A new K-means singular value decomposition method based on self-adaptive matching pursuit and its application in fault diagnosis of rolling bearing weak fault

Hongchao Wang and Wenliao Du

Abstract
Sparse decomposition has excellent adaptability and high flexibility in describing arbitrary complex signals based on redundant and over-complete dictionary, thus having the advantage of being free from the limitations of traditional signal processing methods such as wavelet and fast Fourier transform being imposed by orthogonal basis. Sparse decomposition provides an effective approach for feature extraction of intricate vibration signals collected from rotating machinery. Self-learning over-complete dictionary and pre-defined over-complete dictionary are the two dictionary construction modes of sparse decomposition. Normally, the former mode owns the virtues of much more adaptive and flexible than the latter one, and several kinds of classical self-learning over-complete dictionary methods have been arising in recent years. K-means singular value decomposition is a classical self-learning over-complete dictionary method and has been used in image processing, speech processing, and vibration signal processing. However, K-means singular value decomposition has relative low reconstruction accuracy and poor stability to enhance the desired features. To overcome the above-mentioned shortcomings of K-means singular value decomposition, a new K-means singular value decomposition sparse representation method based on traditional K-means singular value decomposition method was proposed in this article, which uses the sparse adaptive matching pursuit algorithm and an iterative method based on the minimum similarity of atomic structure. The effectiveness and advantage of the proposed method were verified through simulation and experiment.

Keywords
Feature extraction, latent fault, rolling bearing, sparse representation, self-adaptive matching pursuit, improved K-means singular value decomposition

Introduction
Sparse decomposition (SD) is a relative new signal processing method which could capture the essential feature of the analyzed signal without the premise of orthogonal basis expansion, and it has been attracted more and more attention in the area of mechanical fault diagnosis. Dictionary construction is one of the key steps of SD, and there are two ways to construct...
the over-complete dictionary: pre-defined dictionary and self-learning dictionary. Normally, the pre-defined dictionary SD method has the shortcomings of needing prior knowledge of the analyzed signal and the self-learning dictionary SD method has the disadvantage of poor interference robustness. To solve the above-mentioned problems, a large amount of literatures have been arising in recent years. A hierarchical discriminative SD method was proposed in Jiao et al.,\(^1\) which could isolate the interferences effectively. In order to address the same above-mentioned problems, a parametric impulsive dictionary based on Laplace wavelets is designed for impulsive fault feature extraction of rolling bearing in Sun et al.,\(^2\) which has better feature extraction performance. The traditional orthogonal matching pursuit method was improved and an adapted dictionary-free orthogonal matching pursuit (ADOMP) method was proposed in Huang et al.,\(^3\) which has the advantages of flexibility and noise robustness, and the adapted dictionary orthogonal matching pursuit method presented potential advantages in early weak fault feature extraction of rolling bearing. An improved variational mode decomposition (VMD) method was proposed in Li et al.,\(^4\) which overcame the shortcomings of VMD in reasonable selection of algorithm parameters. Besides, the improved VMD was combined with adaptive sparse code shrinkage denoising and used in fault diagnosis of rotating machinery. To overcome the difficulty of fault identification in planet bearing due to the intricate kinematics and multiple modulation effects, a classification method based on dictionary learning was introduced in Zhao et al.,\(^5\) which has the virtues of avoiding the feature design requirement in common intelligent diagnosis method. To alleviate the drawback of pursuit efficiency of optimal atom in SD, a new SD method based on time-frequency spectrum segmentation was proposed in Yan et al.,\(^6\) and the higher decomposition efficiency and better approximation precision of the proposed method were verified by conduction of simulations and experiments. The empirical wavelet transform was improved based on the idea of SD and the sparsity-guided empirical wavelet transform method was proposed and used in fault feature extraction of rolling bearing,\(^7\) and the effectiveness of the proposed method was verified by single and multiple bearing fault signals. A new SD signal processing method based on multi-scale chirplet was proposed in fault diagnosis of gearbox vibration signals,\(^8\) and the feasibility of the proposed method was validated by both simulations and experiments. A novel intelligent fault diagnosis method of rolling bearing was proposed by combining SD with neighborhood preserving deep extreme learning machine in Li et al.\(^9\) A supervised feature extraction method named supervised regularized sparse filtering was proposed in feature extraction of rotating machinery which provided a new way for optimizing the solution of sparsity.\(^10\) Considering the sparsity of impact force, an enhanced sparse regularization method based on reweighted \(l_1\)-norm minimization is developed for reducing the peak force error and improving the identification accuracy of impact force.\(^11\) A novel gear fault diagnosis method based on the structured sparsity time-frequency analysis is proposed.\(^12\) Following the emerging sparse representation theory, a combined method of sparse reconstruction and delay-and-sum (DAS) for high-resolution Lamb wave inspection is proposed.\(^13\)

