
Adaptive Signal Decomposition Methods for Vibration Signals of Rotating Machinery

Wei Guo and Ming J. Zuo

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/67530>

Abstract

Vibration-based condition monitoring and fault diagnosis are becoming more common in the industry to increase machine availability and reliability. Considerable research efforts have recently been directed towards the development of adaptive signal processing methods for fault diagnosis. Two adaptive signal decomposition methods, i.e. the empirical mode decomposition (EMD) and the local mean decomposition (LMD), are widely used. This chapter is intended to summarize the recent developments mostly based on the authors' works. It aims to provide a valuable reference for readers on the processing and analysis of vibration signals collected from rotating machinery.

Keywords: signal processing, empirical mode decomposition, local mean decomposition, fault diagnosis, rotating machinery

1. Introduction

Signal processing methods with adaptive basis functions are more effective in revealing the overlapping components in vibration signals. They are able to adaptively disassemble non-linear and non-stationary signals into some simpler signal components. The empirical mode decomposition (EMD) [1] method and the local mean decomposition (LMD) [2] method have been recognized to be such effective adaptive signal processing methods.

Since the introductions of EMD in year 1998 [1] and LMD in year 2005 [2], many improvements and applications have been reported. In this chapter, we summarize the recent developments mostly based on the authors' works. We hope that it is a valuable reference for readers on the processing and analysis of vibration signals collected from rotating machinery. This chapter

is organized as follows. Section 2 briefly introduces the fundamentals of EMD and LMD. Section 3 summarizes key results on the improvements of EMD and LMD. Section 4 outlines future work and remaining challenges.

2. Fundamentals of EMD and LMD

The EMD method [1] decomposes a nonlinear and non-stationary signal as a sum of some intrinsic mode functions (IMFs). Resembling the popular wavelet transform, EMD can also display the spread of signal energy on available frequencies locally in time [3]. Their key difference is that the EMD method is direct and adaptive, so that some potential and valuable information can be obtained from the data without the influence from *a priori* basis. Hence, it is widely applied in diverse areas of signal processing, especially in the field of mechanical vibration, such as health monitoring and diagnosis, analysis and identification of weak vibration signals, mainly in rotating machinery with critical elements, bearings and gears [3].

Another adaptive signal processing method, the LMD method, was originally used as a time-frequency analysis tool of the electroencephalogram signals [2]. It is an iterative approach to decompose a signal into some product functions (PFs) [2], each of which is an amplitude-modulated and frequency-modulated signal (AM-FM signal) from which the instantaneous amplitude (IA) and the instantaneous frequency (IF) can be derived [4]. Compared with the corresponding IMF in EMD, as shown in **Figure 1**, the calculated IF and IA are not involved in the EMD and require adopting of the Hilbert transform (HT).

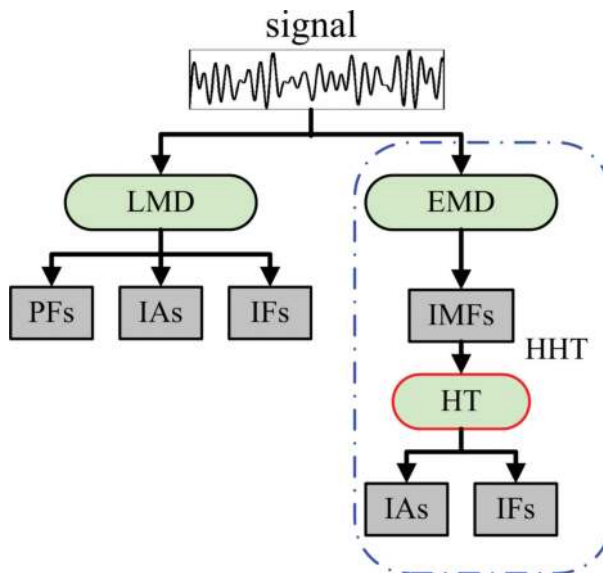


Figure 1. Main differences between EMD and LMD [4].

Although the details of decomposition and the resulted signals are quite different, these two methods share some common advantages, for example, the adaptive property. They also share some common challenges, which will be addressed in Section 3. Ref. [4] provides a comparative study, and Ref. [5] reviews applications of EMD in the field of fault diagnosis.

No matter which of two methods is used, a multi-component signal, $x(t)$, can be adaptively decomposed into k mono-components, $x_p(t)$ ($p = 1, 2, \dots, k$) (IMFs for EMD or PFs for LMD) and a residue, u_k , and can be reconstructed by summing them together, i.e.

$$x(t) = \sum_{p=1}^k x_p(t) + u_k(t). \quad (1)$$

3. Reported improvements in EMD and LMD

The EMD and the LMD methods are proven to be quite versatile in a broad range of applications for adaptively extracting signals of interest from noisy data. This section discusses their main and common challenges, including end effects, mode mixing, feature signal selection and strong noise reduction. After analysing each issue, the corresponding improvement is also shown. Other open issues, such as stopping criterion and envelope function, will be briefly discussed in Section 4.

3.1. End effects

End effects have plagued data analysis from the beginning of any known method [6]. The end effects were first mentioned in the spine fitting of the EMD. This section briefly reviews related improvements and then introduces an adaptive method to eliminate the end effects for the vibration signals collected from rotating machinery.

3.1.1. Improvements for eliminating end effects

Two ways have been proposed to eliminate end effects. One timid way is to use a sliding window [7], as is done routinely in Fourier analysis [6]. The sliding window is successfully applied to Fourier analysis using various windows and continuous wavelet analyses. However, appropriate and reliable windows are often analysis method related but not related to the data themselves. It inevitably leads to sacrifice some precious data near the ends [8]. Furthermore, it would be a hindrance for data processing when the data are short.

The other elimination way is extension or prediction of data beyond their existing range, which is still the best basic solution. Huang et al. [1] first proposed to add characteristic waves to treat the effects, in which the extra points are determined by the average of n -waves in the immediate neighbourhood of the ends. Motivated by this idea, some extension methods tried to extend the temporal waveforms forward and backward by using all available information in data, including feature-based extension, mirror or anti-mirror extension, intelligent prediction, pattern comparison, etc. [9].

It is proven that prediction methods can provide good performance on the extension of data. It is not needed to predict the whole time series, but to predict the value and location of the extrema adjacent to the ends. However, as pointed out by Huang and Shen [6], the data extension or prediction is a risky procedure even for linear and stationary processes. For nonlinear and non-stationary processes, the problems, such as predictable conditions, method and accuracy, are still open at present. Meanwhile, intelligent methods have their own shortcomings, including minima and over-fitting in artificial neural network (ANN) and sensitiveness to parameter selection in both support vector regression and ANN.

No matter which method is developed, their main idea is that newly added points have minimal interior perturbations and extend the signal implicitly or explicitly beyond the existing range. Furthermore, the extending data can well repeat the form or feature of the original signal. The reliability of such extension will sharply decrease as its distance away from the known data set increases, and thus it is necessary to be careful in extending a signal only by adding the extrapolation data to it [10]. Otherwise, the error of such operation would propagate from the end to the interior of the data and even cause severe deterioration of the whole signal [9].

For most of the vibration signals generated by rotating machinery, their non-linear and non-stationary properties are definite, which is quite challenging for data extension. Although the mirror image extension is easier to be put into practice, the real case that the data are mostly from non-stationary stochastic systems must be faced. Fortunately, the vibration signal has an advantage to assist the extension: it is cyclo-stationary [11]. Meanwhile, the extension based on characteristics of the signal waveform seems to be more appropriate to describe such complexity of problems [10]. In the following section, an adaptive waveform extension method [9] is introduced to extend vibration signals and avoid error accumulation.

3.1.2. Adaptive data extension-based spectral coherence

To facilitate applications to condition monitoring and fault diagnosis, the designed extension method should have good extension performance as well as easy operation to implement. An adaptive extension method [9] was designed for vibration signals, mainly including three steps: waveform segmentation, spectral coherence comparison and waveform extension. Its main idea is to automatically search inside waveforms having mostly similar frequency spectrum to ends, and then use their successive segments for signal extension. In this method, a critical point is how to measure the waveform similarity. Although there are some similarity measures, such as correlation coefficient, cross-correlation, waveform similarity, originally used in the field of data fusion, pattern recognition and speech recognition, most of them are susceptible to noise and not suitable for processing vibration signals since their acquisition and transmission often suffer from noise. Therefore, an index measuring the spectral coherence [12] is introduced here. The procedure is described as follows [9]:

Step 1. Waveform segmentation. Identify zero crossings of the analyzed signal and then separate the signal into N segments, $c_i(t)$ ($i = 1, \dots, N$).

