ISA Transactions xxx (xxxx) xxx

Contents lists available at ScienceDirect

ISA Transactions

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

Practice article

Rolling element bearing fault identification using a novel three-step adaptive and automated filtration scheme based on Gini index

Muhammad N. Albezzawy [a,*], Mohamed G. Nassef [b], Nader Sawalhi [c]

[a] Production Engineering Department, Faculty of Engineering, Alexandria University, Egypt
[b] Industrial and Manufacturing Department, School of Innovative Design Engineering, Egypt-Japan University of Science and Technology, Egypt
[c] Vipac Engineers & Scientists Ltd., Australia

ARTICLE INFO

Article history:
Received 6 August 2019
Received in revised form 10 January 2020
Accepted 10 January 2020
Available online xxxx

Keywords:
Rolling bearing
Fault detection
Morlet wavelet
Gini index
Autoregressive filtration
Deconvolution

ABSTRACT

For early detection of rolling element bearings (REBs) faults in contaminated signals, kurtosis-derived indices are involved in the filtration process prior to demodulation. However, they were found either sensitive to impulsive outliers or requiring many input arguments. In this study, a novel three-step adaptive and automated filtration scheme using Gini index (GI) is proposed as an alternative to kurtosis-based techniques to enhance the weak fault features and eliminate noise and interreferences from the raw vibration signal. The proposed approach was tested using experimental signals with different bearing faults. The filtered signals were greatly denoised and the fault impulses were successfully isolated, which indicates the effectiveness of the proposed approach and the superiority of GI over kurtosis-derived indices as a criterion for proper filter design for REBs fault detection.

1. Introduction

Rolling element bearings (REBs) are involved in about every industrial application because they are vital elements in the structure of about every rotating machinery [1]. As a consequence, REBs failure leads to a major machine breakdown in many cases [2]. Early detection of REBs faults and their surveillance are essential in predictive maintenance of critical applications, such as energy systems and aerospace [3]. REBs fault classification (fault location identification) could provide useful information about failure root causes so that they may be eliminated, and more reliable designs may be developed.

Vibration signals are successfully used to indicate the condition of rotating machinery. However, REBs faults at early stages produce weak fault features in the vibration signal. Such weak features are usually overwhelmed with noise of nearby equipment to the extent that they are difficult to extract [4]. The situation becomes harsher in the case of multipart equipment such as in helicopter gearboxes [5]. The raw spectra of such vibration signals are typically crowded and have extremely low signal to noise ratios (SNRs) [6]; therefore, it is hard to identify the REBs fault characteristic frequencies (BCFs) within them.

Each time a rolling element hits a localized defect, an impulse is naturally produced which excites the system resonances. Hence, the success of a diagnostic technique is judged by its ability to isolate these repetitive impulses [1]. There are interferences that disrupt the isolation of these impulses. Examples of such interferences are the impulsive electromagnetic noise, the periodic interference of the shaft rotation and its harmonics, and gear meshing vibrations present in the case of gearboxes [7].

Envelope analysis is the typically used technique to extract these impulses through resonant band demodulation. It is based on the fact that the occurrence frequency of these impulses "BCF" amplitude-modulates the bearing system resonance [6]. For this technique to be successful, the proper resonant frequency band must be initially identified [8]. The resonant band selection becomes difficult due to the effect of the variation in the signal transmission path and due to the existence of heavy noise [9]. Traditionally, the selection of the useful resonant band was done visually by inspecting the difference in the power spectral density between a signal of a normal REB and the signal of a faulty one. This approach requires a historical no-fault data [5]. Through the past decade, the developed approaches were mainly based on the blind selection of the useful resonant band for demodulation through impulsiveness maximization of the filtered signal [5]. Impulsiveness maximization is achieved using sparsity measures of the signal, such as kurtosis and Shannon entropy.

One of the earliest techniques was the Antoni and Randall's kurtogram [10], in which the optimal band-pass filter was determined through maximizing the spectral kurtosis (SK) of the filtered signal. Bozchalooi and Liang [11] pointed out that kurtosis is sensitive to data outliers, so it may falsely indicate impulsiveness. They alternatively proposed the smoothness index

* Corresponding author.
E-mail addresses: m_nabil@alexu.edu.eg (M.N. Albezzawy), mohamed.nassef@alexu.edu.eg (M.G. Nassef), nadersaw@gmail.com (N. Sawalhi).

https://doi.org/10.1016/j.isatra.2020.01.019
0019-0578/

which possesses minimum values for maximum impulsiveness and showed to be less sensitive to outliers. Shannon entropy minimization was successfully used in [12,13] for the same purpose. Tse and Wang [14] conducted a comparative study between the use of kurtosis, the smoothness index, and Shannon entropy. They found that the three measures give similar results based on their studied signals. McDonald et al. [15] found that kurtosis fails to distinguish between impulsive noise and fault repetitive impulses because the kurtosis of a single impulse may be higher than that of a train of impulses. They then proposed the correlated kurtosis (CK), which considers the impulse's periodicity, to solve this problem. Chen et al. [9] formulated an improved correlated kurtosis (ICK) to take the factor of the signal length into consideration. The main drawback of the CK and the ICK is that they require many input arguments, such as the correlation period which must be set equal to the BCF. This step is substantially impractical as information about which BCF, among all the BCFs, exists in the faulty REB's signal is unknown. In case the information is available, then the fault location is already identified and there is no need for filtration.

The former discussion indicates that the reliability of an automated fault detection technique is conditioned by using a proper criterion to select the appropriate resonant band for demodulation. This work exploits Gini index (GI) as a potential alternative to overcome the problems arise from using other indices such as spectral kurtosis in incipient fault detection of REBs in noisy systems, e.g. helicopter gearbox, and wind turbines.

2. Gini index (GI)

GI is an economical wealth inequality measure that is latterly introduced to the field of REBs fault diagnostics [16]. It was first introduced in 1921 by Corrado Gini [17]. GI was also introduced in [18] as a more useful norm than kurtosis for an appropriate resonant band selection. The findings in [18] and [19] revealed the effectiveness of GI in extracting repetitive transients caused by faulty REBs. GI is found to be less sensitive than kurtosis to impulsive outliers caused by external impacts [18,19]. Hurley and Rickard [20] quantitatively compared several common sparsity measures (including kurtosis and Shannon entropy) based on some desirable features and GI was found to have all these features. Hence, GI is adopted in this work as the filter selection criterion. GI is calculated according to the derived formula in [20] as in Eq. (1). The values of GI ranges from 0 to 1 [16].

