A rolling bearing fault diagnosis method based on Compressive sensing and Local characteristic-scale decomposition

Myong-Jin Jo[1], Su-Jong Kim[2], Tong-Chol Choe[3]
Faculty of Mechanics, Kim Il Sung University, Pyongyang, Democratic People's Republic of Korea
[1]Corresponding author
E-mail: [1]*mj.jo0719@ryongnamsan.edu.kp*, [2]*cioc12@ryongnamsan.edu.kp*, [3]*cioc15@ryongnamsan.edu.kp*

Received 29 March 2024; accepted 4 June 2024; published online 9 August 2024
DOI https://doi.org/10.21595/jmeacs.2024.24128

Abstract. Rolling bearings are an important part of the system with rotating parts. In the past, the rolling bearing fault diagnosis was based on the envelope of the bearing vibration waveform and FFT analysis to identify the fault and classify the fault type by the feature frequency. In addition, they proposed a method to perform signal processing by decomposing the envelope EMD to improve the diagnostic accuracy. However, this method is computationally intensive due to the iterative computation based on the reduced averaging method, and the convergence rate is different depending on the signal characteristics, which makes it difficult to perform real-time functions and consume a lot of memory space for data communication. In this paper, a method for diagnosing faults in rolling bearings based on compressive sensing (CS) and local characteristic-scale decomposition (LCD) is proposed and the effectiveness of bearing fault diagnosis method is verified by numerical experiments. In this paper, we propose a method to improve the diagnostic accuracy and shorten the computational time by identifying characteristic frequencies of the bearing fault from the Hilbert envelope spectrum of the components decomposed by the LCD after preprocessing and signal filtering of vibration signals based on CS.

Keywords: compressive sensing, matching pursuit, special frequency, sparsity decomposition, bearing vibration waveform.

1. Introduction

Bearings are mechanical components that are the spindle point, and once bearing failure occurs, the machine cannot operate safely and the efficiency of the machine will also be reduced. The bearing of an operating machine is usually vibratory and its vibration signal contains a lot of state information. Therefore, the measurement and analysis of vibration signals generated by bearings can be used to identify and classify the fault state.

Fault diagnosis by vibration signal analysis is generally performed in the time domain, frequency domain or time-frequency domain. Statistical parameters have been often used to detect bearing faults in the time domain [1]-[2]. This type of method is relatively simple but it cannot identify the fault types with high precision. Especially, the roller bearing vibration signal has nonlinear and nonstationary characteristics so that it is insensitivity to diagnosis the roller bearing fault [3]. Frequency domain methods describe signals more clearly and it can reflect more details of fault degradation degree. Therefore, in recent years, the time-frequency analysis methods have been widely used [4]-[5]. These methods can provide the local characteristics of signals in both time domain and frequency domain. The time-frequency decomposition methods include empirical mode decomposition (EMD), local mean decomposition (LMD), and local characteristic-scale decomposition (LCD) and so on.

The EMD method proposed by Huang et al is an adaptive decomposition method [6]. This method can effectively decompose the local feature of the signal into several components, and has the ability to process

complex nonlinear and nonstationary signal. So it is widely used in fault diagnosis [7]-[8]. The

envelope analysis method and EMD method are combined and applied to the fault diagnosis of roller bearings, and better analytic results are obtained [9]. However, owing to the EMD problems such as the over envelope, the owing envelope, mode mixing, and the end effect [10], the application effect of this method in practical engineering is affected to some degree.

The LMD method proposed by Smith [11] in 2005 is similar to the EMD method, and is also based on the inherent characteristic of vibration signal to decompose complex signals. Because of the use of the division operation to obtain the product function (PF) component, it can reduce the number of iterations and then reduce the end effect. At the same time, the effect of the LMD method overcoming the over envelope and the owing envelope is better than the EMD method [12]. This method can decompose the PF components which can be closely related to fault. In literature [13], by analyzing the spectrum of PF components, the characteristic information of fault frequency is extracted accurately. However, because the traditional LMD method has defects such as a poor ability to extract weak high frequency components, a large amount of iterative calculation, and the frequency confusion [14], it needs to be further optimized.

The LCD method is a new adaptive time-frequency analysis method based on intrinsic time-scale decomposition proposed by Yang et al. [15]. This method utilizes piecewise linear transformation to obtain the baseline signal, which can avoid the EMD problems such as over envelope and owing envelope etc. It can also reduce computation complexity of EMD and LMD. The iteration number is relatively small and the speed is fast, so it is more suitable for online real-time monitoring and processing of signals [16]. This new time-frequency analysis method has been mainly employed in vibration signal processing and fault diagnosis and has made some achievements in the past few years.

On the other hand, the complex and nonstationary vibration data with a large amount of noise make the failure detection very challenging, especially at the early stage. By means of appropriate vibration signal processing, it is feasible to detect the failure to evaluate the operational conditions qualitatively and quantitatively.

The measured vibration signals often contain noises, which may arise due to sensor imperfection, poor surrounding environment, or communication errors. Useful features information is therefore usually too weak to be distinguished. Removing such noises is of great benefit in successive applications.

There are many denoising methods include singular value decomposition [17], wavelet analysis [18], etc. For the SVD method, in order to discriminate against noise, it decomposes the column space of observation matrix into a dominant and a subordinate part, revealing which of its subspaces can be attributed to the noise-free signal and which can be attributed to the noise. It is often assumed that these two subspaces are orthogonal to each other, which implies that signal and noise are independent. For image de-noising, the basic idea is to retain the energy in the signal subspace and discard the energy in the noise subspace. Denoising by wavelet methods has received much interest [19]-[20]. With wavelet transform, one can decompose the original signal into a smooth part (subband in terms of signal processing) and a detail part. For most signals, energy mainly distributes in the smooth subband, and energy in the detail subband is clustered on a few large wavelet coefficients, corresponding to the edge structure of original signal. In contrast, noise energy spreads over both the smooth subband and the detail subband. With orthonormal wavelet, white noise is also white in the transform domain.

These methods are based on the sampling theory that the sample frequency must be twice the maximum frequency to be analysed. This basic principle underlies the majority of digital signal processing applications such as audio, video, radio receivers, radar applications, medical devices and more. The theory indicates that a large amount of data must then be collected, creating an exceptional challenge for signal acquisition, transmission and processing. Compression of large-scale monitoring data to detect fault features directly from sparse samples is one way to address the challenges.

Compressive sensing (CS) that emerged in 2005-2006 [21-23] breeds a new method to extract impulse signal from heavy noises. The core theory of the CS model is the sparse

representations/approximation. CS lies on the assumption that it is possible to reconstruct a sparse signal exactly from only a few samples, namely solving an underdetermined linear system of equations [24]. The application of CS on image de-noising [25] and speech enhancement [26]-[27] etc. had been studied since the CS theory appeared.

With the advancement of wireless sensors and communication technologies, it is now possible to monitor machines remotely. In this application, the use of compressive sensing is very useful because of its high accuracy and low computational cost due to de-noising during signal processing.

2. CS and LCD

2.1. CS

CS theory was newly presented in 2006 in the field of signal processing and has been used in various fields such as image de-noising and speech enhancement.

CS can be divided into two steps [11]. First, combining the sampling with compressing, x original signal with the compressed signal by using an appropriate recovery algorithm.

The original signal $x \in R^N$ can be sampled as a compressed signal with reduced order $y \in R^M$ by following:

$$y = \Phi x, \tag{1}$$

where $\Phi \in \mathrm{R}^{M \times N}$ $(M < N)$ is a compressive matrix.

