A novel adaptive wavelet stripping algorithm for extracting the transients caused by bearing localized faults

Dong Wang\textsuperscript{a,b}, Changqing Shen\textsuperscript{a,b,c,d}, Peter W. Tse\textsuperscript{a,b,*}

\textsuperscript{a} Smart Engineering Asset Management Laboratory (SEAM), Department of Systems Engineering and Engineering Management, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China
\textsuperscript{b} Centre for Systems Informatics Engineering (CSIE), Department of Systems Engineering and Engineering Management, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China
\textsuperscript{c} Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Suzhou, Jiangsu, China
\textsuperscript{d} USTC-CityU Joint Advanced Research Center, Suzhou, Jiangsu, China

\section*{A R T I C L E I N F O}

Article history:
Received 14 May 2012
Received in revised form 20 February 2013
Accepted 26 July 2013
Handling Editor: K. Worden
Available online 29 August 2013

\section*{A B S T R A C T}

Rolling element bearings are widely used in rotating machinery. Its unexpected failure may result in machine breakdown. Whenever a bearing suffers a localized fault, the transients with a potential cyclic characteristic are generated by the rollers striking the localized fault. This phenomenon is an early bearing fault feature. Therefore, the extraction of the transients is beneficial to the identification of the early bearing fault. In this paper, a novel adaptive wavelet stripping algorithm (AWSA) is proposed to extract the simulated transients from an original bearing fault signal. Firstly, the parametric model of anti-symmetric real Laplace wavelet (ARLW) or impulse response wavelet (IRW) is built to approximate the real transients. Then, with the aid of wavelet correlation filtering analysis, the simulated transients with the optimal frequency, damping coefficient and delay time are adaptively peeled from the original bearing fault signal. The spatial reconstruction of the simulated transients reflects the random occurrence of the real transients. In order to boost the computing time of the AWSA, an enhanced AWSA is developed. At last, the bearing fault signals collected from an experimental machine and an industrial machine are utilized to validate the effectiveness of the AWSA. The results show that the AWSA can adaptively peel the simulated transients from the original bearing fault signals. A comparison with a periodic multi-transient model is conducted to show that the AWSA is better to extract the random characteristics of the real transients.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Mechanical fault diagnosis aims to detect the abnormal health condition of a machine and determine its abnormal cause. Rolling element bearings are widely used in rotating machines. The health condition of a rolling element bearing gradually degrades as the accumulated working time increases \cite{1}. Unexpected bearing failure may cause machine breakdown and result in economic loss and even human casualty. As a result, the identification of an early bearing fault is of great significance in practice \cite{2–4}. The rolling element bearing consists of the different components, such as an inner race, an outer race, a number of rolling elements and a cage. Different types of rolling elements, such as roller, ball, etc., are used in various rotating machines. Once a localized fault is formed on...
the surface of the inner race or the outer race, a transient with an exponential decay will be generated by the roller striking the localized fault. Because of the periodic rotation characteristics of the shaft, the transient will be repeated. Due to the influence caused by the random slippage of rolling elements, the occurrence of these transients is not periodic but random. In addition, the resonance frequencies of the structures between the bearings and the transducers are excited to make most of bearing fault signatures exist at high-frequency resonant bands [5,6]. Therefore, nowadays, cyclic spectral analysis [7] and envelope spectrum analysis [8] become two powerful tools to identify bearing fault signatures.

When envelope analysis is used, a band-pass filter is usually employed to retain one of the resonant frequency bands to enhance the bearing fault signal overwhelmed by heavy noise and other vibration components. Then, demodulation with envelope analysis transforms the signal at one of the high-frequency resonant bands into the signal at a low-frequency band to identify bearing fault signals overwhelmed by heavy noise and other vibration components. Then, demodulation with envelope analysis transforms the signal at one of the high-frequency resonant bands into the signal at a low-frequency band to identify bearing fault diagnoses.

Because wavelet transform has a band-pass property, which can be used to retain one of the resonant transforms the signal at one of the high-frequency resonant bands into the signal at a low-frequency band to identify bearing fault diagnoses.

Therefore, nowadays, cyclic spectral analysis [7] and envelope spectrum analysis [8] become two powerful tools to identify bearing fault signatures.
kurtosis to characterize the non-stationary transient signals. Then, the kurtogram was designed for finding the optimal filters [20].

To improve the computing efficiency, the short time Fourier transform (STFT) and the 1/3-binary filters based fast kurtogram were developed to make on-line condition monitoring and fault diagnosis realistic [21]. Recently, Barszcz and JabLoiiski [22] proposed a new concept called Protrugram to identify the optimal frequency band by measuring the kurtosis of the envelope spectrum amplitudes of the demodulated signal, and they compared their method with the fast kurtogram to show the Protrugram could detect the bearing fault signatures in the case of a very low signal to noise ratio. In the work of Lei et al. [23], they pointed out that although the fast kurtogram was proved to be effective in detecting the transients in vibration signals, the filters used in the fast kurtogram were not as precise as the wavelet filters because the wavelet basis with the different scales could better match the real transients. Accordingly, they proposed an improved kurtogram based on the binary wavelet packet transform to detect the transients caused by the bearing localized faults. Schukin et al. [24] discussed the four criteria to evaluate an optimal wavelet function. Based on these criteria, ten wavelets were investigated and the results showed that the impulse wavelet had the best performance among the ten wavelets for the vibration impulse analysis. In the work of Rafiee et al. [25,26], 324 wavelets were investigated to find the best one which could be used for gear and bearing fault diagnosis. Their results showed that the Daubechies 44 (dB44) was the best one to approximate the real transients. Recently, getting the idea from the wavelet correlation filtering analysis [27], Wang et al. [28] proposed a novel method to extend the parametric model of the Laplace wavelet to a parametric wavelet model dictionary that included the parametric models of the Morlet wavelet, the Laplace wavelet and the Harmonic wavelet. Then, one parametric wavelet model having the largest correlation coefficient with one of the real transients was reconstructed and it was used to approximate all real transients in the signal. In the following, a strict periodic multiple-transient model was built to simulate the cyclic characteristics of the bearing fault signal. Their results demonstrated that their method was effective to identify the mean spacing between two successive simulated transients. However, it should be noted that the periodic assumption used in their method may be improper to reflect the random slippage phenomenon of the roller elements as stated in [7]. In other words, the bearing fault signal was not periodic but random.