Step 2. Segment repetition and fast Fourier transform (FFT). Repeat each segment to form a long waveform and then conduct fast Fourier transform (FFT).

Step 3. Spectral coherence comparison. Use Eq. (2) to calculate the revised spectral coherence (RSC) values of the first segment $c_1(t)$ and other segments, and then find the segment, denoted as $c_{\text{back}}(t)$, having the largest value of the RSC. Similarly, search the segment $c_{\text{for}}(t)$ similar to the last segment $c_N(t)$.

$$\gamma_{ij} = \frac{\sum_F C_i(F) \times C_j(F)}{\sqrt{\left(\sum_F C_i(F) \times C_i(F)\right) \left(\sum_F C_j(F) \times C_j(F)\right)}}, \quad (2)$$

where $C_i(F)$ and $C_j(F)$ are frequency spectra of the signal components $c_i(t)$ and $c_j(t)$, respectively.

Step 4. Waveform extension. Use the pervious segment of $c_{\text{back}}(t)$ for extending backward and the next segment of $c_{\text{for}}(t)$ for extending forward.

Based on this, the extended signal can be decomposed by the EMD or LMD method, and extended samples are at last truncated before further analysis. Using the hidden periodicity, a cyclo-stationary signal, for example, a vibration signal, can be easily extended beyond its original range, and its temporal continuity in time domain and spectral coherence in frequency domain can be properly maintained.

3.1.3. Experiment and analysis

A vibration signal collected from an industrial traction motor [13] is shown in **Figure 2**. The specification of the experiment setup is given in **Table 1**. This signal is cyclo-stationary with around three cycles, and its waveform is thus divided into six segments, which are marked in **Figure 2**. **Figure 3** shows frequency spectra of these segments.

To estimate the influence caused by the end effects, a measure of energy change [14] before and after decomposition is defined as

$$\theta = \frac{1}{R_x} \left| \sqrt{\sum_{p=1}^k R_p^2 + R_u^2} - R_x \right|, \quad (3)$$

where R_x , R_p and R_u are root-mean-square (RMS) values of the original signal $x(t)$, the p th product function $PF_p(t)$, and the residue signal $u_k(t)$, respectively. The value of the measure is $\theta \geq 0$. The closer the measure is to zero, the smaller the error between the original signal and decomposition results is; that is to say, the influence caused by end effects is smaller.

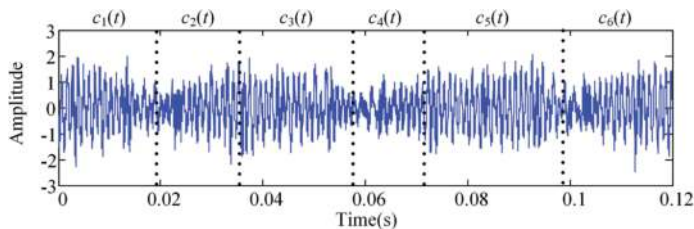


Figure 2. A raw vibration signal collected from a traction motor.

Section	Tested object	Fault type	f_s (kHz)	f_r (Hz)	f_d (Hz)
3.1.3	Bearing	An outer race defect on bearing, and motor eccentric	32.8	25	BPFO = 114
3.2.2					
3.4.3.2					
3.2.1	Bearing	An outer race defect on bearing	80	23.3	BPFO = 135
3.3	Bearing	An inner race defect on bearing	80	23.3	BPFI = 192
3.4.3.1	Bearing	An outer race defect on bearing	51.2	60	BPFO = 183

Note: f_s —sampling frequency; f_r —rotating frequency of the motor; f_d —characteristic defect frequency; BPFO—the ball pass frequency of the outer race; BPFI—the ball pass frequency of the inner race.

Table 1. Specifications of the experiment setup.

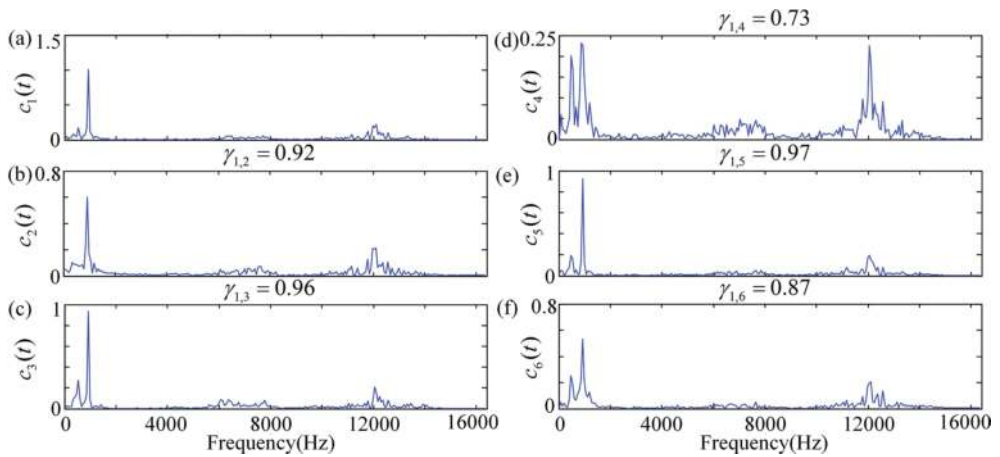


Figure 3. Frequency spectra of six segments $c_1(t)$ – $c_6(t)$ in (a)–(f) in **Figure 2**, and spectral coherence values between $c_1(t)$ and the other segments.

The revised spectral coherence (RSC) values $\gamma_{1,j}$ ($j = 2, \dots, 6$), i.e. RSC values between the segment $c_1(t)$ and one of other segments, are shown above each sub-figure in **Figure 3**. It can be seen that the segment $c_5(t)$ has the largest RSC value of 0.97 and its previous segment $c_4(t)$ is then used for backward extension of $c_1(t)$. In a similar way, $c_2(t)$ has the largest RSC value with the last segment $c_6(t)$, and thus the next one of $c_2(t)$, i.e. $c_3(t)$ is used for forward extension. The extended vibration signal is shown in **Figure 4**, where extended waveforms are shown in red. Its RSC value with the original signal is 0.94, and the measure θ is 0.005. If no extension, the measure θ is 0.106.

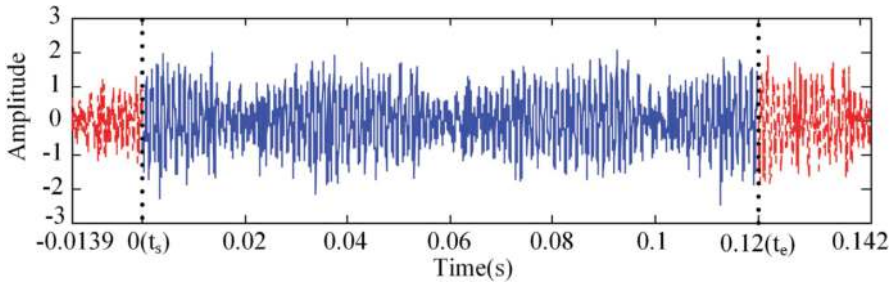


Figure 4. The extended vibration signal obtained by the adaptive waveform extension method [9].

After applying the LMD method to the extended signal and truncating extended parts, five PFs and a residue are obtained, the first three of which have larger correlation coefficient values with the original signal and thus are selected for further analysis. Their waveforms and envelope spectra are shown in Figure 5. In Figure 5(d), the identified characteristic frequency (104 Hz) and its harmonics (around $2 \times$ and $3 \times$ BPFO) can be easily observed. The error between the theoretical value (114 Hz) and the identified one (104 Hz) is mainly caused by inaccurate shaft speed after long use and limited samples (only 0.12 second). In Figure 5(e) and (f), higher impulses are identified at the frequency of 25 Hz, corresponding to the motor rotating frequency. It indicates that PF1 is the signal generated by the inspected bearing with an outer race defect, and PF2-3 is generated by the motor, which turned out to be caused by the eccentric problem after inspection. More cases on bearings and gears can be found in Ref. [9].