According to linear algebra theory, Eq. (1) has infinite number of solutions and when $M < N$, we cannot recover the original signal x uniquely from the compressed signal y. However, if the signal x is sparse, i.e., there only has a few nonzero coefficients in x, it is possible to recover x from y. Generally, the vibration signals should not be sparse, while they can be represented sparsely by some appropriate ways (e.g., orthogonal transformation).

Using some orthogonal basis $\{\psi_i\}$ $(i = 1, N)$, we can expand $x \in R^N$ as:

$$x = \sum_{i=1}^{N} \theta_i \psi_i, \tag{2}$$

where the coefficient $\theta_i = \langle x, \psi_i \rangle = \psi_i^T x$ and $\{\psi_i\}$ is an N-dimensional column vector. Expressing Eq. (2) in matrix form, we can get:

$$x = \Psi \Theta, \tag{3}$$

where $\Psi = [\psi_1, \psi_2, \cdots, \psi_N] \in R^{M \times N}$ is called a dictionary matrix and $\Theta = \{\theta_1, \theta_2, \cdots, \theta_N\}^T$ is a coefficient vector. Substituting Eq. (3) into Eq. (1), and denoting $A = \Phi\Psi$:

$$y = \Phi\Psi\Theta = A\Theta, \tag{4}$$

where A is called the measurement matrix.

Suppose that the coefficient vector Θ is k-sparse on the basis Ψ, i.e. there are k nonzero elements in Θ. That is, unknowns in the above equation will be greatly reduced so that it is possible to recover Θ from y.

We will choose the Orthogonal Matching Pursuit-OMP algorithm for signal recovery. Since Θ has k nonzero components, the vector $y = A\Theta$ is a linear combination of k columns from A. In the language of sparse approximation, we say that y has an k-term reperesnetation over the dictionary A.

Therefore, sparse approximation algorithm can be used for recovering sparse signals. To

identify the ideal signal Θ, we need to determine which columns of A participate in the vector y. At each iteration, we choose the column of A that is most strongly correlated with the remaining part of y. Then we subtract off its contribution to y and iterate on the residual. After k iterations, the algorithm will have identified the correct set of columns. The steps of OMP algorithm are as follows:

Input: measurement matrix $A \in R^{M\times N}$, compressed signal $y \in R^M$ (be obtained by Eq. (1)), the sparsity level of the signal k.

Initialization: the residual $r_0 = y$, the index set $\Lambda_0 = \emptyset$, the matrix $B_0 = \emptyset$.

FOR $t = 1; t = t + 1; t \le k$

Step 1: Find the index that solves the easy optimization problem $\lambda_t = \max_j |\langle r_{t-1}, A_j\rangle|$, where A_j is j-th column of A.

If the maximum occurs for multiple indices, break the tie deterministically.

Step 2: Update the index set and matrix: $\Lambda_t = \Lambda_{t-1} \cup \{\lambda_t\}$, $B_t = B_{t-1} \cup \{A_{\lambda_t}\}$

Step 3: Solve a least squares problem to obtain a new signal estimate: $\hat{x}_t = \min_x \|y - B_t x\|_2$.

Step 4: Calculate the residual: $r_t = y - B_t\hat{x}_t$.

END FOR

Output: An estimate $\hat{x}_k$.

The flowchart of OMP algorithm for signal recovery is shown in Fig. 1.

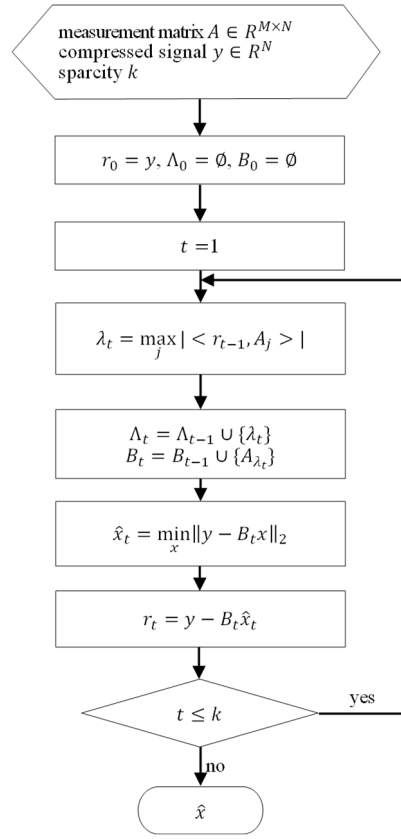


Fig. 1. The flowchart of OMP algorithm

2.2. LCD

Local Characteristic-Scale Decomposition (LCD) is a new adaptive time-frequency analysis

method based on time-scale decomposition.

Theoretically, this method is of great interest in the field of vibration signal processing recently because it decomposes the signal into a space of orthogonal frequency basis, has the unique character of signal decomposition, has the excellent localization properties in both time and frequency domains simultaneously, and has a much superior signal decomposition rate and self-signal adaptation characteristics compared with existing signal decomposition methods, including empirical mode decomposition and local mean decomposition.

Once the signal decomposition is complete, the signal reconstruction can be done according to the requirements of the detailed problem. We give an iterative formula for the ignition of the intrinsic scale component (ISC) based on the principle of the LCD method. By decomposing any signal into a signal of different frequency, it is possible to extract information from the signal of sine or cosine exponential decay.

First of all, we define the ISC and consider the factors that affect the procedure and accuracy of the LCD method to obtain it.

2.2.1. ISC

To use the LCD method, we define the intrinsic scale component (ISC). The intrinsic scale component is a single-component signal expressed as instantaneous frequency that is physically meaningful based on local characteristic parameters of the extrema. The ISC must satisfy the following two conditions to have physical meaning.

1. For the data set, the marks of any two adjacent extrema are mutually different.

2. For the data set, the extrema are L_k, $k = 1, 2, \cdots, M$ with their respective corresponding moment τ_k, $k = 1, 2, \cdots, M$. The section of line formed by connected any two maximum (minimum) extrema (τ_k, X_k), (τ_{k+2}, X_{k+2}) is between them. Therefore, we can obtain the corresponding moment τ_{k+1} of the maximum (minimum) extrema, and the function value is set to:

$$H_{k+1} = L_k + \frac{\tau_{k+1} - \tau_k}{\tau_{k+2} - \tau_k}(L_{k+2} - L_k), \tag{5}$$

of which the ratio to the relative maximum(minimum) extrema L_{k+1} remains unchanged.

These two conditions ensure that there is a single mode between any two adjacent extrema of ISC, and the local curve (curve between extrema and adjacent zero point) approximate to the standard sine curve, so the instantaneous frequency has physical significance.

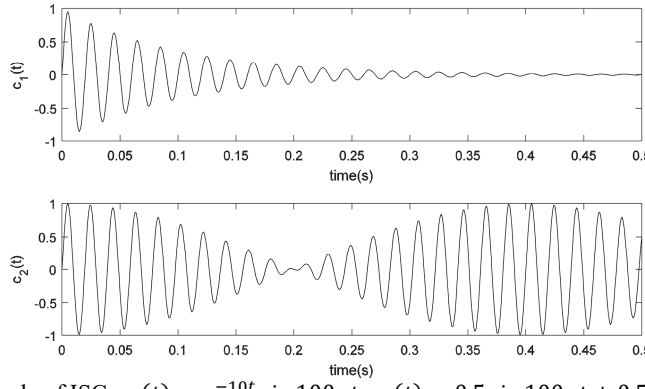


Fig. 2. An example of ISC: $c_1(t) = e^{-10t}\sin 100\pi t$, $c_2(t) = 0.5\sin 100\pi t + 0.5\sin 105\pi t$

The two end points of the envelope (starting and ending points) are obtained by spline interpolation.