Although a large number of rolling bearing fault diagnosis methods based on SD have been arising as stated above, most of them are pre-defined dictionary SD methods which have the drawback of lack of flexibility. Besides, most of the self-learned SD methods own the disadvantage of low computational efficiency. In this article, an improved \(K\)-means singular value decomposition (KSVD) sparse representation method based on traditional KSVD method was proposed using the sparse adaptive matching pursuit algorithm and an iterative method based on the minimum similarity of atomic structure to overcome the above-mentioned problems. Through the verifications of simulation and experiment, the proposed method not only could capture the latent features of analyzed signal successfully, but also has the advantage of higher computational efficiency.

This article is organized as follows. The basic theory of traditional KSVD is presented in section “Traditional KSVD.” The theory and details of the proposed method are presented in section “Improved KSVD.” The simulation and experiment verifying the feasibility and effectiveness of the proposed method are carried out in sections “Simulation” and “Experiment,” respectively. Besides, the analyzed experimental results using the Autogram method are also presented in section “Experiment” to prove the advantage of the proposed method. Conclusions obtained from the above results are presented in section “Conclusion”.

### Traditional KSVD

KSVD could be regarded as the extension of \(K\)-means. Optimal matching pursuit and singular value decomposition (SVD) are used together in KSVD to improve the convergence rate of sparse coding and dictionary atom due to the reason that they could synchronize the updating of sparse coefficients and dictionary atom.

The purpose of KSVD algorithm is to find the dictionary \(D\) that makes the coefficient \(S\) as sparse as possible to represent the object signal sparsely

\[
Y = DS
\]
where \( Y \) is the trained signal, and \( D \) is the over-complete matrix (column number greater than row number) or complete matrix (column number equal to row number). The column vectors \( \{s_1, s_2, \ldots, s_i, \ldots, s_N\} \) in \( S \) are corresponding to \( Y \), which means that \( Y \) is composed linearly of the columns in \( D \) according to \( s_i \), and it could be transformed into mathematical model as the following equation

\[
\min_{D,X} \left\{ \|s_i\|_l_1 \right\} \quad \text{s.t.} \quad \|Y - DS\|_{l_2} \leq \epsilon \tag{2}
\]

Equation (2) could be rewritten as follows

\[
\min_{D,X} \|Y - DS\|_{l_2}^2 \quad \text{s.t.} \forall i, \|s_i\|_l_1 \leq T_0 \tag{3}
\]

Finding the global optimal dictionary \( D \) once is an NP-Hard (non-deterministic polynomial-time hardness) problem giving the trained data \( Y \), and the optimal solution is only approximated step by step. There are mainly two steps in KSVD: sparse coding and dictionary updating.

**Sparse coding**

First, set an initialized dictionary \( D \) and use it to represent the given data \( Y \) sparsely. The sparse coefficients \( S \) of the given trained data \( Y \) on dictionary \( D \) could be obtained using optimal matching pursuit. Now, \( DS \) could be regarded as the sum of the product of each column in \( D \) and the corresponding row in \( S \), which is expressed by the following equation

\[
DS = \sum_{i=1}^{K} d_i s^T_i \tag{4}
\]

where \( d_i \) represents the \( i \)th column of \( D \), and \( s^T_i \) represents the \( i \)th row of \( S \).