Given a vector $x = [x(1), \ldots, x(N)]$ with its elements re-ordered and represented by: $x_{[k]}$ for $k = 1, 2, \ldots, N$, where $|x_{[1]}| \leq |x_{[2]}| \leq, \cdots, \leq |x_{[N]}|$, then:

$$GI(x) = 1 - 2\sum \frac{|x_{[k]}|}{x_1} \times \left(\frac{N - k + \frac{1}{2}}{N}\right) \quad (1)$$

where x_1 is the l_1 norm of x.

2.1. Gini index versus kurtosis

Simulated primary signals were generated in this work to test the response of GI and to compare it with the response of kurtosis to the features in these signals.

The first signal (Sig 1) is a DC signal with constant amplitude which equals unity and is illustrated in Fig. 1. The second signal (Sig 2) contains a single pulse and is illustrated in Fig. 2. The third signal (Sig 3) contains two pulses and is illustrated in Fig. 3. The fourth signal (Sig 4) contains a pulse train consisting of nine pulses and is illustrated in Fig. 4. The fifth signal (Sig 5) contains white Gaussian noise with −15 dB signal to noise ratio (SNR) and is illustrated in Fig. 5. The last signal (Sig 6) is a sinusoid with 5 Hz frequency and 1 kHz sampling frequency and is illustrated in Fig. 6. The value of GI for each of these signals is calculated and listed against the values of kurtosis in Table 1.

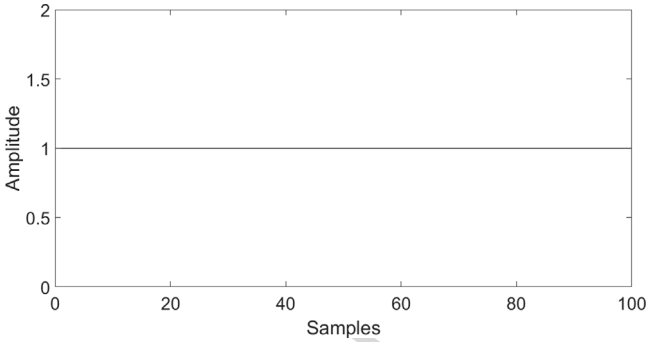

Fig. 1. Sig 1: a DC signal.

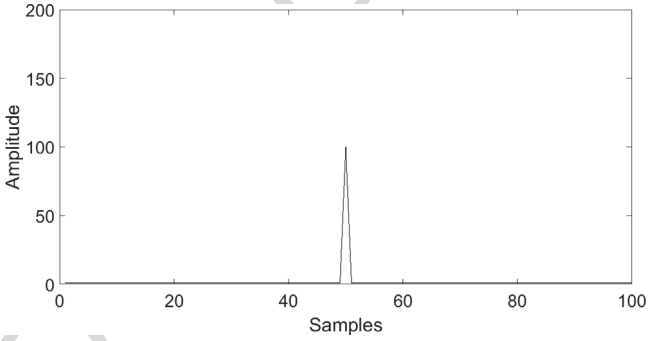

Fig. 2. Sig 2: a single pulse.

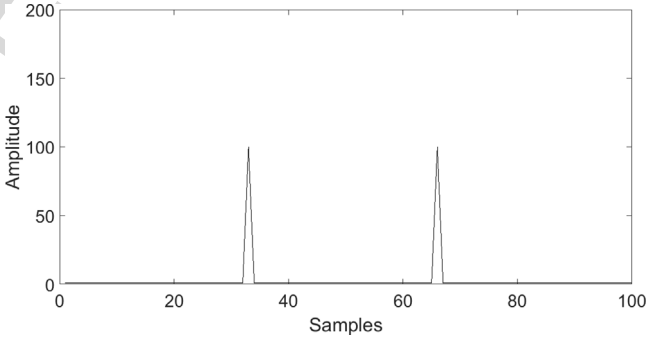

Fig. 3. Sig 3: two pulses.

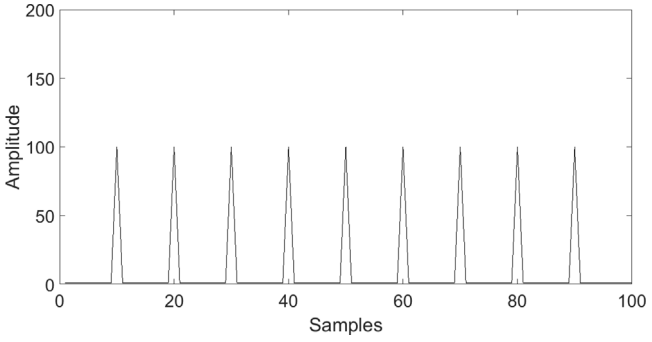

Fig. 4. Sig 4: a pulse train.

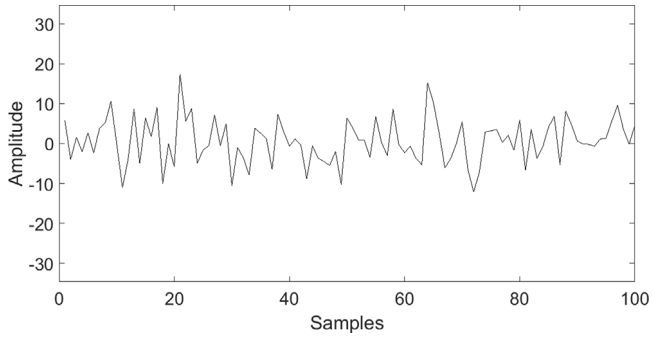

Fig. 5. Sig 5: white Gaussian noise with SNR = −15 dB.

