IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 8, 2020, accepted November 18, 2020, date of publication November 24, 2020, date of current version December 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3040209

Dictionary Learning via a Mixed Noise Model for Sparse Representation Classification of Rolling Bearings

JIALING ZHANG [1], JIMEI WU [1,2], AND BINGBING HU [2]
[1]School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology, Xi'an 710048, China
[2]Faculty of Printing, Packaging and Digital Media Engineering, Xi'an University of Technology, Xi'an 710048, China

Corresponding author: Jimei Wu (wujimei1@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 52075435, and in part by the Natural Science Foundation of Shaanxi Province under Grant 20JY054.

ABSTRACT Rotating machinery contains a great number of rolling bearings, which play an indispensable role. However, bearing vibration signals in complex environments are often mixed with various noises, which makes it difficult to extract fault characteristics from original signals. It is still challenging to identify the fault types of rolling bearings. To address this issue, a dictionary learning method based on a mixed noise model for the sparse representation classification of rolling bearings (DLMN-SRC) is proposed. Our framework constructs a dictionary learning method based on mixed noise, which has better robustness to complex noise pollution than a single noise model. Then, an alternating direction method of multipliers (ADMM) algorithm is used to solve the optimization problem of the proposed model. Eventually, the redundant errors between the detected and reconstructed signals in the dictionary learning model are calculated for sparse representation classification. The results of two examples prove that faults of the rolling bearing are successfully extracted and classified by DLMN-SRC. Compared with traditional diagnosis methods, the performance of this method has obvious superiority and good application prospects.

INDEX TERMS Dictionary learning, mixed noise, sparse representation classification, bearing fault diagnosis.

I. INTRODUCTION

Rolling bearings are commonly applied to modern rotating machinery and are a key component in ensuring the safe and stable operation of the system. However, due to the working characteristics of the rotating machinery, the bearings are often exposed to heavy load, high temperature, variable load and harsh working environments. Bearings have become one of the most vulnerable components [1], [2]. Therefore, the condition monitoring and fault diagnosis (CMFD) of bearings has attracted widespread attention in the scientific and engineering communities [3]. This has significance in reducing operation and maintenance costs and improving the reliability of rotating machinery.

Most fault signals cannot be measured or can only be measured under noisy conditions because of complex mechanical structures and nonstationary operating conditions [4]. When a rolling bearing is damaged by wear and impact, the damping and stiffness are changed, resulting in nonlinear and nonstationary vibration signals. Therefore, it is important to extract the implicit useful state variables through a series of noisy data. The existing literature about signal denoising and fault feature extraction usually includes a time domain analysis, frequency domain analysis and time-frequency analysis, such as, the Kalman filter [5], Fourier transform [6], wavelet transform [7] and empirical mode decomposition [8]. The above methods require experts to accurately identify fault types under the premise of retaining professional knowledge. In recent years, as a derivative of signal processing, sparse representation (SR) based on the sparse modeling of vibration signals has been used. SR constructs the sparsest or nearly sparse signal representation through atomic linear representation in overcomplete dictionaries. SR has been one of the important research areas in signal processing [9], image denoising [10] and face recognition [11].

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

To date, many successful applications have been reported for CMFD. A popular way to design a dictionary is called dictionary learning. The selection of a dictionary is the key part of signal sparse representation and feature extraction. At present, existing dictionary learning methods are generally divided into two categories: analysis dictionary and learning dictionary [12]. The design of an analysis dictionary mainly depends on prior knowledge of the target signal. Nevertheless, the fault characteristics may be unknown and unpredictable. At present, the learning dictionary has become a common dictionary because of its strong adaptability, good sparsity and reconstruction accuracy. Chen *et al.* [13] studied the characteristics of fault signals under strong background noise. A pulse sparse dictionary was proposed for a sparse model based on compression perception. Yang *et al.* [14] developed a sliding-window dictionary learning denoising method with time-domain signals. Zhou *et al.* [2] and Sun *et al.* [15] established an overcomplete dictionary (or overcomplete atomic library) of attended cosines basis and a parameter pulse dictionary that highly match the bearing fault waveform. Jiao *et al.* [16] presented a hierarchical discrimination sparse coding (HDSC) method, which successfully extracted weak fault features under strong noise and environmental interference. Although the above methods have made satisfactory contributions to feature extraction, it is difficult to extract feature signals with different types of noise pollution in actual working conditions [17]. If we directly analyze the original sample as a dictionary, then the noise destroys the structure of the subspace, thereby reducing the performance of the dictionary. It should be noted that the above methods mostly construct dictionaries based on the characteristics of fault signals, but there are few studies from the perspective of noise. Xiao *et al.* [18] proposed a double sparsity approach on the basis of an l_1-l_0 denoising model. This method combined a median filter with K-SVD to recover the image of a Gaussian plus impulse noise. Chen and Wu [19] presented a new robust dictionary learning approach that decomposed the noise into Gaussian noise and Laplacian noise. Zhou *et al.* [20] designed a structured dictionary in which the noise was composed of interference signals and reconstruction residuals. The algorithm adopted ADMM to extract the image feature samples. Most of the above research results were focused on face recognition and image processing, but there are few reports on mechanical equipment fault diagnosis.

In actual working conditions, the noise of bearings is caused by the inherent noise of the bearings, design and processing errors and scars. In addition, the noise also comes from noise produced by other components, the sensor's noise in the transmission process and the noise of the measuring instrument system. Traditional dictionary learning assumes that the noise obeys a Gaussian distribution and Laplace distribution [21], [22]. Nevertheless, industrial process data are often affected by abnormal values due to environmental influences, erroneous measurements and the inherent characteristics of sensors. These outliers make the noise distribution no longer obey the Gaussian distribution, which seriously affects the performance of dictionary learning and fault classification. These phenomena make the noise distribution of the system show heavy tail characteristics, that is, the probability density function distribution has a thicker tail statistic [23], [24]. Fortunately, the T distribution has thicker or higher tail than the Gaussian distribution, so it has better modeling ability for the thick-tailed noise than the Gaussian distribution. T distribution usually describes the noise distribution in the presence of abnormal values [25], [26]. To address these problems, a dictionary learning method based on a mixed noise model is proposed that aims to classify the sparse representation of rolling bearings from another perspective. First, a real noise model based on a Gaussian distribution, Laplace distribution and T distribution is constructed. This makes the sparse representation dictionary more robust to complex noise pollution than single distribution. Then, an effective ADMM algorithm is the solution to the optimization problem of the mixed noise model. The problem is decomposed into several subproblems with closed solutions. Finally, the fault types of bearings are identified by sparse classification according to the constructed model. The main innovations and contributions are as follows:

(1) A novel dictionary learning method that constructs a mixed noise model is proposed. This uses a low rank representation learning dictionary and a mixed noise model based on Gaussian distribution, Laplace distribution and T distribution.
(2) The ADMM algorithm is adopted to solve the optimization problem of the mixed noise model.
(3) The proposed DLMN-SRC method does not require complicated functional design or selection, and the redundant errors between the detected and reconstructed signals in the dictionary learning model are calculated for sparse representation classification, which provides a basis for fault identification.