3.2. Mode mixing

Another open problem for EMD and LMD is the mode mixing. It is originally defined as a single IMF either consisting of signals of widely disparate scales, or a signal of a similar scale residing in different IMF components, which causes serious aliasing in the time-frequency

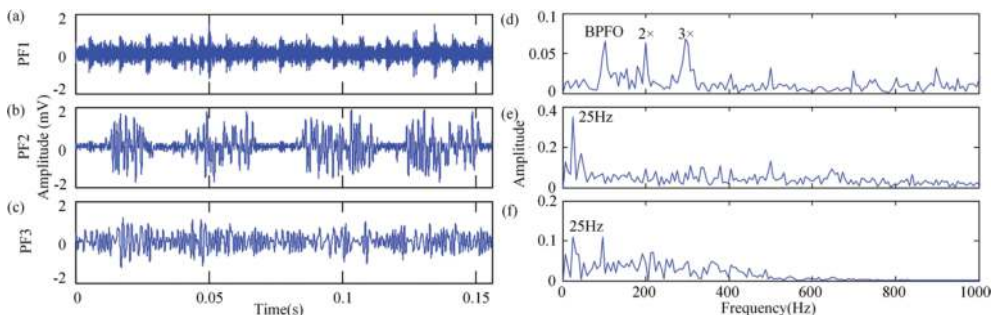


Figure 5. Decomposition results PF1-PF3 in (a)–(c) by applying the LMD to the signal in Figure 4, and their corresponding envelope spectra in the range of 0–1 kHz in (d)–(f).

distribution and makes the meaning of IMF unclear [8]. This section focuses on the solution to the problem of mode mixing.

3.2.1. Separation of disparate components

According to the above definition, there are two possibilities: either completely different components existing in one IMF, or one component appearing in more than one IMF. To remove the former case, Wu and Huang [8] presented a noise-assisted signal processing method, called ensemble EMD (EEMD). In this method, white noise with a pre-setting amplitude is introduced to perturb the analyzed signal and enables the EMD method to visit all possible solutions in the finite neighbourhood of the true final IMF [8]; and the ensemble means of decomposition results help to remove the remaining noise in the results. For the EEMD method, two parameters, the noise amplitude and the ensemble number, are critical, the former of which has more influence on its performance [15]. In order to process signals adaptively, it is ideal to automatically find appropriate parameters for the analyzed signal. A parameter optimization method [13] is designed for the EEMD. In this method, an index termed relative root-mean-square error (RMSE) is first used to evaluate the performance of the EEMD method when fixing a small ensemble number and setting various noise amplitudes, and then the signal-to-noise ratio (SNR) is introduced to evaluate the remaining noise in the results when gradually increasing the ensemble number.

For a signal, $x_o(k)$, it is assumed that it consists of main component(s), background noise and some components having small correlation coefficients with the chief one, which has the largest correlation coefficient with the signal $x_o(k)$ is marked as $c_{\max}(k)$. The desired decomposition is to completely separate the component $c_{\max}(k)$ from others, and the relative RMSE is thus used to evaluate the separation performance when setting various noise amplitudes. Its formulation is

$$\text{Relative RMSE} = \sqrt{\frac{\sum_{k=1}^S (x_o(k) - c_{\max}(k))^2}{\sum_{k=1}^S (x_o(k) - \bar{x}_o)^2}}, \quad (4)$$

where \bar{x}_o is the mean of the signal $x_o(k)$, and S is the number of samples in this signal. The value of this index is in the range of 0–1. The smaller this index is, the closer the component $c_{\max}(k)$ to the original signal. It means that the extracted IMF contains not only the main component of interest but also other components, and thus the objective is not achieved. However, there exists a value for the noise amplitude that maximises the index. At this point, the error between $x_o(k)$ and $c_{\max}(k)$ is from noise and other components, that is to say, the extracted IMF and the other in the original signal share no common component, and the main component of interest is extracted from the original signal. The corresponding value is the optimal noise amplitude. Its procedure is briefly described as follows [13]:

Step 1. Set a small value of the initial ensemble number, for example, $N_E = 10$, and choose a relatively large value as the initial noise level, $L_N = l_0$. The noise amplitude A is to multiply the noise level by the standard deviation of the signal.

Step 2. Perform the signal decomposition using the EEMD method and calculate the relative RMSE of the chief component $c_{\max}(k)$.

Step 3. Decrease the noise level and repeat Step 2 until the change in the relative RMSE is negligible or small enough.

Step 4. Identify the optimal noise level corresponding to the maximal relative RMSE.

Once the optimal noise level is numerically determined, the ensemble number can be determined by comparing the SNR values when gradually increasing the ensemble number from its pre-setting value.

To demonstrate this method, a vibration signal was collected from a small motor [13] and is shown in **Figure 7(a)**. In the experiment, a fault was set on the outer race of the tested bearing. The specification of the experiment is shown in **Table 1**. Initial parameters are set as: a larger value for the noise level, $L_N = 2$, and the ensemble number $N_E = 10$. During the execution of the above program, the noise level is gradually decreased. When $2 \leq L_N \leq 0.1$, the noise level is decreased in the step of 0.1; when $0.1 < L_N \leq 0.01$, its decreasing step is 0.01; when $0.01 < L_N \leq 0.001$, its decreasing step is 0.001.

After applying the EEMD method with the above optimization method to decompose the vibration signal, the relative RMSEs for various noise levels are shown in **Figure 6**. As shown in this figure, the maximal relative RMSE is arrived at the noise level of 0.4, corresponding to the optimal one, and accordingly, the extracted IMF (IMF1) is shown in **Figure 7(b)**. Comparing with the original signal, most of noise and redundant components are separated from IMF1, and its kurtosis value is 26.07.

To compare with this, extracted IMFs when setting any three non-optimal noise levels are also shown in **Figure 7**. **Figure 7(c)** and **(d)** shows the results when setting the noise levels of 2 and 1, respectively. **Figure 7(e)** shows the result when setting a quite small noise level of 0.009. Their kurtosis values of these IMFs are 11.68, 7.26 and 7.65, respectively. It demonstrates that better decomposition results are obtained after setting the optimal noise level.

Having determined the optimal noise level, appropriate ensemble number is then determined. The variation in the SNR is shown in **Figure 8**. As the figure shows, when the ensemble number is less than 80, increasing the ensemble number gently accelerates the increase in the SNR value. When the ensemble number is larger than 120, the SNR value fluctuates smoothly.

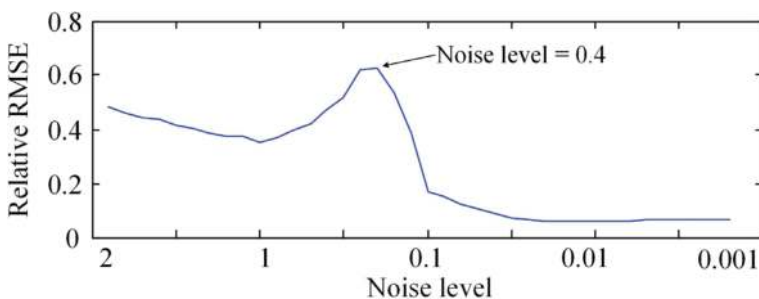


Figure 6. Relative RMSEs when adding white noise with various noise levels to the vibration signal in **Figure 7(a)**, and the optimal noise level for this signal is 0.4.

