Springer

Journal of Mechanical Science and Technology

Journal of Mechanical Science and Technology 23 (2009) 2000~2007

www.springerlink.com/content/1738-494x
DOI 10.1007/s12206-009-0438-9

Sifting process of EMD and its application in rolling element bearing fault diagnosis †

Hongbo Dong[1], Keyu Qi[2], Xuefeng Chen[1], Yanyang Zi[1], Zhengjia He[3] and Bing Li[1,*]

[1]*School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, PR China*
[2]*Institute of Electromechanical Information Technology, Xi'an 710065, PR China*
[3]*State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, PR China*

(Manuscript Received December 24, 2007; Revised December 8, 2008; Accepted February 23, 2009)

Abstract

Among the vibration-based fault diagnosis methods for rolling element bearing, the shock pulse method (SPM) combined with the demodulation method is a useful quantitative technique for estimating bearing running state. However, direct demodulation often misestimates the shock value of characteristic defect frequency. To overcome this disadvantage, the vibration signal should be decomposed before demodulation. Empirical mode decomposition (EMD) can be an alternative for preprocess bearing fault signals. However, the trouble with this method's application is that it is time-consuming. Therefore, a novel method that can improve the sifting process's efficiency is proposed, in which only one time of cubic spline fitting is required in each sifting process. As a consequence, the time for EMD analysis can be evidently shortened and the decomposition results simultaneously maintained at a high precision. Simulations and experiments verify that the improved EMD method, combined with SPM and demodulation analysis, is efficient and accurate and can be effectively applied in engineering practice.

Keywords: Sifting process; EMD; Demodulation; Shock pulse method; Bearing fault diagnosis

1. Introduction

Bearings are one of the most important and frequently encountered components in rotating machinery. Fault identification of rolling element bearings using condition monitoring techniques has been the subject of extensive research for the last two decades. Vibration-based condition monitoring has been the most utilized technique in this context [1].

Vibration-based analysis is established as the most common and reliable method for nondestructive testing. Defects and wears cause impacts at frequencies governed by the operating speed of the unit and the geometry of the bearings, which in turn are modulated by machine natural frequencies. The signature of a damaged bearing consists of an exponentially decaying ringing that occurs periodically at the characteristic defect frequency [2]. Among the vibration-based fault diagnosis methods for rolling element bearing, the shock pulse method (SPM) combined with the demodulation method is a useful quantitative technique for estimating bearing running state. SPM is a bearing monitoring method developed by the SPM Instrument AB Company; the method provides an experiential formula for judging the bearing's status. With this experiential formula, the bearing status can be quantificationally estimated by a corresponding dB value, which can be calculated from the shock value of the bearing. The shock value is often gained by the demodulation of the bearing vibration signal. However, direct demodulation often misestimates the shock value of characteristic defect frequency. To overcome this disadvantage, the vibration signal should be decomposed before demodulation.

† This paper was recommended for publication in revised form by Associate Editor Eung-Soo Shin

*Corresponding author. Tel.: +86 29 8266 7963, Fax.: +86 29 8266 3689
E-mail address: bli@mail.xjtu.edu.cn

From the past decades, wavelet transform has become one of the fast-evolving mathematical and signal processing tools [3, 4]. The basic operation of wavelet transform involves the processes of dilation and translation, which lead to a multi-scale analysis of the signal. Hence, it can effectively extract both the time and the frequency features of the inspected signal. Although wavelet transform is capable of analyzing non-stationary signals and is deemed suitable for vibration-based machine fault diagnosis, many deficiencies have been reported in its application [5, 6]. A non-adaptive nature is one of its disadvantages. Once the basic wavelet is selected, one will have to use it to analyze all the data. Moreover, the number of levers that the signal will be decomposed to must be decided by the user, and the frequency bands of all lever signals should be fixed as well.

Due to the deficiencies of wavelet transform, Huang et al. [7] proposed a new type of signal processing method called empirical mode decomposition (EMD), with which any complicated data set can be decomposed into a finite and often small number of intrinsic mode functions (IMFs). The IMF is a kind of complete, adaptive, and almost orthogonal representation of the analyzed signal. It can determine all the instantaneous frequencies from the nonlinear or non-stationary signal because it is almost a mono-component. And EMD has been widely used in fault diagnosis [8, 9]. High sampling frequency is generally required in bearing fault diagnosis, and with the defined frequency resolution ratio, it means that a large deal of data will be processed. For long signal processing, the boundary distortion [10] of EMD is negligible, but the time for decomposing becomes a factor that must be considered. However, the trouble with this method's application is that it is time-consuming. Hence, an improved sifting process of EMD is proposed in this paper. With this method, EMD can be implemented using less time and, at the same time, with high precision.