Note that the ISC of the difference is the one that should be passed over, but the modal analysis theory only includes a modal signal of some order (not necessarily difference).

The latter include vibration mode amplitude, natural frequency and damping of some order.

The angular mode order signals are independent of each other and decompose the signal into frequency orthogonal basis space. That is, the decomposition of the signal has uniqueness.

Fig. 2 shows an example of two functions that fit the conditions of the ISC. As you can see, the ISCs are functions of local symmetry up and down with respect to the zero axis.

2.2.2. Ignition-iteration formula for ISC calculations

The algorithm for decomposing the signal $x(t)$ using the LCD method is as follows.

Step 1: Calculate extreme points L_k $(k = 1,2,\cdots,j)$ of the signal $x(t)$, and then calculate the fitting points:

$$g_k = \frac{L_k + H_k}{2}. \tag{6}$$

Step 2. On the basis of extending the endpoints in time series, use cubic spline function for fitting all g_k, and at last get the mean curve $B_{g_1}(t)$.

Step 3: Eliminate the mean curve $B_{g_1}(t)$ from the original signal:

$$P_{11}(t) = x(t) - B_{g_1}(t). \tag{7}$$

Eq. (7) is called the "decreasing average" process.

Step 4: Judge whether $P_{11}(t)$ meets the ISC conditions or not. If it is satisfying, take $P_{11}(t)$ as the first ISC, signed as $\mathrm{ISC}_1(t)$. If not, take $P_{11}(t)$ as the original signal and repeat steps (1)-(3):

$$P_{12}(t) = P_{11}(t) - B_{g_2}(t), \tag{8}$$
$$P_{13}(t) = P_{12}(t) - B_{g_3}(t), \tag{9}$$
$$\vdots$$
$$P_{1k_1}(t) = P_{1k_1-1}(t) - B_{g_{k_1}}(t). \tag{10}$$

And then cycle k_1 times until P_{1k_1} meets the ISC contitions, thus P_{1k_1} is the first component $\mathrm{ISC}_1(t)$ of signal $x(t)$:

$$\mathrm{ISC}_1(t) = x(t) - \sum_{i=1}^{k_1} P_{1i}(t), \tag{11}$$

where k_1 is the number of "decreasing average" processes of the signal $P_{11}(t)$.

Step 5: Eliminate $\mathrm{ISC}_1(t)$ from signal $x(t)$, we can get a new signal x_2. Take x_2 as the original signal and repeat steps 1~4, thus we can get the second component $\mathrm{ISC}_2(t)$:

$$\mathrm{ISC}_2(t) = x_2(t) - \sum_{i=1}^{k_2} B_{gi}(t), \tag{12}$$
$$P_{2k_2}(t) = P_{2k_2-1}(t) - B_{g_{k_2}}(t), \tag{13}$$
$$P_{20}(t) = x(t) - \mathrm{ISC}_1(t), \tag{14}$$

Repeat the cycle *n* times until r_n is a monotonic function, then we can get all ISCs of the signal $x(t)$:

$$\mathrm{ISC}_j(t) = x_j(t) - \sum_{i=1}^{k_j} P_{ji}(t), \tag{15}$$

$$P_{2k_j}(t) = P_{2k_j-1}(t) - B_{g_{k_j}}(t), \tag{16}$$

where:

$$x_j(t) = \begin{cases} x(t), & j = 1, \\ x(t) - \sum_{i=1}^{j-1} \mathrm{ISC}_i(t), & j \neq 1. \end{cases} \tag{17}$$

From the above ignition iteration algorithm, any signal $x(t)$ is decomposed into n ISCs and one residual term.

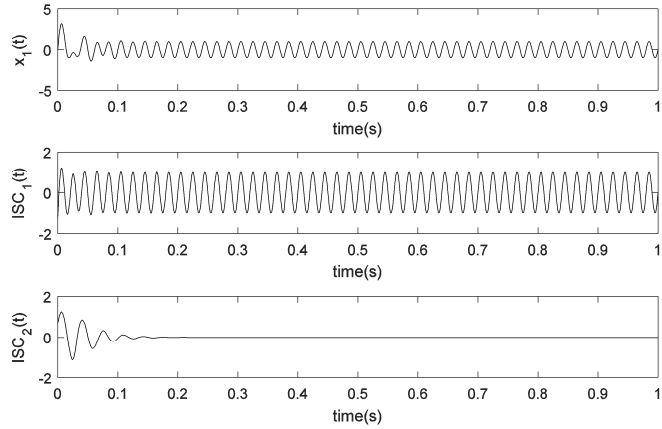


Fig. 3. LCD result of $x_1(t) = \sin(100\pi t) + 3\exp(-10\pi t)\sin(60\pi t)$

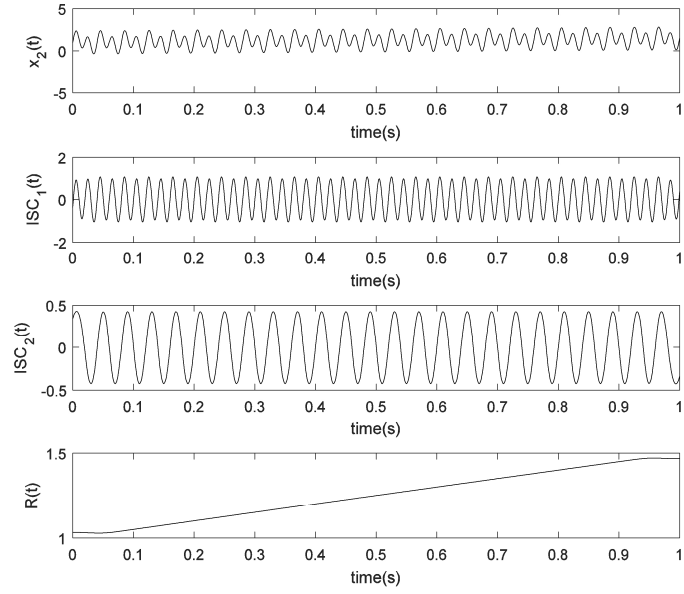


Fig. 4. LCD result of $x_2(t) = \sin(100\pi t) + 0.5\sin(50\pi t) + 0.5t + 1$

Figs. 3-4 show the LCD results for different signals. Fig. 3 is a graph of $x_1(t) = \sin(100\pi t) +$

$3\exp(-10\pi t)\sin(60\pi t)$ and the ISCs decomposed by using LCD. Fig. 4 is a graph of $x_2(t) = \sin(100\pi t) + 0.5\sin(50\pi t) + 0.5t + 1$, the ISCs and residual decomposed by using LCD.

2.2.3. Comparisons between EMD and LCD

We consider the signal shown in Eq. (18):

$$\begin{aligned} x(t) &= x_1(t) + x_2(t), \\ x_1(t) &= [1 + 0.5\sin(10\pi t)]\cos(100\pi t + 30\pi t^2), \\ x_2(t) &= 4\sin(40\pi t). \end{aligned} \tag{18}$$

In order to compare LCD with EMD, we now decompose the signal $x(t)$ using EMD and LCD, respectively.

The waveforms of the signals are shown in Fig. 5. The sampling frequency is 1000 Hz.