From the previous work [10–18,23–26,28], it was found that a unique optimal wavelet, which had the optimal scale or center frequency and bandwidth, was used to match every real transient in a signal. In fact, the shape of these real transients was different from each other even though they were generated by the same localized fault and collected by the same accelerometer. It means that the unique optimal wavelet could not simultaneously match every real transient well. An adaptive signal processing is required to extract the time-varying transients. One recent work was that Makowski and Zimroz [29] used the weighted summation of the derivatives of the reflection coefficients obtained by the adaptive Schur filter to detect the bearing cyclic fault features. The use of the derivatives of the reflection coefficients was easier than the variation of the prediction errors obtained by the adaptive Schur filter. In this paper, a novel adaptive wavelet stripping algorithm (AWSA) is proposed to provide a more accurate method for extracting the transients caused by the bearing localized faults than the method proposed in [28]. The AWSA adaptively peels the simulated transients via the parametric model of the Laplace wavelet (ARLW) [30] or impulse response wavelet [31] from the bearing fault signal. The spatial reconstruction of the simulated transients can accurately reveal the random characteristics of the real transients caused by the localized bearing faults. Bearing fault type can be well identified by the mean spacing between two successive simulated transients.

The rest of this paper is summarized as follows. Section 2 briefly reviews the fundamental theory of the wavelet correlation filtering analysis. A novel adaptive wavelet stripping algorithm for extracting bearing fault transients is reported in Section 3, followed by the case studies using experimental bearing fault signals and an industrial multi-fault signal as stated in Section 4. A comparison with a periodic multi-transient model is conducted in Section 5. Section 6 provides a discussion on boosting the computing time of the AWSA. Finally, the conclusions are drawn in Section 7.

2. Introduction of the wavelet correlation filtering analysis

In the field of mechanical engineering, the occurrence of transients is often indicative of the abnormal health condition of a machine. When a vibration signal is collected from an accelerometer, the transients are often mixed with heavy noise and other vibration components. Therefore, it is not easy to directly identify the transients and recognize the cause of their generation. The idea of wavelet correlation filtering analysis is to calculate Pearson’s correlation coefficient between the collected vibration signal and the parametric models of a wavelet [27]. Pearson’s correlation coefficient is a powerful tool to measure the strength of the linear dependence between two signals. Denote the collected vibration signal as $x(n)$ and the parametric model of the wavelet as $\psi_r(n)$. Assume both of the signals has a zero mean. Pearson’s correlation coefficient $\rho_{x(n),\psi_r(n)}$ between them can be defined as [32]

$$\rho_{x(n),\psi_r(n)} = \frac{\langle x(n), \psi_r(n) \rangle}{\sqrt{\langle x(n), x(n) \rangle \langle \psi_r(n), \psi_r(n) \rangle}} = \frac{\sum_{i=1}^{N} x(i) \times \psi_r(i)}{\sqrt{\sum_{i=1}^{N} x(i)^2} \sqrt{\sum_{i=1}^{N} \psi_r(i)^2}},$$

(1)

where $N$ is the length of the signal and $\langle \cdot, \cdot \rangle$ is the inner product. In terms of Cauchy–Schwarz inequality, Pearson’s correlation coefficient $\rho_{x(n),\psi_r(n)}$ is constrained to

$$-1 \leq \rho_{x(n),\psi_r(n)} \leq 1,$$

(2)

the closer the correlation coefficient is to 0, the weaker the linear dependence relationship between the two signals. The interpretation of the size of Pearson’s correlation coefficient depends on the specific context and purposes. In signal
processing, Pearson's correlation coefficient is affected by the size of the signal to noise ratio of one signal [33]. In vibration analysis, the signal to noise ratio can be defined as the intensity of the vibration signal of interest to the intensity of the other vibration components and the background noise [34].

Wang et al. [28] extended the wavelet correlation filtering analysis by the several parametric models of the wavelets including Morlet wavelet, Harmonic wavelet and Laplace wavelet. The parametric model of the real part of the complex Morlet wavelet is given as

\[
\psi_{\text{Morlet}}(f, \zeta, \tau, n) = e^{-c/\sqrt{1 - \zeta^2} \times (2\pi f \times (n - \tau))} \cos(2\pi f \times (n - \tau)),
\]

where the discrete frequency \(f\), the discrete damping coefficient \(\zeta\) and the discrete delay time \(\tau\). They belong to the subsets \(F\), \(Z\) and \(T\), respectively, as follows:

\[
F \subset R^+, \\
Z \subset (R^+ \cap [0, 1)), \\
T \subset R^+.
\]

the parametric model of the real part of the complex Harmonic wavelet is given as

\[
\psi_{\text{Harmonic}}(f, \zeta, \tau, n) = \frac{\sin(\zeta/\sqrt{1 - \zeta^2} \times 2\pi f(n - \tau)) - \sin(\zeta/\sqrt{1 - \zeta^2} \times \pi f(n - \tau))}{\zeta/\sqrt{1 - \zeta^2} \times \pi f(n - \tau)}.
\]

the parametric model of the real part of the complex Laplace wavelet is given as

\[
\psi_{\text{Laplace}}(f, \zeta, \tau, n) = e^{-c/\sqrt{1 - \zeta^2} \times 2\pi f \times (n - \tau)} \cos(2\pi f \times (n - \tau)),
\]