**Dictionary updating**

The updating of dictionary \( D \) is carried out column by column. First, suppose \( S \) and \( D \) being fixed, and the \( k \)th column \( d_k \) in \( D \) and the \( k \)th row \( s^T_k \) in \( S \) are to be updated, then there is the following equation

\[
\|Y - DS\|_{l_2} = \left\| Y - \sum_{i=1}^{K} d_i s^T_i \right\|_{l_2}^2 = \left\| Y - \sum_{i \neq k} d_i s^T_i \right\|_{l_2}^2 \tag{5}
\]

\[
- \sum_{i \neq k} d_i s^T_i \left\|_{l_2}^2 = \|E_k - d_k s^T_k\|_{l_2}^2
\]

In equation (5), it is only to adjust \( d_k \) and \( s^T_k \) to make the error between \( E_k \) and \( d_k s^T_k \) as small as possible.

The main steps of traditional KSVD are shown as follows:

1. **Initialization**
   Set the initialization matrix \( D^0 \in \mathbb{R}^N \times K \) and the initialization iteration value \( J = 1 \).

2. **Sparse coding**
   Solve \( S \) approximately using optimal matching pursuit algorithm and set the error \( \varepsilon \), which satisfies equation (2), and the sparse coefficient \( s_i \) is obtained to represent \( y_i \).

3. **Dictionary updating**
   - Define the used sample set \( w_k = \{i|1 \leq i \leq N, s^T_k(i) \neq 0\} \).
   - Calculate the residual error \( E_k \) and make it satisfy \( E_k = Y - \sum_{i \neq k} d_i s^T_i \).
   - Restrict \( E_k \) by only selecting the columns corresponding to \( w_k \) and make the non-zero positions in \( E_k \) be the same as \( s^T_k \), then the new \( E_k \) is obtained, which is marked as \( E_k^* \).
   - Decompose \( E_k^* = U \Delta V^T \) using SVD and update dictionary atom \( \delta_k = u_1 \) and \( \delta'_k = \Delta[1,1] \cdot v_1 \), in which \( \Delta \) is the SVD of \( E_k^* \), the biggest singular value of \( E_k^* \) is represented by \( \Delta[1,1] \), \( v_1 \) is the first column of the right vector singular matrix \( V \), and \( u_1 \) is the first column of the left vector singular matrix \( U \).
   - Update \( J = J + 1 \).

4. **Output**
   The learned dictionary \( D^j \) is obtained.

**Improved KSVD**

The traditional KSVD uses optimal matching pursuit to solve the sparse coefficients approximately in the stage of sparse coding, and equation (2) is used as the object function due to the reason that the sparsity of the sparse coefficients is unknown. So the accuracy of the reconstructed signal is determined by the set residual error \( \varepsilon \). Besides, the increase in the iteration number also might cause the non-convergence of \( \varepsilon \), which brings out the poor reconstruction result of the reconstructed signal.

To solve the above-mentioned problems, an improved KSVD method is proposed in this article.

In the stage of sparse coding, the self-adaptive matching pursuit (SAMP)\(^4\) is used instead of optimal matching pursuit in improved KSVD method. Because the sparsity of the sparse coefficients is not known, it could be described adaptively by changing the step length. Besides, the iteration termination conditions of SAMP are stricter than optimal matching pursuit the latter is determined by the setting iteration number and the residual error \( \varepsilon \), and the former is determined by the set iteration number and the updated residual error \( \varepsilon \) or the updated residual error \( \varepsilon \) bigger than the
residual error $\varepsilon$. So, much more accurate dictionary atom and sparse coefficients could be obtained in improved KSVD than traditional KSVD.