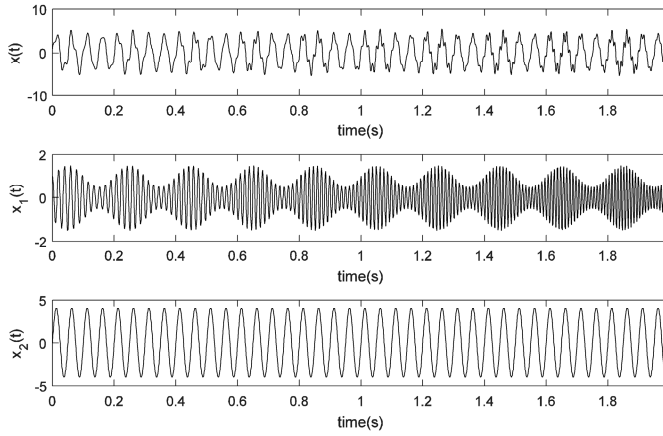


Fig. 5. Time waveform of the simulated signal represented by Eq. (18)

The decomposition results of the simulated vibration signals obtained by EMD and LCD are shown in Figs. 6-7.

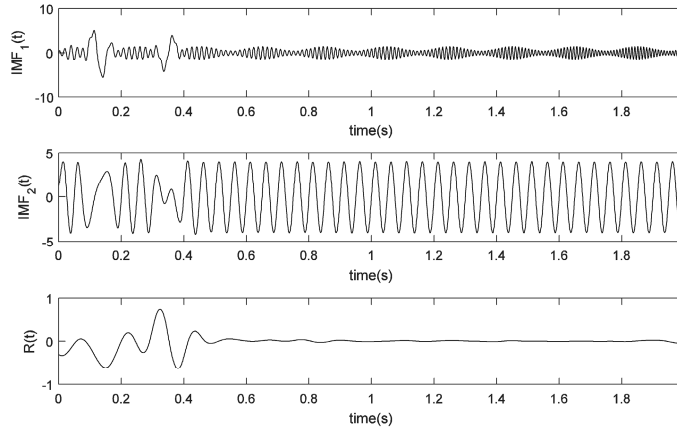


Fig. 6. EMD results of simulated signals

From Figs. 6 and 7 we could find that the two IMFs derived from EMD have the phenomenon

of mode mixing, while the LCD method derives a well result matching the true components extremely.

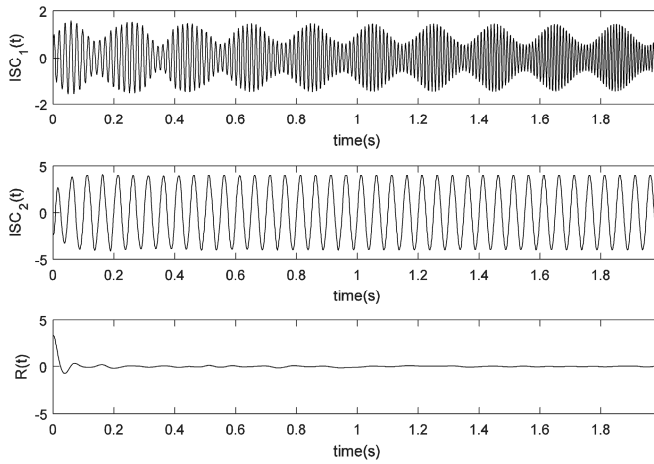


Fig. 7. LCD results of simulated signals

In order to compare them detailed, consuming time and iteration times for getting each IMF and each ISC are considered as evaluation parameters. Besides, since the first two components obtained by EMD and LCD are in accordance with $x_1(t)$ and $x_2(t)$, the correlations of the first IMF and ISC with $x_1(t)$ and the second IMF and ISC with $x_2(t)$ are considered, respectively. The three evaluation parameters are given in Table 1.

Table 1. The comparison parameters of IMF and ISC components

	Time (ms)	Iteration times	Correlation
IMF_1	26	8	0.674 5
IMF_2	3.6	2	0.958 1
ISC_1	1.5	14	0.991 6
ISC_2	2	1	0.999 1

From Table 1, it can be found that the consuming time of the two components obtained by EMD are both more than that obtained by LCD. Besides, the correlations of the two ISC components with the true components are higher than that of IMF components', which means the LCD decomposition results are more close to the true values. Therefore, through this simulation signal, LCD is superior to EMD in certain aspects at least for certain signals.

In EMD, the baseline is defined based on the connection of the whole maxima and minima with cubic spline. The fitting errors generate easily in the boundary for the cubic spline fitting in EMD. In LCD, however, firstly the adjacent maxima (or minima) are connected by straight lines and then the two values at every extrema are averaged respectively. And the baseline is constructed by connecting the obtained values with cubic spline. Hence, in theory LCD has a smaller fitting error and a faster running-time than EMD.

From Eqs. (5)-(6), g_k can be rewritten as:

$$g_k = a_k L_k + a_{k+1} L_{k+1} - a_{k+2} L_{k+2}, \tag{19}$$

where:

$$a_k = \frac{1}{2}\left[1 - \frac{t_{k+1} - t_k}{t_{k+2} - t_k}\right], \quad a_{k+1} = \frac{1}{2}, \quad a_{k+2} = \frac{t_{k+1} - t_k}{2(t_{k+2} - t_k)}.$$

In fact, L_k is an ensemble average of three adjacent extrema, meanwhile, the base line is connecting all the L_k with the cubic spline. Hence, LCD has a well local characteristics and a simple geometrical definition, compared with EMD. Therefore, compared with EMD, LCD at least has certain advantages in running time, reducing fitting errors and restraining end effect and mode mixing.

As can be seen from the flowchart, frequency analysis methods based on sparse decomposition, LCD, Hilbert transform and FFT transform are used for the diagnosis of bearing and gear states of the gearbox. We consider the detailed procedure of bearing fault diagnosis.

3. Bearing fault diagnosis algorithm with CS and LCD

Fig. 8 shows the flowchart of the bearing fault diagnosis algorithm based on CS and LCD.

The calculation and processing procedures of this algorithm are as follows.

Step 1: Measure the vibration from the equipment with the bearing to be tested using a vibration measuring device.

Step 2: Decompose the measured vibration signal into CS and reconstruct it to perform the denoising process.

Step 3: Calculate the RMS value of the denoised vibration signal and calculate the risk determined by the ratio of this value to the allowable vibration RMS value. If the risk is less than the threshold, the bearing is considered normal and finished. If not normal, go to step 4.

Step 4: Decompose the denoised vibration signal obtained in step 2 into LCD to obtain the ISCs.

Step 5: Calculate the instantaneous amplitude by performing a Hilbert transform for the first ISC with lots of fault information.

Step 6: Perform FFT for instantaneous amplitude data.

Step 7: Check whether there is a peak value corresponding to the fault characteristic frequency of the bearing in the power spectrum. If not, it shall be considered normal and finished. If there is a peak value, go to step 8.

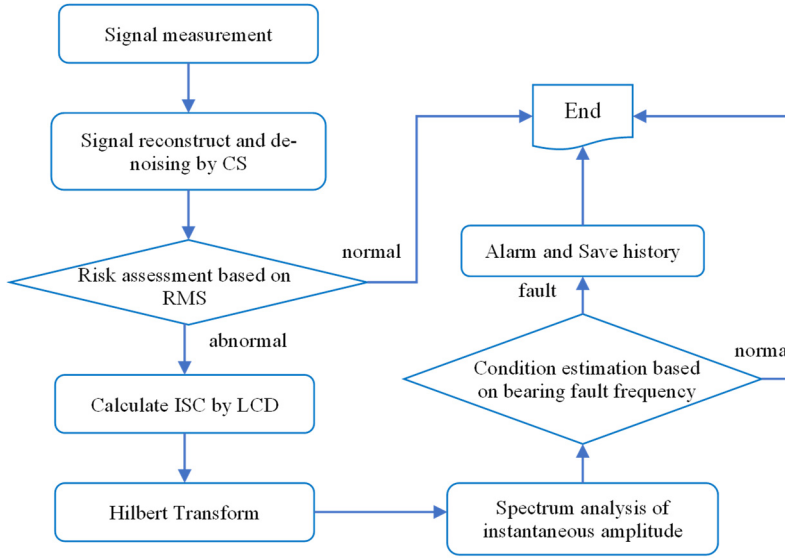


Fig. 8. The flowchart of the bearing fault diagnosis algorithm

Step 8: Ring the alarm, save the fault history, and finished.

