Application of Sparse Representation Based on Novel K-SVD Algorithms in Mechanical Fault Diagnosis

Yutao Lan
School of Mechanical-electronic and Automobile Engineering, Beijing University of Civil Engineering and Architecture
Beijing, China
lanyutao@163.com

Yanxue Wang
School of Mechanical-electronic and Automobile Engineering, Beijing University of Civil Engineering and Architecture
Beijing, China
WYX1999140@126.com

***Abstract*—In order to solve the problems of dictionary redundant, the low calculating speed and the influence, based on the theory of sparse representation, integrating the ideas of SVD (singular value decomposition) and sliding window. A sparse representation diagnosis method for bearing impact fault based on SWD-KSVD (sliding window denoising-K means singular value decomposition) is proposed. Firstly, the K-SVD algorithm commonly used in dictionary training of sparse representation is applied to learn a short section of time-domain signal containing impact and reconstruct this signal. Then, the best mode with only one impact was selected based on variance by sliding window. Finally, the sliding window inner product of the optimal mode is calculated on the whole fault signal to find the time when the impact occurs and reconstruct the impact, and fault characteristic extraction of rolling bearing would be realized. This mentioned algorithm only uses a shorter optimal mode and carries out optimal selection and correlation analysis, thus, the algorithm has a high reduction degree of impact morphology, a more accurate phase restoring, a faster speed of extracting fault feature, and a stronger anti-noise ability. Experiment shows that the proposed algorithm can effectively extract bearing impact fault features.**

Keywords: KSVD; inner product; spare representation; fault diagnosis

I. INTRODUCTION

At present, many scholars have researched mechanical fault diagnosis based on sparse representation algorithm from many aspects, and several kinds signal processing algorithms under sparse representation which has been established. Wang and Cui [1, 2] focused on the method of how to select the right atoms to extract the fault impact component, and they have achieved some works. According to the principle of gearbox bearing and drive system, Ding and He [3] designed a stable dictionary and impact dictionary to sparsely represent fault signals and their efforts yielded good results. Yu [4] applies K-SVD to extract atoms from early fault of bearing and the maximum kurtosis is used as the cut-off condition at the sparse coding stage, the adaptability of algorithm was improved correspondingly. Dong [5] combined K-SVD with MED (Minimum Entropy Deconvolution) for refining the accuracy of fault diagnosis, MED could reduce the influence of strong background noise in the signal. Feng and Liang [6] omitted the step of signal segmentation in K means-SVD algorithm in the method of extracting of time-invariant characteristic of planetary gearbox signals by using Time-invariant K-SVD. Yang and Chen [7] proposed a data-driven bearing fault diagnosis technique based on time-invariant K-SVD, and the method can represent impact signals with the same features at different locations by one basis function. Jiang [8] constructed label consistent K-SVD algorithm in which classification parameters and label parameters are added to the objective function, and both of them can improve the efficiency of K-SVD dictionary learning. A new type proximal decomposition algorithm which likes parallel FISTA was designed by Wang [9] for reestablishment of sparse time-frequency representation. This sparse representation is from the limited noisy observations which is based on the recently modified compressive sensing, consequently the availability of recovering buried sparse signatures is significant improved.

Through the summary of the research status in this field, it is easy to find that there are some shortcomings in the extraction of fault feature of current sparse representation algorithms. For example, these algorithms need prior knowledge of signals to construct analysis dictionaries and require a large amount of acquisition signal to compile a training dictionary. In addition, there are some problems, such as computation speed is slow because of relatively long fault signal and the sparseness of fault features is not strong when dictionaries become over redundant which includes vast amounts of redundant information. Besides, K-SVD is greatly affected by signal phase. When the signal noise is high, the signal characteristics cannot be extracted well and the reconstruction accuracy will decrease. In the aspect of reconstruction algorithm, the classical algorithm can recover the signal which has obvious fault characteristic, nevertheless, it is susceptible to noise and the accuracy of signal reconstruction with small signal-to-noise ratio is not high enough. In other words, the performance of the extraction of sparse feature by K-singular value decomposition largely is decided by the signal segment chosen in the rolling bearing fault. Generally speaking, the fault diagnosis algorithm based on sparse representation needs to be improved urgently. This paper constructed a KSVD algorithm based on sliding window operation. The SW-KSVD

The financial sponsorship from the project of National Natural Science Foundation of China (51875032, 51475098 and 61463010), Guangxi Natural Science Foundation (2016GXNSFFA380008).