Pearson's correlation coefficients \(\rho_{\text{wavelet}}\) between the collected signal and the real parts of Morlet wavelet, Harmonic wavelet and Laplace wavelet can be unified as

\[
\rho_{\text{wavelet}}(f, \zeta, \tau) = \frac{\langle x(n), \psi_{\text{wavelet}}(f, \zeta, \tau, n) \rangle}{\sqrt{\langle x(n), x(n) \rangle} \sqrt{\langle \psi_{\text{wavelet}}(f, \zeta, \tau, n), \psi_{\text{wavelet}}(f, \zeta, \tau, n) \rangle}}.
\]

where \(\psi_{\text{wavelet}}\) could be one of the parametric models of the wavelets shown in Eqs. (3), (5) and (6). Consequently, wavelet correlation filtering and the parameter identification of each parametric wavelet model can be realized by maximizing Pearson's correlation coefficient shown in Eq. (7). The results reported in reference [28] show that the real part of the complex Laplace wavelet was more sensitive to the transients generated by the localized bearing faults and the Morlet wavelet was more suitable to match the transients caused by gear faults because of the symmetry property of the Morlet wavelet.

3. A novel adaptive wavelet stripping algorithm for extracting bearing fault features

The wavelet correlation filtering is able to identify the parameters of a parametric wavelet model. Once its parameters are determined, the parametric wavelet model can be used to approximate the real transient. Recently, Wang et al. [28] used the wavelet correlation filtering to find a suitable parametric wavelet model to match the bearing fault transient. Then, they constructed a strict periodic multiple-transient model to detect the temporal cyclic intervals of bearing localized faults. As previously mentioned in introduction, the bearing fault signal is not periodic but random because of the random slippage of the roller elements. Besides, in their work, one unique optimal parametric wavelet was used in the periodic multiple-transient model to approximate all real transients in the fault signal. In reality, these real transients in the fault signal may have different shapes caused by the uncertainty of outside working factors. Specifically, these transients may have different frequencies, damping coefficients and amplitudes. In order to improve the wavelet correlation filtering based periodic multiple-transient model, a novel adaptive wavelet stripping algorithm (AWSA) for extracting the random characteristics of the real transients is proposed in this section. The advantages of the proposed algorithm are summarized as follows: (1) it is an adaptive signal processing method; (2) every real transient corresponds to or matches with one simulated transient; (3) the spatial reconstruction of the simulated transients reflects the random characteristics of the transients caused by the bearing localized fault. The flowchart of the AWSA is shown in Fig. 1 and the details are described below.

Step 1. Load the original bearing vibration signal. Initialize \(i = 1\). The parametric model of anti-symmetric real Laplace wavelet (ARLW) [30] or impulse response wavelet [31] is employed in this paper because Laplace wavelet is more similar to the transient caused by bearing localized faults (this fact will be demonstrated in Section 4). Compared with the real part of complex Laplace wavelet [28] shown in Eq. (6), the ARLW has a phase delay and the anti-symmetric extension. The parametric model of the ARLW is built as follows:

\[
\psi_{\text{ARLW}}(f, \zeta, \tau, n) = e^{-c/\sqrt{1 - \zeta^2} \times 2\pi f \times (n - \tau)} \sin(2\pi f \times (n - \tau)),
\]

in order to reduce the computing time of scanning the parameter \(f\) from a subset \(F\), the range of the subset \(F\) can be restricted to the frequency range of one of the bearing resonant frequency bands by inspecting the frequency spectrum of the original bearing fault signal.
Step 2. Scan all possible parameters of \( f, \zeta, \text{ and } \tau \) to find the optimal \( f_{opt}^i, \zeta_{opt}^i, \text{ and } \tau_{opt}^i \) by maximizing Pearson’s correlation coefficient \( \rho_{ARLW}(f, \zeta, \tau) \) between the signal \( x(n) \) and the ARLW \( \psi_{ARLW}(f, \zeta, \tau, n) \). The mathematical description of Step 2 is illustrated as:

\[
(f_{opt}^i, \zeta_{opt}^i, \tau_{opt}^i) = \arg\max_{f \in F, \zeta \in \Omega, \tau \in T} \rho_{ARLW}(f, \zeta, \tau), \quad \frac{\langle x(n), \psi_{ARLW}(f, \zeta, \tau, n) \rangle}{\sqrt{\langle x(n), x(n) \rangle} \sqrt{\langle \psi_{ARLW}(f, \zeta, \tau, n), \psi_{ARLW}(f, \zeta, \tau, n) \rangle}}. \tag{9}
\]

Step 3. Use the optimal parameters \( f_{opt}^i, \zeta_{opt}^i, \text{ and } \tau_{opt}^i \) obtained by Step 2 to construct the \( i \)th simulated transient \( \psi_{ARLW}(f_{opt}^i, \zeta_{opt}^i, \tau_{opt}^i, n) \), which is given as follows:

\[
\psi_{ARLW}(f_{opt}^i, \zeta_{opt}^i, \tau_{opt}^i, n) = e^{-i \zeta_{opt}^i} \sqrt{1 - (\zeta_{opt}^i)^2} \sin(2\pi f_{opt}^i \times n - \tau_{opt}^i). \tag{10}
\]

Step 4. Assume that the segment of the signal \( x(n), n = n_1^i, n_1^i + 1, \ldots, n_2^i \) (unit: sample) is associated with \( \psi_{ARLW}(f_{opt}^i, \zeta_{opt}^i, \tau_{opt}^i, n) \). Their Pearson’s correlation coefficient achieves the maximum. The temporal locations \( t_1^i \) and \( t_2^i \) correspond to the maximum amplitude \( A_{max}^i \) and the minimum amplitude \( A_{min}^i \) of \( \psi_{ARLW}(f_{opt}^i, \zeta_{opt}^i, \tau_{opt}^i, n) \), respectively. In order to recover the amplitudes of \( \psi_{ARLW}(f_{opt}^i, \zeta_{opt}^i, \tau_{opt}^i, n) \) from the real signal \( x(n) \), two scale factors (SF) are proposed as follows:

\[
SF_1^i = \frac{x(t_1^i)}{A_{max}^i}, \tag{11}
\]

