Bearings Fault Diagnosis Using Vibrational Signal Analysis by EMD Method

Majid Norouzi Keshtan & Mehrdad Nouri Khajavi

To cite this article: Majid Norouzi Keshtan & Mehrdad Nouri Khajavi (2016) Bearings Fault Diagnosis Using Vibrational Signal Analysis by EMD Method, Research in Nondestructive Evaluation, 27:3, 155-174, DOI: 10.1080/09349847.2015.1103921

To link to this article: http://dx.doi.org/10.1080/09349847.2015.1103921

Accepted author version posted online: 23 Oct 2015.
Published online: 23 Oct 2015.

Submit your article to this journal

Article views: 67

View related articles

View Crossmark data
Bearings Fault Diagnosis Using Vibrational Signal Analysis by EMD Method

Majid Norouzi Keshtan and Mehrdad Nouri Khajavi

ABSTRACT
Studying vibrational signals is one reliable method for monitoring the situation of rotary machinery. There are various methods for converting vibrational signals into usable information for fault diagnosis, one of which is the empirical mode decomposition method (EMD). This article is about diagnosing bearing faults using the EMD method, employing nondestructive test. Vibration signals are acquired by a bearing test machine. The discrete wavelet bases are used to translate vibration signals of a roller bearing into time-scale representation. Then, an envelope signal can be obtained by envelope spectrum analysis of wavelet coefficients of high scales. Local Hilbert marginal spectrum can be obtained by applying the EMD method to the envelope signal from which the faults in a roller bearing can be diagnosed and fault patterns can be identified. The results have shown bearing faults frequencies are easily observable. There is a variant of the EMD method called the ensemble EMD (EEMD), which overcomes the mode mixing problem which may occur when the signal to be decomposed is intermittent. The EEMD method is also applied to the acquired signals, and the two methods were compared. While the outcomes of both methods do not differ much, one important merit of the EMD is that it has much less computational processing time than EEMD.

KEYWORDS
Nondestructive test; empirical mode decomposition; Hilbert transformation; EEMD

1. Introduction
Fault diagnosis techniques have been developed to prevent human and financial losses and to increase productivity and production rate. Rolling element bearings are one of the most important and most used parts in rotating machineries, and they are the focus of nondestructive test for fault diagnosis. Rolling element bearings faults can be due to different factors such as wrong design or wrong mount, improper lubrication, plastic deformation, etc. The most common fault is due to material fatigue after some work period. This phenomenon starts with small cracks developing on the surface.
of bearing elements. Due to fluctuating loads, these cracks grow to the surface and cause the piercing or fracture of the surface. In this research, crack faults in inner race, outer race, and balls are considered. There are many techniques which can be used for bearing fault diagnosis. Vibration analysis is one of the most common methods of fault diagnosis and detection. Vibrations analysis can be classified into three categories, namely, time-domain, frequency-domain, and time-frequency domain classes. Time-domain and frequency domain analysis is used for processing stable and linear signals; these two techniques give incorrect results when applied to nonstationary and nonlinear signals such as mechanical faults [1].

Frequency-time methods show a frequency distribution in different time intervals of signals and, therefore, are capable of detecting local changes in signals. Simultaneous display of signal in two time-frequency domains makes the analysis of unstable signals possible. EMD is one of the time-frequency methods. This method is based on the hypothesis that each signal includes various simple sinusoidal waves.

EMD decomposes any complicated signal into a collection of intrinsic mode functions (IMFs) based on the local time scale characteristic of the signal. The IMFs represent the natural oscillatory mode embedded in the signal. EMD is self-adaptive because the IMFs, working as the basis functions, are determined by the signal itself rather than by what is predetermined. Therefore, the EMD is highly efficient in nonstationary data analysis.