Furthermore, the $K$ atoms in dictionary $D'$ do not need to be updated in the improved KSVD method, and the atoms among the $K$ atoms owing the greatest impact on the reconstructed signal are only to be updated. Calculate the structural similarity\textsuperscript{15} of the residual error $\varepsilon$ corresponding to $D'$, and the structural similarity between signal $x$ and $y$ could be defined as follows

$$C_{SS}(x,y) = \frac{(2u_x u_y + c_1)(2\delta_{xy} + c_2)}{(u_x^2 + u_y^2 + c_1)(\delta_{x}^2 + \delta_{y}^2 + c_2)}$$

where $u_x$ and $u_y$ are the means of $x$ and $y$, respectively; $\delta_{xy}$ and $\delta_{x}^2$ are the variances of $x$ and $y$ respectively; $\delta_{y}$ is the covariance of $x$ and $y$; $c_1$ and $c_2$ are the constants; and $C_{SS}(x,y) \in [0, 1]$.

The $C_{SS}$ of residual error $E_k$ is the $C_{SS}$ of $Y$ and $\sum_{i \neq k} d_k s_i T$, and compare all the $C_{SS}$ corresponding to $E_1, E_2, ..., E$ and select the residual error $E$ with the smallest value of $C_{SS}$. Once the minimum residual error $E$ is determined, the corresponding dictionary atom $d_k$ and sparse coefficients $s_k$ are also updated. Then, calculate the $C_{SS}$ values of all $K$ residual errors corresponding to the updated $d_k$ and $s_k^T$ and compare their values, and the smallest one is selected. The above-stated iteration process is repeated to avoid the blindness of traditional KSVD algorithm in selecting updated atoms.

The main steps of improved KSVD are as follows:

Step 1: Initializaion

Set the initialization matrix $D^0 \in \mathbb{R}^{N \times K}$ and the initialization iteration value $J = 1$.

Step 2: Sparse coding

Solve $S$ accurately using SAMP algorithm, which satisfies equation (3), and the sparse coefficient $s_i$ is obtained to represent $y_i$.

Input: Sensing matrix $\Theta$ with a size of $M \times N$ and the $M$-dimensional observed signal $y$. Set the relevant initial values as follows: the initial residual error $r_0 = y$, supporting set $\Omega = \Phi$, step length $L = S = 1$, index set $T_{add}$, the total number of iterations $I$, and the initial iteration number $t = 1$.

The main steps of sparse coding are as follows:

1. Calculate the correlation coefficients $u = \{u_i | u_i = \| (r, \theta_j) \|, j = 1, 2, ..., N \}$, where $\theta_j$ is the column vector of $\Theta$, and $L$ maximum values in $u$ are selected and composed as set $T_{add} = \max \{ \| u \| \}$.
2. Update the supporting set $\Omega_t = \Omega_{t-1} \cup T_{add}$, $\Theta_{t+1} = \{ \theta_j | \theta_j \in \Omega_t \}$.
3. Calculate and obtain the sparse signal $s_t = (\Theta_{t}^T \Theta_{t+1})^{-1} \Theta_{t}^T y$.
4. Select the largest absolute value among $s_t$ and mark as $s_{tl}$, and mark the corresponding $\Theta_{tl}$ as $\Theta_{t+1}$, and the corresponding supporting set as $\Omega_{tl}$, that is, $F = \Omega_{tl}$.
5. Update the residual error $r_{new} = y - \Theta_{t+1} s_t$.
6. If there is $r_{new} = 0$, then stop the iteration. Otherwise, if there is $\| r_t \| \geq \| r_{t-1} \|$, update iteration step as $L = L + S$ and go back to Step 1 and continue the iteration. If the above two conditions are not satisfied, then $\Omega_t = F_t \cup r_t = r_{new}$ and the iteration number is added by 1.
7. Reconstruct the sparse signal $s$ using the non-zero terms in $\Omega_t$.
8. Output the reconstructed sparse signal $s$.

