Research and Application of Regularized Sparse Filtering Model for Intelligent Fault Diagnosis Under Large Speed Fluctuation

BAOKUN HAN, GUOWEI ZHANG, JINRUI WANG, XIAOYU WANG, SIXIANG JIA, AND JINGTAO HE
College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao 266510, China
Corresponding author: Jinrui Wang (wangjr33@163.com)

ABSTRACT The speed of mechanical rotating parts often fluctuates during the working process. Vibration signals collected under constant speed have a strong correlation with the corresponding fault types. However, the mapping relationship becomes complex under large speed fluctuation, which is an urgent research subject in intelligent fault diagnosis. As an effective unsupervised learning method, sparse filtering (SF) has been successfully used in intelligent fault diagnosis. However, the generalization capability of this method to deal with large speed fluctuation remains poor. To overcome this deficiency, this study adds regularization to the loss function of SF to obtain regularized SF methods. The frequency domain signals under large speed fluctuation are directly input to regularized SF for feature extraction, and softmax regression is used as a classifier for fault type identification. Experimental results of gearbox and bearing datasets show that L1/2 regularized sparse filtering (L1/2-SF) model can solve the problem of large speed fluctuation more effectively than other regularized SF models can.

INDEX TERMS Rotating parts, fault diagnosis, large speed fluctuation, sparse filtering, regularization.

I. INTRODUCTION
Rotating parts are important components of many machines used in spacecraft, high-speed rail, and automobile. Failure of these parts directly affects the operation of machines and results in huge economic losses and even casualties [1]. Thus, precise diagnosis of the health condition of machines is necessary. Intelligent fault diagnosis has achieved remarkable success in recent years [2], [3]. Most of the methods used in fault diagnosis can achieve high accuracy under the general assumption that training and testing datasets are extracted at constant rotating speeds [4], [5]. However, the rotating speed of mechanical transmission varies under different operating conditions and loads [6]. For example, the wheel bearing of high-speed rail is influenced by the running speed and start–stop of the train [7], the speed of the automobile gearbox is regulated by the road condition [8], the speed of the top drive transmission system in an oil drilling machine fluctuates considerably during the working process [9], and the speed of wind turbine blades varies with the wind condition [10]. In summary, speed fluctuation widely exists in many mechanical equipment.

Increased attention has been paid to research on intelligent fault diagnosis under large speed fluctuation. Time–frequency analysis technologies [11], such as short-time Fourier transform, Wigner–Ville distribution [12], and wavelet transform [13], have been widely used in such research. Given that the vibration signal of mechanical equipment is time-varying and strongly coupled in the time–frequency domain, solving the high time–frequency resolution display problem of speed fluctuation signals is difficult when these methods are used. Several researchers [14]–[16] utilized order tracking to process time-varying signals and found that this approach is effective in eliminating the effects of speed fluctuation. However, all of the methods mentioned above require human labor and cause inconvenience.

Unsupervised feature learning could overcome these shortcomings. In essence, unsupervised learning can be regarded as learning a nonlinear function for transforming unlabeled raw data from the original space to a feature space. This technique has been widely applied in audio,
video, image, and other fields [17], [18]. Sparse filtering (SF) [19] is an unsupervised two-layer network that automatically extracts features without label participation. This method needs to adjust only one parameter, thus demonstrating simplicity and easy implementation. Lei et al. [4] utilized SF in bearing fault diagnosis by adopting a two-stage learning method, which greatly reduced human labor and eased the handling of big data in intelligent fault diagnosis. Yang et al. [20] used SF to extract sparse features directly from raw time-domain signals and classified health status using traditional support vector machines based on improved particle swarm optimization. An et al. [21] solved the over fitting problem in bearing fault diagnosis by removing multi-correlation operations in the weight matrix of SF.

Although SF can directly learn from the original vibration signal to the deep discriminative features, its generalization capability in dealing with large speed fluctuation remains poor. In deep learning, many strategies have been explicitly designed to reduce generalization errors, and these strategies are collectively known as regularization [22]. In this study, we add regularization strategy to the objective function of SF to constrain the sparsity of the weight matrix and improve the generalization capability of standard SF. In addition, the initial phase of the time-domain signal is different, and it increases the difficulty of feature extraction under large speed fluctuation. However, the frequency spectra can make the samples exhibit regularity [2], which can eliminate this interference. The frequency spectra of the vibration signal are adopted as the model input. So, in this paper, regularized SF models are proposed to deal with the frequency-domain signal under large speed fluctuation.

