Rolling bearing diagnosing method based on Empirical Mode Decomposition of machine vibration signal

Jacek Dybala, Radosław Zimroz

Rolling element bearings, also known as rolling bearings, are widely used in rotary machinery systems. Rolling bearings fall out of service for various reasons, such as unexpected heavy loads, unsuitable or inadequate lubrication and ineffective sealing. The components that often fail in rolling bearings are the rolling elements, the inner race and the outer race. Rolling bearings’ diagnostics is important for guaranteeing machine safety and production efficiency. The damage of a bearing may cause the breakdown of a rotary machine, leading to serious consequences. One of the key issues in rolling bearing diagnostics is to detect the defect at its early stage and alert the machine operator before it develops into a catastrophic damage. Contrary to oil condition and thermal monitoring, the vibration analysis of the empirically determined local amplitude is used. To validate the proposed method, raw vibration signals generated by complex mechanical systems employed in the industry (driving units of belt conveyors), including normal and fault bearing vibration data, are used in two case studies. The results show that the proposed rolling bearing diagnosing method can identify bearing faults at early stages of their development.

1. Introduction

Rolling element bearings, also known as rolling bearings, are widely used in rotary machinery systems. Rolling bearings fall out of service for various reasons, such as unexpected heavy loads, unsuitable or inadequate lubrication and ineffective sealing. The components that often fail in rolling bearings are the rolling elements, the inner race and the outer race. Rolling bearings’ diagnostics is important for guaranteeing machine safety and production efficiency. The damage of a bearing may cause the breakdown of a rotary machine, leading to serious consequences. One of the key issues in rolling bearing diagnostics is to detect the defect at its early stage and alert the machine operator before it develops into a catastrophic damage. Contrary to oil condition and thermal state monitoring methods that detect damages of bearings at very late stages of their development (close to catastrophic stages), vibroacoustic analysis detect most of the damages yet at much earlier stages of bearings’ technical degradation.

Rolling bearing is a complex vibration system whose components (e.g. rolling elements, outer race, inner race and cage) interact to generate complex vibration signal. When a fault on one surface of a bearing element strikes another surface, an impact is generated. The successive mechanical impacts (which are the result of the passage of the fault through the load zone) produce a series of impulses observed in a bearing signal. These mechanical impacts modulate the bearing signal at characteristic frequencies depending on the localization of the defect, such as; Fundamental Train (Cage) frequency \( fT \), Ball Spin Frequency \( fBF \), Ball Fault Frequency \( fBF = 2 \cdot fBF \), Ball Pass Frequency Outer Race \( fBFO \) and Ball Pass Frequency Inner Race \( fBFI \) [1,2]. Calculations of the characteristic frequencies assume that the rolling elements do not slide, but roll over the race’s surfaces. There is always some slip and real characteristic frequencies differ from calculated characteristic frequencies by about a few percent [3]. There are two main groups of diagnosing techniques using vibration signals: time-domain and frequency-domain analysis techniques. Traditional time-domain analysis calculates characteristic
features from vibration signal waveform, such as root mean square, skewness, kurtosis or crest factor and they have been applied with limited success for rolling bearing diagnosing [4]. Kurtosis of vibration signal can be used to detect bearing faults at early stages of their development [5]. The kurtosis is a statistical parameter based on the fourth and the second moments of a signal, which is close to 3 for Gaussian noise and other stationary signals, but large for impulsive signals containing series of impulses, such as a signal generated by damaged bearing. However, precise nature of the fault cannot be defined by the kurtosis analysis and for such information it is necessary to use a more sophisticated diagnostic method. The advantage of frequency-domain analysis, based on the transformation of a signal in the frequency domain, is its ability to easily identify certain spectral components of the signal. With high frequency resonance analysis (also known as envelope analysis) it is possible to identify not only the occurrence of the bearing’s fault, but also identify this fault, like damage in the outer race or in the rolling element [1]. In short, the conventional Hilbert-transform-based envelope detection is based on amplitude demodulation and consists of band-pass filtering and the Hilbert transform. Defects in rolling bearings can be detected and localized by discovering spectral components of vibration signal with the frequencies (and their harmonics) typical for the fault.