Step 3: Dictionary updating

1. Define the used sample set $w_k = \{ i | 1 \leq i \leq N, s^T_k \theta_j \neq 0 \}$.
2. Calculate the residual error $E_k$ and make it satisfy $E_k = Y - \sum_{i \neq k} d_k s_i$, and calculate the $C_{SS}$ of $E_k$ and then select the smallest $C_{SS}$.
3. Restrict $E_k$ by only selecting the columns corresponding to $w_k$ and make the non-zero positions in $E_k$ be the same as $s^T_k$, and new $E_k$ is obtained, which is marked as $E^R_k$.
4. Decompose $E^R_k = U \Delta V^T$ using SVD and update dictionary atom as $d_k = u_1$, $s_k^T = \Delta[1, 1] \cdot v_1$.
5. Update the iteration number as $J = J + 1$.

Step 4: Output

The learned dictionary $D'$ is obtained.

**Simulation**

First, the outer race fault simulation of rolling bearing is carried out. The local pitting of inner race or outer race is often manifested as the early weak fault of rolling bearing. In the section, the rolling bearing fault model whose mathematical equation can be expressed as equation (7)\textsuperscript{6, 12} is used to verify the feasibility of the proposed method. $\tau_i$ is the tiny fluctuation around mean period $T$. Set the sampling frequency as $f_s = 25, 600$ Hz. The outer race fault characteristic frequencies (FCFs) are set as $f_o = 25$ Hz. Assume the random slide between the rolling element and race is normally distributed whose standard deviation is $2\%$ of the shaft rotation ratio.
The time-domain waveform of the simulation signal of rolling bearing outer race fault is shown in Figure 1 from which the impulsive characteristic is evident, and the time interval between each shock is the reciprocal of FCF of $f_{o}$, that is, 1/25 s. In order to simulate the vibration signal of rolling bearing early weak fault, add white noise into the simulation signal, and the signal-to-noise ratio (SNR) is $-10$ dB. The noised simulation signal is shown in Figure 2 from which the impulsive characteristic of the simulation signal is buried by the strong background noise.

Before applying the improved KSVD as introduced in section “Improved KSVD” on the noised signal, the noised signal should be segmented into several training samples to increase the training efficiency and the length of each training sample is set as 64 point empirically, and the overlap rate between each sample is 0 to avoid the reuse of training signals. The number of the learned dictionary atoms is set as 256 points (256 points are enough to reflect a complete impact characteristic when rolling bearing fails) and the length of each dictionary atom is also set as 256 points for calculation efficiency. When the iteration number is chosen as 4, the matching speed is highest with higher accuracy through repeated verification, so the iteration number is set as 4. Apply the improved KSVD on the signal as shown in Figure 2, and there are a total of 256 dictionary atoms being learned. Here, part of the learned dictionary atoms is shown in Figure 3. Then, reconstruct the analyzed signal using the learned dictionary atoms, and the last reconstructed signal is shown in Figure 4;

\[
\begin{align*}
\chi(t) &= s(t) + n(t) = \sum A_i h(t - iT - \tau_i) + n(t) \\
A_i &= A_0 \cos(2\pi f_i t + \phi_A) + C_A \\
h(t) &= e^{-\frac{t}{\lambda}} \cos(2\pi f_o t + \phi_o)
\end{align*}
\]
it is evident that the impulsive components are extracted successfully by the proposed method. Apply the envelope demodulation spectral analysis on the reconstructed signal, and the result is shown in Figure 5 from which the outer race FCF with its harmonics is extracted successfully. Since the proposed method is mainly developed from the traditional KSVD method, the traditional KSVD and the proposed method are compared. As analyzed above, the early failure of rolling bearing is manifested as periodic impacts. Therefore, the reconstruction accuracy of the fault signal is important. It is the extraction of the occurred impacts which we should focus on. So kurtosis of the reconstructed signal (KRS) could be used as a good indicator reflecting the de-noising effectiveness between the traditional KSVD and the proposed method. Add different levels of white noise into the simulation signal as shown in Figure 1 and the SNR varies from −10 to 2 dB, and the KRS using the traditional KSVD and the proposed method is shown in Figure 6 from which the advantages of proposed method over the traditional KSVD method are evident.