\[
SF_2^i = \frac{x(t_2^i)}{A_{min}^i}. \tag{12}
\]
the aim of the SF\(_i\) and SF\(_j\) are used to scale the positive and negative amplitudes from Eq. (10). The scale operation is described as follows: 
\[
\psi(x(n)) = e^{-\nu_{i,j}^{opt}} \sqrt{1-\nu_{i,j}^{opt}^2} \sin(2\pi f_{i,j}^{opt} \times (n-t_{i,j}^{opt})).
\]
otherwise, 
\[
\psi(x(n)) = e^{-\nu_{i,j}^{opt}} \sqrt{1-\nu_{i,j}^{opt}^2} \sin(2\pi f_{i,j}^{opt} \times (n-t_{i,j}^{opt})).
\]
Step 5. Subtract \(\psi(x(n))\) from \(x(n)\) and a new residual signal \(x'(n)\) is obtained as follows: 
\[
x'(n) = \begin{cases} 
  x(n) - \psi(x(n)), & n \in [t_i, t_i + W - 1] \\
  x(n), & \text{otherwise}
\end{cases}
\]
Step 6. Replace \(x(n)\) with \(x'(n)\) and reconstruct a signal \(y(n)\) which consists of \(L\) \(\psi(x(n))\): 
\[
y(n) = \sum_{i=1}^{L} \psi(x(n)),
\]
where \(L\) is the number of ARLWs. Three stoppage criteria based on the normalized mean squared error (NMSE) are used in this paper. 

**Stoppage criterion 1** is 
\[
\text{NMSE}(L) = \frac{\sum_n (x(n)-y(n))^2}{\sum_n x(n)^2} = \sum_n \frac{(x(n)-y(n))^2}{\sum_n x(n)^2} < \sigma_1,
\]
where \(\sigma_1\) is empirically set as 0.35 in this paper. Stoppage criterion 1 is used based on the idea of the stoppage criterion used in empirical mode decomposition proposed by Huang et al. [35]. In the work of Huang et al., they employed the standard deviation calculated from the two consecutive sifting results to stop the sifting process. They recommended the limitation of the standard deviation value of 0.2–0.3 is good enough. 

**Stoppage criterion 2** is 
\[
0 \leq (\text{NMSE}(L-1)-\text{NMSE}(L)) < \sigma_2,
\]
where \(\sigma_2\) is empirically set as 0.01 in this paper. 

**Stoppage criterion 3** is 
\[
\text{NMSE}(L-1) < \text{NMSE}(L),
\]
if Stoppage criterion 3 is reached, Eq. (16) is revised as follows: 
\[
y(n) = \sum_{i=1}^{L-1} \psi(x(n)),
\]
Step 7. Repeat Steps from 2 to 6 until one of the above stoppage criteria is satisfied and then the adaptive wavelet stripping algorithm stops. Once the reconstructed signal \(y(n)\) is obtained, the mean spacing between two successive simulated transients can be calculated to detect the bearing fault types. The mean spacing can be calculated by the reciprocal of bearing fault characteristic frequency. Outer race fault characteristic frequency \(f_o\), the inner race fault characteristic frequency \(f_i\) and rolling element fault characteristic frequency \(f_b\) are given as \([2,4,6]\) 
\[
f_o = \frac{W \times 5}{2} \left(1 - \frac{d}{D \cos \theta}\right),
\]
\[
f_i = \frac{W \times 5}{2} \left(1 + \frac{d}{D \cos \theta}\right),
\]
\[
f_b = \frac{D \times 5}{d} \left(1 + \frac{d^2}{D^2 \cos^2 \theta}\right),
\]
where \(f_o, d, D, W,\) and \(\theta\) denote the shaft rotation frequency in Hz, the diameter of the rolling element, the pitch diameter of the bearing, the number of rolling elements and the contact angle, respectively.
4. Bearing fault transient extraction by the adaptive wavelet stripping algorithm

4.1. Case study 1: transient extraction from experimental fault signals

The bearing fault signals provided by Case Western Reserve University (CWRU) [36] were used to validate the effectiveness of the proposed method. The physical experimental platform is described in Fig. 2, where it includes a 2 horsepower (hp) motor (left), a torque transducer and a dynamometer (right). The motor shaft was supported by the bearings with the type of 6205-2RS JEM SKF. In the experiments, a single point fault was artificially induced to each normal bearing with electro-discharge machining. The bearing size information is given in Table 1. The motor speed and load were 1750 rev/min and 2 hp. The
sizing frequency was equal to 12,000 Hz. Bearing outer race fault characteristic frequency and bearing inner race fault characteristic frequency were calculated as 104.6 Hz and 157.9 Hz provided in references [37,38]. The mean intervals of the fault impulses caused by the outer race faults and the inner race faults were 114 samples and 76 samples, respectively.

The original outer race fault temporal signal is plotted in Fig. 3(a). Its corresponding frequency spectrum is plotted in Fig. 3(b). The adaptive wavelet stripping algorithm was applied to process the outer race fault signal shown in Fig. 3(a) to extract the fault impulses caused by the outer race fault and the inner race fault and the inner race faults were 114 samples and 76 samples, respectively.

The original outer race fault temporal signal is plotted in Fig. 3(a). Its corresponding frequency spectrum is plotted in Fig. 3(b). The adaptive wavelet stripping algorithm was applied to process the outer race fault signal shown in Fig. 3(a) to extract the fault impulses caused by the outer race fault and the inner race fault impulses caused by the outer race fault and the inner race fault.

Fig. 4. Evolution of NMSE for processing outer race fault signal (Stoppage criterion 3 was automatically reached).