The theories of EMD and SPM are introduced in Section 2. Section 3 proposes the improved sifting process of EMD. Section 4 analyzes the disadvantage of direct demodulation and introduces demodulation based on an improved EMD. Section 5 is an experimental verification of the proposed method, as it is engaged in a bearing fault diagnosis. The conclusions are given in Section 6.

2. Empirical mode decomposition and shock pulse method

Huang et al. [7] presented the use of EMD to decompose any complicated data set into a finite and often small number of IMFs. The IMFs satisfy the following two conditions: (1) in the whole data set, the number of extreme and the number of zero-crossings must either be equal or differ at most by one; and (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima must be zero.

The EMD is developed based on the assumption that any signal consists of many different IMFs. The procedures to decompose a given signal $x(t)$ to different IMFs can be categorized into the following steps. First, identify all the local extrema from the given signal and then connect them with a cubic spline line as the upper envelope. Second, repeat the first step for the local minima to produce the lower envelope. The upper and lower envelopes should cover the entire signal between them. Third, designate their mean as $m_1(t)$ and the difference between the signal $x(t)$ and $m_1(t)$ as the first component $h_1(t)$:

$$h_1(t) = x(t) - m_1(t) \tag{1}$$

Ideally, after the sifting operation of Eq. (1), $h_1(t)$ should be an IMF. The construction of $h_1(t)$ described above seems to have satisfied all the requirements of IMF. However, during the process, overshoots and undershoots may exist and can be classified as the new extrema. A few overshoots and undershoots may shift or exaggerate the existing ones, ultimately distorting the means. Moreover, the envelope mean may be different from the true local mean for the nonlinear signal, which may make $h_1(t)$ asymmetric. To eliminate riding waves and make the wave profiles more symmetric, Huang et al. repeated the sifting process of Eq. (1) as many times as required to reduce the extracted signal to an IMF. Therefore, the fourth step is to repeat the sifting process by treating $h_1(t)$ as the signal and to repeat Eq. (1) as

$$h_{11}(t) = h_1(t) - m_{11}(t) \tag{2}$$

The sifting process will be repeated k times until $h_{1k}(t)$ becomes a true IMF, that is,

$$h_{1k}(t) = h_{1(k-1)}(t) - m_{1(k-1)}(t) \tag{3}$$

then it is designated as

$$c_1(t) = h_{1k}(t) \tag{4}$$

Finally, we obtained the first IMF ($c_1(t)$) from the signal.

Huang et al. also suggested a criterion for stopping the sifting process. This is accomplished by limiting the size of the standard deviation, denoted as SD, which is calculated from two consecutive sifting results as

$$SD = \sum_{t=0}^{N} \left[\frac{\left| h_{1(k-1)}(t) - h_{1k}(t) \right|^2}{h_{1(k-1)}^2} \right] \tag{5}$$

According to Huang et al., the SD value of 0.2-0.3 for the sifting process is a very rigorous limit for the difference between two consecutive siftings.

Generally, $c_1(t)$ should contain a component with the finest scale or the shortest period of the signal. Removing $c_1(t)$ from the rest of the signal by

$$r_1(t) = x(t) - c_1(t) \tag{6}$$

will then give us the residue of signal $r_1(t)$, which contains a component with a longer period than the previous component. By treating $r_1(t)$ as a new signal and repeating the same sifting process as described above, we can then obtain the second IMF ($c_2(t)$). Similarly, we can obtain a series of IMFs $c_i(t)\ (i = 1, 2, \cdots, n)$ and the final residue $r_n(t)$. The sifting process can be stopped by any of the following predetermined criteria: either when the component $c_n(t)$ or the residue $r_n(t)$ becomes less than the predetermined value of substantial consequence, or when the residue $r_n(t)$ becomes a monotonic function from which no further IMFs can be extracted. Summing up all the IMFs and the final residue, we should be able to reconstruct the original signal $x(t)$ by

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{7}$$

SPM is a bearing monitoring method developed by the SPM Instrument AB Company. It provides an experiential formula for judging the status of bearing. According to SPM, a standard dB value is defined as Eq. (8). This value is used to estimate the bearing running state.

$$dB = 20 \log \frac{2000 \times SV}{N \times D^{0.6}} \tag{8}$$

where N represents the rotating speed of the rotor assembled in bearing; D is the inner diameter of the bearing; and SV is the so-called shock value in SPM (Actually, it is used for expressing the severity degree of bearing vibration with velocity unit). In addition, SV can be easily calculated using the acceleration signal of bearing vibration.

