Using Unbiased Autocorrelation to Enhance Kurtogram and Envelope Analysis Results for Rolling Element Bearing Diagnostics

Using Unbiased Autocorrelation to Enhance Kurtogram and Envelope Analysis Results for Rolling Element Bearing Diagnostics

Ali Moshrefzadeh[1], Alessandro Fasana[1], Luigi Garibaldi[1]

[1]DIMEAS - Politecnico di Torino
Corso Duca degli Abruzzi, 24 - 10129 Torino, Italy
Ali.Moshrefzadeh@polito.it

Abstract

Envelope analysis is one of the most advantageous methods for rolling element bearing diagnostics but finding the suitable frequency band for demodulation has been a substantial challenge for a long time. Introduction of spectral kurtosis (SK) mostly solved this problem but in situations where signal to noise ratio is very low or in presence of non-Gaussian noise this method will fail. This major drawback may noticeably decrease the effectiveness of the SK and goal of this paper is to overcome this problem. Vibration signals from rolling element bearings exhibit high levels of 2nd order cyclostationarity, especially in the presence of localised faults. A second-order cyclostationary signal is one whose autocovariance function is a periodic function of time: the proposed method, named Autogram by the authors, takes advantage of this property to enhance the conventional spectral kurtosis. First, a maximal overlap discrete wavelet packet transform (MODWPT) is adopted to split a signal in different frequency bands and central frequencies. Second, unbiased autocorrelation of the squared envelope is calculated to reduce the level of uncorrelated random noise. Third, kurtosis of the autocorrelation is computed and a two dimensional colormap, named Autogram, is presented in order to locate the optimal frequency band for demodulation. The purpose is to increase the detection and characterization of transients in the temporal signal, which contains the bearing defect frequencies as well as appropriate frequency at which the fault impulses are modulated. Finally, the Fourier transform is used to obtain a frequency domain representation of the envelope signal so to identify the defect frequencies of the bearing. The proposed method has been tested on experimental data and compared with literature results so to assess its performances in rolling element bearing diagnostics. The results are very positive, and bearing characteristic frequencies from signals masked by Gaussian and non-Gaussian background noise can be extracted.

1 Introduction

Rolling element bearings (REBs) are one of the most diffused element in rotating machinery and their failure is the most important cause of machinery breakdowns. Thus, correctly detecting and diagnosing bearing faults at stages prior to their complete failure is of vital importance. It avoids potential catastrophic damage not only to the apparatus but also to the personnel.

As a localised defect develops on an inner race, outer race or roller part of a bearing, an impact is generated each time the defect is engaged and consequently the bearing and machine structure are excited, in particular at their resonance frequencies. The corresponding vibration signal will comprise all the harmonics of this impact, which repeats almost periodically at a rate dependent on bearing geometry. Investigation of the generated vibrations is indispensable to detect the faults and many methods have been developed to extract the bearing characteristic frequencies from the measured vibrations. Among them envelope analysis, also called high frequency resonance technique, has been used successfully for a long time: a signal is first bandpass filtered in the excited structural resonance frequency band, and then spectrum of the envelope signal -which contains the desired diagnostic frequencies- is formed. The main challenge has always been finding the most suitable frequency band for demodulation.

Spectral kurtosis (SK) has been a significant step to unravel this problem. It is a method which effectively detects the sequence of impulses in a signal and can be used to determine the proper demodulation frequency band in which a signal has the maximum impulsivity.