Fig. 5. Four ARLWs $\psi_{ARWL}(f_{i, opt}, \phi_{i, opt}, \gamma_{i, opt}, n)$, $i = 1, 2, 3, 4$ and reconstructed signal $y(n)$ in the case of bearing outer race fault signal: (a) $i = 1$; (b) $i = 2$; (c) $i = 3$; (d) $i = 4$; (e) reconstructed signal $y(n)$; (f) frequency spectrum of $y(n)$. The sampling frequency was equal to 12,000 Hz. Bearing outer race fault characteristic frequency and bearing inner race fault characteristic frequency were calculated as 104.6 Hz and 157.9 Hz provided in references [37,38]. The mean intervals of the fault impulses caused by the outer race faults and the inner race faults were 114 samples and 76 samples, respectively.

The original outer race fault temporal signal is plotted in Fig. 3(a). Its corresponding frequency spectrum is plotted in Fig. 3(b). The adaptive wavelet stripping algorithm was applied to process the outer race fault signal shown in Fig. 3(a) to extract the fault impulses caused by the outer race faults and the inner race faults.
transients. In order to enhance the computing efficiency, the subset $F$ of the discrete frequency $f$ was restricted to the frequency range of one of the resonant frequency bands $[3430:1:3470]$ by inspecting the frequency spectrum plotted in Fig. 3(b). The subset $T$ of the discrete delay time $\tau$ was set to $(1:1:4:51)$ according to the vibration signal length. The subset $Z$ of the discrete damping coefficient $\zeta$ was set to $(0.005:0.001:0.03) \cup (0.03:0.01:0.1) \cup (0.1:0.1:0.9)$ because higher resolutions at lower damping ratios were provided [28].

After the adaptive wavelet stripping algorithm was performed on the outer race fault signal, the evolution of the NMSE is depicted in Fig. 4, where Stoppage criterion 3 is automatically reached to stop the adaptive wavelet stripping algorithm. According

**Table 2**

Specifications of the parameters of each ARLW obtained by the proposed method for outer race fault signal.

<table>
<thead>
<tr>
<th>$i$</th>
<th>Frequency $f_{opt}^i$</th>
<th>Damping coefficient $\zeta_{opt}^i$</th>
<th>Delay time $\tau_{opt}^i$</th>
<th>Pearson’s correlation coefficient $\rho_{ARLW(f_{opt}^i,\zeta_{opt}^i,\tau_{opt}^i)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3430</td>
<td>0.04</td>
<td>5</td>
<td>0.8457</td>
</tr>
<tr>
<td>2</td>
<td>3470</td>
<td>0.06</td>
<td>115</td>
<td>0.7954</td>
</tr>
<tr>
<td>3</td>
<td>3430</td>
<td>0.03</td>
<td>233</td>
<td>0.7945</td>
</tr>
<tr>
<td>4</td>
<td>3430</td>
<td>0.04</td>
<td>348</td>
<td>0.7331</td>
</tr>
</tbody>
</table>

**Fig. 6.** Comparison between the original outer race fault signal plotted with the solid lines and the reconstructed signal plotted with the dash lines: (a) the entire signals; from (b) to (e) each segment of the entire signals.
to Eq. (20) and Stoppage criterion 3, 4 ARLWs were adaptively peeled from the original outer race fault signal. Each of the four ARLWs is plotted in Fig. 5(a)–(d), respectively. The frequency, the damping coefficient and the delay time of each ARLW are tabulated in Table 2, where Pearson's correlation coefficient between the real transient and its corresponding simulated transient is given. From the results shown in Table 2, it is found that the different ARLWs have the different frequencies, damping coefficients and delay time even though the four ARLWs are extracted from the same outer race fault signal. These different parameters also indicate that the real transients are not exactly same with each other. Recalling the idea of the traditional wavelet analysis, a wavelet with a fixed optimal scale (or the optimal center frequency and bandwidth) is used to simultaneously simulate all transients in a fault signal. Compared with this idea, each of the ARLWs obtained by the proposed method corresponds to only one real transient. The idea of the proposed method can more accurately match with each of the transients coming from one fault signal. The reconstructed signal \( y(n) \) consisting of the four ARLWs is shown in Fig. 5(e), in which the mean spacing between two successive transients was estimated as 114 samples, which maintains a high degree of consistency with regard to the theoretical bearing outer race fault interval of 114 samples. The frequency counterpart of the reconstructed signal \( y(n) \) is given in Fig. 5(f), in which it is seen that these transients come from the resonant frequency band. Therefore, the transients caused by the outer race localized fault are successfully extracted. In order to compare the original outer race fault signal with the reconstructed signal, the reconstructed signal plotted with the dash lines and the original outer race fault signal plotted with the solid lines are simultaneously shown in Fig. 6(a). Each of the transients is highlighted in Fig. 6(b)–(e), respectively. Besides, Pearson’s correlation coefficient between the original outer race fault signal and the reconstructed signal was calculated as 0.78 that demonstrates the reconstructed signal has a good linear dependence on the original outer race fault signal.

The original inner race fault signal and its corresponding frequency spectrum are shown in Fig. 7(a) and (b). According to the principle of the selection of the subsets illustrated previously, the subset \( F \) of the discrete frequency was set to \((2500:1:3500)\) and the subsets \( T \) and \( Z \) of the delay time and the damping coefficient were the same with those used in the case of the outer race fault signal analysis.

By applying the adaptive wavelet stripping algorithm to process the inner race fault signal, the evolution of the NMSE is depicted in Fig. 8, in which Stoppage criterion 1 was automatically activated to stop the adaptive wavelet stripping algorithm. 4 ARLWs are extracted from the original inner race fault signal and they are shown in Fig. 9(a)–(d). The specifications of the frequency, the damping coefficient and the delay time of each of the ARLWs and Pearson’s correlation coefficients between the original transients and each of the ARLWs are tabulated in Table 3, where it is found that the extracted ARLWs have high
Pearson’s correlation with the original real transients. In addition, each of the ARLWs has different parameters, which indicates that each real transient is somewhat different from each other. The traditional wavelet analysis used the optimal wavelet to simultaneously match all real transients in one fault signal. Therefore, the idea of the traditional wavelet analysis cannot match all transients well. However, the proposed method could peel each simulated transient from the original signal one by one. Each of the simulated transients corresponds to its own real transient. Their match is more accurate than that used in the traditional

![Graphs and Tables](image)

**Fig. 9.** Four ARLWs $\psi_{ARLW}(f_{opt}^{i}, \zeta_{opt}^{i}, \tau_{opt}^{i}, n)$, $i = 1, 2, 3, 4$ and reconstructed signal $y(n)$ in the case of bearing inner race fault: (a) $i = 1$; (b) $i = 2$; (c) $i = 3$; (d) $i = 4$; (e) reconstructed signal $y(n)$; (f) frequency spectrum of $y(n)$.