Therefore, the main contributions of this paper are summarized as follows.

1. To deal with large speed fluctuation problem in fault diagnosis, some regularization strategies are introduced into the penalty term of sparse filtering objective function, so as to improve the feature extraction ability of sparse filtering.

2. We try to explore a physical interpretation of L1/2 regularization [23] in its robust ability on dealing with non-stationary signal.

3. Through the comparison of different regularized SF models. The L1/2 regularized sparse filtering (L1/2-SF) model can get sparser solution and higher precision than other models.

The remainder of this paper is organized as follows. In Section 2, the SF and the regularization strategy in deep learning are introduced. In Sections 3 and 4, the regularization strategy is used for SF, and gear and bearing datasets are investigated using the different regularized SF models, respectively. The conclusion is presented in Section 5.

**II. THEORETICAL BACKGROUND**

**A. SPARSE FILTERING**

Among the many unsupervised algorithms, SF, a simple and effective feature learning algorithm, requires the least parameter adjustment, and it can directly analyze the optimal feature distribution bypassing the estimation of data distribution. As shown in Fig. 1, the inputs and outputs of SF are the sample dataset and the learned features, respectively.

First, the original time-domain data collected from each health condition are divided into $M$ samples, and the training set $\{x_i\}_{i=1}^M$ is formed using fast Fourier transform (FFT), where $x^i = \mathcal{F}[x]$ is a sample containing $N$ Fourier coefficients. Second, weight matrix $W \in \mathbb{R}^{N \times L}$ is obtained by training the SF model with $\{x_i\}_{i=1}^M$. Finally, each sample is mapped into feature vector $f^i = \mathcal{F}[x]$ by $W$. In addition, the rectified linear unit function [24] is used as the activation function. The features of each sample can be calculated as

$$f^i = \text{ReLU}(W^T x^i).$$

The feature matrix is composed of $f^i$, and the L2-norm of the feature matrix is used to normalize each row of samples.

$$\tilde{f} = f / \|f\|_2$$

Then, each column is normalized by its L2-norm, and the features are mapped into the unit L2-sphere.

$$\tilde{f}^i = \frac{\tilde{f}^i}{\|\tilde{f}\|_2}$$

Afterward, the L1 penalty is used to optimize the sparsity of the normalized feature. The cost function of SF is given as

$$\min_W \sum_{i=1}^M \|\tilde{f}^i\|_1 = \sum_{i=1}^M \left(\|\tilde{f}^i\|_2 - 1\right).$$

**B. L1/2 REGULARIZATION**

Regularization is the process of reducing test errors. The ultimate goal of the deep learning model is to achieve satisfactory performance when dealing with new data. However, overfitting occurs when the assumption is complex or the input dataset is small. Regularization terms are often introduced into the learning process to prevent overfitting and improve generalization performance [25].

Many regularization strategies are currently available; several of them add constraints to the learning model to constrain the value of parameters or add items to the objective function to soft-constrain the value of parameters. This action is equivalent to creating a constraint on the original solution space when the regularization term is added to the original objective function and then determining the minimum value under this constraint. In general, the regularization method can be expressed as

$$\min \left\{ \frac{1}{n} \sum_{i=1}^n I(y_i, f(x_i)) + \lambda \|w\|_k \right\}$$

**FIGURE 1. SF structure.**
where $l$ represents the objective function, $\{(x_i, y_i)\}_{i=1}^{n}$ is a known dataset, and $\lambda$ denotes the regularization parameter that aims to control the model complexity of machine learning algorithms. If $f$ is a linear function and its objective function is a square loss, $\|w\|_k$ is generally used to represent the $k$-norm of the coefficient.

Almost all machine learning algorithms can be interpreted as the above mentioned regularization form. If $k = 0$, then L0 regularization is obtained, that is, the number of non-zero coefficients in the model will be punished according to AIC and BIC rules. If $k = 1$, then L1 regularization corresponding to the Lasso algorithm is obtained. If $k = 2$, then L2 regularization corresponding to ridge estimation is present.