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

algorithm trains the fault signal by K-SVD, and the optimal pattern of high frequency oscillation which contains bearing impact signal is selected. Then the feature of weak impact of bearing is enhanced by optimized pattern and sliding window inner product operation of signal. At last the reconstruction of selected inner product signal is finished at the peak points. Experiment proves that the algorithm has a high reduction degree of impact morphology and a stronger anti-noise ability.

II. CONSTRUCTION OF ALGORITHM

A. Sparse representation and K-SVD algorithm

The concept of sparse representation not only enables signal analyzation to not confined to the dimensions of in time-frequency domain, but also parses signal from different dimensions with different sparse dictionaries, and the feature of signal would be more prominent [10]. The concept mentioned above is of great benefit to the identification and diagnosis of faults of rotating machinery. Suppose $\boldsymbol{D} \in R^{n\times m}$ is a dictionary, each column $\boldsymbol{d_1}, \boldsymbol{d_2}, \cdots, \boldsymbol{d_m} \in R^{n\times l}$ is an atom, then signal $\boldsymbol{s} \in R^{n\times l}$ can be expressed by formula (1) as

$$\boldsymbol{s} = \boldsymbol{D\alpha} = \sum_{k=1}^{m} \boldsymbol{d_k}\alpha_k = \boldsymbol{d_1}\alpha_1 + \boldsymbol{d_2}\alpha_2 + \cdots + \boldsymbol{d_m}\alpha_m \tag{1}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_m]^T$ are the coefficient vector of the dictionary, considering sparsity requirements, the most sparse solution can be obtained by solving the problem of norm l_0 in formula (1)

$$\widehat{\boldsymbol{\alpha}} = arg\min_{\alpha} \parallel \boldsymbol{\alpha} \parallel_0 \quad s.t. \boldsymbol{s} = \boldsymbol{D\alpha} \tag{2}$$

noise remainder $\boldsymbol{r} \in R^{n\times l}$ is added to the equation and considering the noise in the real signal, equation (1) can be written as

$$\boldsymbol{s} = \boldsymbol{D\alpha} + \boldsymbol{r} \tag{3}$$

In equation (3) $\parallel \boldsymbol{r} \parallel_2 \leq \varepsilon$, ε is a smaller constant. The previous optimization problem is transformed to

$$\widehat{\boldsymbol{\alpha}} = arg\min_{\alpha} \parallel \boldsymbol{\alpha} \parallel_0 \quad s.t. \parallel \boldsymbol{s} - \boldsymbol{D\alpha} \parallel_2 \leq \varepsilon \tag{4}$$

The key point in sparse representation is to select or construct proper dictionaries. Compare with the fixed dictionary, the learning dictionary takes into consideration the actual characteristics of the signal. Therefore, the information contained in the dictionary is closer to the fault feature, and the problems are solved more pertinently. K-SVD (K means singular value decomposition) is a commonly used dictionary learning algorithm, it continuously updates dictionary atoms by the way of orthogonal matching pursuit or matching pursuit and singular value decomposition. the core concept of K-SVD is renewing the dictionary column's atom one column after another and updating the sparsity coefficients of dictionaries in the meantime. The target optimization function is:

$$arg\min_{D,X}\{\parallel \boldsymbol{Y} - \boldsymbol{DX} \parallel_F^2\} \quad s.t. \parallel \boldsymbol{x}_i \parallel_0 \leq k, i = 1,2 \cdots, N \tag{5}$$