<table>
<thead>
<tr>
<th>$\psi_{ARLW}(f_{opt}^{i}, \zeta_{opt}^{i}, \tau_{opt}^{i}, n)$</th>
<th>Frequency $f_{opt}^{i}$</th>
<th>Damping coefficient $\zeta_{opt}^{i}$</th>
<th>Delay time $\tau_{opt}^{i}$</th>
<th>Pearson’s correlation coefficient $\rho_{ARLW}(f_{opt}^{i}, \zeta_{opt}^{i}, \tau_{opt}^{i}, n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>2855</td>
<td>0.07</td>
<td>158</td>
<td>0.9139</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>2832</td>
<td>0.027</td>
<td>304</td>
<td>0.8739</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>2904</td>
<td>0.05</td>
<td>82</td>
<td>0.8438</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>2938</td>
<td>0.1</td>
<td>226</td>
<td>0.8313</td>
</tr>
</tbody>
</table>

Pearson’s correlation with the original real transients. In addition, each of the ARLWs has different parameters, which indicates that each real transient is somewhat different from each other. The traditional wavelet analysis used the optimal wavelet to simultaneously match all real transients in one fault signal. Therefore, the idea of the traditional wavelet analysis cannot match all transients well. However, the proposed method could peel each simulated transient from the original signal one by one. Each of the simulated transients corresponds to its own real transient. Their match is more accurate than that used in the traditional...
wavelet analysis. In Fig. 9(e), the final reconstructed signal $y(n)$ includes 4 ARLWs. From the result shown in Fig. 9(e), the mean spacing between two successive transients is equal to 74 samples. The frequency spectrum of the reconstructed signal $y(n)$ is plotted in Fig. 9(f), where it is similar with the frequency spectrum of the original inner race signal. Consequently, the transients caused by the bearing inner race fault were extracted.

For further comparison, the reconstructed signal plotted with the dash lines and the original inner race fault signal plotted with the solid lines are simultaneously shown in Fig. 10. Pearson’s correlation coefficient between the original inner race fault signal and the reconstructed signal was calculated as 0.86 that illustrates the transients extracted by the proposed method match with the real transients well.

In conclusion, the adaptive wavelet stripping algorithm not only automatically indicates the random occurrence of the real transients but also matches with the local characteristics of the real transients well.

4.2. Case study 2: transient extraction from an industrial fault signal

An industrial multi-fault signal including the rotor eccentric fault and the bearing outer race fault was used to further validate the adaptive wavelet stripping algorithm. The multi-fault signal was collected from a typical traction motor. The schematic diagram is illustrated in Fig. 11(a) where the gearbox is uncoupled with the motor and two rolling element bearings are used. One was a single row deep groove ball bearing (SKF 6215) at the drive end (denote its position as B). Another was a single row cylindrical roller bearing (NU 210) at the non-drive end (denote its position as A). Four accelerometers from four different directions (positions C, D, E
Fig. 11. Description of the experimental platform: (a) schematic diagram of the traction motor; (b) the real view of the traction motor.

and F) were used to sample the vibration signals. Fig. 11(b) shows a real view. Vibration signals were amplified by a coupler (Kistler 5134), and a digital cassette recorder (Sony PC 204AX) was utilized to record the signals at the sampling rate of 48 kHz for each of the channels. Then, the signals were transmitted to the PC with a data acquisition card at sampling rate of 32.77 kHz. The vibration data captured by the accelerometer closest to the bearing location (position D) was used for the analyses. The shaft rotation speed \( f_r \) was around 1498 rev/min. The outer race fault characteristic periodic interval approximated to 327 samples.

The original multi-fault signal and its frequency spectrum are shown in Fig. 12(a) and (b), where it is obvious to see that the original multi-fault signal is dominated by the rotor eccentric fault. The outer race fault signal is difficult to be identified. References [4,6] suggested that it was better to remove the low-frequency periodic components prior to bearing fault signal analysis. In this paper, the autoregressive (AR) filtering [4,6] was conducted on the original signal. The reasons for the use of the AR filtering are given as follows. Firstly, the aim of the AR filtering is to remove the low-frequency components caused by the rotor eccentric fault. Second, the low-frequency periodic components are deterministic and can be well established by the AR filter. At last, after the AR filtering is performed, the signal to noise ratio of the residual signal is low so that the proposed algorithm can be validated by a fault signal overwhelmed by heavy noise. The order of AR was chosen as 54 by minimizing Akaike information criterion (AIC). The residual signal containing the outer race fault signal is plotted in Fig. 12(c), where even though the low-frequency periodic components were removed by the AR filtering the transients are difficult to be identified. The frequency spectrum of the residual signal is plotted in Fig. 12(d).