Some applications of EMD in mechanical fault diagnosis have been studied, for instance, in structural health monitoring [2,3], beam crack detection [4], and gear fault diagnosing [1,5]. Other corresponding articles are as follows. Bahar and Ramezani [6] proposed a new enhanced Hilbert-Huang transform (HHT) to avoid mathematical limitations of the Hilbert spectral analysis; an additional parameter is employed to reduce the noise effects on the instantaneous frequencies of IMFs. The efficacy of the proposed method is demonstrated based on two case studies: a typical 3-DOF model subjected to a random excitation and a real 15-story building under ambient excitation.

Dong et al. [7] proposed a method based on the EMD and the vector autoregressive moving average (VARMA) model for structural damage detection.

Esmaeel and Taheri [8] introduced an EMD-based nondestructive damage detection methodology for detecting delamination damage in composite beams.

Roveri and Carcaterra [9] developed a novel HHT-based method for damage detection of bridge structures with cracks under a traveling load. The technique uses a single point measurement and is able to identify the presence and the location of the damage along the beam. Theoretical as well as numerical results show that the identification is rather accurate and
detection quality is not very sensitive to the crack depth and ambient noise, while they are sensibly affected by the damage location and by the speed of the moving load as well.

Lin et al. [10] developed a signal processing method for the impact-echo test based on the HHT. Numerical simulations and model tests show that the proposed method is promising in the detection of internal cracks in concrete even when the vibration and noise signals are strong.

Major difficulty of fault detection in the initial stage of fault development in bearing is that vibrations resulting from fault have low amplitude and cannot be distinguished from vibrations of other parts and inherent noise of system. As defects in bearing expand, total energy in the spectrum will increase in a broader frequency range. Although in theory simple fast fourier transform (FFT) analysis or wavelet can detect different bearing faults by detection of peaks at the known bearing elements frequency, it is not possible to distinguish these peaks due to the presence of noise and of very low amplitudes of peaks at the early stages of defect development in practice. To solve this problem, EMD analysis which detects impulse faults has been used. Due to emergence of faults in high frequencies and to remove noises of vibrational signals resulting from low frequencies, discrete wavelet transform has been used. The use of this method leads to detection of fault frequencies of bearings with high resolution, and this capability is the main advantage of this method over other methods.

In the current research, the EMD method has been used as a fault diagnosis tool. This research has been divided into three sections:

- Data acquisition from different bearings including bearing with the outer race, inner race, and ball defects as well as a healthy bearing.
- IMFs can be acquired by applying EMD to the signal.
- Choose some special IMFs to obtain local Hilbert marginal spectrum from which the faults in a roller bearings can be diagnosed and fault patterns can be identified.

1.1. The EMD method

The EMD method is based on the assumption that any signal consists of different simple intrinsic modes of oscillations. In this way, each signal could be decomposed into a number of IMFs, each of which must satisfy the following definition [11]:

1. In the entire data set, the number of extremes and the number of zero crossings must either be equal or differ at most by one;
2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.
An IMF represents a simple oscillatory mode compared with the simple harmonic function. With the definition, any signal \( x(t) \) can be decomposed as follows [11]:

1. Identify all the maxima of the signal, and connect all the local maxima by a cubic spline as the upper envelope. Repeat the procedure on the local minima to produce the lower envelope.
2. Designate the mean of the two envelopes as \( m_1 \), and the difference between the signals \( x(t) \) and \( m_1 \) as the first component, \( h_1 \), i.e.,

\[
x(t) - m_1 = h_1.
\]

If \( h_1 \) is an IMF, take it as the first IMF of \( x(t) \). If \( h_1 \) is not an IMF, take it as the original signal and repeat the steps above until \( h_{1k} \) is an IMF, and designate \( h_{1k} \) as \( c_1 \):

\[
c_1 = h_{1k}.
\]

3. Separate the first IMF \( c_1 \) from \( x(t) \) by

\[
x(t) - c_1 = r_1.
\]

4. Treat residue \( r_1 \) as the original signal, and subject it to the same process as above, so that we can get other IMFs, \( c_2, c_3, \ldots, c_n \), which satisfy

\[
r_1 - c_2 = r_2
\]
\[
\vdots
\]
\[
r_{n-1} - c_n = r_n.
\]

5. By summing up Eqs. (3) and (4), we finally obtain

\[
x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t).
\]

Then, the signal \( x(t) \) is decomposed into \( n \) intrinsic modes and a residue \( r_n \).

