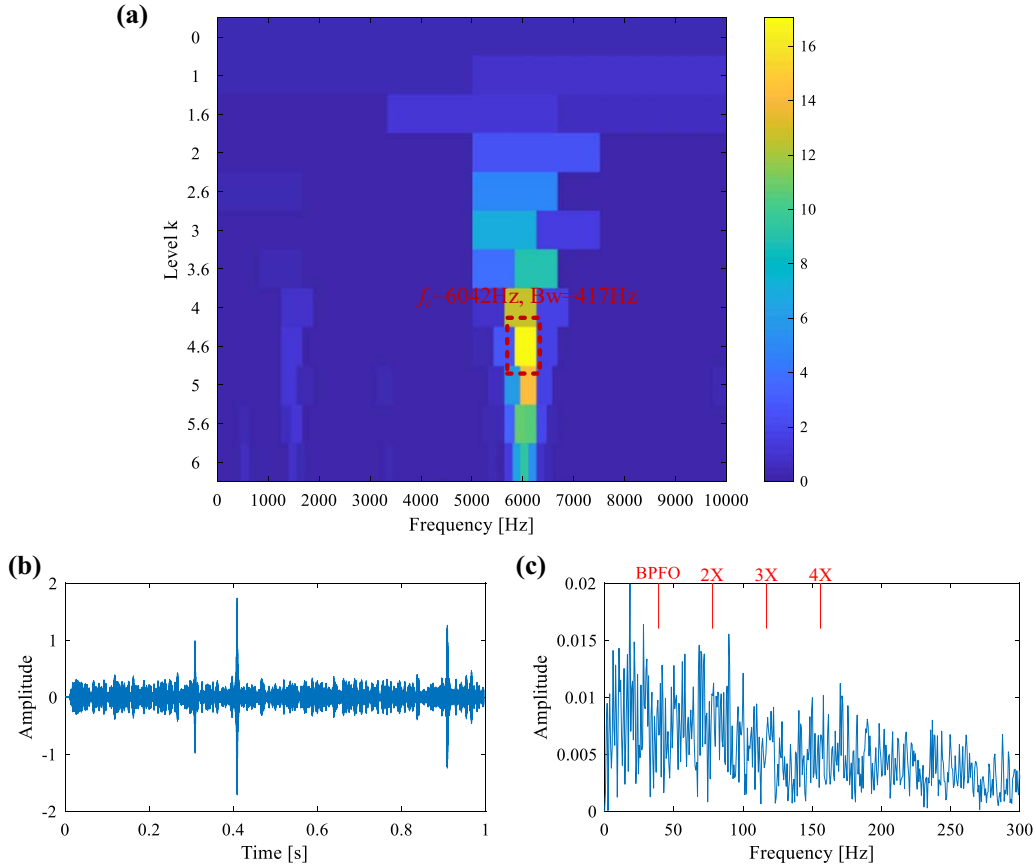

Figure 9. The results obtained by the SK method. (a) Kurtogram, (b) the filtered signal, (c) envelope spectrum.

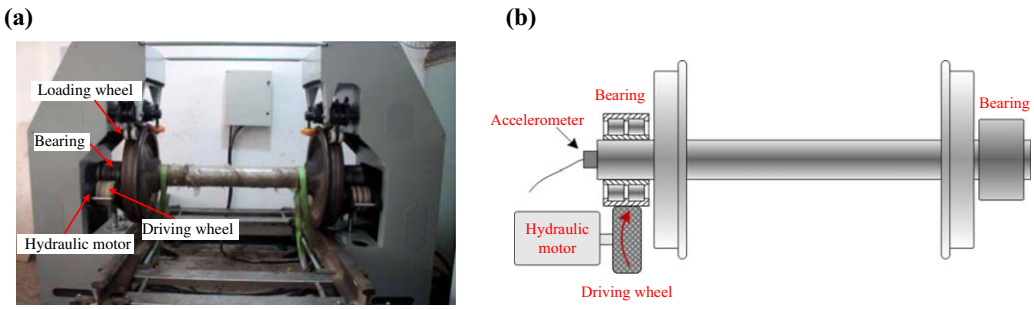

Figure 10. The locomotive bearing test rig. (a) Experiment setup, (b) schematic view.