If too many redundant variables are present in a representation model, the feature extraction process should eliminate as many redundant variables as possible while ensuring that the real variables can be identified, which corresponds to the sparsity problem. We consider the variable selection problem of the following sparse linear model.

$$y = X^T \beta + \varepsilon \quad E \varepsilon = 0 \quad Cov(\varepsilon) = \sigma^2 I, \quad (6)$$

where $y = (y_1, y_2, \ldots, y_n)^T$ is an $n \times 1$ vector; $X = (X_1, X_2, \ldots, X_m)$, $X_i = (x_{i1}, x_{i2}, \ldots, x_{in})$, $i = 1, 2, \ldots, n$; $b = (b_1, b_2, \ldots, b_n)^T$ is an unknown $p \times 1$ parameter; $\varepsilon$ is the random error; and $\sigma^2$ is the error variance. We assume that the real model is $f^*(x, b) = b_1^* x_1 + b_2^* x_2 + \cdots + b_p^* x_p$ and $p_0 \leq p$.

L0 regularization can be expressed as

$$\beta_{L0} = \arg \min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i^T \beta)^2 + \lambda \sum_{i=1}^{p} |\beta_i| \right\}. \quad (7)$$

L0 regularization can produce the sparsest solution. Thus, it is the most ideal variable selection method among all regularization methods. However, L0 regularization is inapplicable to massive or high-dimensional data it needs to solve a combinatorial optimization problem.

Therefore, L2 regularization is developed as

$$\beta_{L2} = \arg \min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i^T \beta)^2 + \lambda \sum_{i=1}^{p} |\beta_i|^2 \right\}. \quad (8)$$

L2 regularization can process high-dimensional and massive data and generate smooth solutions, thereby exhibiting excellent performance in solving the overfitting problem. However, this method does not possess sparsity.

Thus, L1 regularization is developed as

$$\beta_{L1} = \arg \min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i^T \beta)^2 + \lambda \sum_{i=1}^{p} |\beta_i| \right\}. \quad (9)$$

L1 regularization can take advantage of its feature selection ability to produce a relatively sparse solution by solving a convex optimization problem [26]. However, considering that the objective function of the regularization model is not smooth near zero, we cannot directly use the conventional derivation optimization method. Furthermore, the complexity of the solution might increase, and an inconsistent problem in the selection of a certain variable may be generated [27].

Thus, L1/2 regularization based on a non-convex penalty is developed as

$$\beta_{L1/2} = \arg \min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i^T \beta)^2 + \lambda \sum_{i=1}^{p} |\beta_i|^2 \right\}. \quad (10)$$

At this time, the penalty term is not convex, which provides a new idea for dealing with the problem of variable selection and feature extraction. Many studies have reported that the L1/2 regularization model can produce a sparser solution than that of L1 regularization. In addition, it possesses excellent theoretical properties, such as robustness and non-bi-partiality [28].

Fig. 2 shows the corresponding penalty term graphs to analyze the sparsity of the above mentioned regularization models in a two-dimensional case (i.e., only two parameters, namely, $w_1$ and $w_2$, need to be learned). The center of the red circle is the smallest error, and the errors on each red line are the same. The regularized equation adds the extra error on the blue edge, and the errors on the blue edge are also the same. The green point at the intersection of the red line and blue edge can minimize the sum of two errors, which represents the regularized solution for $w_1$ and $w_2$. As shown in Fig. 2(a), the contour line and L2 regularization cannot intersect on the coordinate axis, so the solution is not sparse. In Fig. 2(b), the constraint region of L1 regularization is a square, so the intersection of the contour line and square vertex must be zero, which makes the solution sparse. The constraint region of L1/2 regularization is shown in Fig. 2(c); it is likely to intersect with the contour line on the coordinate axis, thereby increasing the sparsity of the solution.

Only $w_2$ ($w_1 = 0$) is obtainable when L1 and L1/2 regularizations are used. The solution of L1 regularization is unstable when the batch dataset is used for training. Fig. 3 shows that the error curves of each training epoch are
slightly different. The green point does not move far from L2 regularization, as displayed in Fig. 3(a), but the green point of L1 regularization can jump to many different places, as shown in Fig. 3(b), because the total error is almost the same. By contrast, the green point of L1/2 regularization is unmoved, as presented in Fig. 3(c), indicating that this regularization can produce the sparsest and most robust solution among the three regularization models.