Where $\boldsymbol{Y} \in R^{d\times N}$ is the matrix to be trained, which can be acquired by transforming the measuring signal $\boldsymbol{s} = [s_1, s_2, \cdots, s_n]^T \in R^{n\times l}$ to be analyzed into Hankel matrix through formula (6). $\boldsymbol{D} \in R^{d\times m}$ is learning dictionary and M means the number of columns in dictionary $\boldsymbol{D}$. For represent the signal properties to a greater extent and decrease the number of cycles, the first M column of $\boldsymbol{Y}$ performing singular value decomposition is used as the initial dictionary. $\boldsymbol{X} \in R^{M\times N}$ is coefficient matrix where each column should be spare enough, which means $\parallel \boldsymbol{x}_i \parallel_0 \leq k, i = 1,2 \cdots, N$.

$$\mathbf{Y} = \mathbf{hankel}(\boldsymbol{s}) = \begin{pmatrix} s_1 & \cdots & s_{n-d+1} \\ \vdots & \ddots & \vdots \\ s_d & \cdots & s_n \end{pmatrix} \tag{6}$$

K-SVD is divided into two steps: dictionary updating and sparse coding, the brief steps are shown in Table I. The first step is to get $\boldsymbol{X}$ through sparsely coding $\boldsymbol{Y}$ by tracking algorithm. The second step is to calculate the residual matrix, then, select the non-zero columns and decompose them by SVD, the first singular value and its relevant singular vectors are used to update sparse matrices and corresponding columns of dictionary [11].

TABLE I. BRIEF PROCESS OF K-SVD

Input	the column amount of the dictionary M, Hankel matrix $\boldsymbol{Y} \in R^{d\times N}$, the maximum loop time K, error threshold ε_0.
Computation	***initialize:*** take the first M columns of the singular value decomposition result of $\boldsymbol{Y}$. And consider M as the initial dictionary. ***Step 1:*** encode sparse representation. Code $\boldsymbol{Y}$ through Batch-OMP algorithm to obtain $\boldsymbol{X} = batchomp(\boldsymbol{Y}, \boldsymbol{D})$. ***Step 2:*** dictionary renewing. Renew dictionary atom and coefficient matrix one column after another. Firstly, compute error matrix $\boldsymbol{E_t} = \boldsymbol{Y} - \sum_{j\neq t}^{N} d_j x_j^T$, then find the non-zero columns to form $\boldsymbol{E_t^z}$ and perform SVD on it, i.e. $\boldsymbol{E_t^z} = \boldsymbol{U\Lambda V^T}$. Choose $\boldsymbol{U}(1,:)$ to update the first column of dictionary, and $\boldsymbol{\Lambda}(1,:) \times \boldsymbol{V}(1,:)$ to renovate the coefficient matrix's first column. ***Stopping criteria:*** if the loop time attains K or the error $\parallel \boldsymbol{Y} - \boldsymbol{DX} \parallel_2^2 \leq \varepsilon_0$
Output	the relevant coefficients matrix $\boldsymbol{X'}$ and the renovated dictionary $\boldsymbol{D'}$.

B. Extraction of bearing fault feature based on SWD-KSVD algorithm

Sliding window denoising KSVD algorithm chooses the optimal high frequency oscillation pattern including bearing impulse signal through K-SVD training of fault signal. Then, feature enhancement of bearing weak impact is carried out by inner product of sliding window of signal and optimal pattern. Finally, the signal reestablishment is achived at the chosen inner product's peak point. The specific steps are as follows:

(1) High-pass filter. The practical bearing signal usually contains a large number of low frequency interference frequencies, which has a great impact on dictionary learning. Therefore, high-pass filtering is demanded to remove. And the low frequency composition includes frequency doubling, rotational frequency and gear meshing frequency. the filtered cut-off frequency could be set up no more than 2 kHz thanks to the general rolling bearings blocks-up frequency distributes from 2 to 20 kHz [12]. Original signal was denoted as $\boldsymbol{s}$, the filtered signal was represented as $\boldsymbol{s_0}$, and the amount of the points of signal as l_s.