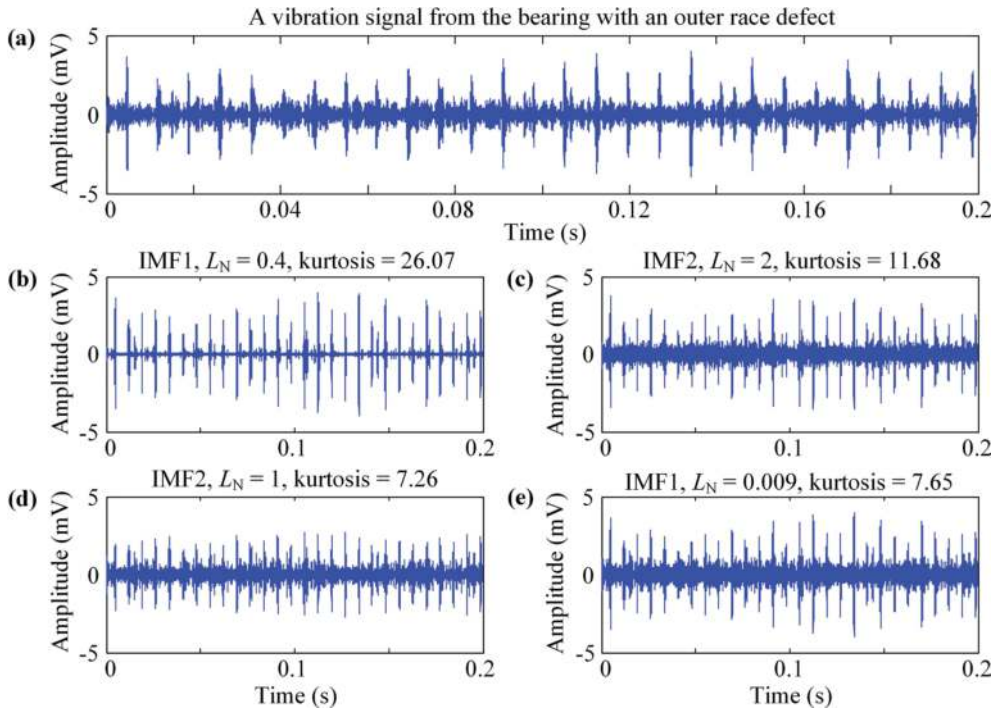


Figure 7. A vibration signal from the bearing with an outer race defect and the corresponding selected IMFs when setting the optimal and three non-optimal noise levels. (a) The signal to be analyzed; (b) IMF1 when setting $L_N = 0.4$; (c) IMF2 when setting $L_N = 2$; (d) IMF2 when setting $L_N = 1$; and (e) IMF1 when setting $L_N = 0.009$.

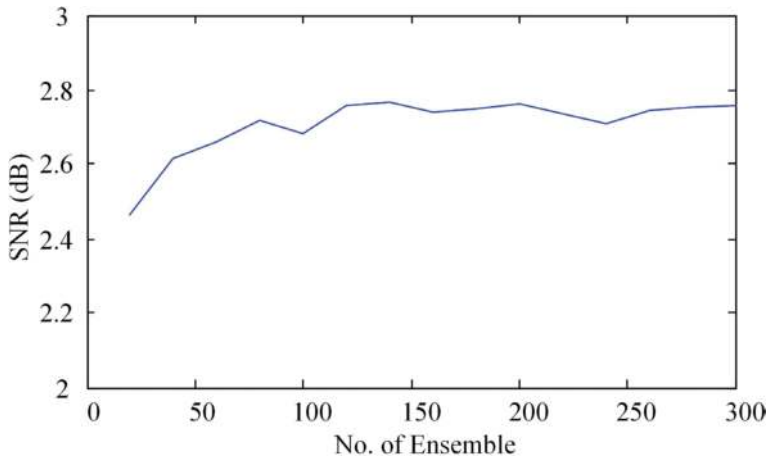


Figure 8. SNR values for various ensemble numbers when setting the optimal white noise ($L_N = 0.4$) to the vibration signal in Figure 7(a).

Further increasing its value contributes to minor increasing of the SNR, but definitely rising computation cost. Therefore, using this optimization method, parameters of the EEMD can be automatically determined according to the signal itself, instead of empirical setting or the trial and error. More cases on bearings can be found in Ref. [13].

3.2.2. Mixing of similar components

Although the EEMD method can successfully separate signal components with different scales, another mode mixing still exists in the decomposition results, that is to say, one component may spread in more than one IMF. This also belongs to the mode mixing and results in energy dispersion and some redundant components without physical significance. It may be caused by repeated sifting process and severe stopping criterion. A simple and convenient solution is to combine the components from the same source. Therefore, the index of spectral coherence in Eq. (2) is used to evaluate the spectral similarity of two successive components and then combine the components with similar spectra into a natural IMF [12].

Using the index of spectral coherence, the similarity criterion of two successive IMFs obtained by the EEMD method is described as:

- (1) If $\gamma_{j,j+1} \rightarrow 1$, it means that the IMFs, c_j and c_{j+1} , have a relationship of similarity in frequency domain, that is to say, they have spectral coherence over the whole frequency range. Thus, these two IMFs should come from the same source and thus are combined to one natural IMF (NIMF).
- (2) If $\gamma_{j,j+1} \rightarrow 0$, they have low, even no spectral coherence and thus are two natural IMFs.
- (3) If $\gamma_{j,j+1}$ is around 0.5, the spectral coherence of two IMFs cannot be determined. Such signal components are also viewed as two natural IMFs and would not be combined together.

The signal in **Figure 2** is used to demonstrate the process of similarity analysis and combination. After applying the EEMD method with the noise level 0.2 and the ensemble number of 30 to the signal, 12 IMFs are obtained, the first four of which have larger correlation coefficient values with the original signal and are shown in **Figure 9**. As shown in the figure, the frequency spectrum of IMF1 is a high-frequency dominated signal and centred at the frequency of 12 kHz, and it indicates that IMF1 corresponds to the signal generated by the faulty bearing in the traction motor; as for IMF3 and IMF4, they share the common frequency of 920 Hz generated by the faulty motor. Furthermore, the revised spectral coherences of all IMFs are calculated and the results are shown in **Table 2**. According to this table, there are three local minimal points, i.e. $\gamma_{2,3}$, $\gamma_{5,6}$ and $\gamma_{7,8}$. The RSC values of IMF3-IMF4 and IMF4-IMF5 are larger than 0.5 and it shows their similarity on frequency domain, and thus these three components are combined to one natural IMF. Between the second and the third local minimal values, IMF6 and IMF7 show the spectral similarity. Similarly, the remaining components, IMF8-IMF12, also show their spectral similarity, and thus are merged into another natural IMF. The RSC value of IMF1 and IMF2 is not close to 1 or 0 and these two IMFs are thus two natural IMFs. Final results are shown in **Figure 10**. The last two components are practically residues. Based on local minima of RSC, a fusion rule [12] was designed to automatically combine

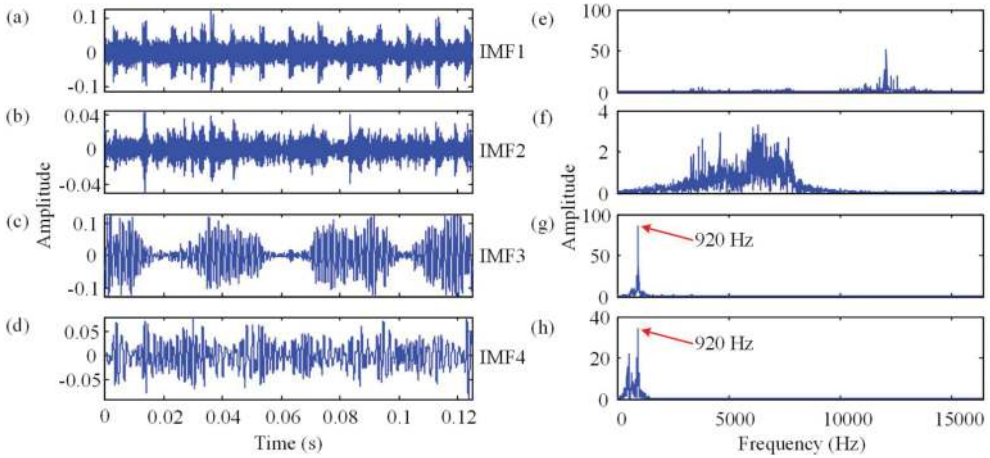


Figure 9. IMF1-IMF4 obtained by applying the EEMD ($N_E = 30, L_N = 0.2$) to the signal in **Figure 2** and frequency spectra of these IMFs (a)-(d) The temporal waveforms of IMF1-IMF4, and (e)-(h) corresponding frequency spectra of IMF1-IMF4 in (a)-(d).

RSC	$\gamma_{1,2}$	$\gamma_{2,3}$	$\gamma_{3,4}$	$\gamma_{4,5}$	$\gamma_{5,6}$	$\gamma_{6,7}$	$\gamma_{7,8}$	$\gamma_{8,9}$	$\gamma_{9,10}$	$\gamma_{10,11}$	$\gamma_{11,12}$
Value	0.279	0.054	0.768	0.542	0.375	0.568	0.398	0.593	0.619	0.993	0.958

Table 2. RSC values of decomposition of the signal in **Figure 2** by using EEMD ($N_E = 30, L_N = 0.2$).

components from the same source and remove the mode mixing in the original EMD method. Other applications on bearings can be found in Ref. [12].