Besides, the Autogram is also used as comparison, which is a new and effective rolling bearing signal processing method. Envelope analysis is one of the most advantageous methods for rolling element bearing diagnostics, but finding a suitable frequency band for demodulation has been a substantial challenge for a long time. Introduction of the spectral kurtosis (SK) and Kurtogram mostly solved this problem, but in situations where SNR is very low or in the presence of non-Gaussian noise, these methods will fail. This major drawback may noticeably decrease their effectiveness, and the goal of Autogram is to solve the above-mentioned problem. Vibration signals from rolling element bearings exhibit high levels of second-order cyclostationarity, especially in the presence of localized faults. The autocovariance function of a second-order cyclostationary signal is periodic and the proposed method, named Autogram, takes advantage of this property to enhance the conventional Kurtogram. The Autogram computes the kurtosis of the unbiased auto-correlation of the squared envelope of the demodulated signal, rather than the kurtosis of the filtered time signal. The analysis result of the simulation signal using the Autogram method is presented to verify the advantage of the proposed method. Figure 7 is the analysis result of the noised simulation signal as shown in Figure 2 using Autogram. The key parameters such as center frequency and bandwidth of band-pass filter could be obtained based on Figure 7: center frequency \( f_c = 1000 \) Hz and bandwidth \( B_w = 800 \) Hz. Construct an optimal band-pass filter using the obtained key parameters to filter the noised signal as shown in Figure 2, and then apply the envelope demodulation spectral analysis on the filtered signal; the analysis results are shown in the bottom chart of Figure 7 from which the FCF of outer race could not be identified, and it shows that the proposed method has the advantage of noise robustness compared with Autogram method.

The simulation of rolling bearing inner race fault model is same as outer race fault: Set the inner race FCF as \( f_i = 40 \) Hz, sampling frequency as \( f_s = 25,600 \) Hz, and rotating frequency as \( f_r = 10 \) Hz. When the pitting occurs in the inner raceway of rolling bearing, its absolute position changes with the rotation speed, and the modulation phenomenon will arise in the vibration signal. Ideally, the FCF inner race is not only extracted, but also the modulation characteristic could be extracted successfully. Same as the above
outer race fault simulation analysis processes, the corresponding analysis results are presented from Figures 8–13, and the last analysis result using the proposed method is shown in Figure 12 from which the FCF inner race is not only extracted, but also the modulation characteristic is extracted perfectly.

**Figure 7.** The analysis result of the signal as shown in Figure 2 using the Autogram method.

**Figure 8.** Simulation signal of rolling bearing inner race fault.

**Figure 9.** Simulation noised signal of rolling bearing inner race fault.

**Experiment**

In most of the relevant literatures relating to rolling bearing fault diagnosis, the defective bearings with typical simulated or seeded damages are used in bearing diagnosis research. These damages are mature faults.
and evident which are less capable to discover the nature degradation of rolling bearing in early stage. In order to validate our proposed method and truly reflect the real natural degradation, accelerated fatigue test of rolling bearing is carried out in this article. Refer to Wang et al.\textsuperscript{19} for the details and processes of the test. In Wang et al.,\textsuperscript{19} the data at 2297th minute are taken as the data of early weak fault stage of rolling bearing. In this article, the proposed method could identify the fault feature much earlier than the method used in Wang et al.\textsuperscript{19} to avoid the occurrence of sudden catastrophic accidents, and the data at 2290th minute are analyzed using the proposed method. Note that because the experiment is conducted in the case of accelerated bearing run-to-failure condition, even 1 min earlier will make significant difference.