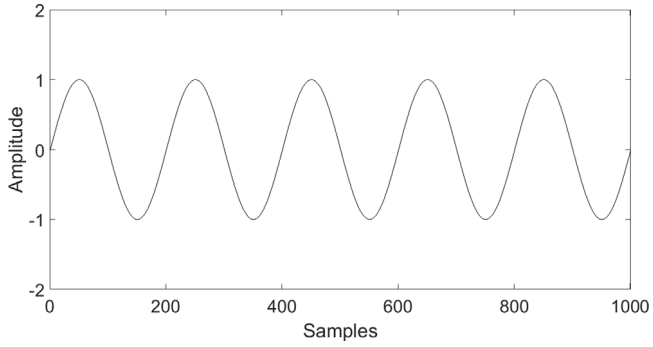

Fig. 6. Sig 6: a sinusoid.

Table 1
GI vs. kurtosis of the six test signals.

Signal	Kurtosis	GI
Sig 1: DC	NaN	3.3307×10^{-16}
Sig 2: a single pulse	98.0101	0.4925
Sig 3: two pulses	48.0204	0.6511
Sig 4: a pulse train	9.21	0.8182
Sig 5: white noise	4.0883	0.4339
Sig 6: a sinusoid	1.934	0.274

It can be seen from Table 1 that GI possesses lower values for DC, sinusoid and white noise signals while peaks at higher values in the cases of impulsive signals (Sig 2–4). The more are the impulses in the signal, the higher GI gets. In other words, GI value is found directly proportional with the number of pulses in the signal. GI reaches its highest value in the case of the pulse train signal (Sig 4), what makes it a perfect index for REBs fault detection since REBs faults produces a pulse train in the time domain. Although, kurtosis maintains lower values for sinusoids and white noise, its value is the highest for the single pulse and is lower for the pulse train. Kurtosis value is inversely proportional with the number of pulses in the signal what makes it an improper indicator for REBs faults. Considering that, SK is the kurtosis of the discrete Fourier transform of a signal [21]. This numerically demonstrates that GI is more appropriate than kurtosis and SK for REBs fault detection. Another merit of GI is that it does not require many input arguments compared to the CK and the ICK.

3. REBs fault detection challenges and the proposed methodology

The reviewed literature highlights the main difficulties encountered during isolating the REBs fault's repetitive impulses. These difficulties can be summarized and listed as follows:

1. Existence of periodic interference and gearmesh signals
2. Masking by white noise and transmission path effects
3. Contamination with impulsive noise and impulsive outliers

In order to tackle these difficulties, Sawalhi et al. [1] proposed a three-step approach combining autoregressive (AR) model-based linear prediction filtering, minimum entropy deconvolution (MED), and Morlet wavelet band-pass filtration using SK. However, the whole approach was based on maximizing the kurtosis of the filtered signal which possesses low immunity to impulsive noise and could lead to misleading diagnostics.

The objective of this manuscript is to establish an automated and adaptive REBs fault identification framework based on successive enhancements of the weak fault feature through three filtration steps analogous to those in [1], but all are based on GI which demonstrated to be more indicative for faulty REBs signals than kurtosis and SK and proved to be immune to impulsive noise and outliers while not requiring many input arguments compared to the CK and the ICK.

The three-step filtration scheme is as follows;

1. Inverse autoregressive (IAR) linear prediction filtration to remove the deterministic parts from the measured vibration signal (such as gears meshing frequencies and periodic interference) [1,22]. In this work, this is achieved innovatively based on maximum GI.
2. Maximum Gini index deconvolution (MGID), as a novel proposed improvement to the well-known MED technique, to remove white noise and the effect of the transmission path [22].
3. GI-guided optimum Morlet wavelet band-pass filtration to adaptively select and filter the optimum pass band for demodulation.

The rest of this paper is structured as follows: Section 4 introduces the IAR linear prediction filtration. Section 5 presents the proposed MGID technique. GI-guided optimum Morlet wavelet filtration procedure is explained in Section 6. Section 7 demonstrates the effectiveness of the proposed scheme in early REB fault detection and identification using three experimental signals for three different fault locations (ball fault, inner and outer-race faults). The three used signals are taken from the gearbox test rig of the University of New South Wales (UNSW) and are contaminated with gearmesh frequencies. Section 8 states the main conclusions.

4. Inverse AR linear prediction filtration

In AR process modeling, the current value of the signal $x(n)$ is presented as a linear combination of its preceding values plus a prediction error (which is a white noise process) $e(n)$ [23], according to Eq. (2).

$$x(n) = -\sum_{k=1}^{p} a(k) z^{-k} x(n) + e(n) \tag{2}$$

where p represents the number of previous terms (the model order); $a(k)$ are weighting coefficients; and z^{-k} is the back-shift operator. Hence:

$$x(n) = e(n) \Big/ 1 + \sum_{k=1}^{p} a(k) z^{-k} \tag{3}$$

It is understood from Eq. (3) that $x(n)$ can be generated by passing white noise through an all-pole filter with transfer function [24].

$$H(z) = 1 \Big/ 1 + \sum_{k=1}^{p} a(k) z^{-k} \tag{4}$$

Then, the prediction-error filter (all-zero) has the transfer function in Eq. (5) and its response is a white noise series $e(n)$. In this configuration, the prediction-error filter (IAR filter) whitens the input signal, $x(n)$, and is called a noise-whitening filter [24].

$$A(z) = 1/H(z) = 1 + \sum_{k=1}^{p} a(k) z^{-k} \tag{5}$$

If $x(n)$ is the vibration signal to be analyzed, the AR model expects the deterministic part of the signal (gearmesh signal and the periodic components). However, it cannot adapt to the sudden impulses caused by a localized fault (e.g. a REB fault). Therefore, the fault impulses will be contained in the prediction error term $e(n)$ of the AR model [1].

Accordingly, the IAR filter removes the deterministic (AR) components of the signal and leaves the residual $e(n)$ which consists of white noise and the impulsive feature of a REB fault of the input signal.

The AR model parameters are obtained through Yule–Walker method using a Matlab® function called "aryule" which uses the Levinson–Durbin recursion (LDR) algorithm. The Akaike information criterion (AIC) [22] and kurtosis maximization of the residual signal [1] were used in the literature as criteria for selecting the optimum model order p. In this paper, GI maximization of the squared envelope (SE) of the residual signal (IAR filtered signal) is used in selecting the optimum model order to enhance impulses separation from the original signal while being less sensitive than kurtosis for impulsive noise and outliers.