The remainder of this paper is organized as follows. Section II introduces the dictionary learning methods based on noise assumptions. In Section III, the proposed diagnosis algorithm via DLMN and sparse-representation-based classification is described in detail. The overall procedures of the DLMN-SRC framework are illustrated in Section IV. Section V presents the effectiveness and superiority in this method, which are verified by two cases and compared with existing diagnosis methods. Conclusions are reached in the last section.

II. RELATED WORK

We use the N n-dimension signal set $Y = [y_1, y_2, \ldots, y_N] \in \mathrm{R}^{n\times N}$ and redundant dictionary $D = [d_1, d_2, \ldots, d_K] \in \mathrm{R}^{n\times K}$. The dictionary learning (DL) method is achieved by learning a distinct overcomplete dictionary D from the original signal set Y, which is robust to various noises in the fault diagnosis. DL is completed under the assumption that the noise obeys the Gaussian distribution [27], [28]. The noise distribution is described by the l_2-norm or l_F-norm. The classic DL model is expressed as follows:

$$\min_{D,A} \|Y-DA\|_F^2 \text{ subject to } \left\|d_{(i)}\right\|_2^2 \leq 1, \text{ for } i=1,2,\ldots,K$$
$$\left\|a_{(j)}\right\|_0 \leq T, \quad \text{for } j=1,2,\ldots,N \quad (1)$$

where $A=[a_1,a_2,\ldots,a_N] \in R^{K\times N}$ is the sparse representation coefficient vector. T represents the sparsity, aiming to make the number of nonzero atoms in the sparsity coefficient $a_{(j)}$ not exceed T. K-SVD [29] and MOD [30] are typical dictionary learning algorithms based on a Gaussian distribution hypothesis. However, in order to solve the NP-hard nature of Eq. (1), Mairal *et al.* [31] used l_1-norm instead of l_0-norm, and Eq. (1) is rewritten as follows:

$$\min_{D,A} \|Y-DA\|_F^2 + \alpha\|A\|_1$$
$$\text{subjuct to } \left\|d_{(i)}\right\|_2^2 \leq 1, \quad \text{for } i=1,2,\ldots,K \quad (2)$$

where α is a positive constant to weigh sparse terms. Inspired by this idea, Zhao *et al.* [22] and Qin *et al.* [32] added constraint terms on the basis of Eq.(1) and (2) to extract and classify bearing fault characteristics. Nevertheless, the method based on Gaussian noise may lose some useful feature information because they are too smooth [27]. To overcome the above shortcomings, scholars proposed DL methods based on Laplace distribution [33]. Eq. (2) can be rewritten as:

$$\min_{D,A} \|Y-DA\|_1 + \alpha\|A\|_1$$
$$\text{subjuct to } \left\|d_{(i)}\right\|_2^2 \leq 1, \quad \text{for } i=1,2,\ldots,K \quad (3)$$

Although the DL methods based on Laplacian noise have achieved clustering and noise reduction effects, the noise in practical application is not only composed of a single noise distribution. In Ref. [19] and [34], Chen and Selesnick constructed the DL method based on the addition of Gaussian noise and Laplacian mixed noise, such that

$$\min_{D,A,B} \|Y-DA-B\|_F^2 + \alpha\|A\|_1 + \beta\|B\|_1$$
$$\text{subjuct to } \left\|d_{(i)}\right\|_2^2 \leq 1, \quad \text{for } i=1,2,\ldots,K \quad (4)$$

where α and β are two the sparse regularization terms. B denotes Laplacian noise. Unfortunately, industrial data are often affected by abnormal noise, and it is not comprehensive to use Gaussian and Laplace distributions to characterize noise, which limits dictionary learning methods.

III. DICTIONARY LEARNING BASED MIXED NOISE MODEL FOR SPARSE REPRESENTATION CLASSIFICATION

As mentioned above in the dictionary learning theory background, a novel fault diagnosis method via dictionary learning of mixed noise and low-rank sparse classification is proposed. The DLMN-SRC framework involves three key issues, as follows:

A. DLMN OF STRUCTURES

In the past, most DL methods have assumed that noise comes from a Gaussian distribution, Laplacian distribution, or a mixture of these types. Nevertheless, a Gaussian-based dictionary is sensitive to abnormal values, resulting in outliers in noise measurement. In addition, original signals are often affected by abnormal values due to environmental influences, erroneous measurements and inherent characteristics of sensors in complex background conditions. The noise distribution of the system becomes a heavy-tailed characteristic that no longer only satisfies the above noise distribution. This may reduce the performance of dictionary learning [35]. The T distribution has a better ability to describe nonlinear and thick-tailed noise [36], [37]. Furthermore, in order to strengthen the correlation between the subspace dictionary and the representation coefficient vectors, the constructed dictionary has low rank. Motivated by this, the mixed noise model can be formulated as follows:

$$\min_{D,A,B,T} \|Y-DA-B-T\|_F^2 + \alpha\|A\|_* + \beta\|B\|_1 + \lambda\|T\|_1 \quad (5)$$

where α, β and λ are the sparse regularization terms. A is the sparse representation coefficient vector of signal set Y's redundant dictionary D. B, T and E=Y-DA-B-T represent the Laplacian, T distribution and Gaussian noise, respectively. $\|A\|_*$ is equal to the sum of the singular values of the sparse representation coefficient vectors.

B. SOLVING DLMN VIA ADMM

The mixed noise model is decomposed into several subproblems by the alternating direction of multipliers (ADMM) method [38]. Eq. (5) can be rewritten as:

$$\min_{D,A,B,T} \|Y-DA-B-T\|_F^2 + \alpha\|H\|_* + \beta\|B\|_1 + \lambda\|T\|_1$$
$$\text{subject to } A=H \quad (6)$$

The augmented Lagrangian function L_μ is defined as:

$$\begin{aligned} &L_\mu(D,A,B,T,H,M) \\ &= \|Y-DA-B-T\|_F^2 + \alpha\|H\|_* + \beta\|B\|_1 + \lambda\|T\|_1 \\ &\quad + \langle M, A-H\rangle + \mu\|A-H\|_F^2 \end{aligned} \quad (7)$$

where $\langle M, A-H\rangle = tr\left\{M^T(A-H)\right\}$. M represents the Lagrangian multiplier. μ represents a penalty parameter.