A batch dataset with the same fault type is expected to demonstrate a huge difference due to the effect of speed fluctuation, which makes network training difficult. In this study, the excellent characteristics of the regularization strategy in machine learning are combined with those of the SF model to address the speed fluctuation problem in fault diagnosis.

The three regularization terms are then added to the objective function of SF.

L2 regularized SF (L2-SF):

\[
\text{minimize}_{W} \sum_{i=1}^{M} \left\| \hat{f}_i \right\|_1 + \lambda \|W\|_2
\]

\[
= \sum_{i=1}^{M} \left\| \hat{f}_i \right\|_1 + \lambda \sum_{i=1}^{C} \sum_{j=1}^{R} |w_{ij}|^2 \quad (11)
\]

L1 regularized SF (L1-SF):

\[
\text{minimize}_{W} \sum_{i=1}^{M} \left\| \hat{f}_i \right\|_1 + \lambda \|W\|_2
\]

\[
= \sum_{i=1}^{M} \left\| \hat{f}_i \right\|_1 + \lambda \sum_{i=1}^{C} \sum_{j=1}^{R} |w_{ij}| \quad (12)
\]

L1/2 regularized SF (L1/2-SF):

\[
\text{minimize}_{W} \sum_{i=1}^{M} \left\| \hat{f}_i \right\|_1 + \lambda \|W\|_2
\]

\[
= \sum_{i=1}^{M} \left\| \hat{f}_i \right\|_1 + \lambda \sum_{i=1}^{C} \sum_{j=1}^{R} |w_{ij}|^{1/2} \quad (13)
\]

where \( C \) and \( R \) represent the number of columns and rows in the weight matrix, respectively.

III. PROPOSED FRAMEWORK

The framework of the proposed method is displayed in Fig. 4. The specific steps are as follows.

IV. CASE 1: FAULT DIAGNOSIS OF A PLANETARY GEARBOX

A. DATA DESCRIPTION

A specially designed planetary gearbox fault test bench is used to collect vibration signals under large speed fluctuation. Fig. 5 presents the test bench, which includes a motor, a planetary gearbox, two shaft couplings, and two bearing seats. As shown in Fig. 6, the following 10 health conditions of gear are designed: normal condition (NC), sun wheel crack (WC), sun wheel pit (WP), sun wheel worn tooth (WW), pinion crack (PC), pinion pit (PP), pinion worn tooth (PW), wheel worn and pinion worn (WPPW), wheel pit and pinion worn (WPPC), and wheel pit and pinion worn (WPPW).

The sampling frequency is set to 12.8 kHz, and the rotating speed ranges from 700 r/min to 1500 r/min. 250 samples are collected for each health condition, and each sample contains 2400 data points; thus, a total of 2500 samples are obtained. After applying FFT to each time-domain sample, 1200 Fourier coefficients are obtained for each sample.

For the speed fluctuation condition, three gear examples (NC, WW, and WPPW) are randomly selected and shown in Figs. 7(a)–7(c). The vibration signals generated by the three gear fault types exhibit different speed fluctuations.

For the speed fluctuation condition, three gear examples (NC, WW, and WPPW) are randomly selected and shown in Figs. 7(a)–7(c). The vibration signals generated by the three gear fault types exhibit different speed fluctuations.
Algorithm 1 Training of Regularized SF, The \textit{iter} Is Present Number of Iteration Regularized SF

\textbf{Input} \hspace{0.5em} training dataset $S \in \mathbb{R}^{N_{in} \times K}$, regularization parameters $\lambda$

\textbf{Output} \hspace{0.5em} The weight matrix $W \in \mathbb{R}^{N_{out} \times N_{in}}$