5. Maximum Gini Index Deconvolution (MGID)

The IAR filter is phase blind, i.e. it cannot differentiate between noise and impulses as both have broadband excitation and both will remain in the filtration result $e(n)$ [22]. Based on the fact that REBs' vibration signal is convolved by the transmission path function [25] MED was utilized in [1,22] to deconvolve the transmission path effect to enhance the impulsive component.

MED is a system identification technique that was initially established by Wiggins [26] to assist in extracting reflectivity information in seismic data. It was proven to be effective in isolating the impulsive features from a combination of signals [26]. MED optimizes the filter coefficients based on maximizing kurtosis of the output signal of an inverse filter. MED is similar to IAR filtration (which uses second-order statistical properties) but with the use of Higher-order statistical properties [22].

The AR model residual signal $e(n)$ obtained by the IAR filter can be modeled as in Eq. (6). The signal $g(n)$ models the fault impulses, $\mathcal{W}(n)$ models the noise, and the finite impulse response (FIR) filter $h(n)$ models the transmission path effect [1].

$$e(n) = (g(n) + \mathcal{W}(n)) * h(n) \tag{6}$$

The objective of the deconvolution is to find the coefficients of the inverse filter $f(n)$ which attains $h(n) * f(n) = \delta(n - l_m)$, where l_m is a delay to make the inverse filter causal [1].

McDonald [27] implemented an iterative procedure based on the work of Wiggins [26] to select the MED filter based on Kurtosis maximization of the MED filtered signal. McDonald et al. [15] proposed the maximum correlated kurtosis deconvolution (MCKD), as an improvement for MED, based on the correlated kurtosis which counts for the periodic nature of the fault's signal. Thus, it is less sensitive to outliers than kurtosis. MCKD encounter some challenges such as: the strict requirement for many input parameters and the intricate resampling process [28].

In this work, we propose a novel improvement for MED, based on maximizing the GI of the filtered signal. The method is named maximum Gini index deconvolution (MGID). GI proved to be more proper for REBs fault detection than kurtosis and SK because it shows less sensitivity to individual impulses and outliers. GI is also an efficient alternative for CK because it does not require for multi-input parameters and resampling.

The MGID filter is implemented by a method similar to the iterative procedures in [22,27]. GI maximization of the squared envelope (SE) of the filtered signal is selected as the fitness function. This is done by changing the values of the coefficients of the MGID filter and computing the GI. The optimization process terminates when the change in the GI reaches a prespecified value. In this work, this value is set to 0.01.

6. The GI-guided adaptive Morlet wavelet filter

The complex Morlet wavelet function is a complex Gaussian-enveloped sinusoid [29], as shown in Eq. (7). Morlet wavelet's shape approximates the vibration response of a mechanical impact [30].

$$\psi(t) = \sqrt{\frac{\pi}{2\ln 2}} \cdot F_b \cdot e^{-\frac{\pi^2 \cdot F_b^2 \cdot t^2}{2\ln 2}} \cdot e^{i2\pi \cdot F_c \cdot t} \tag{7}$$

where F_c and F_b are the center frequency and the bandwidth of Morlet wavelet, respectively. The continuous wavelet transformation can be understood as a band-pass filtration, hence, the Morlet wavelet filter is simply a band-pass filter and its pass-band is defined as $[F_c - F_b/2, F_c + F_b/2]$ [29]. For a successful demodulation process, the filter parameters F_c, F_b must be carefully selected to get a filtered signal with maximum impulsiveness.

In this work, a heuristic inspired by the well-known particle swarm optimization (PSO) algorithm is specially built to blindly optimize the Morlet wavelet filter parameters based on maximizing the GI of the squared envelope (SE) of the wavelet-filtered signal. The ranges of F_c and F_b are constrained by the conditions stated in [29]. In the PSO algorithm, the swarm size was 50, and the number of iterations was 20.

7. Experimental work and results

7.1. Test rig description

The experimental setup under consideration is the spur gear test rig of the UNSW. It was previously used by Sawalhi [5] to validate a simulated model for faulty REBs' response in a gearbox environment. In the setup, several faults of different types were presented to several REBs.

The test was performed under a load of 50 N m, while the shaft speed was set to 10 Hz. An accelerometer was located on the defective bearing housing to gather vibration data with a sampling frequency of 48 kHz. The test rig is depicted in Fig. 7. The test bearing was double row self-aligning ball bearing (Koyo 1205). Three different types of faults are shown in Fig. 8.

7.2. Results and discussion

Three signals of the three different localized faults (ball, inner race and outer race faults) have been used here to validate the proposed method. The BCFs were estimated for the ball spin as (BSF = 26.5 Hz), the inner race as (BPFI = 71.1 Hz), the outer race as (BPFO = 48.9 Hz) and the fundamental train as (FTF = 4.1 Hz). The signals were processed by the proposed filtration scheme and the results are presented and discussed in the following subsections.

Please cite this article as: M.N. Albezzawy, M.G. Nassef and N. Sawalhi, Rolling element bearing fault identification using a novel three-step adaptive and automated filtration scheme based on Gini index. ISA Transactions (2020), https://doi.org/10.1016/j.isatra.2020.01.019.

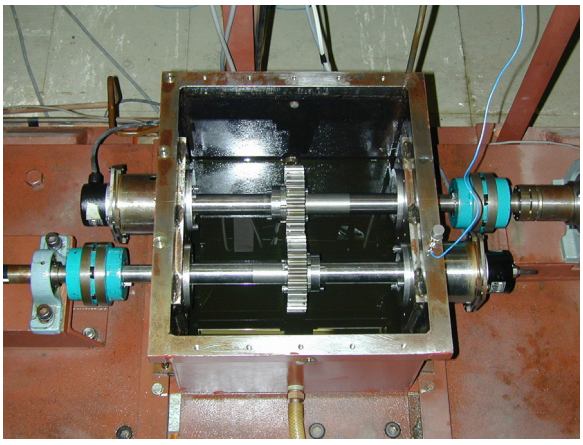

Fig. 7. UNSW gearbox test rig.