3.3. Strong noise reduction

In real rotating machinery, a raw vibration signal generally consists of strong noise and two or more sources. Some vibrations, such as improper installation and surfacing of the installed sensors, random impacts from friction and contact forces and external disturbances [16], are also so strong that the signal of interest is completely overwhelmed. Therefore, the recovery of the feature signal from noise, while preserving its features is a challenging problem. This section introduces a hybrid signal processing method [17] for noisy vibration signals.

3.3.1. Problem analysis

Although the EEMD method improves the scale separation ability of EMD method, both methods are based on extrema to discriminate signals generated by various sources. When the signal of interest is completely overwhelmed by strong noise, there may be a lack of necessary extrema for the EEMD method to separate the real signal from noise. An experimental signal collected from a bearing with an inner race defects is used to illustrate this problem [17]. The specification of the experiments is shown in **Table 1**. To simulate strong noise in real

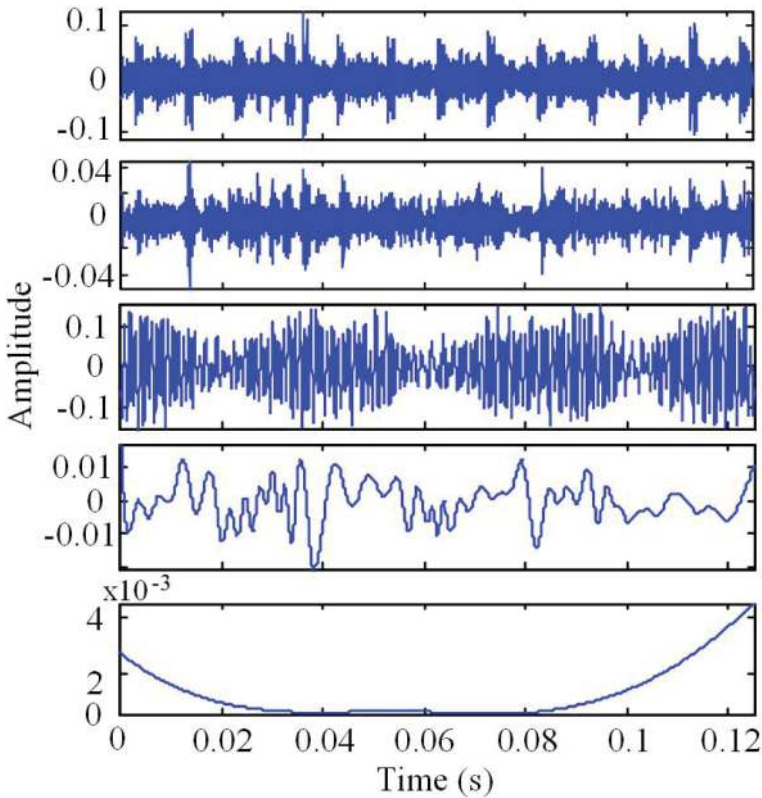


Figure 10. Five natural IMFs after applying similarity criterion to the results obtained by the EEMD.

cases, Gaussian white noise was added to the experimental signal, and the generated noisy signal is shown in **Figure 11**. As shown in the figure, the impulses caused by faulty bearing are completely masked by strong noise. After applying the EEMD method to this signal, 13 IMFs are obtained and the first four having larger correlation coefficient values with the original signal are shown in **Figure 12**, in which impulses are seldom observed and still buried in noise. It is because that the decomposition method lacks necessary extrema generated by the tested faulty bearing.

As for a signal with a relatively low signal-to-noise ratio, it is necessary to design an adaptive filter to extract the weak feature signal of interest from a noisy signal to facilitate further signal decomposition. A possible solution is to use the spectral kurtosis, which is proven to be a powerful tool to identify the existing of bearing faults buried in noise. Its value is large in frequency bands where the impulsive bearing fault signal is dominant, and is effectively zero where the spectrum is dominated by stationary components [16]. Based on this, an SK-based filter [18] was used to pre-process the signal in **Figure 11** and remove part of noise before decomposition. It is a kind of band-pass filter whose parameters, centre frequency and bandwidth,

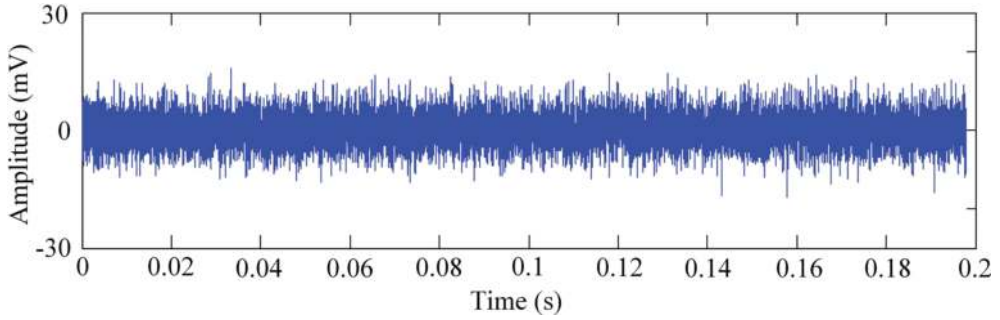


Figure 11. A vibration signal from a bearing with an inner race defect with the added white noise.

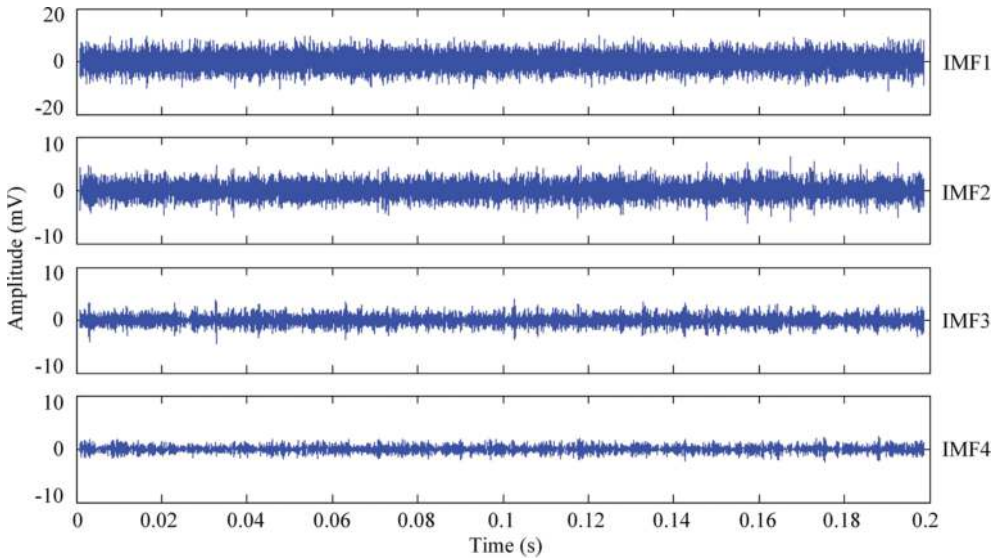


Figure 12. The first four IMFs obtained by applying the EEMD method to the signal in Figure 11.

are optimized by using the kurtogram, a map formed by the STFT-based (short-time Fourier transform based) SK. The filtered result is shown in Figure 13. Although the filtered signal still contains some noise, its impulses are a little clearer than those in the original signal, and its kurtosis value is also increased from 3.07 to 3.97. Consequently, a hybrid method is used to reinforce the performance of noise reduction.