Usually, bearing vibration signal is collected with an accelerometer installed on the bearing housing where the vibration sensor is often subject to collecting active vibration sources from other mechanical components of the machine. The vibration signal from a bearing at an early stage of defect development may be masked by machine noise, making it difficult to detect the fault by vibration analysis techniques [1,6]. Therefore, a method of diagnostic signal extraction is needed to provide useful information regarding the bearing condition. A number of techniques are described for the separation of bearing signals from background signals which mask it [7–10]. For some specific requirements (e.g. time-triggered signal acquisition), not all of them can be always applied in industrial reality. Moreover, the effectiveness of some techniques depends in essential degree on proper values of a given technique’s parameters (e.g. convergence factor, filter order), which must be determined in an empirical study.

There are also more advanced techniques related to time frequency methods [11], especially wavelets [12] and dedicated approaches for signal enhancement using signal modeling [13,14] or deconvolution technique [15]. Relatively new interesting approach is related to algorithms for searching for informative frequency band [31,33]. Diagnostics under non-stationary load and operating speed condition is discussed in recent papers given by different authors [9,11,21,30,32].

Empirical Mode Decomposition (EMD) has attracted attention in recent years due to its ability to self-adaptive decomposition of non-stationary signals. Recent publications on EMD [16–21] show its advantages for non-stationary signals processing and confirm its effective application in many diagnostic tasks.

In this paper, an EMD-based approach for rolling bearing diagnostics is investigated. By using EMD a raw vibration signal is decomposed into a number of Intrinsic Mode Functions (IMFs). Then, a new method of IMFs aggregation into three Combined Mode Functions (CMFs) is applied and finally the vibration signal is divided into three parts of signal: noise-only part, signal-only part and trend-only part. To further bearing fault-related feature extraction from resultant signals, the spectral analysis of the empirically determined local amplitude is used. To validate the proposed method, raw vibration signals generated by complex mechanical system employed in the industry (driving units of belt conveyors), including vibration data of damaged and undamaged bearings, are used in two case studies. The results show that the proposed rolling bearing diagnosing method can identify the bearing faults at early stages of their development.

2. A brief look into Empirical Mode Decomposition (EMD)

Empirical Mode Decomposition (EMD) has been proposed by Huang et al. [22]. This self-adaptive decomposition method decomposes any signal into empirical modes which represent different oscillation modes embedded in the signal. Based on the EMD algorithm, any original signal $x(t)$ can be reconstructed by a linear superposition of empirical modes:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_{n}(t),$$

(1)

where $c_i(t)$ is $i$-th empirical mode and $r_{n}(t)$ is the final residue after the extraction of $n$ empirical modes. Each empirical mode $c_i(t)$, called Intrinsic Mode Function (IMF), fulfills the following two conditions [22]: (1) in the whole empirical mode, the number of mode local extremes and the number of mode zero-crossings are equal or differ at most by one and (2) at any point, the local average of upper and lower envelope is zero.

The algorithm for the extraction of IMFs from original signal $x(t)$ is called sifting process and it consists of the following steps [23]:

Step 1: Define $x(t) = x_0(t)$ and $r_0(t) = x_0(t)$.

Step 2: Define the maximum number of extracted IMFs.

Step 3: Identify all the local extremes (maxima and minima) of $x(t)$.

Step 4: Connect all the local maxima (respectively minima) with a line known as the empirically determined upper envelope $E_{max}(t)$ (respectively the lower envelope $E_{min}(t)$).

Step 5: Construct the mean of empirically determined upper and lower envelope $m(t) = 0.5(E_{max}(t) + E_{min}(t))$.

Step 6: Define the detail (proto-IMF) as $d(t) = x(t) - m(t)$ and replace $x(t)$ by $d(t)$.

Step 7: Repeat steps 3–6 until $d(t)$ meets the IMF conditions and the stoppage criterion of the sifting process is fulfilled, then derive $i$-th IMF from $d(t)$ and replace $x(t)$ by $r_i(t) = r_{i-1}(t) - d(t)$.

Step 8: If the stoppage criterion of the signal’s decomposition is fulfilled then finish the decomposition process; otherwise, go to step 3.