With the dB value, the bearing running state can be estimated as follows [11]:

(1) $0 \le dB < 20$ normal state
(2) $20 \le dB \le 35$ light fault
(3) $35 < dB \le 60$ serious fault

The above three condition zones were established with different dB values. The bounds of the three condition zones were empirically established by testing the bearings under variable operating conditions in the early 1970s. For over 35 years, based on these three condition zones, the SPM has been very successfully used to obtain a fast, easy, and reliable diagnosis of the operating conditions of rolling element bearings [11].

3. Improved sifting process in EMD

Several sifting processes are required to produce each IMF, and two instances of cubic spline fitting will be implemented for each sifting process. However, cubic spline fitting is a time-consuming process that leads to inefficient signal analysis. Therefore, a novel method that can improve the sifting process's efficiency is proposed, with which the time for EMD analysis can be evidently shortened.

First, k extrema in the analyzed signal $x(t)$ is identified and marked as $e(t_i)$, where $i = 1, 2, \cdots, k$, and t_i present the time location of the i th extremum. Then, the mean value between $e(t_i)$ and $e(t_{i+1})$ is calculated by Eq. (9) and named m_i ($i = 1, 2, \cdots, k-1$).

$$m_i(t_{\xi i}) = \frac{1}{t_{i+1} - t_i + \Delta t} \sum_{t=t_i}^{t_{i+1}} x(t) \tag{9}$$

where $t_{\xi i} \in [t_i, t_{i+1}]$ and Δt represent the time interval of the signal.

Obviously, the signal values between two extrema are monotone. Therefore, the mean value m_i intersects with the signal at one point and the corresponding time location is $t_{\xi i}$, as shown in Fig. 1. Generally, suppose that $t_{\xi i}$ is between $x(t_j)$ and $x(t_{j+1})$, then $t_{\xi i}$ can be obtained by Eq. (10).

$$t_{\xi i} = t_j + \frac{\left| m_i - x_j \right| \times (t_{j+1} - t_j)}{\left| x_{j+1} - x_j \right|} \quad \text{and} \quad t_{\xi i} \in [t_j, t_{j+1}] \tag{10}$$

Actually, Eq. (10) is achieved by a similar triangle

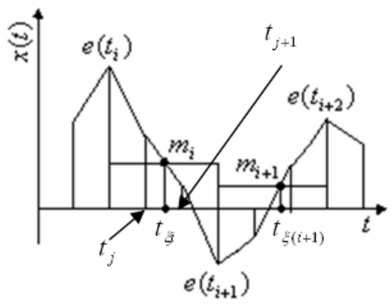

Fig. 1. Signal, extrema, and mean value.

principle, where $x(t_j)$ and $x(t_{j+1})$ are two seriate points between two extrema, and t_j and t_{j+1} are their time locations, respectively.

With the mean value m_i and its time location $t_{\xi i}$ ($i=1,2,\cdots,k-1$), the mean of the analyzed signal can be obtained with one instance of cubic spline fitting. Likewise, the latter process is the same as the primal EMD. Compared with primal EMD, the proposed method only needs one instance of cubic spline fitting in each sifting process. As a consequence, the time consumed by signal decomposition is reduced.

Furthermore, the improved sifting process has an advantage in calculating precise local mean, which reduces the sifting process number for producing each IMF.

Consider the simulative signal shown in Eq. (11).

$$x(t)=x_1(t)+x_2(t)=\sin(2\pi f_1 t)+\sin(2\pi f_2 t) \qquad (11)$$

The simulative signal is composed by two sine waves, $x_1(t)$ and $x_2(t)$, where $f_1=100\text{Hz}$ and $f_2=200\text{Hz}$, respectively. Sampling frequency is 2000 Hz, and 1024 points are sampled. Fig. 2 shows the signal and its spectrum; for the simulative signal, amplitude A does not have a unit. For this simulative signal, the high frequency component $x_2(t)$ should be decomposed as the first IMF, and the low frequency components $x_1(t)$ should be subtracted in the sifting process (i.e., the theoretic local mean of the simulative signal must be the low frequency sine wave $x_1(t)$).