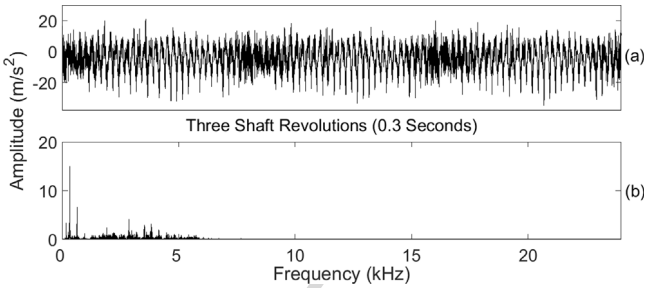

Fig. 9. Signal A: (a) time waveform; (b) spectrum.

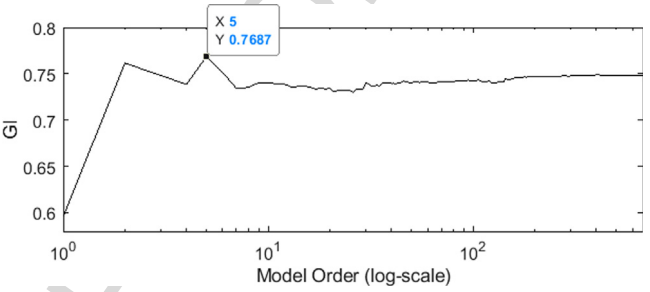

Fig. 10. Optimum IAR filter's model order for Signal A.

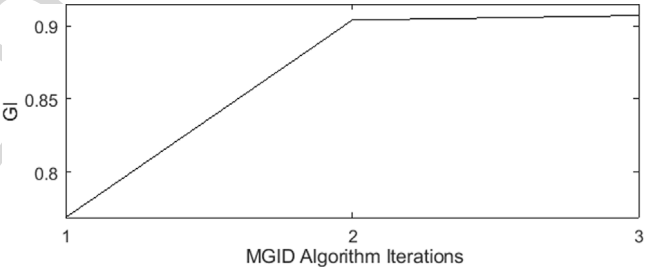

Fig. 11. Convergence curve of the MGID algorithm for Signal A.

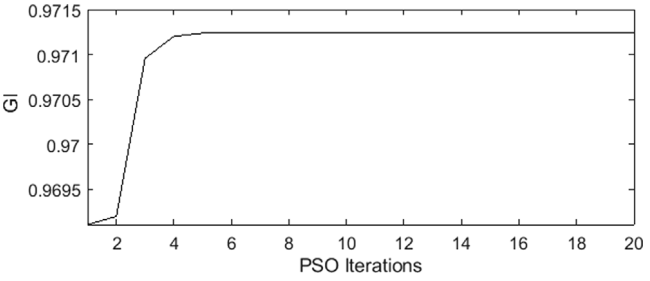

Fig. 12. Convergence curve of PSO of the Morlet wavelet filter for Signal A.

7.2.1. The Ball fault signal (Signal A)

The time waveform and spectrum of the ball fault raw signal (Signal A) are shown in Fig. 9(a, b). The GI of the SE of the raw signal is 0.5029. It can be seen that REBs fault features cannot be detected either in the time domain or in the frequency domain due to masking by the gearmesh frequencies. No impulsive behavior is detected in the time domain, Fig. 9(a). It is known that REBs faults at the very early stages excite high frequency vibration in both the ultrasound spike energy and the bearing resonance frequencies regions of the vibration spectrum [31]. Even though, the spectrum in Fig. 9(b) does not show any frequency content in the high frequency region.

IAR filtration is performed on the raw signal to eliminate the gearmesh frequencies. The optimum model order is selected based on maximizing the GI of the squared envelope (SE) of the IAR filtered signal and it is found to be 5 (Fig. 10). The SE of the IAR filtered signal has a GI of 0.7687. The IAR filtered signal in time domain is shown in Fig. 14(a). The gearmesh frequencies are removed and the impulses of the ball fault are revealed. The squared envelope spectrum (SES) of the IAR filtered signal is shown in Fig. 15(a), which shows a clear peak at twice the ball spin frequency ($2 \times BSF$) and its harmonics and sidebands with the shaft rotation frequency. At first sight, the SES could mislead the diagnosis towards inner race fault because the presence of sidebands with the shaft rotation frequency is a characteristic for inner race fault signals not for ball fault signals. However, the peak at ($2 \times BSF$) and its harmonics indicate the presence of a ball fault.

The IAR filtered signal is then passed through the proposed MGID filter to remove any transmission path effects. The iterative filtration procedure is designed to maximize the GI of the SE of the deconvolved signal. The convergence of the MGID algorithm is shown in Fig. 11. The GI of the SE of the (IAR+MGID) filtered signal is 0.9072.

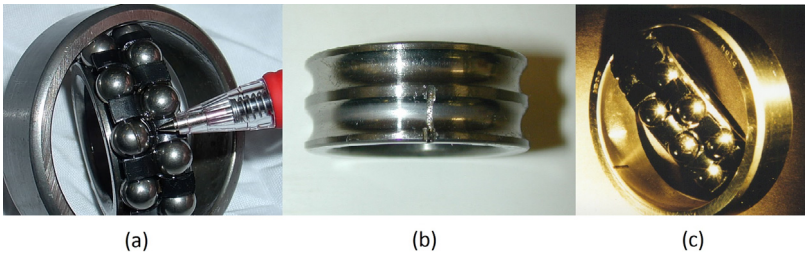

Fig. 8. Three different types of localized faults: (a) ball fault; (b) inner race fault; (c) outer race fault.

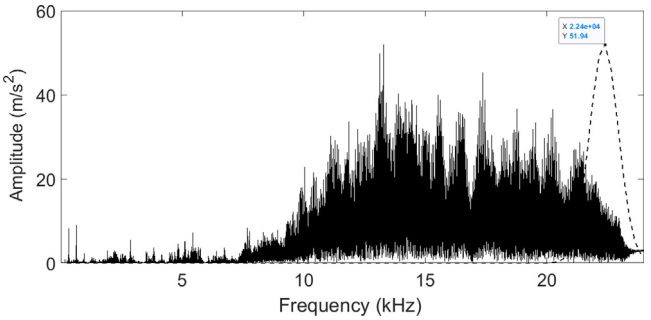

Fig. 13. The optimum Morlet wavelet filter superimposed on the frequency spectrum of the (IAR+MGID) filtered signal.