Antoni proposed two methods, one based on Short Time Fourier Transform (STFT) [1] and another one based on filter banks [2], to calculate the SK. In the STFT based SK, the aim is to find the central frequency f and the window length N_w which maximize the value of the SK over all possible choices. A coloured 2D map called Kurtogram displays the values of SK for each frequency band as a function of f and N_w. Antoni [2] also developed the Fast Kurtogram which is based on the multirate filter-bank structure (MFB) to overcome the rigorous but long computation of full Kurtogram, in order to make the method more efficient and suitable for industrial applications. Fast Kurtogram subdivides the bandwidths into rational ratios that allow the use of fast multirate processing, and then investigates the kurtosis value of the complex filtered signal from selected bandwidths. These two methods almost return the same result. Lei et. al [3] further adopted wavelet packet transform (WPT) as a more precise technique than FIR filters to enhance the original Kurtogram, thus exploiting the good local property of wavelets both in time and frequency domains. Barszcz and Jablonski [4] argue that, when a signal contains relatively strong non-Gaussian noise such as large impulses, the temporal based kurtosis indicator of the Kurtogram would fail. To overcome this drawback they consequently propose a Protrugram, where the kurtosis of the spectrum is displayed, instead of kurtosis of the filtered time signal. However, their proposed method, unlike SK, is not blind and prior knowledge about the defect frequencies is needed. Moreover, Protrugram will fail when defect frequencies are not dominant in the spectrum.

The paper is organised as follows. Section 2 establishes and explains in detail a new method for the optimal band selection for demodulation of bearing signals with localised defects. The experimental validation is carried out in section 3 to examine the performance of the proposed method. In each case, the results are compared with the Fast Kurtogram [5] and literature results [6]. Finally, the conclusions are drawn in section 4.

2 Proposed Method

Using Fast Kurtogram to determine the most impulsive frequency band, followed by envelope analysis of the bandpass filtered signal, has become the benchmark method for bearing diagnostics for years and has accomplished significant results [6]. Kurtogram method is commonly capable of detecting localised hidden non-stationarities, even in presence of strong Gaussian noise, but its performance is limited in several conditions i.e. low signal to noise ratio or strong non-Gaussian noise such as randomly impulsive noise [6]: in these cases Fast Kurtogram was found to be ineffective in seeking the transient signal. These circumstances are common in industrial applications as multiple devices such as gearboxes and bearings work alongside in a complex machine. Also, the acquired signal in harsh environment can be extremely affected by external sources.

This paper proposes a new procedure based on unbiased AutoCorrelation (AC) to overcome the restrictions of imposed by heavy Gaussian and also non-Gaussian background noise. The flowchart of the proposed method is shown in Fig. 1 and the details of each step are described as follows.

Step 1: Wavelets have very good local properties in both time and frequency domains, and wavelet transform (WT) can be used as an effective filter to split a signal in different frequency bands and central frequencies. Moreover, in comparison to FIR filters, they have the capability to filter out noise and also preserve signal characteristics more precisely [3]. Because of the downsampling operations, at higher levels the lengths of the coefficient series are much shorter than the original data and this will cause larger estimation error and limit the ability to investigate the coefficients. Furthermore, the discrete wavelet packet transform (DWPT) can be sensitive to the selection of starting point of the time series as the wavelet and scaling coefficients are not circularly shift equivariant, i.e. a change in the starting point can produce rather different outcomes [7]. These drawbacks can be overcome by using the maximal overlap (undecimated) discrete wavelet packet transform (MODWPT) which removes the downsampling step in DWPT. The details of MODWPT can be found in ref. [7]. Accordingly, when an original time history is loaded, e.g. an acceleration, the MODWPT is applied as a filter and consequently a series of signals, corresponding to different frequency bands and central frequencies, is produced at each level. These coefficients are used as an input for the following steps. In addition to those mentioned benefits, the length of the filtered signal is equivalent to the original one, which is useful to implement the next analysis.