The original simulative signal, theoretic local mean, local mean calculated by original EMD, and improved sifting method are shown in Fig. 3, where a, b, c, and d represent the original simulative signal, theoretic local mean, local mean calculated by original EMD, and improved sifting method, respectively. It is obvious in Fig. 3 that the local mean calculated by an improved sifting method is closer to the theoretic

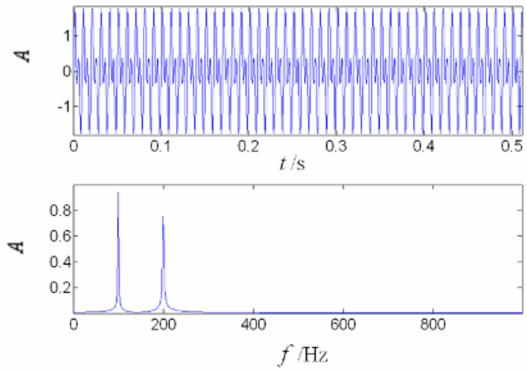

Fig. 2. The simulative signal and its spectrum.

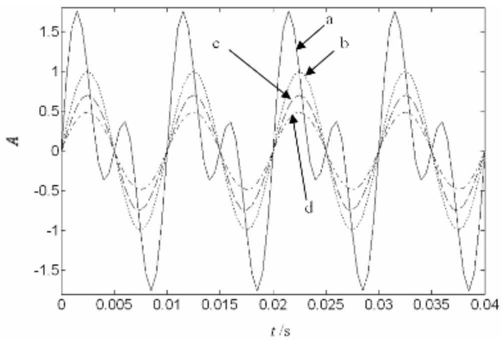

Fig. 3. Local mean calculated by EMD and the improved method.

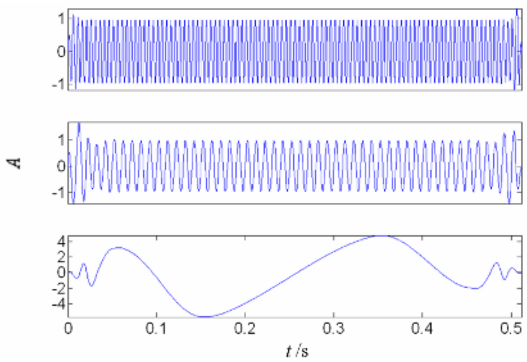

Fig. 4. Results using the original EMD.

local mean, which means that less sifting processes are required for producing the IMF. Less sifting processes will also lead to time saving for signal decomposition.

With the original EMD, the simulative signal is decomposed into two IMFs and a residual as shown in Fig. 4. The decomposing process consumes 2.8 s. Obviously, the two IMFs present the two sine waves in a simulative signal, respectively. Due to the fitting

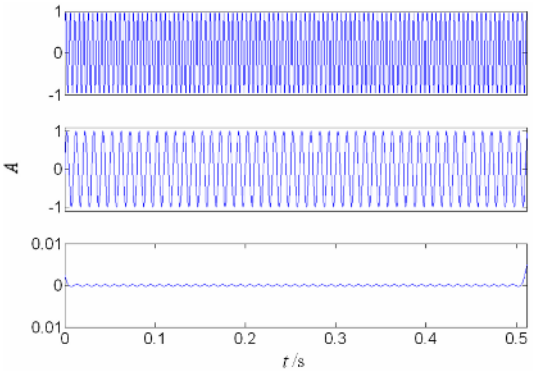

Fig. 5. Results using the improved EMD.

error, an evident boundary distortion appears, and the residual has a large energy.

With the proposed method, the simulative signal is also decomposed into two IMFs and a residual, as shown in Fig. 5. Using the same computer hardware and software, the time consumed by the proposed method is notably shortened to 0.3 s. Boundary distortion barely appears in the obtained IMFs, and the residual has relatively little energy. These results indicate that the proposed method can decompose signals with high efficiency and precision.