The ADMM iteration process is decomposed into subproblems.

1) Updating A:

$$\begin{aligned} A^{k+1} &= \arg\min_A L_\mu\left(D^k, A, B^k, T^k, H^k, M^k\right) \\ &= \left\|Y^k - D^kA - B^k - T^k\right\|_F^2 + \alpha\left\|H^k\right\|_* \\ &\quad + \beta\left\|B^k\right\|_1 + \lambda\|T\|_1 \\ &\quad + M^T\left[A-H^k\right] + \mu\left\|A-H^k\right\|_F^2 \end{aligned} \quad (8)$$

Eq. (8) obtains a closed solution

$$A^{k+1} = \left(D^{kT} D^k + \mu I\right)^{-1} \times \left[D^{kT}\left(X - B^k - T^k\right) + \mu H^k - \frac{M}{2}\right] \quad (9)$$

2) Updating H:

$$H^{k+1} = \arg\min_{H} \left[\alpha \|H\|_* + M^T (A - H) + \mu \|A - H\|_F^2\right] \quad (10)$$

A singular value threshold algorithm is used to solve the problem

$$H^{k+1} = I_{\frac{\alpha}{\mu}} \left(A^{k+1} + \frac{M}{2\mu}\right) \quad (11)$$

where $I_{\frac{\alpha}{\mu}}$ is the shrinkage factor.

3) Updating B:

$$B^{k+1} = \arg\min_{B} \left(\beta \|B\|_1 + \|Y - DA - B - T\|_F^2\right) \quad (12)$$

The solution of Eq. (12) is

$$B^{k+1} = S_\beta \left(Y - D^k A^{k+1} - T^k\right) \quad (13)$$

where S_β is the Laplace's soft-thresholding operator.

4) Updating T:

$$T^{k+1} = \arg\min_{T} \left(\lambda \|T\|_1 + \|Y - DA - B - T\|_F^2\right) \quad (14)$$

The solution of Eq. (14) is

$$T^{k+1} = S_\lambda \left(Y - B^{k+1} - D^k A^{k+1}\right) \quad (15)$$

where S_λ is the T distribution's soft-thresholding operator.

5) Updating D:

$$D^{k+1} = \arg\min_{D} \|Y - DA - B - T\|_F^2 \quad (16)$$

Eq. (16) obtains a closed solution

$$D^{k+1} = \left[Y - \left(B^{k+1} + T^{k+1}\right)\right] A^{k+1} \times \left[A^{k+1}\left(A^{k+1}\right)^T\right]^{-1} \quad (17)$$

C. SPARSE CLASSIFICATION

Another key process of the DLMN-SRC framework is the sparse classification of fault signals. The sparse coefficient vector A and dictionary D are decomposed into a series of sub-vectors $a_{(j)}$ and $d_{(i)}$, respectively. Among them, only the vectors mapped to the subdictionary coordinates are invariant, while the nonzero values are zero elsewhere. The principle of sparse classification is visualized in Fig. 1. The calculation formula for each redundant error of the component part on the subdictionary is as follows:

$$\varepsilon_i = \left\|Y - Da_{(j)}\right\|_2^2, \quad j = 1, 2, \ldots, N \quad (18)$$

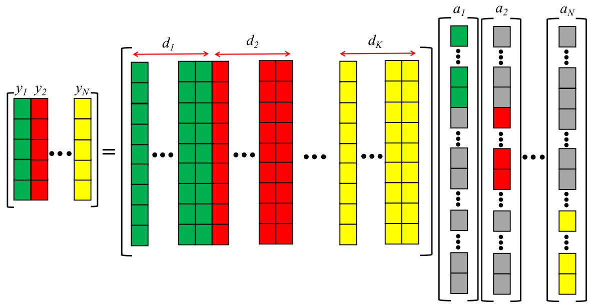


FIGURE 1. Visual sparse classification principle.

The atoms corresponding to the matching subdictionary sparsely represent samples of each category. Therefore, the minimum value $\varepsilon_{i\,\max}$ corresponds to the fault type of the test signal.

IV. OVERALL PROCEDURES OF DLMN-SRC FRAMEWORK FOR FAULT CLASSIFICATION

In this paper, a dictionary learning method based on low-rank representation of mixed noise model is proposed. This is applied to the sparse representation classification of common fault types in rolling bearings. The framework is divided into vibration signal acquisition, mixed-noise model dictionary learning and sparse representation-based classification. The corresponding flowchart is illustrated in Fig. 2.

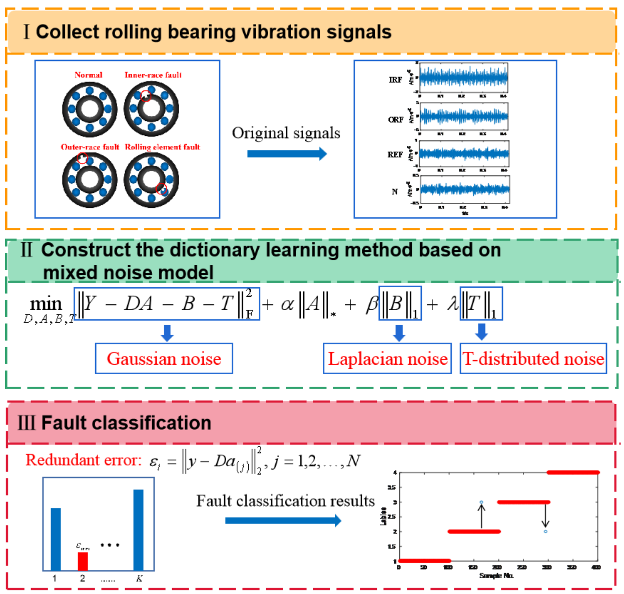


FIGURE 2. DLMN-SRC process.

Step 1: Collect the vibration signals of experimental objects in different health conditions as test and training samples for subsequent experiments.

Step 2: Update the variables A, H, B, T and D in the dictionary model according to Section III-B. Update Lagrange multipliers $M^{k+1} = M^k + \mu^k \left(A^{k+1} - H^{k+1}\right)$ and penalty parameter $\mu^{k+1} = \min\left(10^8, 1.01\mu^k\right)$ until convergence $\max\{\|A - H\|_\infty\} \leq 10^6$ is satisfied. Then, stop the iterations to obtain dictionary D and sparse coefficient A of the training samples.

Step 3: Calculate the redundant errors between the detected and reconstructed samples. The fault types of rolling bearings are identified by judging the minimum redundant error.