Step 2: a signal is assumed to be cyclostationary of order n when its nth order statistics is periodic. Antoni

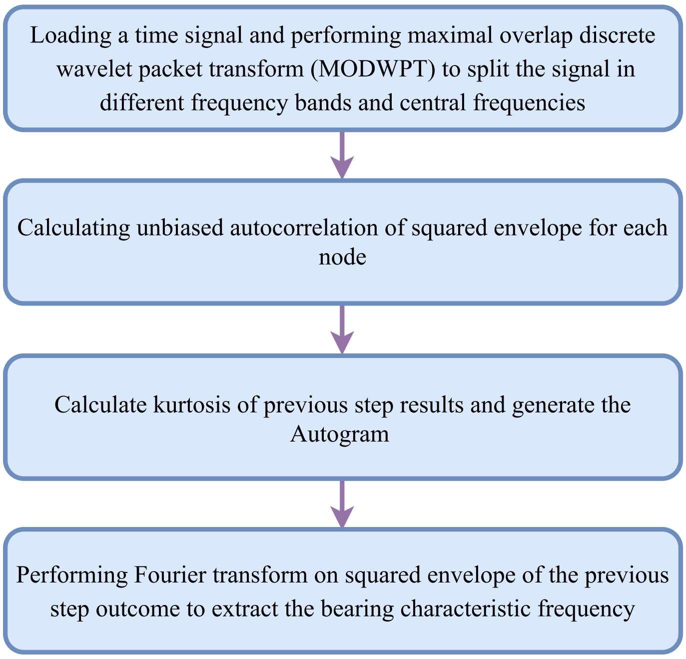

Figure 1: Flowchart of the proposed method

and Randall [8] pointed out that a bearing localized fault may be modelled as a 2nd order cyclostationary process. It determines processes with a periodic autocovariance function in time (eq. 1).

$$R_{xx}(t,\tau) = \mathbf{E}\{x(t-\tau/2)x(t+\tau/2)\} \tag{1a}$$
$$R_{xx}(t,\tau) = R_{xx}(t+T,\tau)\,, \tag{1b}$$

The fundamental concept which motivates this work is to take advantage of this periodicity. Therefore, in this proposed method we calculate the unbiased AC of the (periodic) instantaneous autocovariance of the signal $R_{xx}(t,\ 0)$, where x is the signal filtered by MODWPT at step 1. It can be seen from Eq.1a that the instantaneous autocovariance is computed by the mathematical expectation operator or ensemble average operator. Unfortunately, in many situations we are not able to calculate the expected value since only a single record of data rather than a set of records is available. But -once cyclostationarity in time domain is assumed- other features of the signal, such as the instantaneous power and the envelope function, provides similar information about periodicity as the instantaneous autocovariance function would [9].

Therefore, in this paper the ensemble average is effectively replaced by the squared envelope of a signal, which can be computed much more straightforwardly. Hence, in this step unbiased AC is performed on the squared envelope of the signal instead of its instantaneous autocovariance.

This step has the benefit of removing the noise and the random impulsive content, both unrelated to the specific bearing fault, thus leading to a higher signal to noise ratio. Furthermore, the periodic part of the signal (directly related to the defects) is enhanced, showing an additional virtue of this process.

It also should be mentioned that, as the value of τ increases, the number of data sample available for AC will decrease, and hence the AC result will not have adequate accuracy. As a result, after performing the unbiased AC just the first half of the outcome is chosen for further investigation. Moreover, the first coefficients of filtered signal, that are affected by the transients of filters, should not be included in calculation of the AC.

This step will lead to a more accurate diagnostic process than possible with the original outputs of MODWPT. In fact, impulsive noise, which ineffectively assigns very high kurtosis to a signal and proved to be the main drawback of SK, should largely be removed.

Step 3: our goal in this step is to find the most suitable frequency band for demodulation. A proper selection is substantial to have a successful diagnosis of bearings faults, since fault information cannot be extracted from the demodulated signal if the appropriate frequency band and central frequency is not selected. Fast Kurtogram and Protrugram are two widely known methods for choosing the proper demodulation band and compute the kurtosis of the filtered time signal or of the spectral lines of the envelope respectively. In this step, we introduce an alternative approach for selecting the optimal frequency band of demodulation. The proposed method differs from both mentioned techniques because we calculate the kurtosis value of the signals resulting form step 2, i.e. the unbiased autocorrelation of the squared envelope, for each level and decomposed frequency band. Subsequently, the kurtosis values of all nodes, similar to Kurtogram, are presented in a colormap (the color scale is proportional to kurtosis value) and the vertical and horizontal axis represent level of the MODWPT decomposition and frequency respectively. Since the concept is analogous to Kurtogram, and this proposal is based on autocorrelation, the authors suggest the name "Autogram" for this newly developed approach.