1: \textbf{Initialization} \hspace{0.5em} The iteration step $i = 0$, $W$ is randomly initialized
2: \textbf{for} $i = 1$ \textbf{to} \textit{iter} \textbf{do} \\
3: \hspace{1em} Calculate $W$ based on $S$ using minimizing (11-13) \\
4: \hspace{1em} $i = i + 1$ \\
5: \textbf{end for} \\
6: \hspace{1em} Calculate learning features based on $W$ and $S$ \\
7: \textbf{End}

in the time domain. Fig. 7(d) displays the irregular speed fluctuation information of the three gear fault types between 700 and 1500 r/min and the three curves of the three different gear rotating speeds. Meanwhile, two time-domain samples and their corresponding frequency-domain samples are randomly selected from NC, WW, and WWPW. Fig. 8 shows that distinguishing the different health conditions from the time and frequency domains of the signal is difficult because even the samples under the same health condition have different feature distributions.

\textbf{B. DIAGNOSIS RESULTS}

In this section, the regularized SF is used to deal with the gear dataset under large speed fluctuation. Two feature parameters, namely, $N_{in}$ and $N_{out}$, need to be adjusted. After testing, we set 1200 as the dimension of $N_{in}$ and $N_{out}$. First, we randomly select 50\% of the frequency-domain samples from each health condition. Second, the frequency-domain samples are used to create the training set for training the regularized SF model and obtaining $W$. Third, we use the remaining 50\% samples as the testing set and calculate the learned features using $W$. A 1200-dimensional feature
We investigate the selection of regularization parameters. The diagnosis results of L1-SF are displayed in Fig. 9. The accuracies are stable after $\lambda = 1E^{-5}$, so $1E^{-5}$ is chosen as the regularization parameter in this paper.

The raw SF model without regularization is utilized for comparison. Each experiment is repeated 20 times to reduce the effects of randomness. The testing accuracies of the four methods are displayed in Fig. 10. L1/2-SF has the highest average testing accuracy (99.54%) and the lowest standard deviation (0.07%) among all the methods. The testing accuracies of L2-SF and L1-SF are 98.88% ± 0.28% and 98.63% ± 0.11%, respectively, which are slightly lower than that of L1/2-SF. The raw SF model, with an average testing accuracy of 98.6% ± 0.21%, exhibits the most unsatisfactory performance among the methods. In conclusion, L1/2-SF demonstrates the best performance when dealing with large fluctuations in rotating speed.

To confirm the advantages of L1/2-SF, we use t-distributed stochastic neighbor embedding (t-SNE) [29] to reduce the dimensions of the high-dimensional features obtained by the four methods and provide visualization. The results of the gear dataset processed by t-SNE are shown in Fig. 11. The dimension reduction features of SF and L1-SF are not clustered well, and the scattered points of the gear samples
under different health conditions are interlaced, as shown in Figs. 11(a) and 11(b). Fig. 11(c) illustrates that the features learned by the L2-SF cluster are better than those learned by SF and L1-SF. However, many samples are still separated based on the corresponding health condition. The dimension reduction result of L1/2-SF is displayed in Fig. 11(d), which shows that except for the slight mixing of WC and PC, the other gear samples under different health conditions are well distinguished, and feature points with the same health condition are gathered together. Therefore, L1/2-SF can effectively extract features from the frequency-domain signal and process gear datasets under large speed fluctuation.

V. CASE 2: FAULT DIAGNOSIS OF A MOTOR BEARING

A. DATA DESCRIPTION
The bearing data are measured from the test rig, as shown in Fig. 12(a). The rig includes a motor, three shaft couplings, a bearing seat, and a brake. The bearing dataset contains three fault types, namely, outer race (OF), inner race (IF) and roller faults (RF). The vibration signals of three damage severity levels (0.2, 0.4, and 0.6 mm) are separately collected from the three fault types, and the signals of NC are added to the dataset, resulting in 10 health conditions. The three fault bearings are depicted in Fig. 12(b). The accelerometer is mounted on the bearing box with a sampling frequency...
of 25.6 kHz. 200 samples are collected for each health condition, and 1200 Fourier coefficients are obtained for each sample.

For the speed fluctuation condition, four bearing examples (NC, IF0.2, OF0.4, and RF0.6) are randomly selected. Fig. 13 shows the irregular speed fluctuation information of the four bearing fault types between 500 and 3000 r/min and the four different curves corresponding to the four different gear speeds.