As you can see from the flowchart, we use CS, LCD, Hilbert transform and frequency analysis methods based on FFT transform for the diagnosis of bearing fault. In this section, we consider

ISSN PRINT 2669-2600, ISSN ONLINE 2669-1361

the detailed procedure of bearing fault diagnosis.

3.1. Reconstruction and de-noising of vibration signals using CS theory

In general, the capacity of the vibration data obtained for a certain period of time is very large. Therefore, data compression methods that can reduce the dimensionality of the collected vibration data while preserving the most useful information are necessary. Sparsity decomposition theory is a new kind of data sampling technique, and using this theory, it is possible to reconstruct the original signal with only a few samples less than the Nyquist sampling rate.

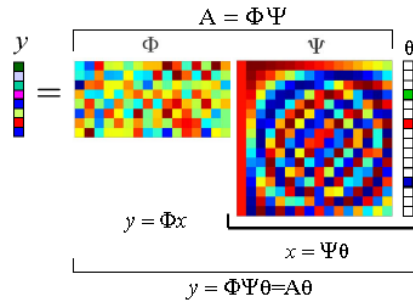


Fig. 9. Compression process of vibration signal

Let us show through the simulation signal that the vibration signal is sparsely represented by the fast Fourier transform matrix. In Fig. 10, the simulated vibration signal, represented by Eq. (20), is shown by a waveform sampled at 1024 Hz:

$$x(t) = \cos(2\pi \times 80t) + 0.05t^2 \cos(2\pi \times 30t + 2). \tag{20}$$

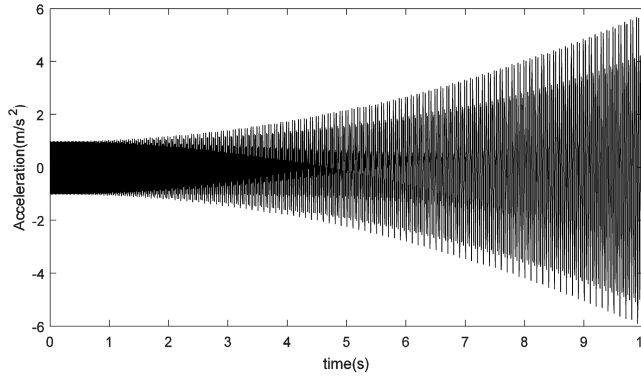
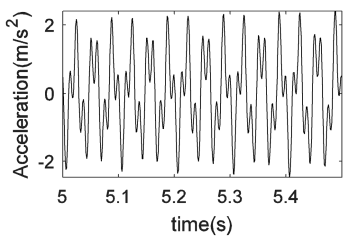


Fig. 10. Simulated vibration signal

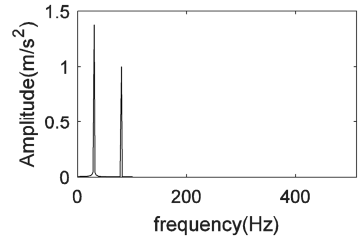
When performing FFT with 5-5.5 s of data, it can be seen that only a few data points are nonzero, and most of the transform coefficients are zero, as shown in Fig. 11(b).

However, FFT of the noise does not result in zero coefficients at almost all data points. That is, the noise is not sparsely represented by the FFT (Fig. 11(d)).

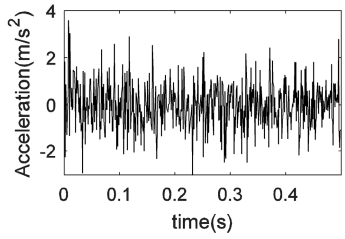
To validate the de-noising of the vibration signal using CS, the simulated vibration signal, denoted by Eq. (20), was used Fig. 12(a) shows the data of 0-0.1 s in the simulated signal represented by Eq. (20), and the simulated signal added 1dB noise (Fig. 12(b)) to the data is shown in Fig. 12(c). The results of de-noising using compressive sensing are shown in Fig. 13. Fig. 13(a) shows uncontaminated signal and Fig. 13(b) shows contaminated signal (See the Fig. 12(a) and Fig. 12(c)). Fig. 13(c) shows the result of de-noising using CS.



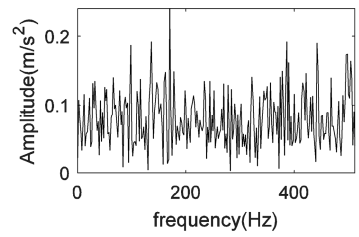
a) The waveform of the vibration signal



b) FFT coefficients of the vibration signal

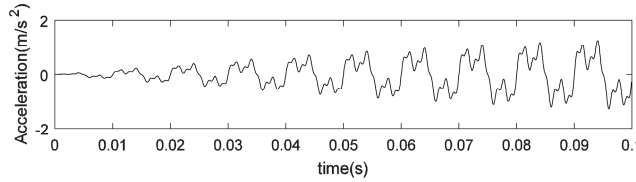


c) The waveform of the noise

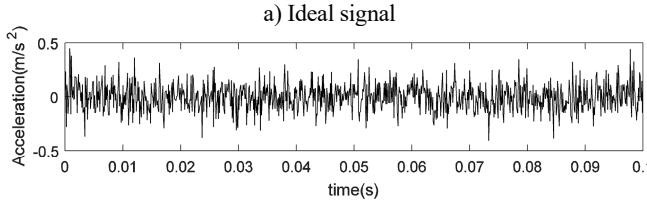


d) FFT coefficients of the noise

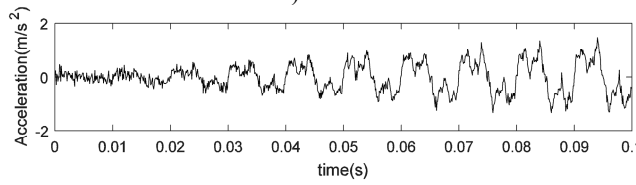
Fig. 11. Waveforms of vibration signal and noise and its FFT coefficients



a) Ideal signal



b) Noise



c) Noise-contaminated signal

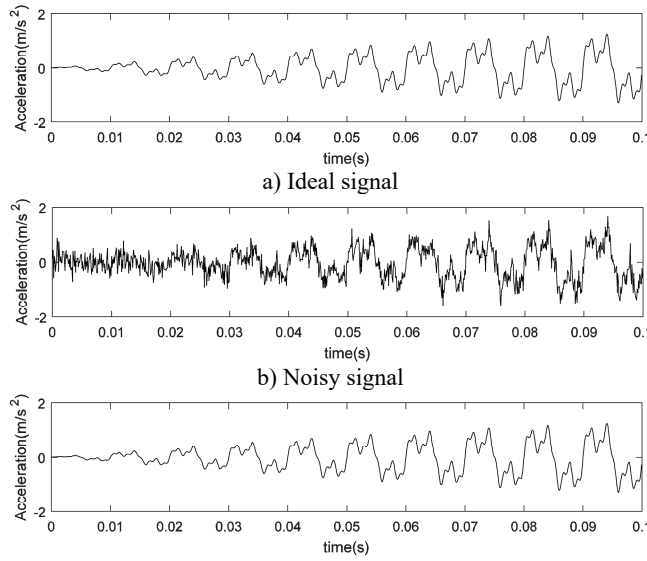
Fig. 12. Simulated signal

To quantitatively evaluate the de-noising result, let us calculate the signal-to-noise ratio (SNR) of the de-noising signal. The SN ratio of the de-noising signal can be calculated as follows:

$$\mathrm{SNR} = 20\log\left(\frac{\|x\|_2}{\|\hat{x} - x\|_2}\right), \tag{21}$$

where $\hat{x}$ is the de-noising signal.