(2) Cut out a small section of signal containing which contains one impact at least for the dictionary training of K-SVD. The length of the Intercepted segment could be defined through the longest passing period in rolling bearing impact faults. In the bearing's rolling elements, the longest passing period was corresponding to the smallest fault feature frequency. Through the above measures, all potential defects could be sought out. In practice, in order to improve the similarity between learning dictionary and impulse signal, a segment of signal containing impact impulse is intercepted as far as possible in time domain as training samples. Furthermore, in order to avoid the change of impacts intervals which was caused by the slide of the location of impact and the undulation of rotational speed, the practical Intercepted segment is denoted as $\boldsymbol{s_p}$ with the point number l_p. And the length of segment mentioned above ought to be expanded base on the longest passing period.

(3) K-SVD with Sliding window denoising. There are two facts that the inherent frequency of rolling bearings distributes from 2 to 20 kHz and the steel structure's damping ratio is usually less than 0.2. In the light of the fact mentioned above, an impact pattern's general length can be determined. Supposing that the impact pattern is intact when the amplitude of the impact signal η is attenuated to 10^{-3}, so the impact pattern's time interval Δt_w could be defined on the basis of equation (7).

$$\eta = e^{2\pi\xi f_g \Delta t_w} \quad (7)$$

the inherent frequency f_g =2000 Hz and the damping ratio ξ was presumed as 0.1, a rather conservative and reasonable damping time interval Δt_w could be acquired which is 0.0027s. The relevant point number $lw_0 = f_s \times \Delta t_w$, in this equation the sampling frequency was represented as f_s. On the whole, the pattern gained by K-SVD embodies transition bands both in front of and behind itself, at the same time the, it should make every effort to ensure impact's integrity, therefore the practical number points which is expressed as l_w should be expanded slightly.

(4) Choose the optimal impact pattern. When K-SVD algorithm is served to dictionary learning, the signal's phase affects the validity of dictionary greatly, so $\gamma_i\ (i = 1, 2, \cdots, l)$ have different numerical value, it means that a pattern which is including the most obvious impact feature ought to be cautiously chosen. In consideration of an ideal impact pattern's mean value is zero. In addition, according to equation (8), the impact pattern's variance could be seen as the sum which is the squares of amplitude. Square can enlarge the impact component whose original amplitude is larger, while the square of the non-impact component which is close to white noise becomes smaller, hence the variance of the pattern is larger than that of other others. And the variance is close to the impact component.

$$V = \frac{\sum_{j=1}^{l_w} (\gamma(j) - \bar{m})^2}{l_w} = \frac{\sum_{j=1}^{l_w} \gamma^2 - 2\bar{m} \sum_{j=1}^{l_w} \gamma(j) + l_w \bar{m}^2}{l_w} \approx \frac{\sum_{j=1}^{l_w} \gamma^2(j)}{l_w} \quad (8)$$

where $\bar{m}$ and V denote the mean value of the pattern γ and variance respectively. The optimal pattern was chosen with the largest variance, represented as γ_0.

(5) inner product operation of sliding window. γ_0 is normalized to be $\gamma_0{}'$. In the light of equation (9), the inner product is implemented to get $l_s - l_w + 1$ inner products. And the inner product value is between $\gamma_0{}'$ and $\boldsymbol{s}_0$. Obviously, with the bigger inner product, the closer the signal of sliding window length at the beginning of the moment is to the shape of $\gamma_0{}'$.

$$p_i = \sum_{j=1}^{l_w} \left(\gamma_0{}'(j) \cdot \boldsymbol{s}_0(\boldsymbol{i} + \boldsymbol{j}) \right), \quad i = 1, 2, \cdots, l_s - l_w + 1 \quad (9)$$

(6) The signal Reestablishment of inner products at local peak. Firstly, seek the local peaks in the inner products, secondly, remove too close peaks in accordance with the shortest passing period and the impact undulation. Thus, the K_p local peak points $p_i(\mathrm{i} = 1, 2, \cdots, K_p)$ could be acquired. This operation could restrain noises actually. The signal is reestablished at these local peak points with $\gamma_0{}'$ so as to get $\boldsymbol{s}'$. fault feature frequency could be sought out by Hilbert demodulation analysis of $\boldsymbol{s}'$ more clearly.