(When computing upper envelope and lower envelope functions, the type of interpolation scheme is usually important. In this article, cubic spline has been used for interpolation on sampled data, as was proposed in [11].) To guarantee that the IMF components retain enough physical sense, we must determine a criterion for the sifting process to stop. The criterion is as follows:
\[ SD_i = \sum_{k=0}^{T_i} \frac{|h_{i-1}(k) - h_i(k)|^2}{h_{i-1}(k)^2} \]  \[ (6) \]

In this article, we take \( SD_i < 0.3 \).

### 1.2. The EEMD method

When a signal contains intermittency, the EMD algorithm described above may encounter the problem of mode mixing. Frequent appearance of mode mixing is 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. The intermittence could not only cause serious aliasing in the time-frequency distribution, but also make the individual IMF devoid of physical meaning \[12\]. To overcome this limitation, a new noise assisted data analysis (NADA) method is proposed, i.e., the EEMD.

EEMD algorithm is as follows \[12\]:

(a) Initialize the number of ensemble I.
(b) Generate \( X^i[t] = X[t] + W^i[t] (i = 1, \ldots, I) \) are different realization of white Gaussian noise.
(c) Each \( X^i[t] (i = 1, \ldots, I) \) is fully decomposed by EMD getting their modes \( IMF^i_k[t] \), where \( k = 1, 2, \ldots, K \) indicates the modes.
(d) Assign \( \overline{IMF}_k \) as the kth mode of \( X[t] \) obtained as the average of the corresponding

\[ \overline{IMF}_k[t] = \frac{1}{I} \sum_{i=1}^{I} IMF^i_k[t]. \]  \[ (7) \]

Just like in the EMD method, the given signal, \( X(t) \), can be reconstructed according to the following equation:

\[ X(n) = \sum_{k=1}^{K} \overline{IMF}_k(t) + \bar{r}(t), \]  \[ (8) \]

where

\[ \overline{IMF}_k[t] = \frac{1}{I} \sum_{i=1}^{I} IMF^i_k[t] \]

\[ \bar{r}(t) = \frac{1}{I} \sum_{i=1}^{I} r_i(t) \]  \[ (9) \]

The EEMD described here employs all the important characteristics of noise. Its principle is simple: when a collection of white noises are added to the target signal, they cancel each other out in a time space ensemble mean.
The reason is, obviously, that the added white noise would populate the whole time-frequency space uniformly with the constituting components of different scales separated by the filter bank.

2. The local Hilbert marginal spectrum

For one IMF $c_i(t)$ in Eq. (5), we can always have its Hilbert transform as

$$H[c_i(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{c_i(t')}{t - t'} dt'.$$

(10)

With this definition, we can have an analytic signal as

$$z_i(t) = c_i(t) + jH[c_i(t)] = a_i(t)e^{j\varphi_i(t)},$$

(11)

in which

$$a_i(t) = \sqrt{c_i^2(t) + H^2[c_i(t)]}$$

(12)

$$\varphi_i(t) = \arctan \frac{H[c_i(t)]}{c_i(t)}.$$  

(13)

From Eq. (13), we can have the instantaneous frequency as

$$\omega_i(t) = \frac{d\varphi_i(t)}{dt}.$$  

(14)

After performing the Hilbert transform to each IMF component, the original signal can be expressed as the real part (RP) in the following form:

$$x(t) = RP \sum_{i=1}^{n} a_i(t)e^{j\varphi_i(t)} = RP \sum_{i=1}^{n} a_i(t)e^{j\int^{\omega_i}(t)}dt.$$  

(15)