V. EXPERIMENTS AND DISCUSSION

A. CWRU CASE

The experimental data of the bearing vibration signals were obtained from CWRU. As shown in Fig. 3, the experimental facility consists of a 2-hp induction motor, torque sensor, electric motor and control electronic (not shown). The test bearings are installed on the driving end of the motor to support the motor shaft, and the acceleration sensor is fixed above the bearing seat to collect the acceleration data. The sampling frequency is 12 kHz. The test object is an SKF6205 deep-groove ball bearing. SFK bearings status types include normal (N), outer race fault (ORF), inner race fault (IRF) and rolling element fault (REF). The single point faults produced by electrodischarge machining on the rolling bearings are 0.007, 0.014 and 0.021 inches, respectively.

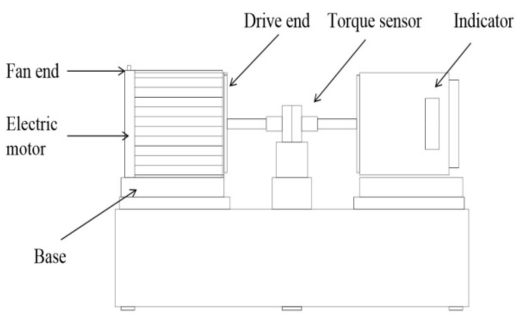


FIGURE 3. Schematic diagram of CWRU signal acquisition device.

In this subsection, the vibration signals of each type are divided into 160 time series without overlapping sampling points as the experimental samples. Without loss of generality, we classify the samples of the four health states in the experiment according to the failure types of IRF, ORF, REF and N status. The sample set for each fault contains 100 samples. A total of 800 samples are randomly selected as training and test samples according to a proportion of 1:1. In the next case, the experimental samples are selected according to this rule. A detailed description of the CWRU dataset is shown in Table 1. The corresponding waveform of the test samples are shown in Fig. 4 (e.g., 0.007 inches). Although there are some differences between the fault impulse characteristics of the four fault signals in the time waveform and Fourier spectrum, the periodic characteristics related to the fault signals are difficult to extract because of the coupling and noise interference among the components.

The algorithm initialization parameters of the DLMN-SRC framework are set as follows. The sparse regularization terms $\alpha, \beta, \lambda > 0$ are used to balance the weights of different regularization terms, which have little influence on the proposed method. The penalty parameter $\mu = 10^{-5}$. The initial values of other parameters of the mixed noise dictionary learning model are set to zero $A = B = T = H = M = 0$. As in Section IV, all kinds of fault samples are trained to obtain four types of subdictionaries $d_{(1)} \sim d_{(4)}$, where each type of subdictionary contains 300 atoms. The trained subdictionaries are combined to form the final dictionary D, which contains 1200 atoms. The atoms learned from the dictionary represent almost all of the samples of fault types sparsely. Fig. 5 shows the sparse representation coefficients of the CWRU bearing on the basis of the mixed noise model dictionary learning. It can be seen from the coefficient distribution diagram that the amplitude of the sparse coefficients in each region is related to the type of fault, and most of the coefficients come from the matching group. As shown in Fig. 5-(a), the higher amplitude of sparse coefficient is distributed in the first 300 items of the coefficient diagram, and their corresponding weight is the largest, which indicates

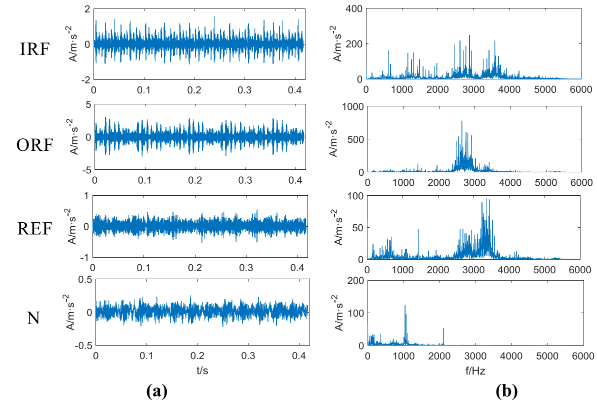


FIGURE 4. (a) Time waveform of 0.007-inch CWRU bearing data. (b) Corresponding Fourier spectrum.

TABLE 1. CWRU rolling bearing data set.

Fault location	Fault size (inches)	Training sample number	Test sample number	Sample length	Class-label
Inner race	0.007	100	100	160	1
Outer race	0.007	100	100	160	2
Rolling element	0.007	100	100	160	3
Normal	0	100	100	160	4
Inner race	0.014	100	100	160	1
Outer race	0.014	100	100	160	2
Rolling element	0.014	100	100	160	3
Inner race	0.021	100	100	160	1
Outer race	0.021	100	100	160	2
Rolling element	0.021	100	100	500	3

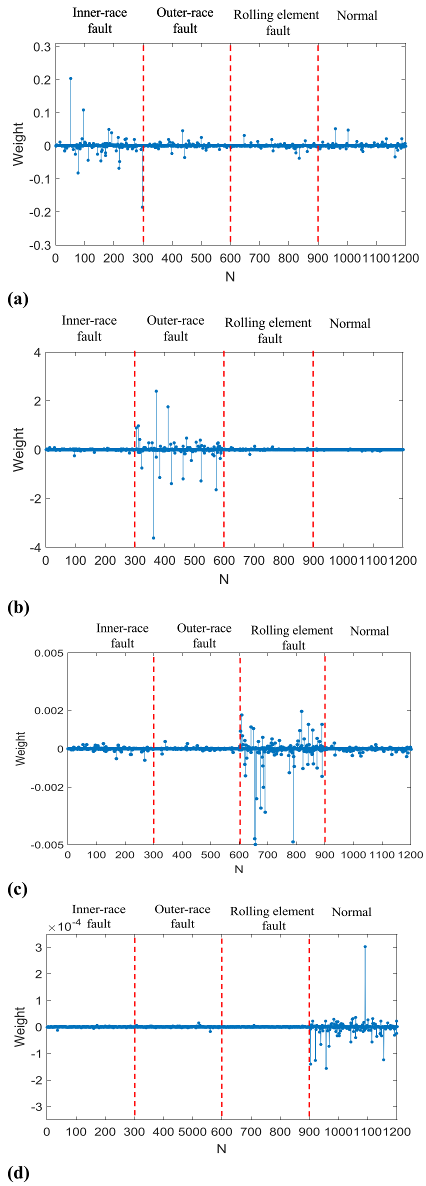
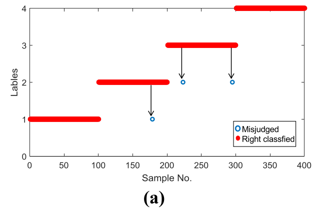


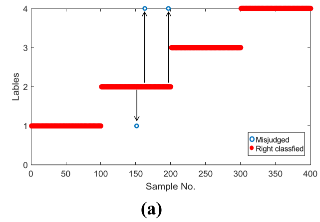
FIGURE 5. Sparse representation coefficients of CWRU bearing. (a) IRF (b) ORF (c) REF (d) N.