4. Demodulation analysis-based improved EMD

Hilbert demodulation is used in this paper to process the signal and because the amplitude modulation phenomenon existing in the signal can be obviously represented by this technique.

Consider the simulative signal defined by Eq. (12).

$$x(t) = [1 + 0.2\sin(2\pi f_1)]\sin(2\pi f_{c1}) + [1 + 0.2\sin(2\pi f_2)]\sin(2\pi f_{c2}) \tag{12}$$

where two modulation frequencies are defined as $f_1 = 60\text{Hz}$ and $f_2 = 80\text{Hz}$. The two modulation waves are modulated by two carry waves, respectively, and are defined as $f_{c1} = 2000$ Hz and $f_{c2} = 4000$ Hz. Obviously, the simulative signal is composed of two amplitude modulation components. This signal is defined to represent the case where more than one fault appears in a bearing system. Sampling frequency is 12800 Hz, while 8K points are sampled. The modulation signal and its spectrum are shown in Fig. 6

The direct Hilbert demodulation results are shown in Fig. 7. The top figure shows the demodulation signal, while the bottom figure presents the demodulation spectrum.

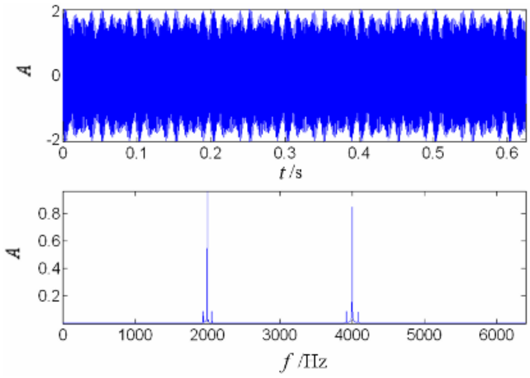

Fig. 6. The modulation signal and its spectrum.

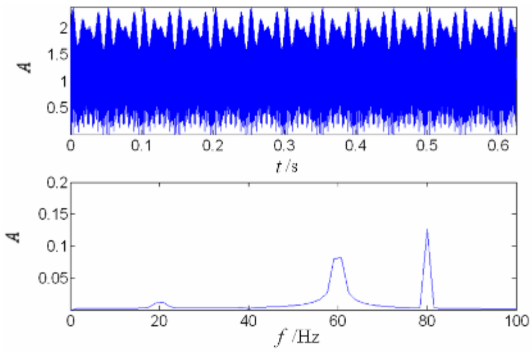

Fig. 7. Direct Hilbert demodulation results.

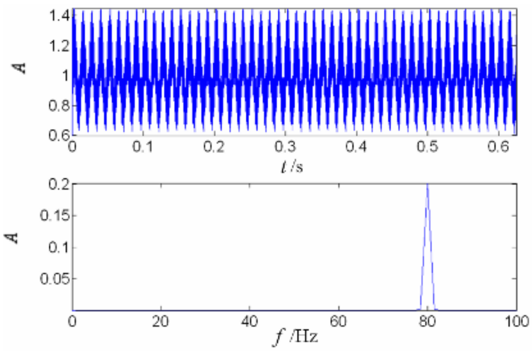

Fig. 8. Demodulation results of the first IMF.

Theoretically, the demodulation spectrum should have two peaks with values that are equal to 0.2; however, the peak values in Fig. 7 are all less than 0.15. This indicates that in the case of more than one modulation frequency appearing in the signal, the demodulation results directly using Hilbert demodulation is not satisfied.

To overcome this disadvantage, the signal should be decomposed with EMD before Hilbert demodulation. With EMD, two IMFs are obtained representing

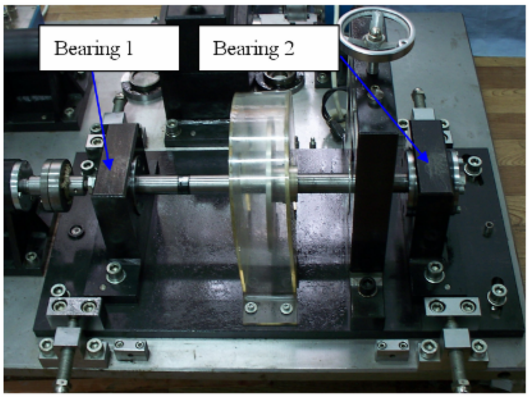

Fig. 9. Experimental setup.