### B. DIAGNOSIS RESULTS

The network structure and parameter set are similar to those in Case 1. Fifty percent of the samples are used for training, and the rest is used for testing. The classification accuracies of the 20 trials using the four methods are shown in Fig. 14. The second trial testing accuracy of L1/2-SF is 99.99%, and that of the rest are all 100%. L2-SF, which has 99.93% ± 0.10% testing accuracy, ranks a very close second. By contrast, the average accuracies of SF and L1-SF are low and have considerably large standard deviations. This finding indicates that the L1/2-SF method achieves the highest accuracy and robustness among the four methods in terms of diagnosing bearing fault types under large speed fluctuation.

The results of the bearing dataset processed by t-SNE are shown in Fig. 15. Fig. 15(a) indicates that the dimension reduction features of the SF model are the most unsatisfactory among the four methods, with only four bearing types separated and the other types of samples still mixed. The performance of L1-SF is slightly better than that of SF, as shown in Fig. 15(b). However, a large area of mixing is observed, which indicates failure to achieve fault classification. Figs. 15(c) and 15(d) present the respective feature visualization maps obtained by L2-SF and L1/2-SF methods. The feature points of the same health condition are gathered together. The L1/2-SF method has a better effect compared with L2-SF, and the mapped features between different fault types are obviously separated. This result reveals the superior performance of L1/2-SF in dealing with bearing fault diagnosis under large speed fluctuation.

### VI. COMPARING WITH OTHER METHODS

In order to prove the effectiveness of the proposed L1/2-SF model, it is compared with other intelligent fault diagnosis methods, as summarized in Table 1. The gearbox experiment shows that the L1/2-SF method can achieve the highest testing accuracy of 98.6%. Meanwhile, the L1/2-SF achieves the best testing accuracy of 100% when classifying ten different bearing faults. Two experiments show that L1/2-SF is more efficient and accurate than the other three classical methods.

![Figure 13. Large speed fluctuation information of the four bearing fault types.](image)

![Figure 14. Diagnostic results of 20 trials on bearing datasets using the four methods.](image)

| Method | No. of training samples (%) | Testing accuracy of gearbox dataset (%) | Testing accuracy of bearing dataset (%) |
|--------|-----------------------------|----------------------------------------|----------------------------------------|
| SAE    | 50                          | 95.47                                  | 96.55                                  |
| SDA    | 50                          | 94.56                                  | 96.23                                  |
| CNN    | 50                          | 94.89                                  | 94.46                                  |
| L1/2-SF| 50                          | 98.60                                  | 100                                    |
FIGURE 15. Feature visualization map of the t-SNE result for the bearing dataset: (a) SF, (b) L1-SF, (c) L2-SF, and (d) L1/2-SF.

VII. CONCLUSION

(1) The different regularized SF models are proposed to deal with frequency-domain signals under large speed fluctuation. The results of case studies on gearbox and bearing datasets show that the L1/2-SF model can adaptively extract features from the frequency spectra signal under large speed fluctuation and is superior to other regularized SF models.

(2) An intelligent fault diagnosis framework based on regularized SF is constructed for the problem of large speed fluctuation. The frequency spectra of the vibration signal are used as the input of the framework to overcome the problem of different initial phases of the time-domain signal and obtain high diagnosis accuracy.

(3) Given that the signal exhibits variability under large speed fluctuation, massive data are required to train the model. However, data are insufficient in many cases. Our future study will utilize data enhancement techniques to overcome this shortcoming.

(4) Two experiments show that the algorithm can be applied in the range of 500-3000r/min. We will study the maximum speed range of the algorithm in the future work.

REFERENCES

[1] X. M. Zhao, Q. H. Hu, Y. G. Lei, and M. J. Zuo, “Vibration-based fault diagnosis of slurry pump impellers using neighbourhood rough set models,” Proc. Inst. Mech. Eng. C: J. Mech. Eng. Sci., vol. 224, no. 4, pp. 995–1006, Apr. 2010.

[2] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, “Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data,” Mech. Syst. Signal Process., vol. 72–73, pp. 303–315, May 2016.

[3] J. Wang, S. Li, B. Han, Z. An, X. Yin, W. Qian, and Q. Wu, “Construction of a batch-normalized autoencoder network and its application in mechanical intelligent fault diagnosis,” Meas. Sci. Technol., vol. 30, no. 1, Dec. 2018, Art. no. 015106.