The second IMF condition is too rigid to use, so it is necessary to change it to implement the EMD. The local average of upper and lower envelope must be close to zero according to some criterion. The evaluation (how small it is) of the amplitude of the local average may be done in comparison with the amplitude of the corresponding mode. In [24] authors introduce a new criterion based on the local mode amplitude $a(t) = 0.5(E_{max}(t) - E_{min}(t))$ and the evaluation function $Q(t) = |m(t)/a(t)|$. In this paper, $d(t)$ meets the second IMF condition, when $max(Q(t)) < \theta$ (the coefficient $\theta$ was equal to 0.2).

A critical part of the EMD procedure is the stoppage criteria of the sifting process and decomposition process. The stoppage criterion of the sifting process determines the point when sifting is complete and a new IMF has been found. Two different stoppage criteria of the sifting process were considered.

The first stoppage criterion of the sifting process is determined by using a Cauchy type of convergence test [22]. If the two details (proto-IMFs) from successive iterations are close enough to each other, it is assumed that the last extracted detail is an IMF. The normalized squared difference $SD$ between two successive details $d_{k-1}(t)$ and $d_k(t)$ during $k$-th iteration is defined as:
If this squared difference $SD_k$ is smaller than a predetermined $T_D$ value, the sifting process will be stopped. In this paper, $T_D$ was equal to $10^{-5}$. The used $T_D$ value was determined experimentally.

The second stoppage criterion of the sifting process is based on the agreement of the number of zero-crossings and extremes. The sifting process is stopped when the numbers of detail zero-crossings and detail extremes are the same for $S$ successive siftings. Typical values for $S$ are in the 3–5 range [25]. In this paper, $S$ was equal to 5.

The sifting process stops when the replications of sifting procedure exceed the predefined maximum number. Selecting a maximum number of siftings prevents the sifting procedure from locking in a never-ending loop. This number should be set large enough to guarantee that IMF is extracted. In this paper, the maximum number of siftings was 750. The sifting process also stops when $x(t)$ has less than two extremes (the signal must have at least two extremes, one maximum and one minimum, to successfully decompose the signal into IMFs).

![EMD algorithm flow chart](image-url)
The stoppage criterion of the decomposition process determines how many components will be extracted from the signal. The decomposition process can be stopped finally by any of the following predetermined criteria: (1) the maximum absolute value of the remaining residue $r_i(t)$ is smaller than tolerance level $\beta_i$ should be a meaningful component of the original signal regarding value level; (2) the predefined number of empirical modes has been extracted (here 20). Here the first stoppage criterion of the decomposition process is described by the following relationship:

$$\max(|r_i(t)|) < \beta_i \cdot \left(\max(x_i(t)) - \min(x_i(t))\right), \quad (3)$$

where $r_i(t)$ is i-th remaining residue, $x_i(t)$ is the original signal (the object of decomposition) and $\beta_i$ is the tolerance coefficient (here: $\beta_i = 0.01$). The decomposition process is also stopped when next IMF cannot be extracted.

In order to clarify the decomposition process, Fig. 1 shows the flow chart of the applied EMD algorithm.

3. Method of IMFs identification and aggregation

Empirical Mode Decomposition is an iterative process of separating complicated signal into a finite number of IMFs. The successive IMFs include signal components from different frequency bands ranging from high to low frequency. Therefore, EMD corresponds to an adaptive (data-driven) filtering [26].

In its assumption, the EMD method decomposes the signal into a set of orthogonal IMFs [22]. In practice, the degree of orthogonality among the IMFs is average and the energy leakage between IMFs is severe. One of the major problems in EMD is the mode mixing, by which IMFs will lose their physical meaning. Mode mixing indicates that a single IMF contains several intrinsic oscillation modes, or that a single intrinsic oscillation mode resides in several neighboring IMFs [27]. Mode mixing makes that the analysis of EMD results is difficult. Some method of combining neighboring IMFs into the so-called Combined Mode Function (CMF) may be an effective way to increase EMD efficiency [19].