Table 2. The parameters of the locomotive bearings.

Number of rollers	Pitch diameter (mm)	Roller diameter (mm)	Contact angle (degree)
20	180	23.775	9

Table 3. Bearing characteristic frequencies.

f_r	BPFO	BPFI	BSF
4.44 Hz	38.64 Hz	50.24 Hz	16.53 Hz

can be found that the rotating frequency of bearing outer race is 4.44 Hz and ball pass frequency inner race (BPFI) is 50.24 Hz.

Figure 11 plots the collected raw signal, in which the fault features are severely disturbed by noise. In order to provide more useful diagnostic information for fault detection, NNM and SVNNM are first employed to process the raw signal. As mentioned above, both NNM and SVNNM extract low-rank components by retaining large SVs. In particular, SVNNM further emphasizes large singular components due to the non-decreasing weighting strategy. The results obtained by NNM and SVNNM are given in figures 12 and 13. It can be detected that NNM and SVNNM fail to extract fault features from noisy signal. Moreover, in the envelope spectrum, BPFI and its harmonics are totally indistinct, while the rotating frequency and its multiples are obvious. The above results further imply that NNM and SVNNM are easily affected by the harmonic

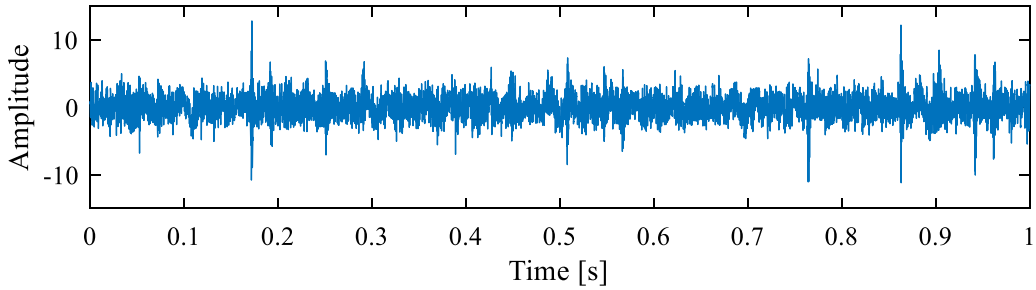

Figure 11. Raw vibration signal.

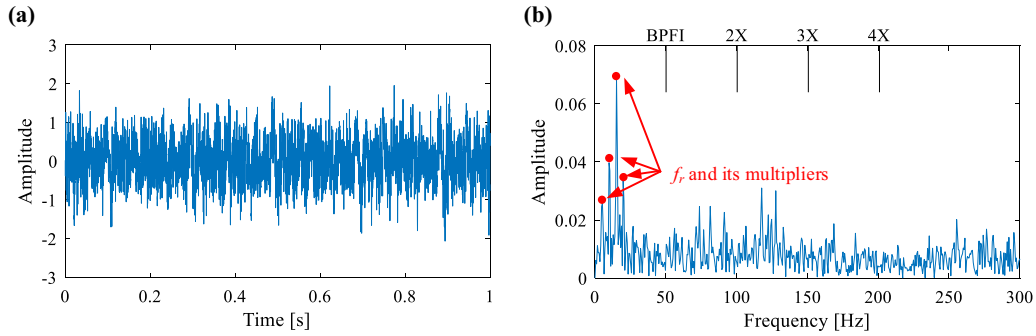

Figure 12. The results obtained by NNM. (a) Time domain waveform, (b) envelope spectrum.