The flow process of the whole algorithm is shown in Table II.

TABLE II. ALGORITHM FLOW OF SWD-KSVD

Input	sampling frequency f_s, filtering cut-off frequency f_c, the smallest fault feature frequency f_m, the length of $\boldsymbol{s}$ l_s, original signal $\boldsymbol{s}$, length of the sliding window l_w, the internal of sliding window patterns *q*, extension parameter $c > 1$, other parameters K-SVD needs: number of dictionary columns *M*, maximum number of cycles *K*.
Computation	$\boldsymbol{s}_0 = highpass(\boldsymbol{s})$; $l_p = length(\boldsymbol{s}_p) = c \times f_s/f_m$; for $i = 1{:}\,1{:}\,l$ $\gamma_\mathrm{i} = KSVD(\boldsymbol{s}_p(q \times (i-1) + 1{:}\, q \times (i-1) + l_w))$; end $\gamma_0 = maxvar(\gamma_i)$; $\gamma_0{}' = \mathrm{normal}(\gamma_0)$; $p_i = \sum_{j=1}^{l_w} \left(\gamma_0{}'(j) \cdot \boldsymbol{s}_0(\boldsymbol{i} + \boldsymbol{j}) \right), \quad i = 1, 2, \cdots, l_s - l_w + 1$; $(L, P) = findpeaks(p_i), \quad i = 1, 2, \cdots, l_s - l_w + 1$; $\boldsymbol{s}'(L_i + j - 1) = p_i \times \gamma_0{}'(j), \quad i = 1, 2, \cdots, K_p, \quad j = 1, 2, \cdots, l_w$; $\boldsymbol{S} = Hilbert(\boldsymbol{s}')$;
Output	the reconstructed signal $\boldsymbol{s}'$, and its demodulation spectrum $\boldsymbol{S}$.

III. VALIDATION OF EXPERIMENTAL SIGNALS

This experiment is mainly carried out on comprehensive simulation experimental platform for mechanical faults: MFS-Magnum, which is manufactured by SpectraQuest Company, USA. Its structure is shown in Figure 1. The device is driven by 1HP motor and adopts VQ data acquisition system (including computer, data acquisition instrument and acquisition card). Vibration data of faulty bearing was acquired by piezoelectric accelerometer mounted on bearing base. During the experiment, the vibration acceleration data of the outer ring fault bearing were collected. The rotational speed is 1790 rpm, the sample frequency $f_s = 12.8kHz$.

First of all, the vibration signal acquired by accelerometer is high-pass filtered with the cut-off frequency f_c=2000Hz, fault feature frequency f_r, and $10000/f_r \times 1.5 \approx 6100$ points is truncated to carry out sliding window operation. l_w is set as 400. Conduct K-SVD operation every 50 points to gain patterns and compute their variances. Figure 2 shows the relationship between variance and the starting point of the sliding window. It can be seen that seen the pattern with the starting point of 2400 points reaches maximum variance. Therefore, the pattern starting from that point is treated as the optimal one, represented as γ_0, which is shown in Figure 3. According to the algorithm, the reconstructed signal, the time-domain signal which is high-pass filtered (Figure 4) and their demodulation spectrum (Figure 5) are obtained. From the reconstructed signal in Figure 4, the interval between impacts can be clearly seen, the interval time of impact point generally matches the feature frequency, except for a few weak shocks are not reconstructed.