3.3.2. A hybrid method for strong noise reduction

By comparing individual performances of the foregoing two methods, a hybrid signal processing method that combines the EEMD and the SK-based filter [17] is introduced. First, an optimal band-pass filter based on SK is employed to remove part noise so that local extrema

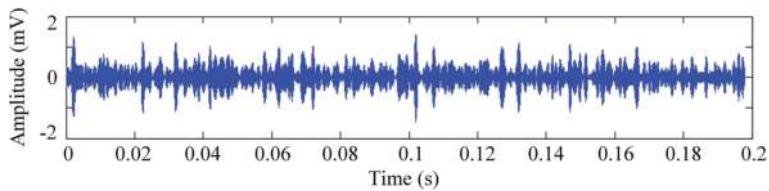


Figure 13. The filtered signal using the SK-based filter, and its kurtosis value is 3.97.

of the signal would not be completely concealed by noise. Then, the EEMD method with parameter optimization is applied to further decompose the filtered signal. As a result, the final signal can be separated from strong noise, which allows good detection of the defects but at the same time minimizes the distortion of the impulses. The main procedure is as follows:

- Step 1.** *Pre-processing.* Filter the raw signal using an optimal band-pass filter based on SK and obtain the filtered signal.
- Step 2.** *Signal decomposition.* Use the EEMD method to decompose the filtered signal into some IMFs.
- Step 3.** *Selection of feature signal.* Calculate the correlation coefficients between the obtained IMFs and the filtered signal, and select the IMF having the largest values of correlation coefficient (CC) as the resultant signal for further analysis.

3.3.3. Experiment and comparison

In this sub-section, the filtered signal in **Figure 13** is decomposed into 13 IMFs, the first three of which have CC values of 0.88, 0.76 and 0.18 with the filtered signal, and the rest of which have CC values close to zero. To save space, only the first four IMFs are shown in **Figure 14**. According to the calculation results of CC, IMF1 has a larger correlation coefficient (0.88) than the other signal components and contains the main component in the filtered signal, and it is thus viewed as the bearing signal recovered from the noisy experimental signal. This result is also verified by the identified BPF1 and its multiples as shown in **Figure 15**. There is also an error between the theoretical and identified values of BPF1, which is caused by the same reason mentioned in Section 3.1.3.

Compared with the filtered signal in **Figure 13**, the extracted bearing signal (IMF1 in **Figure 14**) is much cleaner than the original signal, and the remaining noise in the filtered signal is almost completely separated and resides in IMF2. The kurtosis values of the raw signal, the filtered signal and IMF1 are 3.07, 3.97 and 11.29, respectively, as observably increasing. It indicates that this hybrid method successfully reveals temporal impulses from a noisy signal while preserving its important feature for accurate fault diagnosis. **Figure 16** also shows the filtered signal by applying the normal wavelet threshold denoising to the same noisy signal, and its impulses are not as clear as those in **Figure 14**. More cases on faulty machine components, such as an outer race and a rolling ball, are given in reference [17].

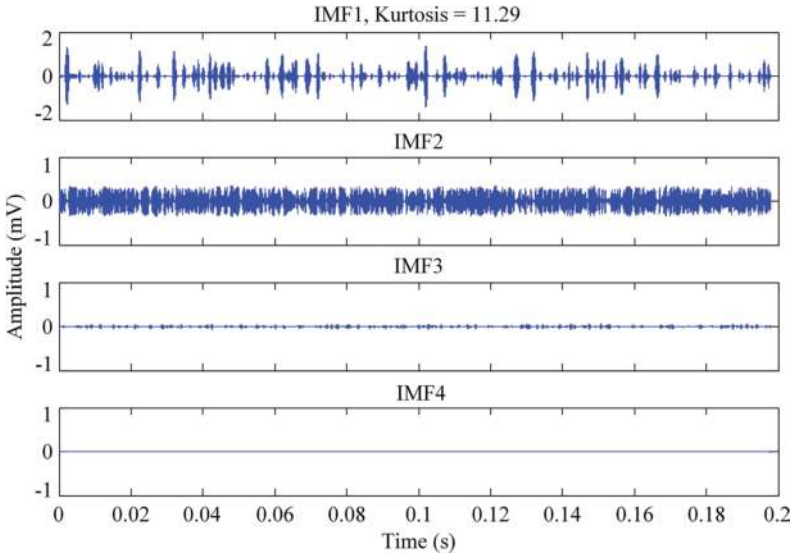


Figure 14. The first four IMFs obtained by applying the hybrid method to the signal in Figure 11.

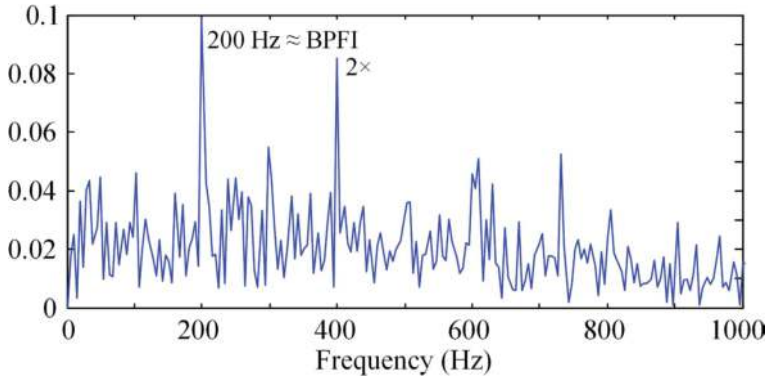


Figure 15. The envelope spectrum of IMF1 shown in Figure 14 in the range of 0–1 kHz.

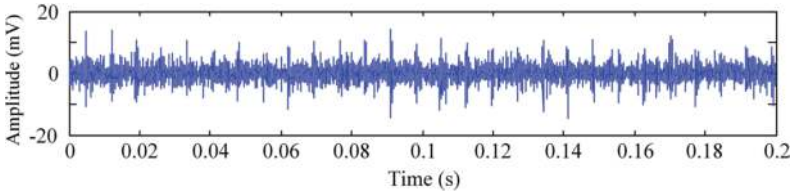


Figure 16. The filtered signal obtained by applying the normal wavelet threshold denoising to the same signal in Figure 11.

3.4. Feature signal component selection

After using the EMD or LMD method, many signal components are disassembled from the original signal. How to effectively select feature signals from many components is critical for further signal processing and analysis. This section primarily discusses the selection method of feature signal components.

3.4.1. Selection based on cluster analysis

For the feature signal selection, a popular solution is to calculate statistical indicators of the signal, for example, correlation coefficient (CC). Dybała and Zimroz [19] used this indicator to divide IMFs into three classes: noise-only IMFs corresponding to low indices and low CC values, signal-only IMFs and trend-only IMFs corresponding to high indices and low CC values. However, it is possible that an impact signal caused by a damaged bearing is wrongly categorized into the class of noise [19]. Similar results can be found when only using single measurement. A more sophisticated diagnostic method is needed to avoid the misdiagnosis. Referring to the idea of the cluster analysis, an adaptive selection method based on multiple statistic indicators is designed for selecting the feature signal of interest from many signal components [20].

In the anomaly detection, a branch of cluster analysis, a detector is designed to detect any object that deviates from the known state (usually the healthy state) [21]. Referring to this, the decomposed signal components are classified into two groups: feature signals and unrelated signals. The former is used for further analysis, and the latter is viewed as useless signals.

The key of this selection is how to evaluate useful content in the analyzed signal. If the feature signal is wrongly classified into the useless part, the state of the monitored object may be misjudged. If an unrelated signal is wrongly marked as the feature signal, the conclusion based on the analyses of feature signals may be conflicting. To correctly classify them, some statistic indicators commonly used in the anomaly detection and feature extraction are introduced here. They are indicated by many literatures to be good at representing hidden features of the analyzed signal. Therefore, these indicators are jointly used to determine the classes of decomposed components, not to determine the fault types of the tested object. In addition, the strategy of using multiple indicators is very common in pattern recognition to combine various experts with the aim of compensating the weakness of each single expert [22]. This combination can be viewed as a kind of ensemble learning and can improve the classification accuracy in machine learning. What is interesting is that the idea of combining individuals' opinions in order to reach a final decision is humans' second nature before making any crucial decision [23].

As for a large number of indicators, the distance evaluation technique (DET) [24] is introduced to quickly organize the classification result of each indicator (or expert). For more than one expert, their conclusions may not always coincide, and thus the principle of minority obeying majority [22, 23] is introduced to solve their conflicts. The detailed selection is described in the following sub-section.

3.4.2. Adaptive feature signal selection

The process of the adaptive feature signal selection can be divided into two stages: classification of each expert and decision of all experts. Its procedure is described as follows:

- Step 1.** Calculate some statistics indicators in time and frequency domains for all decomposed signal components. The indicators include peak-to-peak (P-P), mean, absolute mean, max, root mean square (RMS), standard deviation (SD), skewness, kurtosis, crest factor (CF), shape factor (SF), impulse factor (IF), energy and correlation coefficient (CC).
- Step 2.** Normalize and sort in a descending order for each indicator.
- step 3.** Classify using the DET. For each indicator, the DET makes the distance within a class shorter and the distance between classes longer, and then the components are classified into two groups.
- Step 4.** Vote by all 'experts'. For each signal component, summary how many 'experts' (indicators) classify it into the same class.
- Step 5.** Draw a conclusion. Following the principle of minority obeying majority, the classification results of signal components can be finally determined.