Calculated using Eq. (21), the SN ratio of the de-noising signal was 5.96 dB. It can be seen that the CS theory can reduce the traffic and memory capacity by compressing the signal, and, in addition, the reconstruction process can be further de-noising to extract the signal features more accurately.

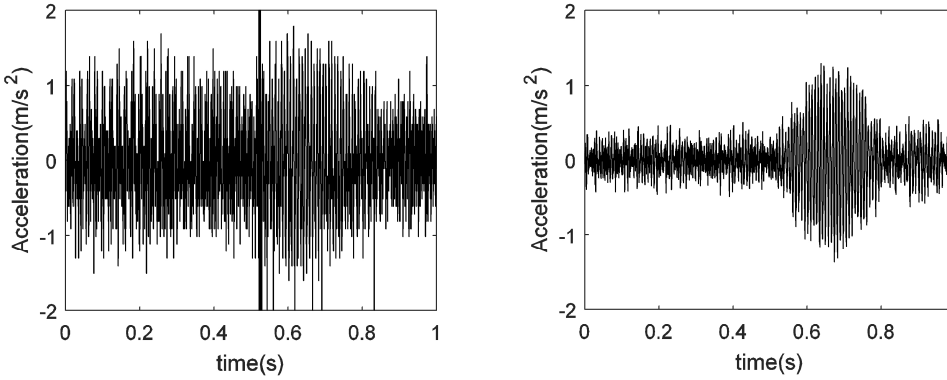


a) Ideal signal

b) Noisy signal

c) De-noising by CS

Fig. 13. Original signal and recovered signal



a) The original signal

b) The signal recovered by CS

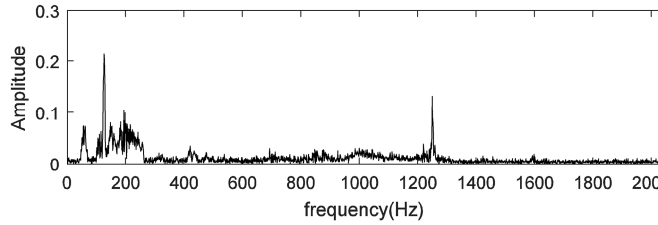
Fig. 14. Bearing vibration signal

Let us verify the effectiveness of de-noising by compressive sensing theory with field measurement signals. Fig. 14(a) shows the time waveform of bearing vibration signal of the raw material reduction gearbox of the Cement Factory. The de-noising of the signal using the CS theory is shown in Fig. 14(b).

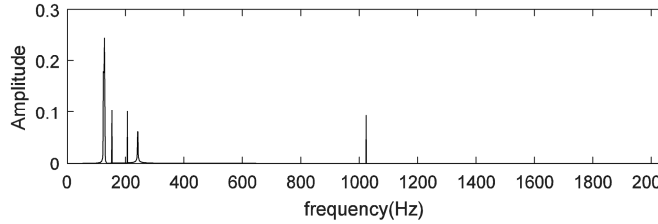
To validate de-noising results, the FFT is shown in Fig. 15. Fig. 15(a) and Fig. 15(b) are FFT of the signal shown in Fig. 14(a) and Fig. 14(b), respectively. As you can see, the CS gives good results for the in field measurement signals.

3.2. Instantaneous amplitude and instantaneous frequency analysis based on LCD and Hilbert transform

When a fault occurs in a rotating machine element such as a gear or rolling bearing, the corresponding vibration signals exhibit amplitude modulation and frequency modulation. To extract the fault feature of the modulated signal, demodulation analysis is widely used, among which the Hilbert transform is one of the most popular demodulation methods because it can effectively extract the envelope of the faulty vibration signal.



a) Fourier transform result of the original signal



b) Fourier transform result of the signal recovered by CS

Fig. 15. Fourier transform result

According to the Hilbert transform and analytic signal theory, if for a signal, has relatively low frequency, and has a relatively high frequency, and uncoupled spectrum, then the analytic signal can be expressed in terms of instantaneous amplitude and instantaneous phase as follows:

$$F(t) = f(t) + \mathrm{j}\bar{f}(t) = A(t)e^{i\varphi(t)}. \tag{22}$$

Here, and is the Hilbert transform of:

$$\bar{f}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{f(\tau)}{t-\tau}d\tau. \tag{23}$$

The instantaneous frequency is given by the derivative of the time of the instantaneous phase, that is:

$$f(t) = \frac{1}{2\pi}\frac{d\varphi(t)}{dt}. \tag{24}$$

The condition for and is a kind of narrowband signal condition that the signal must satisfy, and, therefore, according to the analytic signal theory, the envelope and phase of the narrowband signal are the absolute value and phase of the analytic signal, i.e., the instantaneous amplitude and the instantaneous phase.

Generally, the rotary machine fault vibration signal is a multi-component amplitude modulation-frequency modulation signal, which does not satisfy the above conditions.

To solve this problem, a LCD is used to decompose the multi-component signal into single-component ISCs and the fault feature is extracted by applying the Hilbert envelope spectrum to the first few ISCs containing the major fault component.

3.3. Bearing fault diagnosis algorithm

In this paper, using CS theory, LCD and Hilbert transform for fault diagnosis of super-large moderator for mill, the following vibration diagnosis algorithm is proposed.

First, the measured signal is compressed using CS theory. The compressed data is essential for data transmission and data storage, and reconstructing the compressed data results in a de-noising signal.

Then, the channel-by-channel oscillation value is used to estimate the risk and, if not normal, the signal features are extracted using the LCD and the Hilbert transform.

Then, the bearing failure components are compared with the values of the component.

When a local fault occurs in a rolling bearing, the vibration signal contains a periodic impulse component that is referred to as the fault feature frequency. This impulse component, although present periodically, also has the characteristics of an unsteady signal. Compared with the energy of vibration caused by other factors generated during the operation of a rotating machine, e.g. by gear tooth engagement, the energy of vibration caused by bearing defects is small, and it is generally difficult to distinguish early when the fault is not severe. Therefore, in many cases, the fault spectrum components are clearly separated using various signal decomposition methods and envelope processing to estimate bearing failure states.

Local defects that exists in the parts of the bearing causes a periodically generated impact force. The frequencies generated from these periodic pulses can be calculated theoretically. According to the bearing fault type, the feature frequencies are as follows:

$$f_o = \frac{1}{2} z f_r \left(1 - \frac{d}{E}\cos\alpha\right), \tag{25}$$

$$f_i = \frac{1}{2} z f_r \left(1 + \frac{d}{E}\cos\alpha\right), \tag{26}$$

$$f_b = \frac{E z f_r}{2d}\left(1 - \left(\frac{d}{E}\cos\alpha\right)^2\right), \tag{27}$$

$$f_m = \frac{f_r}{2}\left(1 - \frac{d}{E}\cos\alpha\right), \tag{28}$$

where f_r is the rotational frequency of the shaft, α is the contact angle, E is the pitch circle diameter, d is the ball diameter, z is the number of bearing balls, and f_o, f_i, f_b, f_m are the frequency of outer race fault, inner race fault and rolling element fault, respectively.