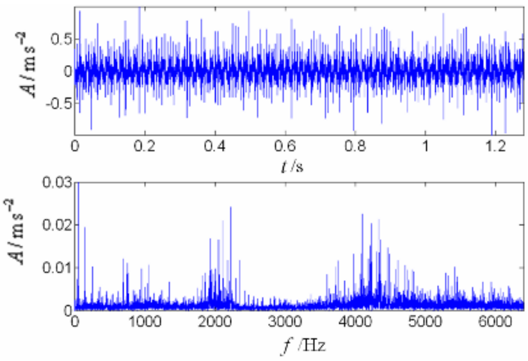

Fig. 10. Vibration signal and spectrum of a bearing with an inner race fault.

the two modulation components in a simulative signal, respectively. The decomposed results using an improved EMD are similar with those of the original EMD, except that the improved EMD consumes less time. The first IMF obtained by improved EMD is demodulated, and the results are shown in Fig. 8. The obtained modulation frequency is 80 Hz, and the peak value is equal to 0.2. Furthermore, the demodulation results of the second IMF are similar to those of the first IMF, except that the modulation frequency is 60 Hz. Through EMD processing, the demodulation results are uniform with the theoretical results. For a long signal, original EMD is a time-consuming procedure; nevertheless, with improved EMD, the time for decomposition is acceptable.

5. Experimental verification

To verify the utility of the proposed method, an experiment is implemented on the bearing test rig. Fig. 9 shows the experimental setup. Tow bearings named

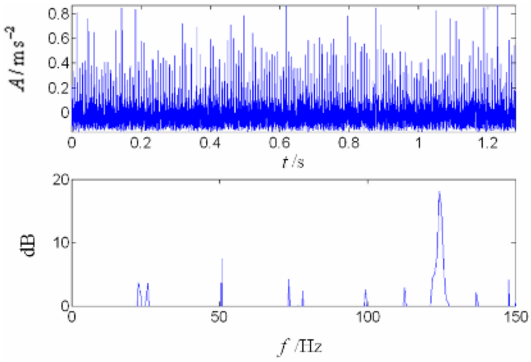

Fig. 11. Direct demodulation results.

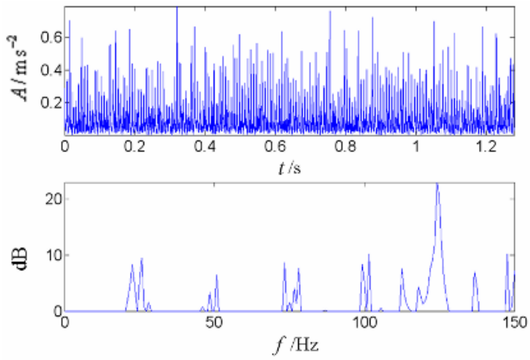

Fig. 12. Demodulation results based on the improved EMD.

bearing 1 and bearing 2 support the rotor. Bearing 2 has a slight damaged on its inner race, while bearing 1 has an outer race fault. The vibration of the bearing with an inner ring fault is sampled by acceleration sensors with a sampling frequency of 12800 Hz. A total of 16 K points is sampled. The inner race fault's characteristic frequency is 122.11 Hz. Fig. 10 shows the vibration signal and its spectrum.

Demodulation results directly using Hilbert demodulation are shown in Fig. 11. The shock value of this bearing can be gained with demodulation, and with Eq. (8) this shock value has been converted into dB values. The dB value will reach its peak in the corresponding frequency if any faults exist in the bearing. It is obvious that the peak value presents itself in 122 Hz, and it is just equal to the inner fault characteristic frequency, which means that this peak value corresponds to the inner fault. The obtained peak dB value is 18.15 dB, which is less than the alarm limit for the light fault (20 dB). Obviously, with direct Hilbert demodulation, the light fault existing in the bearing is improperly ignored. This indicates that

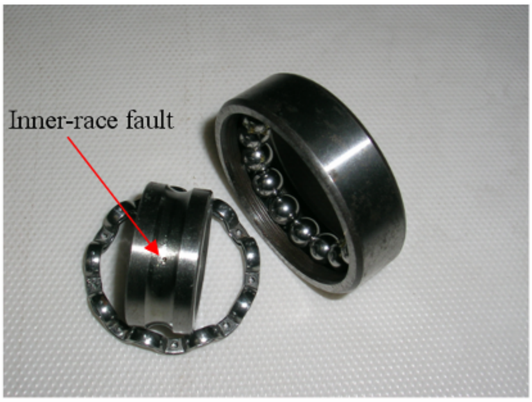

Fig.13. Inner race fault of the bearing.

in the case of more than one fault appearing in a bearing system, the demodulation results directly using Hilbert demodulation are not satisfied.