Here we left out the residue $r_n$ on purpose, for it is either a monotonic function or a constant. Equation (15) gives both amplitude and frequency of each component as functions of time. This frequency-time distribution of the amplitude is designated as the Hilbert spectrum $(\omega,t)$:

$$H(\omega,t) = RP \sum_{i=1}^{n} a_i(t)e^{j\int^{\omega_i}(t)}dt.$$  

(16)

With the Hilbert spectrum defined, we can also define the marginal spectrum, $h(\omega)$, as

$$h(\omega) = \int_0^T H(\omega,t)dt,$$

(17)

where $T$ is the total data length. The Hilbert spectrum offers a measure of amplitude contribution from each frequency and time, while the marginal spectrum offers a measure of the total amplitude contribution from each frequency.
After performing the Hilbert transform on some of the IMF components that we are interested in, we can also define the local Hilbert spectrum by

\[
H'_{0}(\omega, t) = \text{Re}\left(\ldots + a_i(t)e^{j\int_0^t \omega_i(t)dt} + \ldots + a_k(t)e^{j\int_0^t \omega_k(t)dt} + \ldots \right),
\]

and the local Hilbert marginal spectrum by

\[
h'_0(\omega) = \int_0^T H'(\omega, t)dt.
\]

The $H'(\omega, t)$ offers a measure of the amplitude contribution from each time and some frequencies, while the local marginal $h'_0(\omega)$ spectrum offers a measure of the total amplitude contribution from some frequencies that we are interested in.

### 3. Experimental data acquisition

**Figure 1** shows the setup used in data acquisition. Our data is in the form of vibration velocity signals. Our experimental equipment constitute of the following components:

1. An electric motor with 0.5 hp.
2. A fixture for connecting electric motor to shaft with bearing.
3. A load mechanism which puts force on the shaft.
4. Pulse 4 channel data acquisition system manufactured by B&K Company.
5. LS 600 inventor to control motor revolution.

The pulse multichannel data acquisition system has four input channels and two output channels. By connecting one of its input channels to laser Vibrometer, vibration velocity signals are recorded. For simulation of working condition of the bearing to be more realistic, the bearings on the test are subjected to load by the load mechanism shown in **Fig. 1**. The loading mechanism consists of a pneumatic actuator capable of inserting 176.58 N load on bearing. The bearings used in this test are double row bearing model 1206 K, with the spec given in **Table 1**. The bearing component fault frequencies at 1,800 rpm given in **Table 2**. The bearing faults frequencies are given by the following equations [13]:

\[
f_c = \frac{1}{2} f_s \left( 1 - \frac{D_b \cos \theta}{D_c} \right)
\]
In the above equations, \( f_s \) is the shaft rotational frequency, \( f_c \) is the cage rotational frequency, \( f_{BPI} \) is the inner race frequency, and \( f_{BPO} \) is the outer race frequency.
race frequency. Also, $D_b$ is ball diameter, $D_c$ is the cage diameter (the distance from the center of the one ball to the center of the opposite ball). $Z$ is the number of balls, and $\theta$ is the contact angle of the balls with the race. The number of balls is 14, and $\theta$ is zero degree.

Specific faults are made deliberately on the races. In the outer race, a groove with 3 mm depth and 2 mm width is made on the entire width of outer race. In the second bearing, a groove with 3 mm depth and 2 mm width is made on the inner race. For making a fault on the balls, since usually a group of balls gets faulty at the same time, some holes are developed simultaneously on four adjacent balls, as shown in Fig. 2.

Vibration signals from a 1,800 RPM rotating shaft with a sampling frequency of 16 kHz are gathered for a period of 2 min with constant load. It should be noted that during the test the outer race of the bearing was fixed and rotation was applied to the inner race. A sample of acquired signals is shown in Fig. 3.