[4] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, “An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data,” IEEE Trans. Ind. Electron., vol. 63, no. 5, pp. 3137–3147, May 2016.

[5] X. Jiang, S. Li, and Y. Wang, “A novel method for self-adaptive feature extraction using scaling crossover characteristics of signals and combining with LS-SVM for multi-fault diagnosis of gear box,” J. Vibroeng., vol. 17, no. 2015, pp. 1861–1878.

[6] W. Huang, G. Gao, N. Li, X. Jiang, and Z. Zhu, “Time-frequency squeezing and generalized demodulation combined for variable speed bearing fault diagnosis,” IEEE Trans. Instrum. Meas., vol. 68, no. 8, pp. 2819–2829, Aug. 2019.

[7] H. Liu, W. Huang, S. Wang, and Z. Zhu, “Adaptive spectral kurtosis filtering based on morlet wavelet and its application for signal transients detection,” Signal Process., vol. 96, pp. 118–124, Mar. 2014.

[8] X. Jiang, S. Li, and Q. Wang, “A study on defect identification of planetary gearbox under large speed oscillation,” Math. Problems Eng., vol. 2016, Jan. 2016, Art. no. 5289698.

[9] J. Urbanek, T. Barszcz, R. Zimroz, and J. Antoni, “Application of averaged instantaneous power spectrum for diagnostics of machinery operating under non-stationary operational conditions,” Measurement, vol. 45, no. 7, pp. 1782–1791, Aug. 2012.

[10] J. Wang, R. X. Gao, and R. Yan, “Multi-scale enveloping order spectrogram for rotating machine health diagnosis,” Mech. Syst. Signal Process., vol. 46, no. 1, pp. 28–44, May 2014.
[11] A. Roshan-Ghias, M. B. Shamsollahi, M. Mobed, and M. Behzad, “Estimation of modal parameters using bilinear joint time–frequency distributions,” *Mech. Syst. Signal Process.*, vol. 21, no. 5, pp. 2125–2136, Jul. 2007.

[12] B. Tang, W. Liu, and T. Song, “Wind turbine fault diagnosis based on Morlet wavelet transformation and Wigner–Ville distribution,” *Renew. Energy*, vol. 35, no. 12, pp. 2862–2866, Dec. 2010.

[13] R. Yan, R. X. Gao, and X. Chen, “Wavelets for fault diagnosis of rotary machines: A review with applications,” *Signal Process.*, vol. 96, pp. 1–15, Mar. 2014.

[14] R. Randall and J. Antoni, “Rolling element bearing diagnostics—A tutorial,” *Mech. Syst. Signal Process.*, vol. 25.2, no. 2011, pp. 485–520.

[15] P. Borghesani, R. Ricci, S. Chatterton, and P. Pennacchi, “A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions,” *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 23–35, Jul. 2013.

[16] P. D. Mcfadden and M. M. Toozhy, “Application of synchronous averaging to vibration monitoring of rolling element bearings,” *Mech. Syst. Signal Process.*, vol. 14, no. 6, pp. 891–906, Nov. 2000.

[17] K. B. Raja, R. Raghavendra, V. K. Vemuri, and C. Busch, “Smartphone based visible iris recognition using deep sparse filtering,” *Pattern Recognit. Lett.*, vol. 57, pp. 33–42, May 2015.

[18] W. E. Hahn, S. Lewkowitz, D. C. Lacombe, and E. Barenholtz, “Deep learning human actions from video via sparse filtering and locally competitive algorithms,” *Multimedia Tools Appl.*, vol. 74, no. 22, pp. 10997–11010, Aug. 2015.

[19] J. Ngiam, “Sparse filtering,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 1125–1133.

[20] Y. Yang, P. Xiao, Y. Cheng, and X. Zhang, “Sparse filtering based intelligent fault diagnosis using IPSO-SVM,” in *Proc. 36th Chin. Control Conf. (CCC)*, Jul. 2017, pp. 7388–7393.

[21] Z. An, S. Li, J. Wang, W. Qian, and Q. Wu, “An intelligent fault diagnosis approach considering the elimination of the weight matrix multi-correlation,” *Appl. Sci.*, vol. 8, no. 6, p. 906, Jun. 2018.