It is noted that the AC of the squared envelope is a positive function, whose minimum value is arguably different from zero. As a consequence, a different definition of kurtosis is implemented as follows:

$$\text{Kurtosis}_x = \frac{\sum_{i=1}^{N} (x_i - min(x))^4}{\left[\sum_{i=1}^{N} (x_i - min(x))^2\right]^2} \tag{2}$$

where *min* function indicates the minimum value of the AC data.

The signal related to the node, frequency band and center frequency, with the highest kurtosis is eventually considered for further computation. In case of compound signals, i.e. signals containing multiple frequency bands with high kurtosis, more than one resonant frequency may exist, and the local maxima in addition to absolute maximum should be examined.

Step 4: Fourier transform is performed on the outcome of step 3 to extract the fault characteristic frequency and also diagnose the type of the faults. It is worth mentioning that pre-whitening of the signal is usually conducted on the raw signal in order to increase the bearing signal to noise ratio and also to separate the random and deterministic (discrete frequency) components such as shaft rotating frequency [6]. Nevertheless, in this paper pre-whitening of bearing fault signal is not performed prior to the analysis because of two reasons: first, the AC process should remove the uncorrelated components of signal for each node of the MODWPT. Second, the parameters of the pre-whitening filters should be set in advance, but if they are not selected properly the bearing fault signature in the signal might be cancelled.

3 Result

The data set provided by the Case Western Reserve University (CWRU) Bearing Data Center [10] has become a standard reference for diagnosis of bearings. For example, a detailed benchmark study has been provided by Smith and Randall [6] in which three diagnostic methods such as envelope analysis of the raw signal, cepstrum prewhitening and discrete/random separation followed by SK were applied to the data sets. Therefore, in this section these data records will be used to examine the performance of the proposed method and in each case the results will be compared with the benchmark study and Fast Kurtogram, whose code has been provided by [5]. In the following, two bearing signals with defect on the outer race and roller element will be studied. Also, it should be mentioned that to decompose the signal in step 1 of the proposed method, Daubechies wavelet with the vanishing moment 12 (db12) is utilised.

Case 1: in this case the record 176 DE with an inner race defect will be examined. The vibration signal of the record is plotted in Fig. 2a. Many defect repetitive transients are visible in the time waveform and the fault is categorised as "potentially" and "clearly" diagnoseable by methods 1 and 2 of the benchmark study [6].

Fast Kurtogram detects the center frequency 5625 Hz, with bandwidth 750 Hz as the frequency band with the highest kurtosis. Also the proposed method is applied to the same signal. The Autogram is shown in Fig. 2c and the maximum value is assigned to the node (3,8), with center frequency 22500 Hz and bandwidth of 3000 Hz. The squared envelope spectrum of the filtered signals, selected by the Fast Kurtogram and Autogram, are displayed in Fig. 2d and Fig. 2e respectively. Green dash-dot line cursor is depicted at the nominal shaft frequency, a red dashed line at the first three harmonics of the expected fault frequency, and dotted lines demonstrating the first order modulation sidebands around the fault frequency and its harmonics. Although the frequency band of demodulation is different for the Fast Kurtogram and Autogram, inner race ballpass frequency (BPFI) and its harmonics together with their sidebands at the shaft frequency, can clearly be spotted. It indicates that more than one resonance frequency of the system has been excited by the bearing defect impulses. Fast Kurtogram selected frequency band for demodulation also have high kurtosis in the Autogram (see Fig. 2c) while frequency band of demodulation selected by Autogram is not clearly detectable in the Fast Kurtogram.