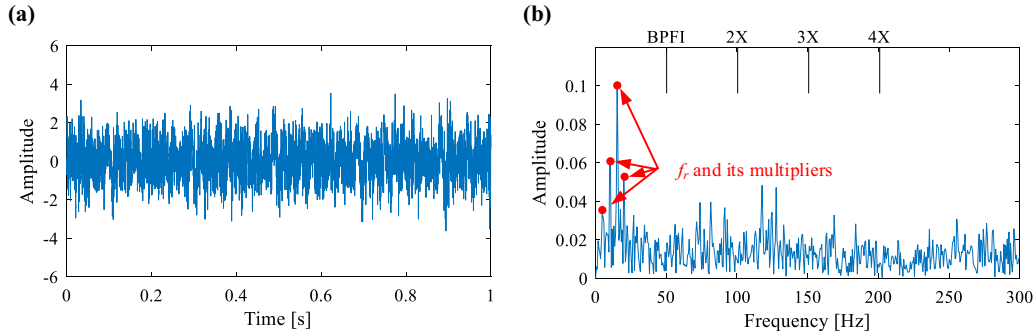

Figure 13. The results obtained by SVNNM. (a) Time domain waveform, (b) envelope spectrum.

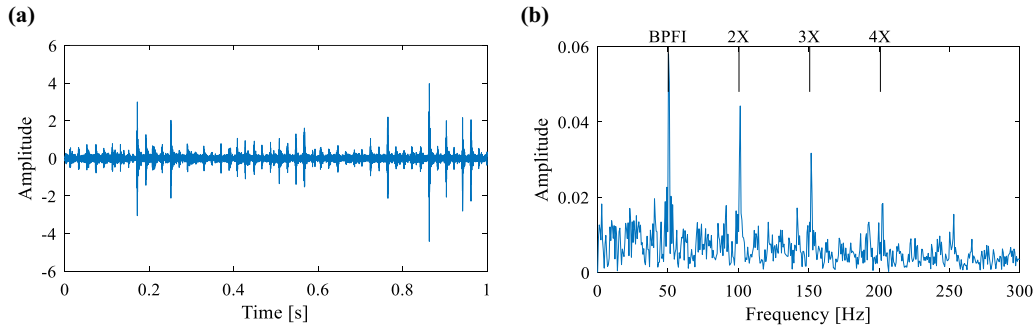

Figure 14. The results obtained by the proposed FIWNNM. (a) Time domain waveform, (b) envelope spectrum.

components and not very suitable for the real vibration signal analysis.

Then, the proposed FIWNNM is adopted to process the same signal. Figure 14 describes the result obtained by FIWNNM. As displayed in figure 14(a), clear repetitive impulses demonstrate that FIWNNM effectively extracts fault features from noisy signal. Meanwhile, the BPFI and its harmonics are very obvious in the envelope spectrum, as shown in figure 14(b), which gives solid evidence of the presence of inner race fault. To validate this result, the tested bearing is

Figure 15. The bearing with a defect on the inner race.

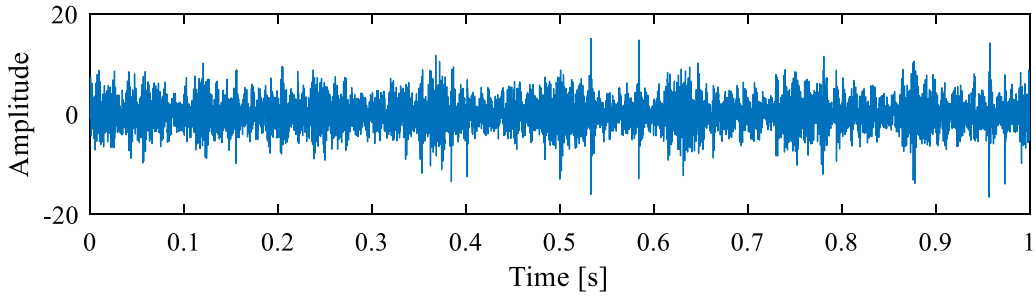

Figure 16. Raw vibration signal.