Taking the computing efficiency into account, the subset \( F \) of the frequency was set to \{11770:50:12370\} by inspecting the frequency spectrum of Fig. 12(b). The subset \( T \) of the delay time was set to \{1:1:1661\} in terms of the signal length. The subset \( Z \) of the damping coefficient was the same with those used in the previous case studies. The adaptive wavelet stripping algorithm was applied to the residual signal obtained by the AR filtering. Fig. 13 shows that the adaptive wavelet stripping algorithm is terminated when Stoppage criterion 3 is automatically reached. Then, 5 ARLWs were obtained. Each of them is depicted in Fig. 14(a)–(e), respectively. The specifications of each of the ARLWs are tabulated in Table 4, where Pearson's correlation coefficients are not high. The reason for the low Pearson's correlation coefficient is that the real transients caused by the outer race localized fault were overwhelmed by the heavy noise introduced by the AR filtering. The reconstructed signal is plotted in Fig. 14(f), in which the mean spacing around 327 samples is estimated so that the transients caused by the outer race localized fault is extracted. In order to further validate this conclusion, the frequency spectrum of the reconstructed signal \( y(n) \) is plotted in Fig. 14(g). In order to distinguish the difference between the residual signal containing the outer race fault signal and the reconstructed signal, both of them are plotted in Fig. 15(a) with the different line styles and their local characteristics are zoomed in Fig. 15(b)–(f). Pearson's correlation coefficient between the residual signal containing the outer race fault signal and the reconstructed signal is 0.27, which is lowered by the heavy noise. Nevertheless, from the local zoom of the results, it is seen that each of the ARLWs is well matched with the real transients corrupted by heavy noise.

5. Comparison with the periodic multi-transient model

In this section, a comparison with the periodic multi-transient model proposed by Wang et al. [28] was conducted to show the superiority of the proposed method. The same parametric model of anti-symmetric real Laplace wavelet was used for fair comparison. Pearson's correlation coefficients between the residual signal shown in Fig. 12(c) and the parametric model of
The first five largest Pearson's correlation coefficients are highlighted by the solid circles from A to E in Fig. 16(a), where the locations of the solid circles A to E are the 1373th sample, the 1050th sample, the 393th sample, the 67th sample and the 719th sample. Recalling the delay time given in Table 4, the locations of the solid circles from A to E are completely matched with the delay time obtained by the proposed method. Then, the anti-symmetric real Laplace wavelet corresponding to the largest Pearson's correlation coefficient at the location of the circle A is reconstructed in Fig. 16(b). In the work of Wang et al. [28], they supposed that the bearing fault signal was absolutely periodic. Therefore, a periodic multi-transient model was introduced based on the wavelet filtering analysis.

In order to establish the periodic value $T$ (unit: sample), Pearson's correlation coefficients between the residual signal shown in Fig. 12(c) and the periodic multi-transient model at different periodic values are shown in Fig. 16(c), where the periodic interval (323 samples) corresponding to the largest Pearson's correlation coefficient highlighted by the solid circle F is found and the largest Pearson's correlation coefficient is equal to 0.15. The periodic multi-transient model is plotted in Fig. 16(d). Although the periodic multi-transient model approximately indicates the possible occurrence of the real transients, the random characteristics of these real transients can not be accurately described. It means that the accurate locations of the real transients can not be simultaneously represented by the locations of the simulated transients of the periodic multi-transient model. The reasons are explained in details below: (1) the bearing fault signal is not periodic but random due to the random slippage of the roller elements [4,6,7]. In order to clearly see the real transients in Fig. 17, the real transients are highlighted by the dash circles, where it is found that only two real transients are well matched with the simulated transients; (2) only a Laplace wavelet with the fixed optimal frequency and damping coefficient is used in the periodic multi-transient model. In other words, except the amplitudes...
and the delay time of the simulated transients used in the periodic multi-transient model, the other parameters, such as the frequency and the damping coefficient, of the simulated transients used in the periodic multi-transient model are same. This is not true for the real transients because all transients are just the realizations of the outside random process. The above two points can also be used to clarify the reason why Pearson’s correlation coefficient of 0.15 obtained by the periodic multi-transient model is lower than that of 0.27 obtained by the proposed method.

Fig. 14. Five ARLWs $\psi_{ARLW}(f_{opt}^{i}, \zeta_{opt}^{i}, \tau_{opt}^{i}, n)$, $i = 1, 2, 3, 4$ for the residual signal containing the industrial outer race fault signal: (a) $i=1$; (b) $i=2$; (c) $i=3$; (d) $i=4$; (e) $i=5$; (f) reconstructed signal $y(n)$; (g) frequency spectrum of $y(n)$.
6. Discussion on improving the computing efficiency of the adaptive wavelet stripping algorithm

The limitation of the adaptive wavelet stripping algorithm is that the computing time is extensive. In order to boost the computing efficiency, a strategy is proposed in this section. The original adaptive wavelet stripping algorithm with the proposed strategy is called an enhanced adaptive wavelet stripping algorithm. The details of the strategy is illustrated as follows. Assume one of the stoppage criterion are not satisfied. It means that Steps 2 to 6 of the proposed algorithm will be repeated until one of the stoppage criteria is reached. Recalling Eq. (15), it is found that only part of the signal $x(n)$ is

Table 4
Specifications of the parameters of each ARLW obtained by the proposed method for the residual signal containing the industrial outer race fault signal.

<table>
<thead>
<tr>
<th>$\Psi_{\text{ARLW}}(f_{\text{opt}}^{(i)}, \zeta_{\text{opt}}^{(i)}, \tau_{\text{opt}}^{(i)}, n)$</th>
<th>Frequency $f_{\text{opt}}^{(i)}$</th>
<th>Damping coefficient $\zeta_{\text{opt}}^{(i)}$</th>
<th>Delay time $\tau_{\text{opt}}^{(i)}$</th>
<th>Pearson's correlation coefficient $\rho_{\text{ARLW}}(f_{\text{opt}}^{(i)}, \zeta_{\text{opt}}^{(i)}, \tau_{\text{opt}}^{(i)})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i=1$</td>
<td>11770</td>
<td>0.2</td>
<td>1373</td>
<td>0.5875</td>
</tr>
<tr>
<td>$i=2$</td>
<td>11770</td>
<td>0.1</td>
<td>1050</td>
<td>0.5825</td>
</tr>
<tr>
<td>$i=3$</td>
<td>11770</td>
<td>0.1</td>
<td>393</td>
<td>0.5825</td>
</tr>
<tr>
<td>$i=4$</td>
<td>11770</td>
<td>0.2</td>
<td>67</td>
<td>0.4791</td>
</tr>
<tr>
<td>$i=5$</td>
<td>12220</td>
<td>0.4</td>
<td>719</td>
<td>0.4791</td>
</tr>
</tbody>
</table>

Fig. 15. Comparison between the residual signal plotted with the solid lines and the reconstructed signal plotted with the red lines: (a) the entire signals; from (b) to (f) each segment of the entire signals.