The method of IMFs aggregation proposed herein is based on the assumption that the IMFs derived by EMD will be divided generally into three classes of IMFs: noise-only IMFs, signal-only IMFs and trend-only IMFs. The problem is to assign each IMF to the appropriate IMFs class. Typically, the noise is captured by IMFs of low indices and the trend is captured by IMFs of high indices. Some methods of identification of noise-only and trend-only IMFs are presented in the literature [28,29]. They are based on the empirically observed energy and mean of signal components. The proposed method of IMFs identification is based on Pearson correlation coefficient of each IMF and the empirically determined local mean of the original signal. The empirically determined local mean of the signal is defined as:

$$m(t) = 0.5 \cdot (E_d(t) + E_u(t)) \quad \text{(4)}$$

where $E_d(t)$ is the empirically determined lower envelope of the signal and $E_u(t)$ is the empirically determined upper envelope of the signal. The IMFs of low indices with low value of Pearson correlation coefficient are identified as the noise-only IMFs. The IMFs of high indices with low value of Pearson correlation coefficient are identified as the trend-only IMFs. Remaining IMFs are identified as the signal-only IMFs.

A noise-only part of signal and a signal-only part of signal are created as the sum of the noise-only and as the sum of the signal-only IMFs, respectively. A trend-only part of signal is created as the sum of the trend-only IMFs and the final residue.

4. Application of proposed rolling bearing diagnosing method – two case studies

4.1. Machine and experiment description

Mining machines seem to be a special class of machines with complex structure, high-power, time-varying load, etc. Photographs of investigated machine working in the mining company are presented in Fig. 2. Depends on the design (required power for driving of belt conveyor), belt conveyor driving station might consist of one up to four drive units with 630 or 1000 [kW] power each. In case discussed here, two drive units are used (note blue housing of electric motors in Fig. 2). In Fig. 3 the scheme of the drive unit for a belt conveyor is shown. The drive unit consists of an electric motor, a coupling and two stage gearbox, that are connected with a pulley. The pulley (Fig. 3b) consists of a shaft, two bearings and the coating covered by rubber (to increase friction between the pulley coating and the belt). Often between the gearbox and the pulley a rigid coupling is used (see Fig. 3c).

The purpose of the diagnostic experiment was to acquire vibration signal from the pulley and the assessment of the condition of the pulley' bearing (see Fig. 3a). The location of accelerometer is shown in Fig. 3d: the sensor has been mounted using screw, in horizontal direction. Based on the bearing geometry and the shaft’s rotational speed, the characteristic defect frequencies of rolling bearings were calculated, namely: $f_{BPFO} = 0.51$ Hz, $f_{BSFI} = 4.45$ Hz, $f_{BSFI} = 8.90$ Hz, $f_{BSPF} = 12.34$ Hz, $f_{BPS} = 16.06$ Hz. Several signal acquisition sessions have been performed. For each measurement the signal was acquired with the following parameters: sampling frequency $f_s = 19,200$ Hz, duration $T = 2.5$ s. More information about machine and diagnostic experiment can be found in other papers concerning diagnostics of these machines [9,13].

Two vibration signals generated by the drive unit, including vibration data of undamaged and damaged bearings, are used in two case studies. Due to rigid connection between gearbox and pulley, a serious participation of gearbox vibration in acquired vibration signal has been noticed. Energy of signal components related to gear meshing frequency completely masks the signal of interest -- the signal from the rolling bearing. Amplitude spectra of acquired vibration signals are presented in Fig. 4. Fig. 5 presents amplitude spectra of Hilbert-transform-based envelopes of vibration signals. Mean values were removed from the envelopes.

4.2. Decomposing of vibration signals

First, the EMD method is used to decompose the vibration signals. The decomposition results are presented in Figs. 6 (the first case – undamaged bearing) and 7 (the second case – damaged bearing). The results of IMFs identification are presented in Figs. 8 (the first case – undamaged bearing) and 9 (the second case – damaged bearing). Figs. 10 (the first case – undamaged bearing) and 12 (the second case – damaged bearing) present the waveforms of the
Fig. 3. Diagnosed object: (a) scheme, (b) pulley with bearing housing mounted on shaft, (c) view on joint of output shaft in gearbox with pulley, and (d) view on sensor location on pulley [9].