[22] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, Cambridge, MA, USA: MIT Press, 2016.

[23] Z. Xu, X. Chang, F. Xu, and H. Zhang, “L1/2 regularization: A thresholding representation theory and a fast solver,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, Jul. 2012.

[24] V. Nair and G. Hinton, “Rectified linear units improve restricted Boltzmann machines,” in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.

[25] W. Wu, Q. Fan, J. M. Zurada, J. Wang, D. Yang, and Y. Liu, “Batch gradient method with smoothing regularization for training of feedforward neural networks,” *Neural Netw.*, vol. 50, pp. 72–78, Feb. 2014.

[26] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *J. Roy. Stat. Soc. B Methodol.*, vol. 58, no. 1, pp. 267–288, Dec. 2018.

[27] H. Zou, “The adaptive lasso and its oracle properties,” *J. Amer. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, Dec. 2006.

[28] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, “Iteratively reweighted least squares minimization for sparse recovery,” *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, Jan. 2010.

[29] L. van der Maaten and G. Hinton, “Visualizing data using t-SNE,” *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

**BAOKUN HAN** received the Ph.D. degree from the School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China, in 2003.

He is currently a Professor with the Shandong University of Science and Technology (SDUST). His current research interests include signal processing and fault diagnosis, vehicle dynamics, and noise and vibration analysis and control.

**GUOWEI ZHANG** received the B.S. degree from Zaozhuang University, Zaozhuang, China, in 2018.

He is currently working as a Graduate Student at the Shandong University of Science and Technology, Qingdao, China. His current research interests include intelligent fault diagnosis of machines, and noise and vibration analysis and control.

**JINRUI WANG** received the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, China, in 2019.

He is currently working as a Postdoctoral Researcher with the Shandong University of Science and Technology, Qingdao, China. His current research interest focuses on intelligent fault diagnosis of machines.

**XIAOYU WANG** received the B.S. degree from the Liren College of Yanshan University, Qinhuangdao, China, in 2017.

He is currently working as a Graduate Student at the Shandong University of Science and Technology, Qingdao, China. His current research interests include intelligent fault diagnosis of machines, and noise and vibration analysis and control.

**SIXIANG JIA** received the B.S. degree from the Shandong University of Science and Technology, Qingdao, China, in 2018.

He is currently working as a Graduate Student at the Shandong University of Science and Technology. His current research interests include intelligent fault diagnosis of machines, and noise and vibration analysis and control.

**JINGTAO HE** received the B.S. degree from Zaozhuang University, Zaozhuang, China, in 2018.

He is currently working as a Graduate Student at the Shandong University of Science and Technology, Qingdao, China. His current research interests include intelligent fault diagnosis of machines, and noise and vibration analysis and control.

**BAOKUN HAN** received the Ph.D. degree from the School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China, in 2003.

He is currently a Professor with the Shandong University of Science and Technology (SDUST). His current research interests include signal processing and fault diagnosis, vehicle dynamics, and noise and vibration analysis and control.

**GUOWEI ZHANG** received the B.S. degree from Zaozhuang University, Zaozhuang, China, in 2018.

He is currently working as a Graduate Student at the Shandong University of Science and Technology, Qingdao, China. His current research interests include intelligent fault diagnosis of machines, and noise and vibration analysis and control.

**JINRUI WANG** received the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, China, in 2019.

He is currently working as a Postdoctoral Researcher with the Shandong University of Science and Technology, Qingdao, China. His current research interest focuses on intelligent fault diagnosis of machines.

**XIAOYU WANG** received the B.S. degree from the Liren College of Yanshan University, Qinhuangdao, China, in 2017.

He is currently working as a Graduate Student at the Shandong University of Science and Technology, Qingdao, China. His current research interests include intelligent fault diagnosis of machines, and noise and vibration analysis and control.

**SIXIANG JIA** received the B.S. degree from the Shandong University of Science and Technology, Qingdao, China, in 2018.

He is currently working as a Graduate Student at the Shandong University of Science and Technology, Qingdao, China. His current research interests include intelligent fault diagnosis of machines, and noise and vibration analysis and control.