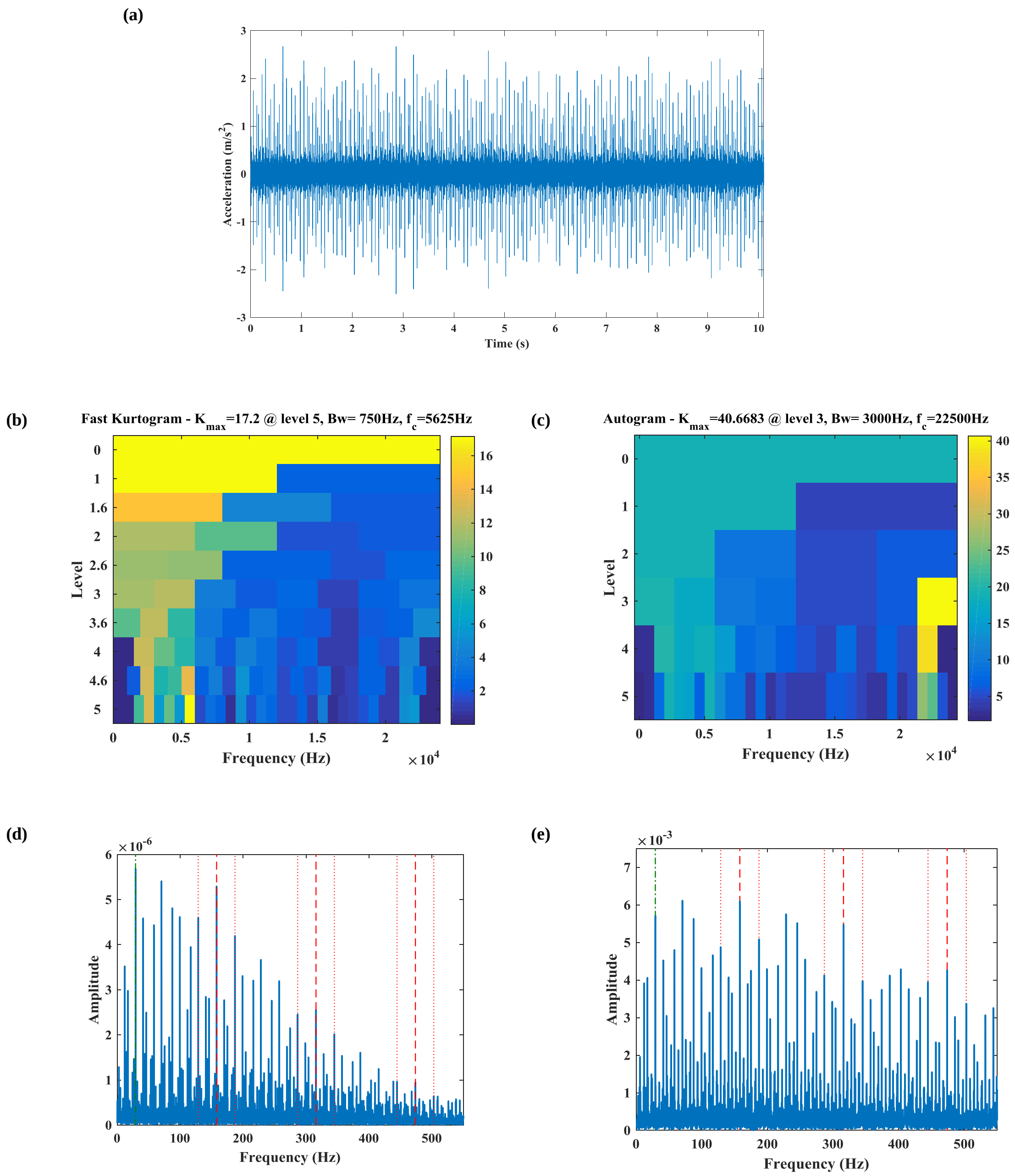

Figure 2: (a) Time domain signal: 176 DE (b) Fast Kurtogram (c) Autogram (d) squared envelope's spectrum of the signal related to node with highest kurtosis in Fast Kurtogram (e) squared envelope's spectrum of the signal related to node with highest kurtosis in Autogram (Green dash-dot line: nominal shaft frequency, red dashed lines: first three harmonics of the BPFI, red dotted lines: first order modulation sidebands at shaft speed around the BPFI and its harmonics)