IRF	100 25.00%	0 0.00%	0 0.00%	0 0.00%	100% 0.00%
ORF	1 0.20%	99 24.75%	0 0.00%	0 0.00%	99% 1.00%
REF	0 0.00%	2 0.50%	98 24.5%	0 0.00%	98% 2%
N	0 0.00%	0 0.00%	0 0.00%	100 25.00%	100% 0.00%
	99.01% 0.99%	98.02% 1.98%	100% 0.00%	100% 0.00%	99.25% 0.75%
	IRF	ORF	REF	N	

(b)

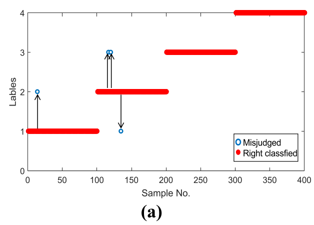
FIGURE 6. (a) Recognition results of 0.007-inch bearing health state. (b) Confusion matrix of 0.007-inch bearing health state detailed classification.



IRF	100 25.00%	0 0.00%	0 0.00%	0 0.00%	100% 0.00%
ORF	1 0.25%	97 24.25%	0 0.00%	2 0.50%	97% 3.00%
REF	0 0.00%	0 0.00%	100 24.75%	0 0.00%	100% 0.00%
N	0 0.00%	0 0.00%	0 0.00%	100 25.00%	100% 0.00%
	99.01% 0.99%	100% 0.00%	100% 0.00%	98.04% 1.96%	99.25% 0.75%
	IRF	ORF	REF	N	

(b)

FIGURE 7. (a) Recognition results of 0.014-inch bearing health state. (b) Confusion matrix of 0.014-inch bearing health state detailed classification.



IRF	99 24.75%	1 0.25%	0 0.00%	0 0.00%	99% 1.00%
ORF	1 0.25%	97 24.25%	2 0.5%	0 0.00%	97% 3.00%
REF	0 0.00%	0 0.00%	100 25.00%	0 0.00%	100% 0.00%
N	0 0.00%	0 0.00%	0 0.00%	100 25.00%	100% 0.00%
	99% 1.00%	98.98% 0.00%	98.04% 1.96%	100% 0.00%	99% 1.00%
	IRF	ORF	REF	N	

(b)

FIGURE 8. (a) Recognition results of 0.021-inch bearing health state. (b) Confusion matrix of 0.021-inch bearing health state detailed classification.

that the test data represent the inner race fault. Likewise, the sparse coefficients in Fig. 5-(b)∼Fig. 5-(d) respectively indicate the ORF, REF and N states.

Moreover, the relationship between the training sparse matrix and the original sample is represented by the l_2-norm of $y - y_i$. The minimum redundancy error of the test sample is calculated and denoted as ε_i in Eq. (18). The sparse classification results are shown in Fig. 6∼Fig. 8. It can be seen from Fig. 6-(a) that there are only three misclassified target categories in the 400 test samples. In the experiment, one outer race fault is considered an inner race fault, and two rolling element faults are considered as outer race faults. Fortunately, the identification results of the inner race fault and the normal are all correct.

For further analysis, we combine Fig. 6-(a) with Fig. 6-(b). On the one hand, the 100 inner test samples are correctly identified with a recognition accuracy of 100%. Among the outer race test samples, 99 samples are identified correctly, and only one sample is misdiagnosed as an inner race fault. Therefore, the accuracy of the outer race fault test sample is 99%. Among the rolling element test samples, 98 samples are identified correctly, and two samples are misdiagnosed as outer race faults. The accuracy of the rolling element fault test sample is 98%. All 100 normal test samples are identified correctly, and the test accuracy of the normal sample is 100%. Overall, 397 samples are correctly identified and 3 samples are incorrectly identified in the total of 400 test samples. Therefore, the comprehensive accuracy of bearing fault identification is 99.25%. On the other hand, there are 101 samples diagnosed as inner race faults in total, of which 100 samples are identified correctly and the other error sample comes from an outer race fault. There are 101 samples diagnosed as outer race faults, of which 99 samples are correct, and the other two samples are misdiagnosed from rolling element faults. The 98 test samples are diagnosed as rolling element faults, all of which are correctly identified. The 100 test samples are diagnosed as normal state, all of which are correctly identified. Therefore, the same average

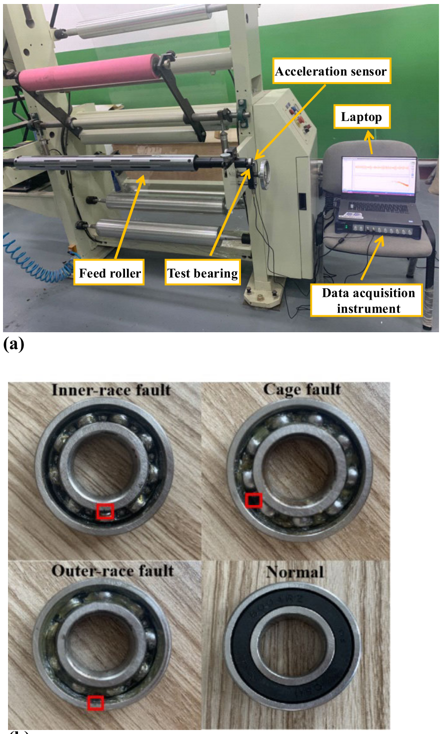


FIGURE 9. (a) Printing press bearing experimental platform. (b) Rolling bearings with four different health states.

testing accuracy of fault identification on the basis of the DLMN-SRC framework is 99.25%. In like manner, Fig. 7 and Fig. 8 show the sparse classification results with fault depths of 0.014 and 0.021 inches, respectively. Fortunately, their average correct rates are 99.25% and 99%, respectively, and the error rate did not exceed 1%. These quantitative results prove the correctness of the proposed DLMN-SRC in the detection of the bearing fault type under a complex noise background.

B. PRINTING PRESS BEARING CASE

Printing equipment contains massive bearings, which are crucial to ensure the accurate operation of the equipment. The paper feed roller of the printing press serves as a feed module to provide a continuous substrate for the printing components. Bearings mounted on the paper feed roller are chemically corroded by ink, alcohol and moistening liquid during operation. Failure of the bearing will lead to two or more sheets of paper, paper skew, paper wrinkling and other abnormal phenomena, which seriously affect the whole printing process. Experimental data of the rolling bearing are collected from the feed spindle of the FR400 gravure printing machine. As shown in Fig. 9-(a), the experimental facility is mainly composed of an AVANT data acquisition instrument, gravure printing press, laptop, feed roller and EA-YD-186 acceleration sensor. The vibration signal of the test bearing is captured by the EA-YD-186 acceleration sensor with a sensitivity of 9.78 mV/ms^{-2}. Four kinds of single-point park damage bearings [i.e., inner race fault (IRF),

TABLE 2. Bearing parameters.

