Measurement 94 (2016) 554–560

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Demodulation analysis based on adaptive local iterative filtering for bearing fault diagnosis

Xueli An [a,*], Hongtao Zeng [b], Chaoshun Li [c]

[a] China Institute of Water Resources and Hydropower Research, Haidian District, Beijing 100038, China
[b] School of Power and Mechanical Engineering, Wuhan University, Wuhan, Hubei Province 430072, China
[c] College of Hydroelectric and Digitalization Engineering, Huazhong University of Science and Technology, Wuhan 430074, Hubei Province, China

ARTICLE INFO

Article history:
Received 31 August 2015
Received in revised form 26 August 2016
Accepted 31 August 2016
Available online 1 September 2016

Keywords:
Adaptive local iterative filtering
Roller bearing
Modulation features
Fault diagnosis

ABSTRACT

The vibration signals from defective rolling bearings are multi-component amplitude modulation (AM)-frequency modulation (FM) signals. Traditional envelope analysis method is based on a filter. The centre frequency and bandwidth of the filter are set according to experience. So the filter method will induce a demodulation error. This research proposes a rolling bearing fault diagnosis method based on adaptive local iterative filtering (ALIF) and envelope spectrum. The ALIF method is a new method for the analysis of non-stationary signals. It uses an iterative filters strategy together with an adaptive, data-driven filter length selection to achieve the necessary decomposition. Smooth filters with compact support from the solutions of the Fokker–Planck equations are used within the ALIF method. The ALIF method offers good performance in obtaining more accurate components of non-stationary signals and in suppressing mode mixing. The ALIF method can decompose a multi-component AM-FM signal into a number of stationary components. The envelope demodulation method is used to analyse those components containing fault information which, in turn, can reveal bearing fault features. The vibration signals from a rolling bearing with an outer race fault and an inner race fault are used to verify the proposed method. The results show that this method can effectively extract bearing fault features.

1. Introduction

When faults occur in a rolling bearing, its vibration signal is often modulated [1–6]. This is because fatigue, or stress concentration, in the rolling bearing surface will damage the surface. Due to this damage, periodic shock vibration signals are generated during operation. The signal is easily modulated by the pulsation excitation force. So fault vibration signals from rolling bearings often exhibit amplitude modulation (AM)-frequency modulation (FM) features [1,2]. For these AM-FM signals, envelope analysis is an effective method [7,8]. Traditional envelope analysis methods often use band-pass filtering to decompose multi-component AM-FM signals into a number of single-component AM-FM signals. Then the Hilbert transform is used to analyse each single-component AM-FM signal to calculate its instantaneous frequency and amplitude. However, in engineering practice, it is difficult to determine the number of carrier frequency components and the size of the carrier frequency. So the centre frequency and bandwidth of the band-pass filter is selected on the basis of a priori knowledge alone. It was apparent that subjectivity in this choice inevitably caused a demodulation error. Therefore, it cannot accurately extract the fault characteristics from a vibration signal.

Antoni and Randall [9] introduced the concept of the kurtogram, from which optimal band-pass filters can be deduced, for instance as a prelude to envelope analysis. The kurtogram is used for fault diagnosis in rotating machines. The result shows that the kurtogram is a powerful tool that can detect the transient components in a complex vibration signal. For the non-linear, and non-Gaussian, characteristics of the vibration signals from rotating machinery, the bispectrum is useful when analysing the signals. Due to modulation or smearing, it is difficult to extract accurate frequency-based features from the bispectrum. Jiang et al. [10] developed a bispectral distribution for machinery fault diagnosis. In the method, the binary images are extracted, from the bispectra, and used as features to construct the target templates. The roller bearing and gear fault signals are used to verify the feasibility of this proposed method. Empirical mode decomposition (EMD) can adaptively decompose a complex multi-component AM-FM signal into a number of single-component AM-FM signals. Therefore, envelope demodulation based on EMD method has been widely used in mechanical fault diagnosis [11]. However, EMD has some

* Corresponding author.
E-mail address: an_xueli@163.com (X. An).

http://dx.doi.org/10.1016/j.measurement.2016.08.039
0263-2241/

defects, such as an excessively large envelope, occasionally an insufficiently wide envelope, mode mixing, and end effects. Cicone et al. proposed an adaptive local iterative filtering (ALIF) method [12] which can adaptively filter a complex signal into several stable components. The ALIF method, for different frequency bands, uses an iterative filtering strategy together with an adaptive, data-driven filter length selection protocol to achieve decomposition. Smooth filters with compact support from solutions of the Fokker–Planck equations are used within the ALIF method. These filters satisfy the derived sufficient conditions for the convergence of the iterative filtering algorithm. The ALIF method performs well when in that it is able to obtain more accurate data about the components of a non-stationary signal while it also suppresses mode mixing.

A rolling bearing fault diagnosis method, based on ALIF and envelope demodulation, was developed. Firstly, ALIF is used to decompose a vibration signal into several stable components. Then, those components which contain key fault characteristics are analysed by the envelope demodulation method to extract modulation information from the fault signal. It can effectively extract the fault characteristics of a rolling bearing vibration signal, and thus distinguish the working states and fault types therein.