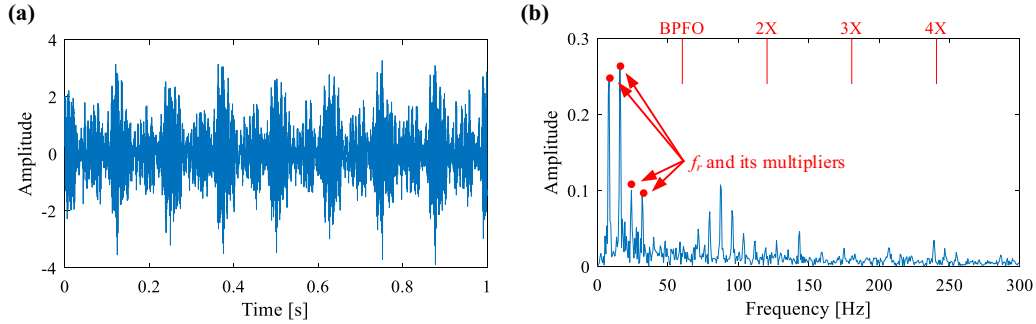

Figure 17. The results obtained by NNM. (a) Time domain waveform, (b) envelope spectrum.

disassembled and shown in figure 15, where the inner race fault is clearly visible. By comparing the above results, the excellent performance of FIWNNM in fault feature extraction under strong harmonic component interferences is illustrated in detail.

Table 4. Bearing characteristic frequencies.

f_r	BPFO	BPFI	BSF
6.92 Hz	60.20 Hz	78.27 Hz	25.76 Hz

5.2. Experiment 2: the bearing with defect on outer race

In experiment 2, vibration signal with outer race defect of the locomotive bearing is analyzed. Table 4 lists the fault characteristic frequencies of the bearing, from which BPFO is found to be 60.20 Hz. The waveform diagram of collected vibration signal is shown in figure 16, in which fault information is difficult to observe. Therefore, traditional NNM, SVNNM and the proposed FIWNNM are all adopted for fault feature extraction.

Figures 17 and 18 display the results obtained by NNM and SVNNM, respectively. It can be found that harmonic components are distinct in the time domain waveforms, and the rotating frequency and its multipliers are also very evident in the corresponding envelope spectrums. However, the repetitive impulses with outer race fault period indicating the existence of outer race fault are not extracted by NNM and SVNNM, and BPFO and its multipliers are submerged by other interference frequencies. According to the above analysis, it can reasonably concluded NNM and SVNNM fail to extract fault features from the original signal that are disturbed by harmonic components and environmental noise.

Then, the same signal is processed by the proposed FIWNNM and the results are shown in figure 19. In the waveform, clear defect impulses are effectively identified. Moreover, BPFO and its harmonics are very obvious in the envelope spectrum, and sidebands around the BPFO and its harmonics also appear as expected. The above results provide conclusive evidence for the presence of a localized defect on

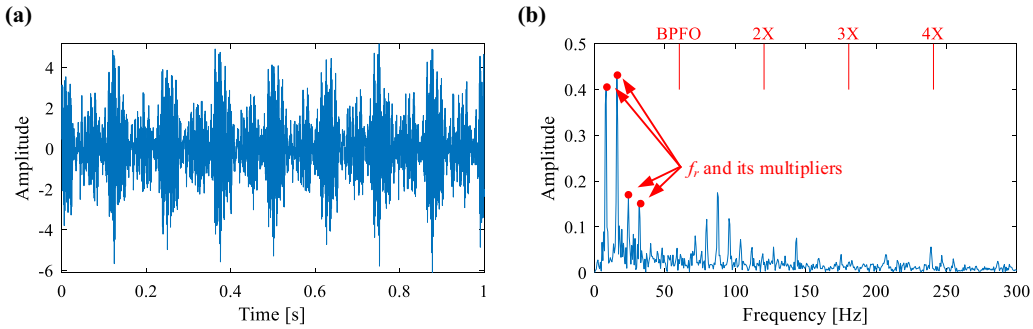

Figure 18. The results obtained by SVNNM. (a) Time domain waveform, (b) envelope spectrum.