Bearing type	Inner diameter (mm)	Outer diameter (mm)	Width (mm)	Number of rollers
JYB6004	20	42	12	9

TABLE 3. Printing press bearing data set.

Fault location	Fault size (mm)	Training sample number	Test sample number	Sample length	Class-label
Inner race	0.4	200	200	300	1
Outer race	0.4	200	200	300	2
Cage	0.4	200	200	300	3
Normal	0	200	200	300	4

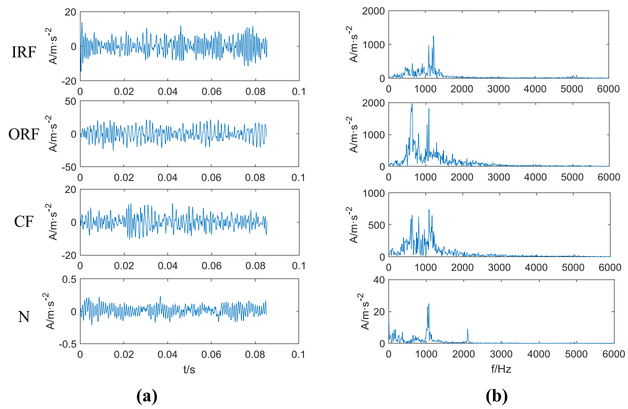


FIGURE 10. (a) Time waveform of printing press bearing data. (b) Corresponding Fourier spectrum.

outer race fault (ORF), cage fault (CF) and normal (N)] were machined by electrical discharge machining (EDM), as illustrated in Fig. 9-(b). The specific parameters are listed in Table 2. During the experiment, the spindle rotation frequency is 40 Hz, and the experimental data are collected by the acceleration sensor mounted on the vertical direction of the analysis bearing. The sampling frequency is 12 kHz.

As in Section V-A, 800 samples are randomly selected as test and training data from ORF, IRF, CF and N status, respectively. Each sample contains 300 points, and there is no overlap between these data. Thus, 800 training samples and 800 test samples (1600 samples in total) are formed. In essence, this is a four-type fault identification issue that needs to be solved. Detailed descriptions of the experimental data are presented in Table 3. Fig. 10 shows the time waveform and frequency domain of the four health states.

The algorithm initialization parameters have little influence on the DLMN-SRC method. The parameter value is the same as that in Section V-B. The training sample dictionary D is a redundant dictionary matrix of 300 × 2000, and four types of subdictionaries $d_{(1)} \sim d_{(4)}$ of four different health states are constructed. Fig. 10 shows the sparse representation coefficients of the printing press bearings. From the coefficient distribution diagram, it can be seen that a majority of the coefficients come from the corresponding type, and the amplitude of the sparse coefficient in each region is related to

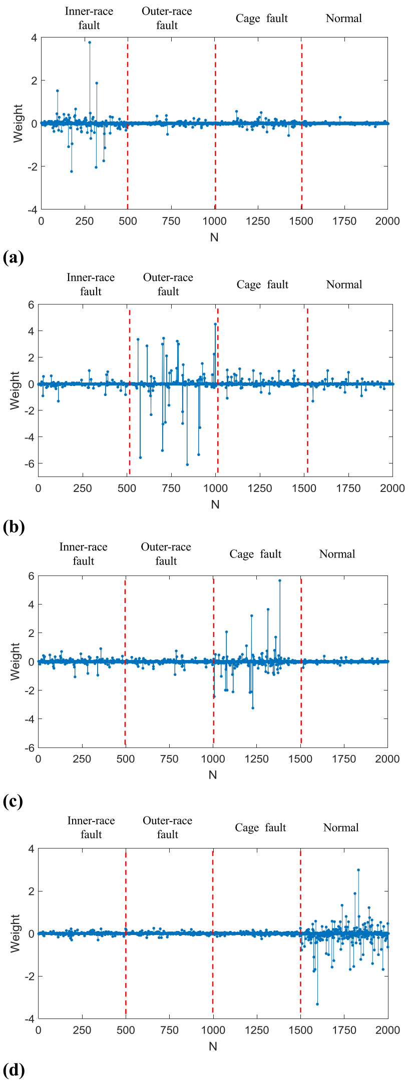
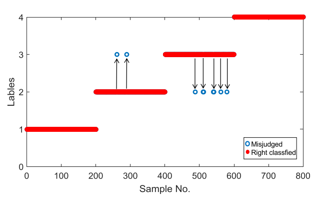


FIGURE 11. Sparse representation coefficients of printing press bearing. (a) IRF (b) ORF (c) CF (d) N.



(a)

IRF	200 25%	0 0.00%	0 0.00%	0 0.00%	100% 0.00%
ORF	0 0.00%	198 24.75%	2 0.25%	0 0.00%	99% 1.00%
CF	0 0.00%	5 0.63%	195 24.37%	0 0.00%	97.50% 2.50%
N	0 0.00%	0 0.00%	0 0.00%	200 25%	100% 0.00%
	100% 0.00%	97.54% 2.46%	98.90% 1.02%	100% 0.00%	99.13% 0.88%
	IRF	ORF	CF	N	

(b)

FIGURE 12. (a) Recognition results of printing press bearing health state (b) Confusion matrix of printing press bearing health state detailed classification.

the fault type. Fig. 11-(a)∼(d) indicate that almost all of the coefficients come from matching groups, and the weight of the sparse coefficients correspond to each fault type.

In addition, the printing press bearing states can be identified by calculating the minimum redundancy error of the test sample on the dictionaries. Fig. 12 presents the sparse classification results. As we can see from Fig. 12-(a) and Fig. 12-(b), there are only seven misclassified target categories in the 800 test samples. All 200 normal test samples are identified correctly, and the identification accuracy of the normal sample is 100%. Among the inner race test samples, 200 samples are correctly identified. The accuracy of the inner race fault test sample is 100%. Among the outer race fault test samples, 198 samples are identified correctly, and only two samples are misdiagnosed as cage faults. The accuracy of the outer race fault test sample is 99%. There are five misclassified target categories in the cage fault test samples. The classification accuracy for detecting rolling bearings with cage faults in 200 samples is 97.5%. Therefore, the comprehensive accuracy of bearing fault identification is 99.13%.