Furthermore, the indicators that win in the voting are viewed as sensitive ones. After comparing the values of any sensitive indicator between the current state and the healthy one, signals in the class having obvious changes can be determined as feature signals.

3.4.3. Experiments and analyses

3.4.3.1. Case 1: a vibration signal collected from a bearing with single defect

One of experimental signals was collected from a small motor that involves a bearing with an outer race defect [17]. The specification of experiments is shown in **Table 1**. After applying the LMD method to this signal, five PFs were obtained, and then indicator values of these five signal components are calculated by 13 indicators in time domain and another 13 indicators in frequency domain. **Figure 17** shows the indicator values after normalization. As shown in this figure, for the first indicator P-P (peak-to-peak), using the DET, PF3 and other PFs are classified into two groups; while, for the indicator of Mean, PF5 and other PFs belong to different groups. The classification results for all indicators in time and frequency domains are shown in the columns 'Case 1' of **Table 3**. Based on the majority principle, PF1 and PF2 are finally classified into one class, and the rest of PFs are classified into the other class. Comparing the energy value with that of a healthy bearing, PF1 and PF2 are determined as feature signals of interest. To verify this conclusion, envelope spectra of PF1-PF3 are shown in **Figure 18**. The characteristic defect frequency f_a and its multiples are only identified in the spectra of PF1 and PF2, which demonstrates the right selection of feature signals.

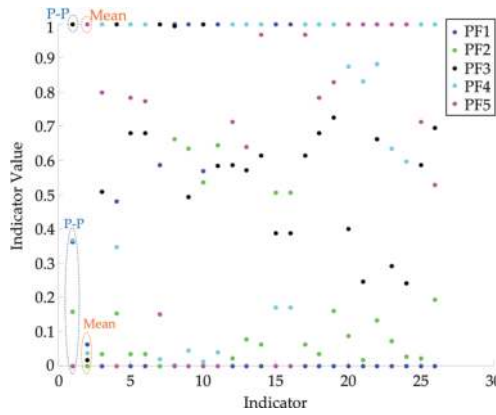


Figure 17. Indicator values of the signal in Case 1 after normalization.

Indicator	Case 1 (T)		Case 1 (F)		Case 2 (T)		Case 2 (F)	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
P-P	3	12 45	345	12	123	4567	123	4567
Mean	5	12 34	2 345	1	2	13 4567	12	3 4567
Absolute mean	345	12	2 345	1	123	4567	12	3 4567
Max	3	12 45	345	12	123	4567	123	4567
RMS	345	12	345	12	123	4567	123	4567
SD	345	12	345	12	123	4567	123	4567
Skewness	1 3	2 45	45	12 3	123 467	5	123 45	67
Kurtosis	45	12 3	45	12 3	23	1 4567	123 45	67
CF	45	12 3	345	12	12 5	3 467	123 67	45
SF	45	12 3	5	12 34	2 45	1 367	1234 57	6
IF	45	12 3	5	12 34	12 5	3 467	1234 57	6
Energy	345	12	345	12	123	4567	123	4567
CC	345	12	345	12	123	4567	123	4567

Notes:

(1) Title line: T—time domain and F—frequency domain.

(2) Numerical values listed above represent the indices of signal components PFs. For example, in the row of Max, the value '3' in the column of Case 1 (T) means that PF3 is classified into Class 1, and the values '1245' in the column of Case 1 (F) means that PF1, PF2, PF4 and PF5 are classified in to Class 2.

(3) Numerical values in bold indicate that the corresponding PFs are correctly classified after the final voting. And numerical values in italic indicate that the corresponding PFs are wrongly classified. For example, if using the index of Max in Case 1(T), PF4 and PF5 should be classified into Class 1, not Class 2, and thus the numerical values '45' in the column of Class 2 are italicized.

Table 3. Classification results of two experimental vibration signals.

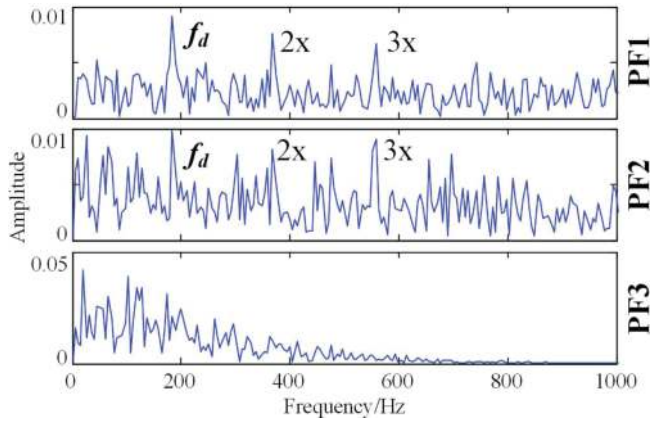


Figure 18. Envelope spectra in the range of 0-1 kHz of part components in Case 1.

3.4.3.2. Case 2: a vibration signal collected from a machine with two defects

Another vibration signal was collected from a traction motor, which involves two faulty machine components, i.e. a faulty motor and a bearing with an outer race defect [13]. Its specification is also shown in **Table 1**. This signal was decomposed into seven PFs by using the LMD. Its classification results are shown in the columns 'Case 2' of **Table 3**. As a result, PF1, PF2 and PF3 are classified into one class, and others belong to another class. **Figure 19**

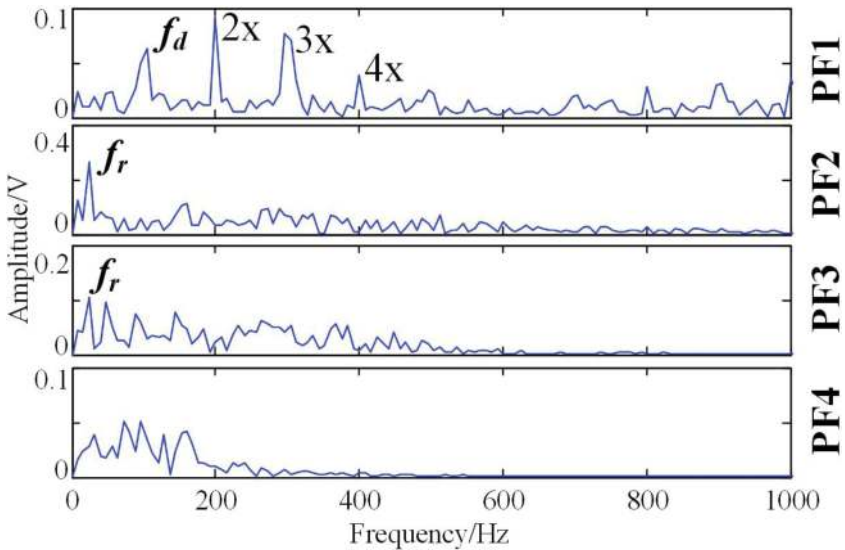


Figure 19. Envelope spectra in the range of 0-1 kHz of part components in Case 2.

shows envelope spectra in the range of 1 kHz of the first four PFs. As shown in this figure, the characteristic defect frequency f_d and its multiples of the faulty bearing are identified in PF1, and the rotating frequency f_r is identified in PF2 and PF3. No specific characteristic frequency can be identified in PF4. These results also match with real condition of the tested machine.

3.4.3.3. Discussion

Above results also indicate that statistic indicators have varying degrees on sensitivity to abnormal states. Some of them are sensitive and closely related to any faults, but others are not sensitive or stable. For above experiments, sensitive indicators include absolute mean, SD, RMS, energy, correlation coefficient in time domain, and max, peak-to-peak, SD, RMS, energy, correlation coefficient in frequency domain. The commonly used indicator, kurtosis in time and frequency domains does not show its sensitivity to feature signals. Although the indicator Energy is one of sensitive features, its values of five PFs in Case 1 are 0.048, 0.053, 0.192, 0.293 and 0.223, the latter three of which corresponding to useless components are much larger. Therefore, single measure is not suitable for fault detection. Further work on assessment of feature signals is necessary for online monitoring and diagnosis.