2. Adaptive local iterative filtering method

The ALIF method is based on the iterative filtering (IF) [12] technique. The main differences between the ALIF method and the IF technique are that the ALIF method can adaptively compute the filter length and the filters can be chosen from solutions to the Fokker-Planck equations to compute the moving average of the signals. The ALIF method is presented as follows:

ALIF method IMF = ALIF(f)
  IMF = {}
  while the number of extrema of $f \geqslant 2$ **do**
    $f_1 = f$
    while the stopping criterion is not satisfied **do**
      compute the filter length $l_n(x)$ for $f_n(x)$
      $f_{n+1}(x) = f_n(x) - \int_{-l_n(x)}^{l_n(x)} f_n(x+t) w_n(x,t) dt$
      $n = n + 1$
    end while
    $\mathrm{IMF} = \mathrm{IMF} \cup \{f_n\}$
    $f = f - f_n$
  end while
  $\mathrm{IMF} = \mathrm{IMF} \cup \{f\}$

In the presented method, $w_n(x,t)$, $t \in [-l_n(x), l_n(x)]$, is the filter at point x for the signal $f_n(x)$, its length is $2l_n(x)$. The ALIF method has two iterations: one capturing a single IMF and another producing all the IMFs. The former is called the inner iteration: the latter is called the outer iteration. The updating step of the inner iteration is as follows:

$$f_{n+1}(x) = f_n(x) - \int_{-l_n(x)}^{l_n(x)} f_n(x+t) w_n(x,t) dt \tag{1}$$

An equivalent formulation for the updating step is used to establish the convergence theorem of the ALIF method. The scaling function can be linear when expressed as $g_n(x,y) = l_n(x)y/L$, or cubic as $g_n(x,y) = l_n(x)y^3/L^3$. The function $g_n(x,y)$ is used to change the right-hand side in Eq. (1) to:

$$\int_{-l_n(x)}^{l_n(x)} f_n(x+t) w_n(x,t) dt = \int_{-L}^{L} f_n(x + g_n(x,y)) W(y) dy$$

So Eq. (1) can be rewritten as

$$f_{n+1}(x) = f_n(x) - \int_{-L}^{L} f_n(x + g_n(x,y)) W(y) dy \tag{2}$$

where $W(y)$, $y \in [-L, L]$, is a fixed filter.

A new operator T is defined as $T_{w,l}(f) := \int_{-l(x)}^{l(x)} f(x+t) w(x,t) dt$. The convergence theorem for the inner iteration of ALIF method is:

Let $f(x)$, $x \in R$, be continuous and $f(x) \in L^{\infty}(R)$. Let

$$\varepsilon_n = \frac{\|T_{w_{n+1},l_{n+1}}(f_{n+1})\|L^{\infty}}{\|T_{w_n,l_n}(f_n)\|L^{\infty}}, \qquad \delta_n = \frac{\|T_{w_{n+1},l_{n+1}}(|f_{n+1}|)\|L^{\infty}}{\|T_{w_n,l_n}(|f_n|)\|L^{\infty}} \tag{3}$$

If $\prod_{i=1}^{n} \varepsilon_i \to 0$, $\prod_{i=1}^{n} \delta_i \to c > 0$, as $n \to \infty$, Then $\{f_n(x)\}$ converges to an IMF.

The convergence theorem for the outer iteration of ALIF method is:

Let $f(x)$, $x \in R$, be continuous and differentiable and let $f(x)$ have a finite number of extreme points in any compact interval. So $f(x)$ has at most countable extreme points. Let x_i, $i = 1, 2, \ldots, k$, be the extreme points of $f(x)$. Assume that $f(x)$ is strictly monotonic in $[x_i, x_{i+1}]$, $i = 1, 2, \ldots, k-1$. The functions $c_n^{(1)}(x)$ and $c_n^{(2)}(x)$ are defined using $f_n(x)$ as:

$$\begin{aligned} c_n^{(1)}(x) &= \int_{-L}^{L} \left[f_n'(x) - f_n'(g_n(x,y) + x)\right] W(y) dy \\ c_n^{(2)}(x) &= \int_{-L}^{L} \left[f_n'(x) - f_n'(g_n(x,y) + x)\right] h(y) W(y) dy \end{aligned} \tag{4}$$

Function $f(x)$ is assumed to be a differentiable function with the properties described above. Eq. (2) is used, if the scaling function is separable, i.e. $g_n(x, y) = l_n(x)h(y)$ and for every $n \in N$:

$$\begin{aligned} c_n^{(1)}(x) + l_n'(x) c_n^{(2)}(x) > 0 \quad \text{when } f_n'(x) > 0 \\ c_n^{(1)}(x) + l_n'(x) c_n^{(2)}(x) < 0 \quad \text{when } f_n'(x) < 0 \end{aligned} \tag{5}$$

Then the number of extreme points of $f(x) - \lim_{n\to\infty} f_n(x)$ is at most the number of extreme points of $f(x)$ if $\lim_{n\to\infty} f_n(x)$ exists.

A detailed introduction to the ALIF method is given elsewhere [12]. The ALIF algorithm was implemented using MATLAB™ software.

3. Simulation signal analysis

To verify the effectiveness of the proposed method, a mixed signal $x(t)$ is studied:

$$x(t) = x_1(t) + x_2(t) \tag{6}$$

where $x_1(t) = [1 + 0.5\sin(5\pi t)]\sin[122\pi t + 21\pi t^2]$, $x_2(t) = \sin(40\pi t)$, $t \in [0, 1]$.

This signal $x(t)$ is composed of an AM-FM signal $x_1(t)$ and a sinusoidal signal $x_2(t)$. Fig. 1 shows the time-domain waveform of signal $x(t)$. ALIF is used to decompose the signal $x(t)$, the decomposition results are shown in Fig. 2. It can be seen that two components were obtained by use of the ALIF method: f1 corresponds to the AM-FM signal $x_1(t)$, f2 corresponds to the sinusoidal signal $x_2(t)$. The two components were more accurate and thus the ALIF method was shown to have worked.

Fig. 3 shows the decomposition results from signal $x(t)$ based on the EMD method: six components were obtained, but there were significant end effects (compared with ALIF), and the EMD method produced more false components.

4. Case study

When bearing faults occur in the inner, or outer, race, during rotation of the bearing, due to the modulation characteristics of