From another perspective, there are 200 samples diagnosed as inner race faults in total, and they are all correct. There are 203 samples diagnosed as outer race faults, of which 198 samples are correct, and the other five samples are misdiagnosed from cage faults. A total of 197 test samples are diagnosed as cage faults, of which two error samples came from outer race faults. In addition, 200 test samples are diagnosed as normal state, all of which are correctly identified. In general, the average accuracy of the proposed DLMN-SRC can reach 99.13% for bearing type fault identification. In addition, each test is repeated ten times to acquire the average accuracy and standard deviation in Table 4. In terms of stability, the standard deviation of all test accuracies with DLMN-SRC is 0.0019. Among them, the Normal type has the highest recognition accuracy and is the most stable. Thus, DLMN-SRC retains a stable diagnosis behavior.

TABLE 4. Accuracy rates of different fault classification.

	IRF	ORF	CF	N	Overall
Average accuracy	99.90%	99.20%	97.45%	100%	99.03%
Standard deviation	0.0022	0.0045	0.0112	0	0.0019

C. COMPARISON WITH DIFFERENT DIAGNOSIS METHODS

In this subsection, three popular and representative methods, namely, K-SVD, D-KSVD and SRC, are used to further prove the superiority of the DLMN-SRC method in bearing fault type identification under a complex background. As a classical dictionary learning algorithm, K-SVD uses a sparse representation dictionary update interactive iterative two-step method to obtain an adaptive redundant dictionary [39]. On the basis of K-SVD, the D-KSVD method inserts a discriminant item into the target function during dictionary learning to obtain a classifier and dictionary that contain both representation ability and discriminant ability. Wright proposed a sparse representation classification (SRC) algorithm and applied it to face recognition [40]. To ensure a fair comparison between the three algorithms, the experiments are run in MATLAB R2015b. The comparison data is from the CWRU case, which contains four different health states with fault depths of 0.007, 0.014 and 0.021 inches under a

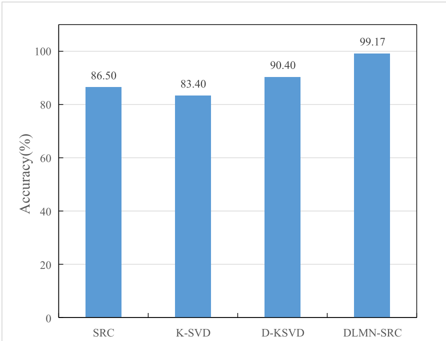


FIGURE 13. Classification performance of the four methods.

0-hp load. We randomly selected 400 samples (100 for each state) as test and training data. In addition, the atomic number K, sparsity threshold T and iteration number I are set as 1200, 20 and 30, respectively.

Fig. 13 presents the recognition results of different classification algorithms based on sparse representation after ten repeated experiments. The comparison results show that the average identification accuracy of DLMN-SRC is 99.17%, which is higher than the three traditional methods. The classification accuracy of SRC and K-SVD are less than 90%, while the recognition rate of D-KSVD is 90.40%. The main reason for the unsatisfactory classification performance of the traditional K-SVD algorithm is that dictionary learning is susceptible to the interference of complex noise and the mixing with false atoms, which affects the performance of dictionary learning. Although the identification accuracy of SRC is not low in the CWRU case, SRC uses all of the training samples to form the dictionary, which is disadvantageous to solving the problem of the sparse coefficient, especially for a larger dataset. The D-KSVD algorithm is sensitive to the initial dictionary, which affects the representation and distinguishing ability of the dictionary.

VI. CONCLUSION

A low-rank representation dictionary learning method on the basis of the mixed noise model was proposed and applied to the sparse representation classification of fault signals to effectively identify rolling bearing fault types. Within the proposed method, a real noise model based on a Gaussian distribution, Laplace distribution and T distribution was constructed in the dictionary learning process. The mixed noise model for complex noise expression had stronger robustness than the single distribution model. The ADMM algorithm was adopted to solve the optimization problem of the mixed noise model. To distinguish different rolling bearing faults, the query signal should belong to the state corresponding to the minimum redundancy error, without additional classifiers or preprocessing. The experimental results showed that the average fault identification accuracy of the two experimental cases was over 99%, among which the single fault type identification accuracy was up to 100%. In addition, compared with the traditional SRC, K-SVD and D-KSVD methods, the DLMN-SRC method has better identification accuracy for rolling bearing faults. The DLMN-SRC method not only extracts the fault feature information in a complex noise environment but also accurately distinguishes the fault types. In the future, we need to improve the calculation efficiency of the proposed DLMN-SRC method when analyzing large-scale data.

ACKNOWLEDGMENT

The authors would like to thank the CWRU and the Shaanxi Key Laboratory of Printing and Packaging Engineering for providing the bearing data.

REFERENCES

[1] C. Mishra, A. K. Samantaray, and G. Chakraborty, "Rolling element bearing fault diagnosis under slow speed operation using wavelet de-noising," *Measurement*, vol. 103, pp. 77–86, Jun. 2017.
[2] H. Zhou, H. Li, T. Liu, and Q. Chen, "A weak fault feature extraction of rolling element bearing based on attenuated cosine dictionaries and sparse feature sign search," *ISA Trans.*, vol. 97, pp. 143–154, Feb. 2020.
[3] Z. Hameed, Y. S. Hong, Y. M. Cho, S. H. Ahn, and C. K. Song, "Condition monitoring and fault detection of wind turbines and related algorithms: A review," *Renew. Sustain. Energy Rev.*, vol. 13, no. 1, pp. 1–39, Jan. 2009.
[4] M. Zhang, Z. Jiang, and K. Feng, "Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump," *Mech. Syst. Signal Process.*, vol. 93, pp. 460–493, Sep. 2017.
[5] C. Gadzhiev, "Dynamic systems diagnosis based on Kalmanl filter updating sequences," *Autom. Remote Control*, vol. 2, no. 1, pp. 50–147, Jan. 1992.
[6] J.-Y. Lee, "Variable short-time Fourier transform for vibration signals with transients," *J. Vib. Control*, vol. 21, no. 7, pp. 1383–1397, Jul. 2013.
[7] M. Kang, J. Kim, and J.-M. Kim, "Reliable fault diagnosis for incipient low-speed bearings using fault feature analysis based on a binary bat algorithm," *Inf. Sci.*, vol. 294, pp. 423–438, Feb. 2015.
[8] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
[9] W. Huang, S. Li, X. Fu, C. Zhang, and Z. Zhu, "Transient extraction based on minimax concave regularized sparse representation for gear fault diagnosis," *Measurement*, vol. 151, pp. 1–11, Feb. 2020.
[10] M. Ghasemi, M. Kelarestaghi, F. Eshghi, and A. Sharifi, "FDSR: A new fuzzy discriminative sparse representation method for medical image classification," *Artif. Intell. Med.*, vol. 106, pp. 1–12, Jun. 2020.
[11] S. Liu, L. Li, M. Jin, S. Hou, and Y. Peng, "Optimized coefficient vector and sparse representation-based classification method for face recognition," *IEEE Access*, vol. 8, pp. 8668–8674, 2020.
[12] J. Guo and P. Zheng, "A method of rolling bearing fault diagnose based on double sparse dictionary and deep belief network," *IEEE Access*, vol. 8, pp. 116239–116253, 2020.
[13] X. Chen, Z. Du, J. Li, X. Li, and H. Zhang, "Compressed sensing based on dictionary learning for extracting impulse components," *Signal Process.*, vol. 96, pp. 94–109, Mar. 2014.
[14] H. Yang, H. Lin, and K. Ding, "Sliding window denoising K-singular value decomposition and its application on rolling bearing impact fault diagnosis," *J. Sound Vib.*, vol. 421, pp. 205–219, May 2018.
[15] R.-B. Sun, Z.-B. Yang, Z. Zhai, and X.-F. Chen, "Sparse representation based on parametric impulsive dictionary design for bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 122, pp. 737–753, May 2019.
[16] J. Jiao, M. Zhao, J. Lin, and K. Liang, "Hierarchical discriminating sparse coding for weak fault feature extraction of rolling bearings," *Rel. Eng. Syst. Saf.*, vol. 184, pp. 41–54, Apr. 2019.
[17] J. Cheng, Y. Peng, Y. Yang, and Z. Wu, "Adaptive sparsest narrow-band decomposition method and its applications to rolling element bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 85, pp. 947–962, Feb. 2017.
[18] Y. Xiao, T. Zeng, J. Yu, and M. K. Ng, "Restoration of images corrupted by mixed Gaussian-impulse noise via l1–l0 minimization," *Pattern Recognit.*, vol. 44, no. 8, pp. 1708–1720, Aug. 2011.