Figure 2. The figure of the ball bearing with the (a) outer race defect, (b) inner race defect, and (c) ball fault.
In this section, acquired signals are processed by the EMD method. The process of the EMD method is as follows:

1. In order to decrease the influence of lower frequency noise, acquired signals are preprocessed using discrete wavelet analysis. Signal decomposition by wavelet analysis is done by two filters, namely: low pass filter and high pass filter. In this technique, the signal passes these two filters and will be decomposed into two signals, one constituting high frequencies (detail) and the other constituting low-frequency (approximation) of the original signal. The same procedure is applied to the signal with low-frequency (approximation). Filtering is done with the convolution of the signal and filter. Then the data in the decomposed signal is down sampled:

\[
f(t) = \sum_{i=1}^{i=j} D_i(t) + A_j(t),
\]

in which \(D_i(t)\) is the details and \(A_j(t)\) is the approximation. Choice of the wavelet function depends on the specific problem being analyzed. In fault diagnosing and monitoring, Daubechies functions (DBN), in which N is the
order of the DBN, has often been used in research. In this article, DB10 is chosen as the wavelet function for signal decomposition.

(1) According to Eq. (10), $H[D_i(t)]$ is acquired by applying Hilbert transform to $D_i(t)$ (details signal), and then the envelope signal $B(t)$ based on wavelet coefficients is obtained as

$$B(t) = \sqrt{D_i^2(t) + H^2[D_i(t)]}.$$  \hspace{1cm} (25)

(2) $B(t)$ is decomposed by using the EMD method, and each signal is decomposed into a sum of IMFs. Figures 4–7 demonstrate the decomposed signals.

Figure 4. The empirical mode decomposition of the signal of a roller bearing with inner-race fault.
(3) Hilbert transform is applied to all IMF components according to Eqs. (10)–(14). Figure 8 shows the flow chart of the process of local Hilbert marginal spectrum approach based on the envelope signal of wavelet coefficients.

(4) According to the bearing faults frequencies, useful IMFs are separated, and bearings fault will be detected by obtaining marginal spectrum of separated IMFs components which are shown in Figs. 9–12.

When there is fault over the inner race of bearing, it has negligible amplitude in spectrum because vibration caused by this defect should be transmitted through roller elements and then outer race, and it will be
attenuated by travelling this path. Figures 9, 10, and 11 show that, by employing the proposed method, the peak of the different bearing elements fault are detected very clearly.

5. EEMD results

The procedure explained in Section 4 is implemented on acquired signals with the exception that the decomposition method was changed from EMD to EEMD. Results from this section are showed in Figs. 13–18. As can be seen from these figures, the results from the two methods are very close, but the advantage of the EMD method over the EEMD is its less computational time.
5. Conclusion

In this research, the analysis of vibrational signals by the EMD and EEMD methods was used to detect bearing faults. In the proposed method, the vibration signal of a roller bearing with faults is translated into time-scale representation by using the discrete wavelet bases. Hilbert transform is then used to make an envelope analysis of wavelet coefficients of high scales that represent the high-frequency components. By applying EEMD, EMD, and Hilbert transform to signal, we can obtain local marginal spectrum by which the faults in a roller bearing can be diagnosed and fault patterns can be identified. The results
revealed that both signal decomposition methods have an appropriate performance in bearings fault diagnosis, but one advantage of EMD over EEMD is its much less computation time.

Figure 8. The flow chart of the proposed method.
Figure 9. The local marginal spectrum of the signal of a roller bearing with inner-race fault.

Figure 10. The local marginal spectrum of the signal of a roller bearing with outer-race fault.

Figure 11. The local marginal spectrum of the signal of a roller bearing with ball fault.
Figure 12. The local marginal spectrum of the signal of a good roller bearing.

Figure 13. The ensemble empirical mode decomposition of the signal of a roller bearing with inner-race fault.
Figure 14. The ensemble empirical mode decomposition of the signal of a roller bearing with outer-race fault.
Figure 15. The ensemble empirical mode decomposition of the signal of a roller bearing with ball fault.

Figure 16. The local marginal spectrum of the signal of a roller bearing with inner-race fault (EEMD).
References