6. Discussion on improving the computing efficiency of the adaptive wavelet stripping algorithm

The limitation of the adaptive wavelet stripping algorithm is that the computing time is extensive. In order to boost the computing efficiency, a strategy is proposed in this section. The original adaptive wavelet stripping algorithm with the proposed strategy is called an enhanced adaptive wavelet stripping algorithm. The details of the strategy is illustrated as follows. Assume one of the stoppage criterion are not satisfied. It means that Steps 2 to 6 of the proposed algorithm will be repeated until one of the stoppage criteria is reached. Recalling Eq. (15), it is found that only part of the signal $x(n)$ is
changed. As a result, only Pearson's correlation coefficients $\rho_{ARLW}(f, \zeta, \tau)$ at the samples $\tau \in [\tau_i^{opt} - W + 1, \tau_i^{opt} + 2W - 2]$ are needed to be recalculated. The samples for the recalculation of Pearson's correlation coefficients is shown in Fig. 18. In other words, not all delay time must be recalculated again. In order to evaluate the computing efficiency, a simple efficient index (EI) is proposed as follows:

$$EI = \frac{CT_O - CT_E}{CT_O} \times 100\%,$$

(24)

where $CT_O$ and $CT_E$ mean the computing time for the original and enhanced adaptive wavelet stripping algorithm, respectively. The comparison of the computing time for the original and enhanced adaptive wavelet stripping algorithm is tabulated in Table 5, where those signals used for the analyses are the same signals used in Section 4. The subsets are also the same with those used in Section 4. All the calculations were done using MATLAB installed in a desktop with Intel(R) Core(TM) i3 CPU, 3.2 GHz and 4 GB (2.99 GB usable) RAM. From the results shown in Table 5, it is seen that the enhanced adaptive wavelet stripping algorithm could significantly reduce the computing time.

Fig. 16. Results obtained by the periodic multi-transient model: (a) Pearson's correlation coefficients between the residual signal containing the industrial outer race fault signal and the parametric model of anti-symmetric real Laplace wavelet; (b) anti-symmetric real Laplace wavelet with the largest Pearson's correlation coefficient; (c) the correlation coefficients between the residual signal and the periodic multi-transient model based the wavelet shown in (b); (d) the periodic multi-transient model.
Fig. 17. Comparison between the residual signal plotted by the solid lines and the periodic multi-transient model plotted by the dash lines: (a) the entire signals; (b) to (f) each segment of the entire signals (note: real transients caused by the outer race localized fault are highlighted by the dash circles).

Fig. 18. Samples used for the recalculation of Pearson’s correlation coefficient.

Table 5
Comparison of the computing time for the original and enhanced adaptive wavelet stripping algorithm.

<table>
<thead>
<tr>
<th></th>
<th>CT₀ (s)</th>
<th>CTₑ (s)</th>
<th>EI (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer race fault signal</td>
<td>309.997</td>
<td>165.619</td>
<td>46.6</td>
</tr>
<tr>
<td>Inner race fault signal</td>
<td>4921.94</td>
<td>3038.183</td>
<td>38.27</td>
</tr>
<tr>
<td>Industrial fault signal</td>
<td>558.456</td>
<td>391.974</td>
<td>29.80</td>
</tr>
</tbody>
</table>
7. Conclusions

In this paper, a novel adaptive wavelet stripping algorithm was proposed to extract the simulated transients from the bearing fault signals. First, the parametric model of anti-symmetric real Laplace wavelet or impulse response wavelet was built to simulate the real transients caused by the bearing localized faults. Then, after one of the stoppage criteria was satisfied, all the simulated transients could be adaptively peeled from the bearing fault signals. The spatial reconstruction of the simulated transients revealed the random characteristics of the real transients embedded in the bearing fault signals. The mean spacing between two successive simulated transients could be used to identify the bearing fault types. At last, an enhanced adaptive wavelet stripping algorithm was developed to boost the computing efficiency.

To illustrate how the adaptive wavelet stripping algorithm could be used to extract the transients, two case studies including the bearing fault signals collected from an experimental machine and an industrial machine were investigated. For the experimental data, because the signal to noise ratio of the data was high, it was clearly seen that the real transients could be well matched with the simulated transients so that Pearson’s correlation coefficient between them was high. For the industrial data, the AR filtering was used to remove the low-frequency components. The residual obtained by the AR filtering had a very low signal to noise ratio. The benefit contributed by the proposed method was that the simulated transients could be adaptively extracted even though the real transients were overwhelmed by heavy noise. However, because the very low signal to noise ratio, Pearson’s correlation coefficient between the simulated transients and the real transients became low. It was concluded that the size of Pearson’s correlation coefficient depended on the signal to noise ratio.

This paper has five main contributions. First, the parametric model of anti-symmetric real Laplace wavelet is well matched with the real transients caused by bearing localized faults. Second, the proposed algorithm adaptively peels the simulated transients from the bearing fault signal. Third, the spatial reconstruction of the simulated transients is better than the periodic multi-transient model proposed by Wang et al. [28] to reflect the random characteristics of the real transients which are associated with the random slippage of the rolling elements. Fourth, the case studies show that the proposed method can extract the simulated transients from the bearing fault signals with a high or low signal to noise ratio. At last, the enhanced adaptive wavelet stripping algorithm can reduce the computing time of the original one.

Acknowledgments

The work described in this paper was partly supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project no. CityU 122011) and by a grant from City University of Hong Kong (Project no. 7008187). We wish to thank Professor K.A. Loparo for his permission to use the bearing fault signals for validating the proposed method. At last, we would like to express our deepest appreciation for the valuable comments provided by two anonymous reviewers to improve this paper.

References