The time-domain waveform of the data at 2290th minute is shown in Figure 14 from which no impulsive characteristic could be identified due to the interference of strong background noise. Apply the envelope demodulation spectral analysis on the signal as shown in Figure 14, and the corresponding result is presented in Figure 15; the spectral lines are chaotic and the inner race FCF could not be extracted. These further prove that the traditional envelope demodulation spectral analysis is not fit for early weak fault diagnosis of rolling bearing. Before applying improved KSVD on the signal as shown in Figure 14, divide the raw signal into segments with 256 points per each segment first. Then, the dictionary is initialized by choosing 256 segments randomly as columns, which is the same as the simulation case.

Parts of the learned dictionary atoms are presented in Figure 16. Same as the ideology presented in simulation, construct the signal using the learned dictionary

![Figure 10](image10.png)

**Figure 10.** Part of the learned dictionary atoms using the improved KSVD method on the signal as shown in Figure 9.

![Figure 11](image11.png)

**Figure 11.** The reconstructed inner race fault signal using the learned dictionary atoms.

![Figure 12](image12.png)

**Figure 12.** The envelope demodulation spectral analysis result of the signal as shown in Figure 11.
atoms, and the corresponding reconstructed bearing early weak fault signal is shown in Figure 17. Finally apply the envelope demodulation spectral analysis method on the signal as shown in Figure 17, and the obtained result is shown in Figure 18 from which the inner race FCF of the test bearing is extracted successfully. To further illustrate the high effectiveness of the improved KSVD in fault feature extraction of rolling bearing early weak fault, the analysis results of the original signal displayed in Figure 14 using other methods are presented.

1. First, same as simulation signal, the analysis result of signal as shown in Figure 14 using
Autogram is presented in Figure 19, from which the inner race FCF could not be identified, and the advantage of the proposed method over Autogram is further verified.

2. The second method being used as comparison is the traditional KSVD method. Sparse extraction of impulse by adaptive dictionary (SpaEIAD) being proposed by Chen et al.\textsuperscript{20} is a successful application of traditional KSVD-based denoising method, and the analysis result of signal as shown in Figure 14 using SpaEIAD is shown in Figure 20. It could be observed that although impulses are extracted and noise level is decreased, the noise is still strong by being compared with Figure 17. Envelope demodulation spectral analysis is also applied on the de-noised signal as shown in Figure 17. Envelope demodulation spectral analysis is also applied on the de-noised signal as shown in Figure 20 to validate whether the inner race FCF could be detected or not, and the result is shown in Figure 21 from which the advantage of the improved KSVD over SpaEIAD is verified.

**Conclusion**

Early weak fault diagnosis of rolling element bearing is a difficult and hot study area. Aiming at the difficult problem of extracting early weak fault features of rolling bearings and the potential of sparse representation self-learning dictionary method in early weak fault diagnosis of rolling bearings, this article proposes an improved KSVD method based on traditional KSVD method, in which the matching pursuit algorithm used in the construction of traditional KSVD sparse dictionary is changed to sparse adaptive matching pursuit algorithm.
The effectiveness and advantage of the proposed method are verified through simulation first. The improved KSVD method could extract the fault feature of rolling early weak fault successfully, and this virtue is verified by the rolling bearing simulation signals interfered by strong noise. Besides, the accuracy of signal reconstruction based on the proposed method is higher than that based on traditional KSVD method, and it is verified by comparing the kurtosis of the signals obtained by traditional KSVD and improved KSVD methods.

Subsequently, the accelerated fatigue test of rolling bearing was carried out and the early weak fault signal of the test was used to testify the effectiveness and advantage of the proposed further. Besides, the traditional KSVD and Autogram method are also applied
on the early weak fault signal of the test and used as comparison to verify the advantage of the proposed method.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was supported by the National Natural Science Foundation (Approved Grant No. U1804141) and the Key Science and Technology Research Project of the Henan Province (Approved Grant No. 192102210105).