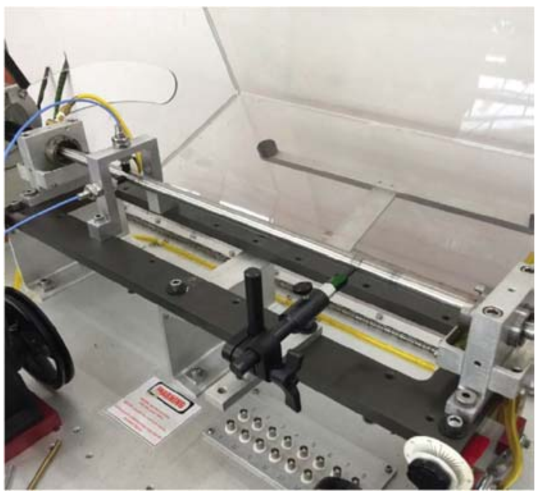

Figure 1. The test rig of rubbing experiment

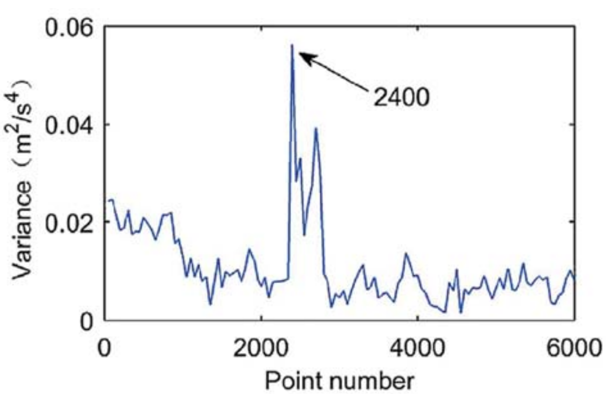

Figure 2. Sliding window patterns' variance of outer ring fault

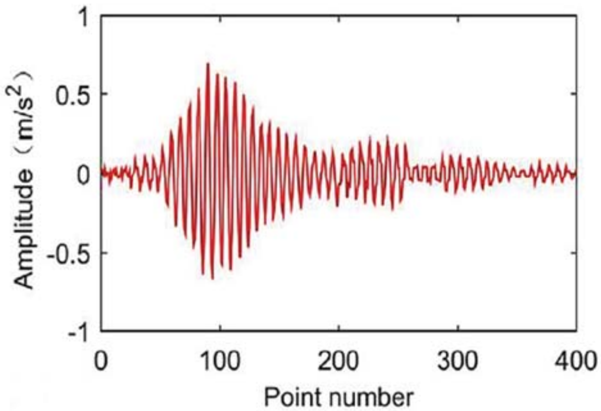

Figure 3. Optimal pattern γ_0 of outer ring fault

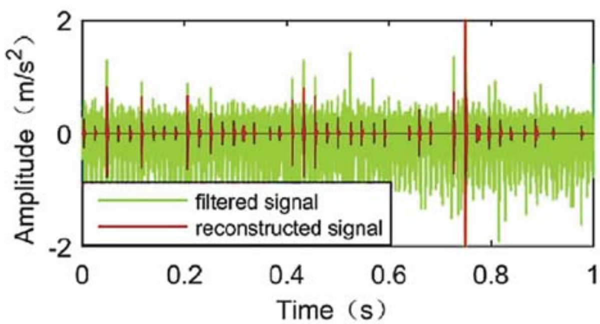

Figure 4. Time domain comparison of reconstructed signal and filtered signal of outer ring fault.

From Figure 3, it can be obviously seen the feature frequency $f_0 = 44Hz$ and its frequency doubling components of outer-loop faults. but there is no apparent modulation component of rotating frequency, which is the spectrum characteristic of outer-loop faults. Therefore, it can be determined that the outer ring of bearing has impact type fault.