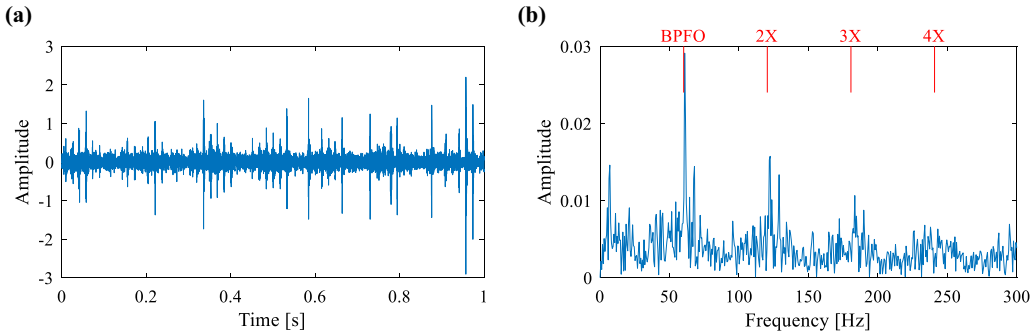

Figure 19. The results obtained by the proposed FIWNNM. (a) Time domain waveform, (b) envelope spectrum.

Figure 20. The bearing with a defect on the outer race.

the outer race. To verify the accuracy of the diagnosis, the test bearing is disassembled, and a defect on the outer race can be distinctly recognized from the photograph displayed in figure 20. By analyzing and comparing the processed results of the three methods, the effectiveness of FIWNNM in processing practical signals is further verified.

6. Conclusion

Traditional NNM and SVNNM approximate the low-rank component by keeping the large singular value, which is very inappropriate for a vibration signal that contains harmonic interference components. In view of this problem, this paper proposes a FIWNNM for fault feature extraction from noisy signal and bearing fault diagnosis. In the proposed FIWNNM, an adaptive weighted strategy based on correlated kurtosis is introduced, which is significant for highlighting the fault features of interest while suppressing the interference components. Meanwhile, an iterative optimization algorithm is established to solve the optimization problem. From the results of simulation analysis and experimental verification, the effectiveness of the proposed FIWNNM is confirmed when NNM, SVNNM and SK are selected as the comparison methods. The results indicate that the proposed FIWNNM has a strong feature extraction ability and great potential in industrial applications. In future work, other advanced weighting strategies will be further investigated for enhancing the performance of weighted NNM.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (Grant No. 51575424), the National Key Research and Development Program of China (Grant No. 2019YFB1311903) and the Joint Foundation of the Ministry of Education of China (Grant No. 6141A02022113).

ORCID iDs

Yuhe Liao https://orcid.org/0000-0001-7948-9934
Xining Zhang https://orcid.org/0000-0003-4432-0471

References

[1] Xia Y and Lu S 2018 Convolutional sparse coding with periodic overlapped group sparsity for rolling element bearing fault diagnosis *Meas. Sci. Technol.* **29** 115103

[2] Xu Y, Tian W, Zhang K *et al* 2018 Application of an enhanced fast kurtogram based on empirical wavelet transform for bearing fault diagnosis *Meas. Sci. Technol.* **29** 035001

[3] Wang Y, Yang L, Xiang J *et al* 2017 A hybrid approach to fault diagnosis of roller bearings under variable speed conditions *Meas. Sci. Technol.* **28** 125104

[4] Ding C, Zhao M, Lin J *et al* 2019 Multi-objective iterative optimization algorithm based optimal wavelet filter selection for multi-fault diagnosis of rolling element bearings *ISA Trans.* **88** 199–215

[5] Wang B, Liao Y, Ding C *et al* 2020 Periodical sparse low-rank matrix estimation algorithm for fault detection of rolling bearings *ISA Trans.*

[6] Antoni J 2006 The spectral kurtosis: a useful tool for characterising non-stationary signals *Mech. Syst. Signal Process.* **20** 282–307

[7] Antoni J 2007 Fast computation of the kurtogram for the detection of transient faults *Mech. Syst. Signal Process.* **21** 108–24

[8] Qiao Z and Pan Z 2015 SVD principle analysis and fault diagnosis for bearings based on the correlation coefficient *Meas. Sci. Technol.* **26** 085014

[9] Zhao M and Jia X 2017 A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery *Mech. Syst. Signal Process.* **94** 129–47