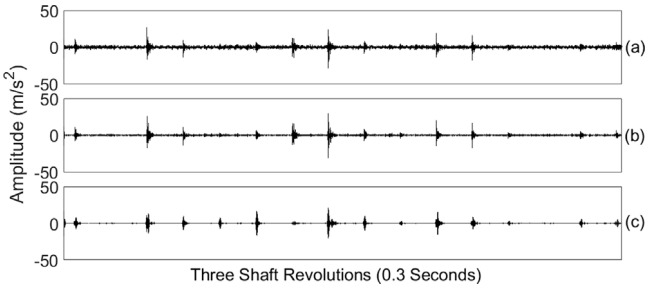

Fig. 14. Signal A filtration output: (a) IAR filtration output of the raw Signal A; (b) MGID filtration output of the signal in "a"; (c) Optimum Morlet wavelet filtration output of the signal in "b".

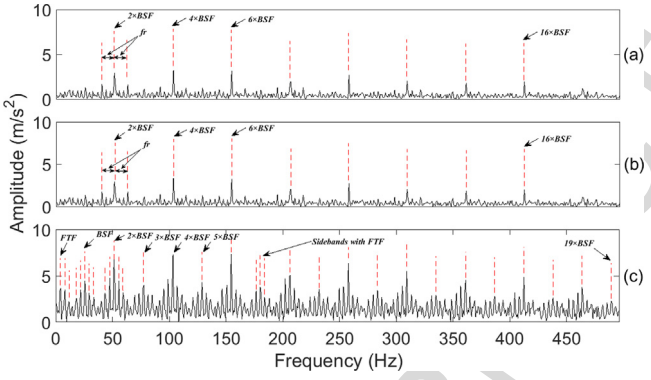

Fig. 15. The SESs of: (a) the signal in Fig. 14(a), (b) the signal in Fig. 14(b), (c) the signal in Fig. 14(c).

Fig. 13 shows the spectrum of the (IAR+MGID) filtered signal, which demonstrates a tremendous amplification of the amplitudes at the high frequency region (10–20 kHz). This conforms with the results of the power spectral analysis in [5], which showed a rise in the power spectral density corresponding to this frequency range in the case of the ball fault. The filtered signal in the time domain is shown in Fig. 14(b). The signal is denoised and the impulses clarity has been enhanced. The SES of the filtered signal is shown in Fig. 15(b). It does not show any further improvement beyond the resulted SES of the IAR filter alone because the fault impulses were originally well separated in the raw signal due to the relatively low shaft rotational speed.

Finally, the output of the previous step is passed through an optimum Morlet wavelet filter whose parameters were optimized based on the maximization of the GI of the SE of the filtered signal. The convergence curve of PSO of the Morlet wavelet filter is shown in Fig. 12. The center frequency and bandwidth of the filter were 22.24 kHz and 980.36 Hz, respectively. Fig. 13 shows

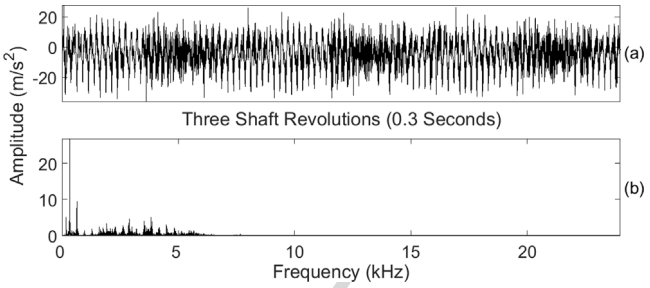

Fig. 16. Signal B: (a) time waveform; (b) spectrum.

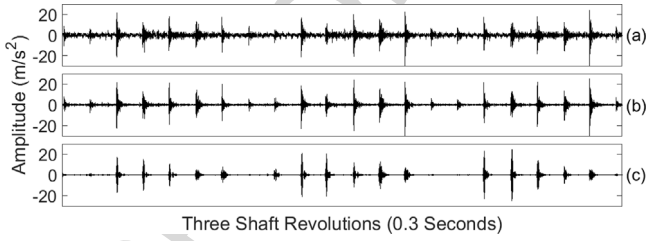

Fig. 17. Signal B filtration output: (a) IAR filtration output of the raw Signal B; (b) MGID filtration output of the signal in "a"; (c) Optimum Morlet wavelet filtration output of the signal in "b".

the optimum Morlet wavelet filter superimposed on the spectrum of the output of the previous steps. The GI of the SE of the output signal (after passing through the three filters) reaches 0.9712 which is a very high value compared to that of the raw signal. The time waveform of the output signal is shown in Fig. 14(c). The signal is highly denoised and the impulses are completely isolated. The SES of the output signal is shown in Fig. 15(c) and it is completely different from that of the two previous steps. It resembles the characteristic shape of the SES of the ball fault signals [32] as: (a) it has a peak at the fundamental train frequency (*FTF*) and its harmonics, (b) it has a peak at the ball spin frequency (*BSF*) and its harmonics and sidebands with the (*FTF*), (c) the even harmonics of the *BSF* are dominant. The previous results indicate the effectiveness of the proposed filtration scheme in the detection of REBs incipient faults, fault location identification, and hence the surveillance of REBs faults. For validation, the filtration scheme was applied on an inner and outer race fault signals and the results are presented and discussed in the following sections.

7.2.2. The inner race fault signal (Signal B)

The time waveform and spectrum of the inner race fault raw signal (Signal B) are shown in Fig. 16(a, b). The GI of the SE of the raw signal is 0.4948. There is no sign of REBs faults, and the low frequency vibration is dominant. Fig. 17(a, b, c) demonstrates the progressive enhancement in the impulses clarity as the raw signal was passed through the three consecutive filters. The GI of the SE of the signals in Fig. 17(a, b, c) are 0.7914, 0.9217, 0.9585, respectively. The time waveform of the output signal in Fig. 17(c) has the impulses modulated by the shaft rotation frequency which is a characteristic for the inner race faults.