4. Validation

Bearing vibration data are obtained from Case Western Reserve University [28]. As shown in Fig. 16, the bearing experimental apparatus consists of a 2 hp three-phase motor (left), a torque transducer(middle), a dynamometer (right) and control electronics (not shown). Drive-end bearings of the motor used in the experiments are deep groove ball bearings of type 6205-2RS JEM SKF.

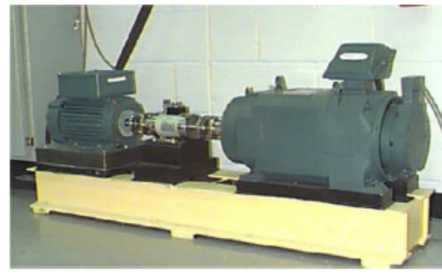


Fig. 16. Experimental setup

To validate the capability of the proposed method, experimental data analysis on rolling bearing with different categories and severities was conducted. The experimental data are bearing fault data which were derived from the Electrical Engineering Laboratory of Case Western Reserve University (CWRU) [25]. The bearing type is 6205-2RS JEM SKF deep groove ball bearings.

The vibration signals of bearing were collected under four conditions including the normal condition, inner race fault condition (IRF), the outer race fault condition (ORF) (at the 6 o'clock position) and the ball fault condition (BF). The fault bearings are using the electro-discharge machining with fault diameters of 0.1778 mm, 0.3556 mm, 0.5334 mm, and 0.7112 mm to simulate different fault severities. The associated rotating speed of driving motor is set to be 1730, 1750, 1772 and 1797 rpm, respectively. The sampling rate of the data acquisition device is set to 12,000 Hz and the sampling time is 1s. In this work, two different bearing datasets are provided. The details about these two bearing datasets are shown in Table2.

Table 2. The bearing datasets which are used in this work

Fault type	Severity (mm)	Rotation speed (rpm)	Characteristic frequency (multiples of shaft speeds)
Outer race fault	0.533 4	1 797	3.585
Inner race fault	0.533 4	1 797	5.415

The time domain waveforms of vibration signals under outer race fault are shown in Fig. 17, respectively. In order to account for field conditions, the Gaussian random noise was added.

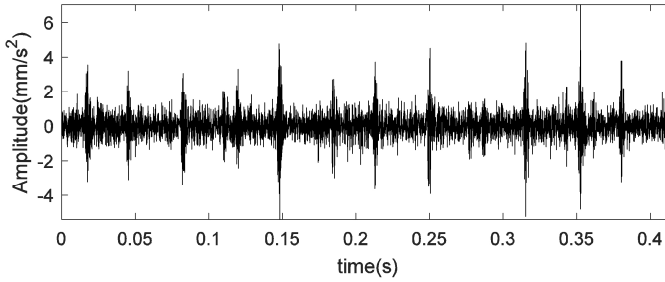


Fig. 17. The time domain waveforms of vibration signals under outer race fault

When a fault occurs in a rolling bearing, the vibration signal is usually a multi-component signal containing noise. Thus, first of all, the CS theory is used to remove noise and compress the data, and the LCD is used to decompose the signal into ISCs. The first three ISCs are shown in Fig. 18.

The corresponding Hilbert envelope spectrum obtained by Hilbert transform of the first ISC is shown in Fig. 19. In Fig. 19, we can see spectrum lines at the characteristic frequency of the outer race fault (107.31 Hz) and its harmonic frequency.

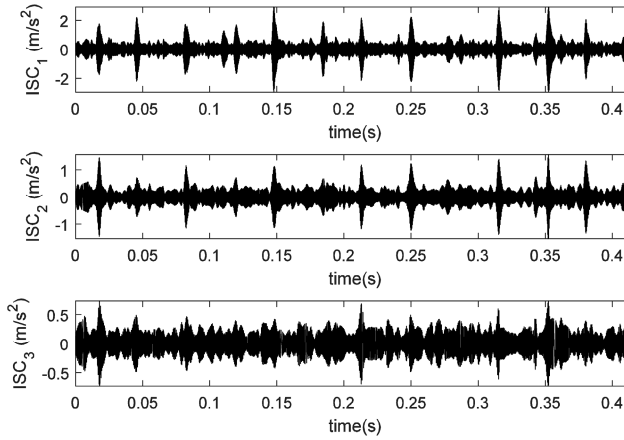


Fig. 18. LCD result of bearing vibration signal shown in Fig. 17

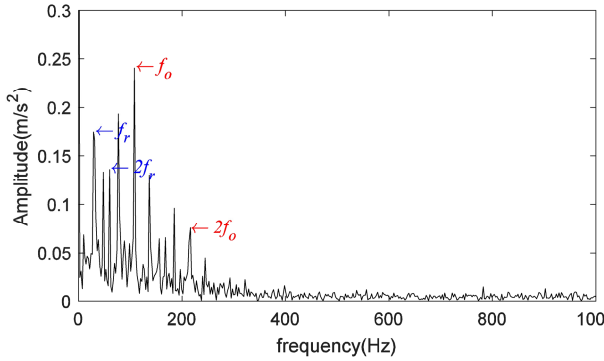


Fig. 19. Hilbert envelope spectrum of the first ISC component

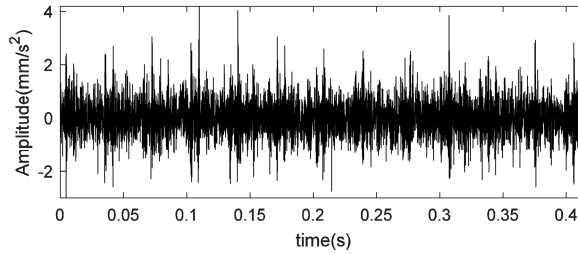


Fig. 20. The time domain waveforms of vibration signals under inner race fault

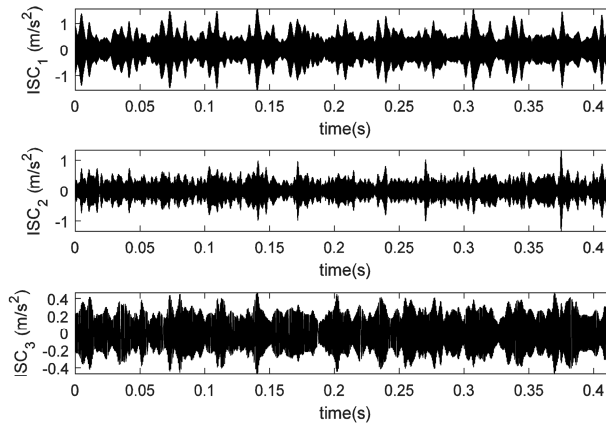


Fig. 21. LCD result of bearing vibration signal shown in Fig. 20

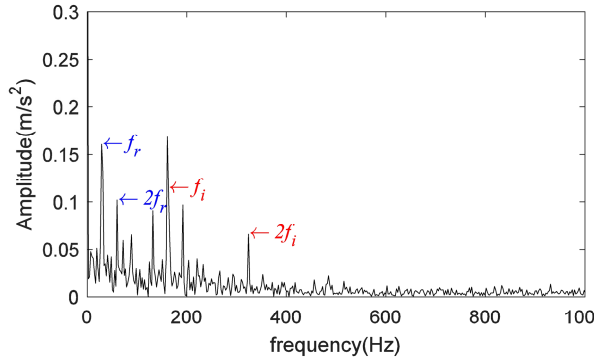


Fig. 22. Hilbert envelope spectrum of the first ISC component

The same procedure was repeated for the bearing vibration signal with the inner race fault. The time domain waveforms of vibration signals under inner race fault and its first three ISCs are shown in Figs. 20-21. And the Hilbert envelope spectrum of the first ISC is shown in Fig. 22. From the spectrum, we can also realize spectrum lines at the characteristic frequency of the inner race fault (162.09 Hz) and its harmonic frequency.