[10] Wang L, Xiang J and Liu Y 2019 A time–frequency-based maximum correlated kurtosis deconvolution approach for detecting bearing faults under variable speed conditions *Meas. Sci. Technol.* **30** 125005

[11] Ding C, Zhao M, Lin J *et al* 2019 Sparsity based algorithm for condition assessment of rotating machinery using internal encoder data *IEEE Trans. Ind. Electron.*

[12] Cai G, Chen X and He Z 2013 Sparsity-enabled signal decomposition using tunable Q-factor wavelet transform for fault feature extraction of gearbox *Mech. Syst. Signal Process.* **41** 34–53

[13] Candès E J and Recht B 2009 Exact matrix completion via convex optimization *Found. Comput. Math.* **9** 717

[14] Parekh A and Selesnick I W Enhanced low-rank matrix approximation *IEEE Signal Process. Lett.* 2016 **23** 493–7

[15] Lu C, Tang J, Yan S *et al* 2014 Generalized nonconvex nonsmooth low-rank minimization *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 4130–7

[16] Lu C, Zhu C, Xu C *et al* 2015 Generalized singular value thresholding *29th AAAI Conf. on Artificial Intelligence*

[17] Xin G, Qin Y, Jia L M *et al* 2018 Low-rank and sparse model: a new perspective for rolling element bearing diagnosis *2018 Int. Conf. on Intelligent Rail Transportation (ICIRT)* (IEEE) pp 1–5

[18] Gu S, Zhang L, Zuo W *et al* 2014 Weighted nuclear norm minimization with application to image denoising *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition* pp 2862–9

[19] Gu S, Xie Q, Meng D *et al* 2017 Weighted nuclear norm minimization and its applications to low level vision *Int. J. Comput. Vis.* **121** 183–208

[20] Lu C, Tang J, Yan S *et al* 2015 Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm *IEEE Trans. Image Process.* **25** 829–39

[21] Du Z, Chen X, Zhang H *et al* 2017 Weighted low-rank sparse model via nuclear norm minimization for bearing fault detection *J. Sound Vib.* **400** 270–87

[22] Achlioptas D and McSherry F 2007 Fast computation of low-rank matrix approximations *J. ACM (JACM)* **54** 9–es

[23] Fazel M, Pong T K, Sun D *et al* 2013 Hankel matrix rank minimization with applications to system identification and realization *SIAM J. Matrix Anal. Appl.* **34** 946–77

[24] Hansson A, Liu Z and Vandenberghe L 2012 Subspace system identification via weighted nuclear norm optimization *2012 IEEE 51st IEEE Conf. on Decision and Control (CDC)* (*IEEE*) pp 3439–44

[25] Markovsky I 2008 Structured low-rank approximation and its applications *Automatica* **44** 891–909

[26] Donoho D L 1995 De-noising by soft-thresholding *IEEE Trans. Inf. Theory* **41** 613–27

[27] Cai J F, Candès E J and Shen Z 2010 A singular value thresholding algorithm for matrix completion *SIAM J. Optim.* **20** 1956–82

[28] Ding Y, He W, Chen B *et al* 2016 Detection of faults in rotating machinery using periodic time-frequency sparsity *J. Sound Vib.* **382** 357–78

[29] Zhao X and Ye B 2011 Selection of effective singular values using difference spectrum and its application to fault diagnosis of headstock *Mech. Syst. Signal Process.* **25** 1617–31

[30] Antoni J and Randall R B 2006 The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines *Mech. Syst. Signal Process.* **20** 308–31

[31] Zhao M, Jiao J and Lin J 2018 A data-driven monitoring scheme for rotating machinery via self-comparison approach *IEEE Trans. Ind. Inf.* **15** 2435–45

[32] Xu X, Zhao M, Lin J *et al* 2015 Periodicity-based kurtogram for random impulse resistance *Meas. Sci. Technol.* **26** 085011

[33] McDonald G L, Zhao Q and Zuo M J 2012 Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection *Mech. Syst. Signal Process.* **33** 237–55