One can diagnose an inner race fault just by looking at the SES of the output signal in Fig. 18 as it shows: (a) peaks at the shaft rotational frequency (f_r) and its harmonics, (b) peaks at the ball passing inner race frequency *BPFI* and its harmonics up to the sixth harmonic, and (c) clear sidebands with the shaft rotational frequency. This confirms the case of an inner race fault because in the final envelope analysis the detection of the modulating effects (sidebands) are important to the diagnosis [2]. Hence, the proposed filtration scheme successfully characterizes the inner race fault and facilitates its detection.

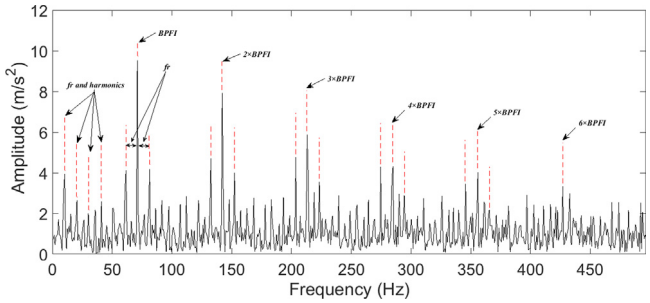

Fig. 18. The SES of the signal in Fig. 17(c).

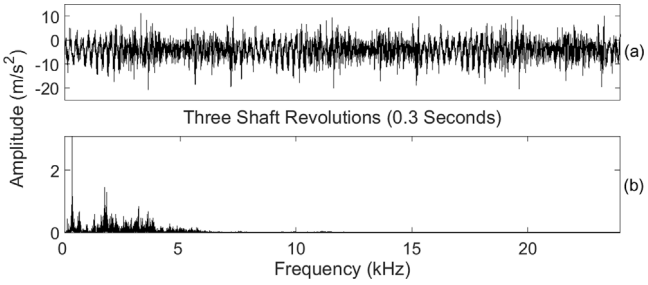

Fig. 19. Signal C: (a) time waveform; (b) spectrum.

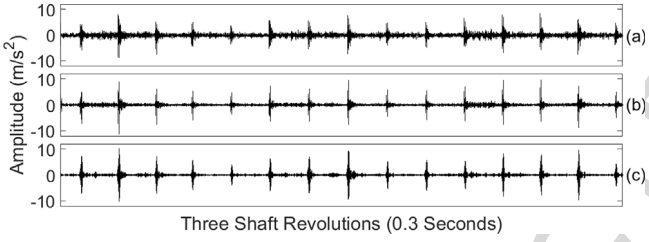

Fig. 20. Signal C filtration output: (a) IAR filtration output of the raw Signal C; (b) MGID filtration output of the signal in "a"; (c) Optimum Morlet wavelet filtration output of the signal in "b".

7.2.3. The outer race fault signal (Signal C)

The time waveform and spectrum of the outer race fault raw signal (Signal C) are shown in Fig. 19(a, b). The GI of the SE of the raw signal is 0.4989. There is no sign of REBs faults, and the low frequency vibration is dominant. Fig. 20(a, b, c) demonstrates the progressive enhancement in the impulses clarity as the raw signal was passed through the three consecutive filters. The GI of the SE of the signals in Fig. 20(a, b, c) were 0.7516, 0.8668, 0.9127, respectively. The resulted SES, as shown in Fig. 21, indicates the presence of an outer race fault as it shows peaks at the ball passing outer race frequency *BPFO* and its harmonics up to the sixth harmonic. So, the proposed filtration scheme helped to clearly identify the outer race fault.

8. Conclusions

Early REBs fault identification is an important practice in machine health monitoring. Faults in early stage leave only weak features in the vibration signal so their detection is a hard task. Envelope analysis is considered as the typical technique to help solve this problem. Vibration signals are always interfered with components such as: periodic and gearmesh signals, white noise, impulsive noise and the transmission path effect. These interferences obstruct a proper band selection for demodulation; thus, envelope spectra become irrelevant.

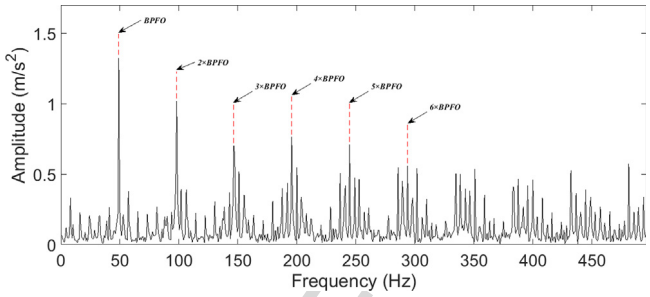

Fig. 21. The SES of the signal in Fig. 20(c).

An automatic, three- step, GI-based, adaptive approach is proposed in this paper to solve the mentioned problems to blindly select the best demodulation band. GI is a recently introduced impulsiveness measure in the field of REBs fault detection. The results of a simulated study proved that GI is more proper for REBs fault detection than kurtosis and SK. The three applied consecutive filters are; IAR filtration, MGID and optimum Morlet wavelet filtration. The three filters are guided and optimized by GI. The effectiveness of the scheme is tested using three experimental signals for different bearing fault locations; all are contaminated with a gearmesh signal. Time waveforms aside to SESs for the filtered signals are shown and carefully discussed. The resulted filtered signals were greatly denoised and the fault related impulses were successfully isolated. Bearing fault locations were clearly identified and the resulted SESs showed the specific characteristic shapes for the corresponding fault locations. The results prove the high effectiveness of the proposed filtration scheme in early REBs fault detection, location identification, and surveillance. They also confirm the suitability of GI as a powerful criterion to guide the blind filtration process accompanying envelope analysis in REBs fault detection, and its superiority over kurtosis-derived indices.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Sawalhi N, Randall RB, Endo H. The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. Mech Syst Signal Process 2007;21:2616–33. http://dx.doi.org/10.1016/j.ymssp.2006.12.002.
[2] Randall RB, Antoni J. Rolling element bearing diagnostics—A tutorial. Mech Syst Signal Process 2011;25:485–520. http://dx.doi.org/10.1016/j.ymssp.2010.07.017.
[3] Ding C, Zhao M, Lin J, Jiao J. Multi-objective iterative optimization algorithm based optimal wavelet filter selection for multi-fault diagnosis of rolling element bearings. ISA Trans 2018. http://dx.doi.org/10.1016/j.isatra.2018.12.010.
[4] Li H, Liu T, Wu X, Chen Q. Application of EEMD and improved frequency band entropy in bearing fault feature extraction. ISA Trans 2018. http://dx.doi.org/10.1016/j.isatra.2018.12.002.
[5] Sawalhi N. Diagnostics, prognostics and fault simulation for rolling element bearings [Doctor of philosophy thesis], Australia: The University of New South Wales; 2007.
[6] Abdusslam SA. Detection and diagnosis of rolling element bearing faults using time encoded signal processing and recognition [Doctor of philosophy thesis], UK: University of Huddersfield; 2012.
[7] Miao Y, Zhao M, Lin J. Identification of mechanical compound-fault based on the improved parameter-adaptive variational mode decomposition. ISA Trans 2019;84:82–95. http://dx.doi.org/10.1016/j.isatra.2018.10.008.
[8] Weatherwax SE. Use of the continuous wavelet tranform to enhance early diagnosis of incipient faults in rotating element bearings [Master's thesis], Texas A & M University; 2009.