4. Future work

Although EMD and LMD methods are quite simple in principle, they also depend on a number of user-controlled tuning parameters and still lack an exact theoretical foundation. Feldman has given some theoretical analyses of the EMD method in Refs. [3, 25]. However, the following issues remain to be further addressed.

4.1. Stopping criterion

No matter which method, EMD or LMD, you use, the adaptive signal decomposition is a 'sifting' process, and you need to choose a criterion to stop it at the right time, which is critical for signal processing. The more times sifting is taken, the closer to zero the average will be [26], that is to say, by sifting as many times as possible, it is more likely to eliminate the riding waves and make the wave profiles symmetric. However, too many repetitions would result in the obliteration of the amplitude variation and the loss of physical meanings. Therefore, it is not an easy task to define an appropriate criterion that makes the definition of IMFs satisfied while retaining enough physical sense of amplitude and frequency modulations.

Standard stopping criterion is very rigorous and difficult to implement in practice. The most commonly used criterion is three-threshold criterion [27], and the recommended setting for the three thresholds is applicable for most of the cases. Many modifications on this criterion are also reported, and their wide verifications are not yet finished. Since most of stopping criteria are the summations over the global domain, an undesired feature is that

the decomposition is sensitive to the local perturbation and to the addition of new data [8]. Therefore, an open problem is to eliminate extra sifting processes caused by local changes.

4.2. Connection between local extrema

In the sifting process of the EMD method, a spline interpolation function is needed to connect the identified local extrema. Commonly used spline functions include linear spline, quadratic spline, cubic spline and cubic Hermite spline (third-order polynomial). Generally, the higher-order spline function can provide better fitting performance for the original signal, whereas, they require additional subjectively determined parameters and take considerable time for computation. The selection of spline function should satisfy the least interferences and maximum smoothness.

Similarly, smoothed connecting between successive extrema is also required to form a smoothly varying continuous function in the LMD method, and the parameter selection of the moving averaging is still explored. Although modifications based on single connection method or a hybrid method are sporadically reported, an appropriate criterion on the selection of connection methods receives little attention and remains an open problem.

Considering that the EMD and the LMD are data-driven analysis methods, they are essentially algorithmic in nature and, hence, suffer from the drawback that there is no well-established analytical formulation on the basis of theoretical analysis and performance evaluation [28]. Accordingly, relevant modifications mainly come from case-by-case comparisons conducted empirically. In spite of this, as adaptive signal processing methods, the EMD and the LMD methods are proven to be useful and adaptive signal processing tools for vibration-based fault diagnosis and detection.

Acknowledgements

The work described in this chapter is fully supported by the National Natural Science Foundation of China (Nos. 51405063, 51375078 and 51537010), and China Postdoctoral Science Foundation Funded Project (No. 2016M590874).

Author details

Wei Guo^{1*} and Ming J. Zuo^{1,2}

*Address all correspondence to: guo.w@uestc.edu.cn

1 School of Mechatronics Engineering, University of Electronic Science and Technology of China, Chengdu, China

2 Department of Mechanical Engineering, University of Alberta, Edmonton, Canada

References

- [1] Huang N.E., Shen Z., Long R., Wu C., Shih H., Zheng Q., Yen C., Tung C., Liu H. The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. *Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineer Sciences*. 1998; 454: 903–995.
- [2] Smith J.S. The local mean decomposition and its application to EEG perception data. *Journal of the Royal Society Interface*. 2005; 2(5): 443–454.
- [3] Feldman M. Analytical basics of the EMD: two harmonics decomposition. *Mechanical Systems and Signal Processing*. 2009; 23(7): 2059–2071.
- [4] Wang Y.X., He Z.J., Zi Y.Y. 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—Transactions of the ASME*. 2010; 132(2): 021010 (10 pages).
- [5] Lei Y., Lin J., He Z.J., Zuo M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*. 2013; 35(1–2): 108–126.
- [6] Huang N.E., Shen S.S.P. *Hilbert-Huang Transform and Its Applications*. World Scientific, Singapore; Hackensack, NJ; London; 2005.
- [7] Pai P.F., Palazotto A.N. Detection and identification of nonlinearities by amplitude and frequency modulation analysis. *Mechanical Systems and Signal Processing*. 2008; 22(5): 1107–1132.
- [8] Wu Z., Huang N.E. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*. 2009; 1(1): 1–41.
- [9] Guo W., Huang L.J., Chen C., Zou H.W., Liu Z.W. Elimination of end effects in local mean decomposition using spectral coherence and applications for rotating machinery. *Digital Signal Processing*. 2016; 55: 52–63.
- [10] Lin D.C., Guo Z.L., An F.P., Zeng F.L. Elimination of end effects in empirical mode decomposition by mirror image coupled with support vector regression. *Mechanical Systems and Signal Processing*. 2012; 31(8): 13–28.
- [11] Antoni J. Cyclic spectral analysis in practice. *Mechanical Systems and Signal Processing*. 2007; 21(2): 597–630.
- [12] Wang D., Guo W., Tse P.W. An enhanced empirical mode decomposition method for blind component separation of a single-channel vibration signal mixture. *Journal of Vibration & Control*. 2015; 22(11): 2603–2618.
- [13] Guo W., Tse P.W. A novel signal compression method based on optimal ensemble empirical mode decomposition for bearing vibration signals. *Journal of Sound and Vibration*. 2013; 332(2): 423–441.

- [14] Ren D.Q., Yang S.X., Wu Z.T., Yan G.B. Research on end effect of LMD based time-frequency analysis in rotating machinery fault diagnosis. *China Mechanical Engineering*. 2012; **23**: 951–956.
- [15] Guo W., Tse P.W. Enhancing the ability of ensemble empirical mode decomposition in machine fault diagnosis. In: *Proceedings of Prognostics & Health Management Conference 2010*; 12–14 Jan. 2010; Macau. p. 1–6.
- [16] Antoni J., Randall R.B. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*. 2006; **20**(2): 308–331.
- [17] Guo W., Tse P.W., Djordjevich A. Faulty bearing signal recovery from large noise using a hybrid method based on spectral kurtosis and ensemble empirical mode decomposition. *Measurement*. 2012; **45**(5): 1308–1322.
- [18] Antoni J. Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing*. 2007; **21**(1): 108–124.
- [19] Dybała J., Zimroz R. Rolling bearing diagnosing method based on Empirical Mode Decomposition of machine vibration signal. *Applied Acoustics*. 2014; **77**(3): 195–203.
- [20] Guo W., Huang L.J., Chen C. Adaptive selection of feature signals in local mode decomposition for rotating machinery. In: *Proceedings of The 7th Asia-Pacific International Symposium on Advanced Reliability and Maintenance Modelling (APARM 2016)*; 24–26 August 2016; Seoul, Korea. 2016. p. 1–8.
- [21] Georgoulas G., Loutas T., Stylios C.D., Kostopoulos V. Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition. *Mechanical Systems and Signal Processing*. 2013; **41**(1–2): 510–525.
- [22] Markou M., Singh S. Novelty detection: a review—part 1: statistical approaches. *Signal Processing*. 2003; **83**(12): 2481–2497.
- [23] Polikar R. Ensemble based systems in decision making. *IEEE Circuits & Systems Magazine*. 2006; **6**(3): 21–45.
- [24] Lei Y.G., He Z.J., Zi Y.Y. Application of an intelligent classification method to mechanical fault diagnosis. *Expert Systems with Applications*. 2009; **36**(6): 9941–9948.
- [25] Feldman M. Theoretical analysis and comparison of the Hilbert transform decomposition methods. *Mechanical Systems and Signal Processing*. 2008; **22**(3): 509–519.
- [26] Cheng J.S. Yu D.J. Yang Y. Research on the intrinsic mode function (IMF) criterion in EMD method. *Mechanical Systems & Signal Processing*. 2006; **20**(4): 817–824.
- [27] Rilling G., Flandrin P., Goncalves P. On empirical mode decomposition and its algorithms. In: *Proceedings of IEEE EURASIP Workshop on Nonlinear Signal and Image Processing, Grado (I)*; June; 2003. p. 1–5.
- [28] Delechelle E., Lemoine J., Niang O. Empirical mode decomposition: an analytical approach for sifting process. *IEEE Signal Processing Letters*. 2005; **12**(11): 764–767.