ORCID iD
Hongchao Wang https://orcid.org/0000-0002-4061-802X

References
1. Jiao JY, Zhao M, Lin J, et al. Hierarchical discriminat-
ing sparse coding for weak fault feature extraction of rolling bearings. Reliab Eng Syst Safe 2019; 184: 41–54.
2. Sun RB, Yang ZB, Zhai Z, et al. Sparse representation based on parametric impulsive dictionary design for bearing fault diagnosis. Mech Syst Signal Pr 2019; 122: 737–753.
3. Huang WT, Sun HJ, Luo JN, et al. Periodic feature oriented adapted dictionary free OMP for rolling element bearing incipient fault diagnosis. Mech Syst Signal Pr 2019; 126: 137–160.
4. Li JM, Yao XF, Wang H, et al. Periodic impulses extraction based on improved adaptive VMD and sparse code shrinkage denoising and its application in rotating machinery fault diagnosis. Mech Syst Signal Pr 2019; 126: 586–589.
5. Zhao C, Feng ZP, Wei XK, et al. Sparse classification based on dictionary learning for planet bearing fault identification. Exp Syst Appl 2018; 15: 233–245.
6. Yan BK, Wang B, Zhou FX, et al. Sparse decomposition method based on time-frequency spectrum segmentation for fault signals in rotating machinery. ISA Trans 2018; 83: 142–153.
7. Wang D, Zhao Y, Yi C, et al. Sparsity guided empirical wavelet transform for fault diagnosis of rolling element bearings. Mech Syst Signal Pr 2018; 15: 292–308.
8. Peng FQ, Yu DJ and Luo JS. Sparse signal decomposition method based on multi-scale chirplet and its application to the fault diagnosis of gearboxes. Mech Syst Signal Pr 2011; 25: 549–557.
9. Li K, Li FC, Su L, et al. A novel fault diagnosis algorithm for rotating machinery based on a sparsity and neighborhood preserving deep extreme learning machine. Neurocomputing 2019; 350: 261–270.
10. Qian WW, Li SM, Wang JR, et al. A novel supervised sparse feature extraction method and its application on rotating machine fault diagnosis. Neurocomputing 2018; 320: 129–140.
11. Qiao BJ, Liu JJ, Liu JX, et al. An enhanced sparse regularization method for impact force identification. Mech Syst Signal Pr 2019; 126: 341–367.
12. Sun RB, Yang ZB, Chen XF, et al. Gear fault diagnosis based on the structured sparsity time-frequency analysis. Mech Syst Signal Pr 2018; 102: 346–363.
13. Xu CB, Yang ZB, Tian SH, et al. Lamb wave inspection for composite laminates using a combined method of sparse reconstruction and delay-and-sum. Compos Struct 2019; 223: 110973.
14. Bi X, Chen XD, Zhang Y, et al. Variable step size stage-wise adaptive matching pursuit algorithm for image compressed sensing. In: Proceedings of the IEEE international conference on signal processing, communication and computing, Kunming, China, 5–8 August 2013, pp.1–4. New York: IEEE.
15. Wang Z, Bovik AC, Sheikh HR, et al. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004; 13: 600–612.
16. McFadden PD and Smith JD. Model for the vibration produced by a single point defect in a rolling element bearing. J Sound Vib 1984; 96: 69–82.
17. McFadden PD and Smith JD. The vibration produced by multiple point defects in a rolling element bearing. J Sound Vib 1985; 98: 263–273.
18. Moshrefzadeh A and Fasana A. The Autogram: an effective approach for selecting the optimal demodulation band in rolling element bearings diagnostic. Mech Syst Signal Pr 2018; 105: 294–318.
19. Wang HC, Chen J and Dong GM. Feature extraction of rolling bearing’s early weak fault based on EEMD and tunable Q-factor wavelet transform. Mech Syst Signal Pr 2014; 48: 103–119.
20. Chen X, Du Z, Li J, et al. Compressed sensing based on dictionary learning for extracting impulse components. Signal Process 2014; 96: 94–109.