[9] Chen X, Zhang B, Feng F, Jiang P. Optimal resonant band demodulation based on an improved correlated kurtosis and its application in bearing fault diagnosis. Sensors 2017;17. http://dx.doi.org/10.3390/s17020360.
[10] Antoni J, Randall RB. The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines. Mech Syst Signal Process 2006;20:308–31. http://dx.doi.org/10.1016/j.ymssp.2004.09.002.
[11] Bozchalooi IS, Liang M. A smoothness index-guided approach to wavelet parameter selection in signal de-noising and fault detection. J Sound Vib 2007;308:246–67. http://dx.doi.org/10.1016/j.jsv.2007.07.038.
[12] Su W, FW Ã, Zhu H, Zhang Z, Guo Z. Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement. Mech Syst Signal Process 2010;24:1458–72. http://dx.doi.org/10.1016/j.ymssp.2009.11.011.
[13] Jiang Y, Tang B, Qin Y, Liu W. Feature extraction method of wind turbine based on adaptive Morlet wavelet and SVD. Renew Energy 2011;36:2146–53. http://dx.doi.org/10.1016/j.renene.2011.01.009.
[14] Tse PW, Wang D. The automatic selection of an optimal wavelet filter and its enhancement by the new sparsogram for bearing fault detection Part 2 of the two related manuscripts that have a joint title as "Two automatic vibration-based fault diagnostic methods using the". Mech Syst Signal Process 2013;40:520–44. http://dx.doi.org/10.1016/j.ymssp.2013.05.018.
[15] McDonald GL, Zhao Q, Zuo MJ. Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection. Mech Syst Signal Process 2012;33:237–55. http://dx.doi.org/10.1016/J.YMSSP.2012.06.010.
[16] Wang Y, Yang L, Xiang J, Yang J, He S. A hybrid approach to fault diagnosis of roller bearings under variable speed conditions. Meas Sci Technol 2017;28. 125104. http://dx.doi.org/10.1088/1361-6501/aa9460.
[17] Gini C. Measurement of inequality of incomes. Econ J 1921;31:124–6. http://dx.doi.org/10.2307/2223319.
[18] Miao Y, Zhao M, Lin J. Improvement of kurtosis-guided-grams via gini index for bearing fault feature identification. Meas Sci Technol 2017;28. 125001. http://dx.doi.org/10.1088/1361-6501/aa8a57.
[19] Wang D. Some further thoughts about spectral kurtosis, spectral L2/L1 norm, spectral smoothness index and spectral Gini index for characterizing repetitive transients. Mech Syst Signal Process 2018;108:360–8. http://dx.doi.org/10.1016/J.YMSSP.2018.02.034.
[20] Hurley N, Rickard S. Comparing measures of sparsity. IEEE Trans Inform Theory 2009;55:4723–41. http://dx.doi.org/10.1109/TIT.2009.2027527.
[21] Vrabie V, Granjon P, Serviere C. Spectral kurtosis: from definition to application. In: 6th IEEE int. work. nonlinear signal image process. 2003. p. xx.
[22] Endo H, Randall RB. Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter. Mech Syst Signal Process 2007;21:906–19. http://dx.doi.org/10.1016/J.YMSSP.2006.02.005.
[23] Ifeachor EC, Jervis BW. Digital signal processing: a practical approach. Addison-Wesley; 1993.
[24] Proakis JG, Manolakis DG. Digital signal processing: principles, algorithms, and applications. Prentice Hall; 1996.
[25] McFadden PD, Smith JD. Model for the vibration produced by a single point defect in a rolling element bearing. J Sound Vib 1984;96:69–82. http://dx.doi.org/10.1016/0022-460X(84)90595-9.
[26] Wiggins RA. Minimum entropy deconvolution. Geoexploration 1978;16:21–35. http://dx.doi.org/10.1016/0016-7142(78)90005-4.
[27] McDonald GL. Vibration signal-based fault detection for rotating machines [Master's thesis], The University of Alberta; 2011, http://dx.doi.org/10.7939/R3SK98.
[28] Miao Y, Zhao M, Lin J, Lei Y. Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings. Mech Syst Signal Process 2017;92:173–95. http://dx.doi.org/10.1016/J.YMSSP.2017.01.033.
[29] He W, Jiang Z, Feng K. Bearing fault detection based on optimal wavelet filter and sparse code shrinkage. Measurement 2009;42:1092–102. http://dx.doi.org/10.1016/j.measurement.2009.04.001.
[30] Qiu H, Lee J, Lin J, Yu G. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. J Sound Vib 2006;289:1066–90. http://dx.doi.org/10.1016/J.JSV.2005.03.007.
[31] STI. Rolling element bearings. F Appl Note, REB, Sales technol leag city, Tex, USA. 2012.
[32] Smith WA, Randall RB. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. Mech Syst Signal Process 2015;64–65:100–31. http://dx.doi.org/10.1016/j.ymssp.2015.04.021.

Please cite this article as: M.N. Albezzawy, M.G. Nassef and N. Sawalhi, Rolling element bearing fault identification using a novel three-step adaptive and automated filtration scheme based on Gini index. ISA Transactions (2020), https://doi.org/10.1016/j.isatra.2020.01.019.