5. Conclusions

In this paper, we proposed a method for detecting rolling bearing faults based on CS theory and LCD and verified its effectiveness. LCD is a new signal processing method, which uses simulated signals to decompose signals, resulting in a 17.9 % higher accuracy compared to the widely used EMD. We also applied CS theory to verify that the fault can be detected by only a few samples without completely reconstructing the vibration signal. Further research on the determination of sparsity k and number of compressed samples M and the computational time reduction of signal processing algorithms for signal reconstruction is needed to improve the efficiency of equipment fault diagnosis using CS theory.

Acknowledgements

The authors have not disclosed any funding.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

Myong-Jin Jo: conceptualization, investigation, methodology, writing-original draft preparation. Su-Jong Kim: data curation, formal analysis, software, visualization, validation. Tong-Chol Choe: project administration, resources, supervision.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] P. D. McFadden and J. D. Smit, "Vibration monitoring of rolling element bearings by the high-frequency resonance technique-A review," *Tribology International*, Vol. 17, No. 1, pp. 3–10, 1984.

[2] N. Tandon and A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribology International*, Vol. 32, pp. 469–480, 1999.

[3] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mechanical Systems and Signal Processing*, Vol. 23, No. 3, pp. 724–739, Apr. 2009, https://doi.org/10.1016/j.ymssp.2008.06.009

[4] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Fault diagnosis of rolling element bearing using cyclic autocorrelation and wavelet transform," *Neurocomputing*, Vol. 110, pp. 9–17, Jun. 2013, https://doi.org/10.1016/j.neucom.2012.11.012

[5] K. Fu, J. Qu, Y. Chai, and T. Zou, "Hilbert marginal spectrum analysis for automatic seizure detection in EEG signals," *Biomedical Signal Processing and Control*, Vol. 18, pp. 179–185, Apr. 2015, https://doi.org/10.1016/j.bspc.2015.01.002

[6] X. Zhao, T. H. Patel, and M. J. Zuo, "Multivariate EMD and full spectrum based condition monitoring for rotating machinery," *Mechanical Systems and Signal Processing*, Vol. 27, pp. 712–728, Feb. 2012, https://doi.org/10.1016/j.ymssp.2011.08.001

[7] Cong Wang, Meng Gan, and Chang'An Zhu, "Non-negative EMD manifold for feature extraction in machinery fault diagnosis," *Measurement*, Vol. 70, pp. 188–202, 2015.

[8] B. Cui and X. Chen, "Improved hybrid filter for fiber optic gyroscope signal denoising based on EMD and forward linear prediction," *Sensors and Actuators A: Physical*, Vol. 230, pp. 150–155, Jul. 2015, https://doi.org/10.1016/j.sna.2015.04.021

[9] D. Yu, J. Cheng, and Y. Yang, "Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings," *Mechanical Systems and Signal Processing*, Vol. 19, No. 2, pp. 259–270, Mar. 2005, https://doi.org/10.1016/s0888-3270(03)00099-2

[10] L. Lin and J. Hongbing, "Signal feature extraction based on an improved EMD method," *Measurement*, Vol. 42, No. 5, pp. 796–803, Jun. 2009, https://doi.org/10.1016/j.measurement.2009.01.001

[11] J. S. Smith, "The local mean decomposition and its application to EEG perception data," *Journal of The Royal Society Interface*, Vol. 2, No. 5, pp. 443–454, Dec. 2005, https://doi.org/10.1098/rsif.2005.0058

[12] Y. Wang, Z. He, and Y. Zi, "A comparative study on the local mean decomposition and empirical mode decomposition and their applications to rotating machinery health diagnosis," *Journal of Vibration and Acoustics*, Vol. 132, No. 2, pp. 613–624, Apr. 2010, https://doi.org/10.1115/1.4000770

[13] Y. X. Wang, Z. J. He, and Y. Y. Zi, "A demodulation method based on improved local mean decomposition and its application in rub impact fault diagnosis," *Measurement Science and Technology*, Vol. 20, No. 2, pp. 1–10, 2009.

[14] J. Cheng, J. Zheng, and Y. Yang, "A nonstationary signal analysis approach-the local characteristic-scale decomposition method," *Journal of Vibration Engineering*, Vol. 25, No. 2, pp. 215–220, 2012.

[15] Y. Yang, M. Zeng, and J. Cheng, "Research on local characteristic-scale decomposition and its capacities," *Journal of Vibration Engineering*, Vol. 25, No. 5, pp. 602–608, 2012.

[16] J. S. Cheng, Y. Yang, and Y. Yang, "Local characteristic-scale decomposition method and its application to gear fault diagnosis," *Journal of Mechanical Engineering*, Vol. 48, No. 9, pp. 64–71, 2012.

[17] Cong et al., "Short-time matrix series based singular value decomposition for rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, Vol. 34, No. 1, pp. 218–230, 2013.

[18] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Processing*, Vol. 96, pp. 1–15, Mar. 2014, https://doi.org/10.1016/j.sigpro.2013.04.015

[19] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, Vol. 38, pp. 617–643, 1992.

[20] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, Vol. 81, No. 3, pp. 425–455, 1994.

[21] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, Vol. 59, No. 8, pp. 1207–1223, 2006.

[22] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, Vol. 52, No. 2, pp. 489–509, 2006.

[23] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, Vol. 52, No. 4, pp. 1289–1306, 2006.

[24] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, Vol. 51, No. 1, pp. 34–81, Feb. 2009, https://doi.org/10.1137/060657704

[25] J. Zou and F. Li, "A method of image denoising based on compressive sensing," *Journal of North China University of technology*, Vol. 24, No. 1, pp. 11–7, 2012.

[26] J.-S. Cheng, "Application of compressed sensing theory in speech denoising," *Modern Electronics Technique*, Vol. 35, No. 7, pp. 884–88, 2012.

[27] X. Zhou, A. Wang, H. Sun, and H. Yang, "Speech enhancement based on compressed sensing," *Journal of Tsinghua University (Science and Technology)*, Vol. 51, No. 9, pp. 1234–1238, 2011.

[28] Case Western Reserve University Bearing Data center," http://csegroups.case.edu/bearingdatacenter/pages/welcome.

Myong-Jin Jo received a Ph.D. degree in mechanics from DPR Korea in 2023. Now he works at Kim Il Sung University. His research interests are signal processing, condition monitoring and fault diagnosis of the rotating machinery.

Su-Jong Kim received an M.S. degree in mechanics from DPR Korea in 2020. Now she works at Kim Il Sung University. Her research interests are rotordynamics, condition monitoring and fault diagnosis of the rotating machinery.

Tong-Chol Choe received a Ph.D. degree in mechanics from DPR Korea in 2019. Now he works at Kim Il Sung University. His research interests are signal processing, condition monitoring and fault diagnosis of the rotating machinery.

ISSN PRINT 2669-2600, ISSN ONLINE 2669-1361