Fig. 4. Amplitude spectra of vibration signals (the first case/undamaged/ – top, the second case/damaged/ – bottom).

Fig. 5. Amplitude spectra of envelopes of vibration signals (the first case/undamaged/ – top, the second case/damaged/ – bottom).
Fig. 6. Decomposition of the first original signal (the first case – undamaged bearing)/19 empirical modes (IMFs) $c_i(t)$ [m/s²] and final residue $r(t)$ [m/s²].

Fig. 7. Decomposition of the second original signal (the second case – damaged bearing)/18 empirical modes (IMFs) $c_i(t)$ [m/s²] and final residue $r(t)$ [m/s²].

Fig. 8. IMFs identification of the first original signal (the first case – undamaged bearing).

Fig. 9. IMFs identification of the second original signal (the second case – damaged bearing).
vibration signals and the waveforms of the noise-only, signal-only and trend-only parts of the vibration signals. Figs. 11 (the first case – undamaged bearing) and 13 (the second case – damaged bearing) present the amplitude spectra of the vibration signals and the amplitude spectra of the noise-only, signal-only and trend-only parts of the vibration signals.

4.3. Analysis of vibration signals and their parts

Kurtosis analysis of the raw vibration signals does not deliver any diagnostic information. The kurtosis values of the raw vibration signals (3.44 in the first case and 3.10 in the second case) are similar and their low level does not indicate any bearing fault.
The kurtosis values of the noise-only signals parts are significantly different (2.59 in the first case and 27.44 in the second case). High value of the noise-only signal's part indicates that in the second case some bearing fault occurs. The precise nature of the fault cannot be defined by the kurtosis analysis and for such information it is necessary to use a more sophisticated diagnostic method.

In order to perform a detailed fault-related analysis of signals, the spectral analysis of the empirically determined local amplitude of a signal is used. The empirically determined local amplitude of the signal is defined as:

$$a(t) = 0.5 \cdot (E_u(t) - E_l(t)),$$

where $E_u(t)$ is the empirically determined upper envelope of the signal and $E_l(t)$ is the empirically determined lower envelope of the signal. In order to conduct the spectral analysis, mean value was removed from the empirically determined local amplitude. Amplitude spectra of the empirically determined local amplitudes of the noise-only signals parts are presented in Fig. 14.

The discovery of high-amplitude spectral components of the empirically determined local amplitude indicates that in the second case some bearing fault occurs. The basic frequency of those spectral components equals 12.7 Hz and corresponds (with 3% tolerance) to Ball Pass Frequency Outer Race ($f_{BPFO}$). The significant coincidence between these frequencies enables, with high probability, the identification of this defect as the bearing outer race defect.

5. Conclusions

The paper presents the rolling bearing diagnosing method based on Empirical Mode Decomposition, a new method of IMFs aggregation into three parts of raw vibration signal and the analysis of the noise-only signal's part. The analysis of the noise-only signal's part provided herein is a two-stage process that involves the kurtosis analysis and the spectral analysis of the empirically determined local amplitude of this signal's part.

Two case studies on the raw vibration signals generated by complex mechanical systems employed in the industry were conducted and the analysis demonstrated that the proposed rolling bearing diagnosing method can identify the bearing faults. The bearing fault at early stage of its development was detected by using the kurtosis analysis of the noise-only signal's part even when the bearing vibration signal was completely masked by machine noise. This fact showed that the proposed method of the noise-only signal's part creation is very useful and important from the diagnostic point of view.
The precise nature of the bearing fault was defined by the spectral analysis of the empirically determined local amplitude of the noise-only signal’s part. The discovery of high-amplitude spectral components of the empirically determined local amplitude enabled the identification of this defect, because the basic frequency was typical for the defined bearing fault. Therefore, it has been demonstrated that the presented method of the empirical determination of the local amplitude is diagnostically useful and equivalent to the Hilbert-transform-based envelope detection method.

Acknowledgements

This paper was financially supported by Polish State Committee for Scientific Research 2010–2013 as research Project no. N504 147838 (R. Zimroz).

References