Next, the original signal is processed by the improved EMD. The demodulation results of the first IMF are shown in Fig. 12. The obtained peak dB value is 22.12, which is greater than the alarm limit. Thus, the existing fault is identified successfully.

6. Conclusions

In the EMD method, several sifting processes are required to produce each IMF, with two instances of cubic spline fitting to be implemented for each sifting process. However, cubic spline fitting is a time-consuming process that leads to an inefficient signal analysis. Therefore, a novel method that can improve the sifting process's efficiency is proposed, with which the time for EMD analysis can be evidently shortened while maintaining high precision results at the same time. With the improved EMD method and demodulation, the amplitude modulation phenomenon existing in the bearing fault signal can be accurately represented. Simulations and experiments verify that the improved EMD method, combined with SPM and demodulation analysis, is efficient and accurate in bearing fault diagnosis. The proposed method can also be effectively applied in engineering practice.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (50875195, 50805114) and the Foundation for the Author of National Excellent Doctoral Dissertation of China (2007B33). The authors are grateful for their financial support.

Reference

[1] V. Purushotham, S. Narayanan and A. N. Suryanarayana, Prasad, Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition, *NDT E Int.* 38 (2005) 654-664.

[2] N. G. Nikolaou and I. A. Antoniadis, Rolling element bearing fault diagnosis using wavelet packets, *NDT E Int.* 35 (2002) 197-205.

[3] N. H. Abu-Zahra and G. Yu, Gradual wear monitoring of turning inserts using wavelet analysis of ultrasound waves, *Int. J. Mach. Tools Manuf.* 43 (2003) 337-343.

[4] Cheng Junsheng, Yu Dejie and Yang Yu, Time–energy density analysis based on wavelet transform, *NDT E Int.* 38 (2005) 569-572.

[5] P. Tse, W.X. Yang and H.Y. Tam, Machine fault diagnosis through an effective exact wavelet analysis, *J. Sound Vibr.* 277 (2004) 1005-1024.

[6] Z. K. Peng, Peter W. Tse and F. L. Chu, An improved Hilbert–Huang transform and its application in vibration signal analysis, *J. Sound Vibr.* 286 (2005) 187-205.

[7] N. E. Huang, Z. Shen, S. R. Long, et al, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. Lond.* 454 (1998) 903-995.

[8] Dejie Yu, Junsheng Cheng and Yu Yang, Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, *Mech. Syst. Signal Proc.* 19 (2005) 259–270.

[9] Guanghong Gai, The processing of rotor startup signals based on empirical mode decomposition, *Mech. Syst. Signal Proc.* 20 (2006) 222-235.

[10] Keyu Qi, Zhengjia He, Yanyang Zi and Xuefeng Chen, Cosine Window Based Boundary Processing Method for EMD and Its Application in Rubbing Fault Diagnosis, *Mech. Syst. Signal Proc.* 21 (2007) 2750-2760.

[11] D. E. Butler, The shock pulse method for the detection of damaged rolling bearing, *NDT E Int.* 6 (1973) 92-95.

Hongbo Dong was born in Chaoyang, China, in 1979. He received the B.E. and M.E. degree from Northwestern Polytechnical University in Mechanical Engineering in 2002 and 2005 respectively and received the Ph.D degree from Xi'an Jiaotong University in Mechanical Engineering in 2009. His research interests include fault diagnosis of rotor and bearing system.

Bing Li was born in Xuzhou, China, in 1976. He received the B.E. and M.E. degree from Northwestern Polytechnical University in Mechanical Engineering in 1999 and 2002 respectively and received the Ph.D degree from Xi'an Jiaotong University in Mechanical Engineering in 2005. After graduating from Xi'an Jiaotong University, he works as a lecturer in Xi'an Jiaotong University. His research interests include wavelet finite element theory and its application in fault diagnosis.