[19] Z. Chen and Y. Wu, "Robust dictionary learning by error source decomposition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 2216–2223.

[20] T. Zhou, Y. Li, and G. Gui, "Noise learning based discriminative dictionary learning algorithm for image classification," *J. Franklin Inst.*, vol. 357, no. 4, pp. 2492–2513, Mar. 2020.

[21] Z. Zhang, X. Yu, and W. Lei, "Signal and noise analysis," in *Communication Theory*. Beijing, China: PHEI, 2018, pp. 20–66.

[22] Z. Zhao, S. Wang, B. An, Y. Guo, and X. Chen, "Hierarchical hyper-Laplacian prior for weak fault feature enhancement," *ISA Trans.*, vol. 96, pp. 429–443, Jan. 2020.

[23] Y. Huang, Y. Zhang, N. Li, and J. Chambers, "Robust student's t based nonlinear filter and smoother," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 5, pp. 2586–2596, Oct. 2016.

[24] Y. Huang, Y. Zhang, N. Li, and J. Chambers, "A robust Gaussian approximate fixed-interval smoother for nonlinear systems with heavy-tailed process and measurement noises," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 468–472, Apr. 2016.

[25] H. Nurminen, T. Ardeshiri, R. Piche, and F. Gustafsson, "Skew-*t* filter and smoother with improved covariance matrix approximation," *IEEE Trans. Signal Process.*, vol. 66, no. 21, pp. 5618–5633, Nov. 2018.

[26] H.-M. Kim, D. Ryu, B. K. Mallick, and M. G. Genton, "Mixtures of skewed Kalman filters," *J. Multivariate Anal.*, vol. 123, pp. 228–251, Jan. 2014.

[27] P. Zhou, C. Fang, Z. Lin, C. Zhang, and E. Y. Chang, "Dictionary learning with structured noise," *Neurocomputing*, vol. 273, pp. 414–423, Jan. 2018.

[28] Z. Feng, M. Yang, L. Zhang, Y. Liu, and D. Zhang, "Joint discriminative dimensionality reduction and dictionary learning for face recognition," *Pattern Recognit.*, vol. 46, no. 8, pp. 2134–2143, Aug. 2013.

[29] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 661–677, Feb. 2013.

[30] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[31] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th Annu. Int. Conf. Mach. Learn. (ICML)*, 2009, pp. 689–696.

[32] Y. Qin, J. Zou, B. Tang, Y. Wang, and H. Chen, "Transient feature extraction by the improved orthogonal matching pursuit and K-SVD algorithm with adaptive transient dictionary," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 215–227, Jan. 2020.

[33] S. Wang, Q. Liu, Y. Xia, P. Dong, J. Luo, Q. Huang, and D. D. Feng, "Dictionary learning based impulse noise removal via L1–L1 minimization," *Signal Process.*, vol. 93, no. 9, pp. 2696–2708, Sep. 2013.

[34] I. W. Selesnick, "The estimation of laplace random vectors in additive white Gaussian noise," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3482–3496, Aug. 2008.

[35] D. Xu, C. Shen, and F. Shen, "A robust particle filtering algorithm with non-Gaussian measurement noise using Student-t distribution," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 30–34, Jan. 2014.

[36] Y. Huang and Y. Zhang, "Robust Student's t-based stochastic cubature filter for nonlinear systems with heavy-tailed process and measurement noises," *IEEE Access*, vol. 5, pp. 7964–7974, 2017.

[37] Y. Huang, Y. Zhang, Z. Wu, and N. Li, "A novel robust student's t based Kalman filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 3, pp. 1545–1554, Jun. 2017.

[38] T. M. D. Tran and A. Y. Kibangou, "Distributed estimation of graph Laplacian eigenvalues by the alternating direction of multipliers method," *IFAC Proc. Vols.*, vol. 47, no. 3, pp. 5526–5531, 2014.

[39] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[40] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Apr. 2009.

JIALING ZHANG received the M.S. degree from the Xi'an University of Technology, Xi'an, China, in 2018, where she is currently pursuing the Ph.D. degree with the School of Mechanical and Precision Instrument Engineering. Her research interests include mechanical equipment intelligent fault diagnosis theory and method, and signal processing.

JIMEI WU was born in 1963. She received the M.S. degree in printing engineering and the Ph.D. degree in mechanical engineering from the Xi'an University of Technology, China, in 1999 and 2009, respectively. She is currently a Professor with the School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology. Her research interests include mechanical equipment fault diagnosis, structural dynamics analyses of printing equipment, and simulation of machinery.

BINGBING HU was born in 1983. He received the M.S. degree from the Lanzhou University of Technology, in 2009, and the Ph.D. degree from Xi'an Jiaotong University, China, in 2017. His research interests include intelligent maintenance and control technology of complex equipment, fault diagnosis, and stochastic resonance.