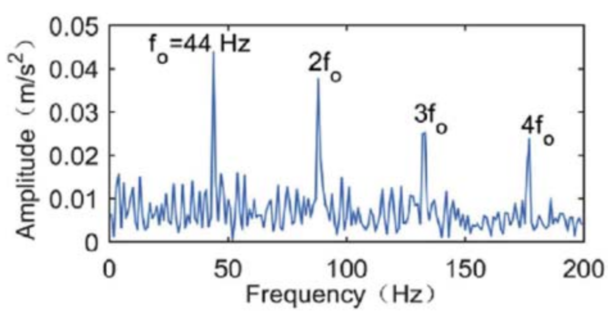

Figure 4. Reconstructed signal demodulation spectrum of experimental Signal of Bearing Outer Ring Fault

IV. CONCLUSION

This paper presents a SWD-KSVD algorithm for extracting impact fault feature of rolling bearings. By carrying out sliding window K-SVD operation and optimal selection operation on a short signal, a pattern which is including high frequency oscillation information is acquired, then, the accurate location of occurring time of impact fault is gained by sliding window inner product operation. Experiments show the effectiveness of the proposed algorithm.

ACKNOWLEDGMENT

The financial sponsorship from the project of National Natural Science Foundation of China (51875032, 51475098 and 61463010), Guangxi Natural Science Foundation (2016GXNSFFA380008).

REFERENCES

[1] L. Cui, J. Wang, and S. Lee, "Matching pursuit of an adaptive impulse dictionary for bearing fault diagnosis," *Journal of Sound and Vibration,* vol. 333, no. 10, pp. 2840-2862, 2014.

[2] S. Wang, W. Huang, and Z. Zhu, "Transient modeling and parameter identification based on wavelet and correlation filtering for rotating machine fault diagnosis," *Mechanical systems and signal processing,* vol. 25, no. 4, pp. 1299-1320, 2011.

[3] G. He, K. Ding, and H. Lin, "Fault feature extraction of rolling element bearings using sparse representation," *Journal of Sound and Vibration,* vol. 366, pp. 514-527, 2016.

[4] Y. Fa-jun, Z. Feng-xing, and Y. Bao-kang, "Initial fault feature extraction via sparse representation over learned dictionary," in *The 27th Chinese Control and Decision Conference (2015 CCDC)*, 2015, pp. 1693-1696: IEEE.

[5] G. Dong, J. Chen, and F. Zhao, "Incipient bearing fault feature extraction based on minimum entropy deconvolution and K-SVD," *J. Manuf. Sci. Eng.,* vol. 139, 2017.

[6] Z. Feng and M. Liang, "Complex signal analysis for planetary gearbox fault diagnosis via shift invariant dictionary learning," *Measurement,* vol. 90, pp. 382-395, 2016.

[7] B. Yang, R. Liu, and X. Chen, "Fault diagnosis for a wind turbine generator bearing via sparse representation and shift-invariant K-SVD," *IEEE Transactions on Industrial Informatics,* vol. 13, no. 3, pp. 1321-1331, 2017.

[8] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 35, no. 11, pp. 2651-2664, 2013.

[9] Y. Wang, J. Xiang, Q. Mo, and S. He, "Compressed sparse time¨Cfrequency feature representation via compressive sensing and its applications in fault diagnosis," *Measurement,* pp. 70-81, 2015.

[10] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation(Article)," *IEEE Transactions on Signal Processing,* no. No.11, pp. 4311-4322, 2006.

[11] D. Zhang, X. Li, J. Yang, Y. Xu, and Z. Zhang, "A Survey of Sparse Representation: Algorithms and Applications," *IEEE Access.*

[12] B. J. Dobson, "Vibration problems in engineering : (5th edition) 1990, by W. Weaver, Jr., S. P. Timoshenko and D. H. Young. Chichester: John Wiley & Sons Limited. Price £39.80; pp. 610 + xiii. ISBN 0 471 632287," *Journal of Sound & Vibration,* vol. 143, no. 3, pp. 537-538, 1990.