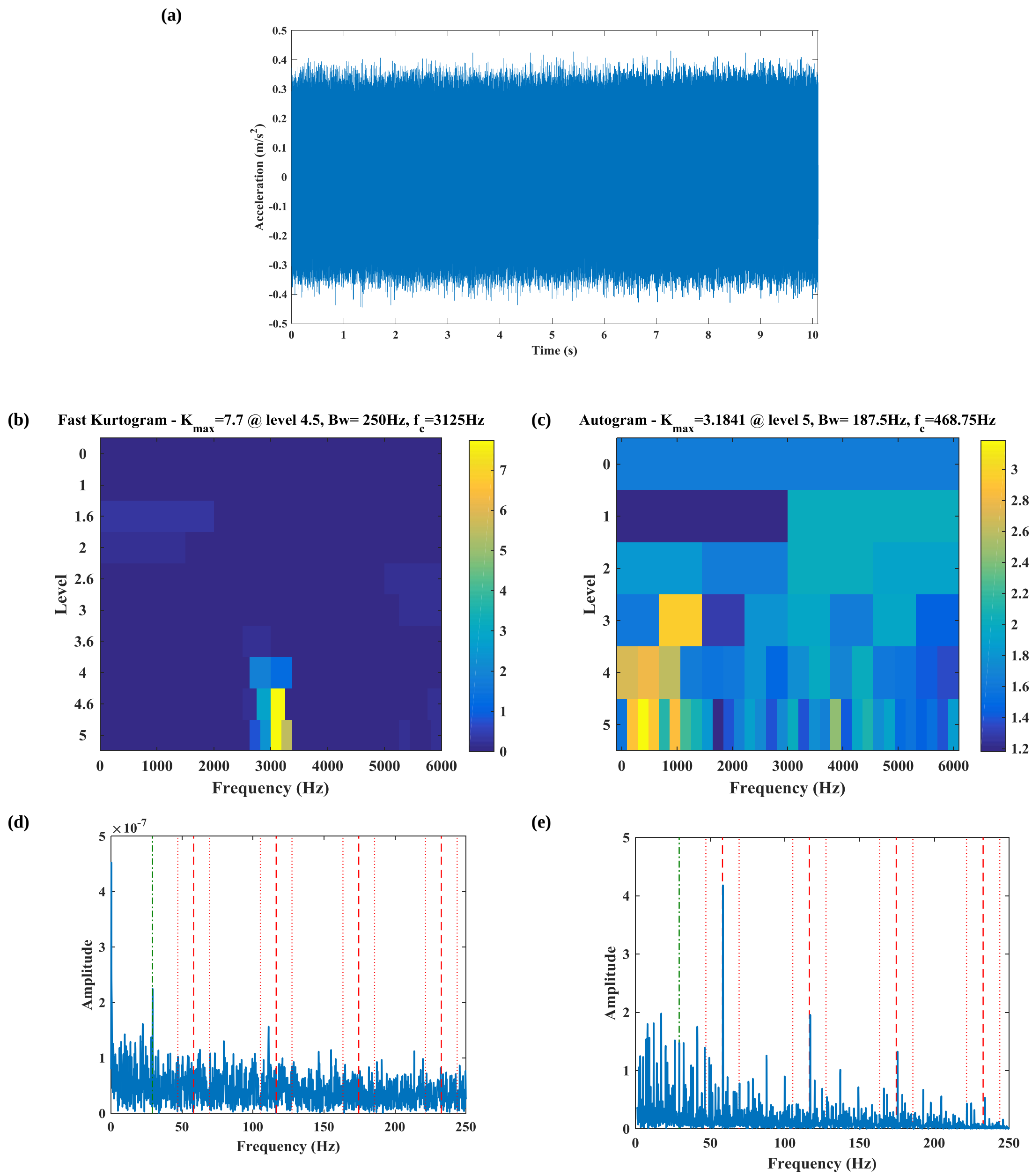

Figure 3: (a) Time domain signal: 284 BA (b) Fast Kurtogram (c) Autogram (d) squared envelope's spectrum of the signal related to node with highest kurtosis in Fast Kurtogram (e) squared envelope's spectrum of the signal related to node with highest kurtosis in Autogram (Green dash-dot line: nominal shaft frequency, red dashed lines: first four harmonics of the BSF, red dotted lines: first order modulation sidebands at cage speed around the BSF and its harmonics)

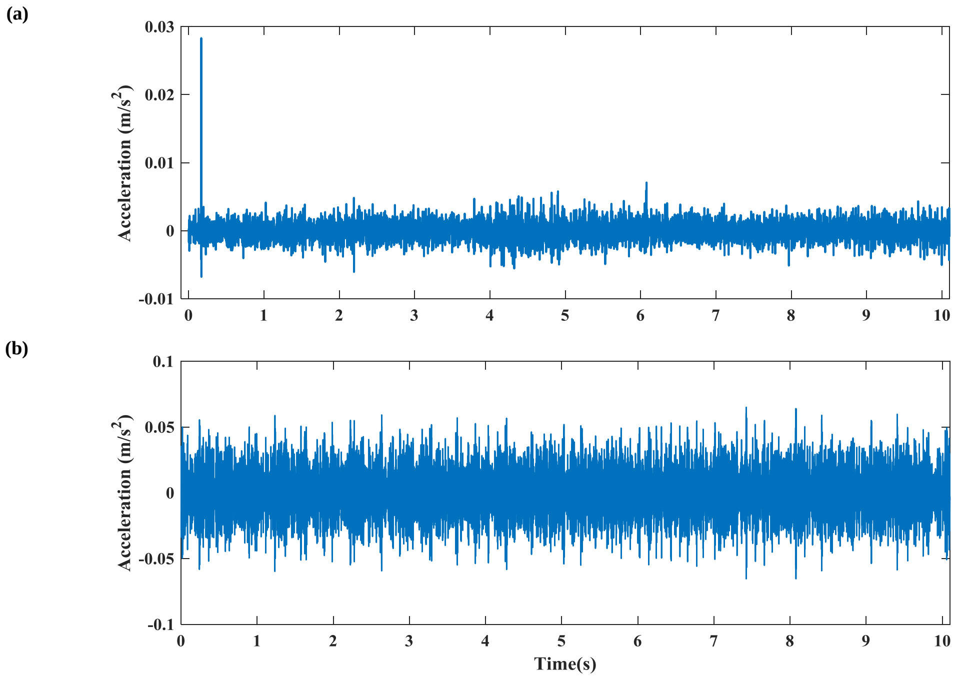

Figure 4: Filtered signal associated with the node with the highest kurtosis in the (a) Fast Kurtogram (b) Autogram

Case 2: in this case the record 284 BA which has a defect on its rolling element will be examined. Bearings with a ball fault has been by far the most problematic cases to diagnose [6]. The vibration signal of the record is plotted in Fig. 3a. Although no defect transients are visible in the time waveform, the fault is diagnoseable by methods 1 and 2 of the benchmark study [6] but method 3 does not determine the bearing fault. This is one of the major drawbacks of Kurtogram (Method 3) which is its vulnerable to impulsive noise, since it tends to highlight the presence of individual impulses rather than sequences of transients. Fig. 3b illustrates the Fast Kurtogram in which the frequency band with center frequency 3125 Hz, bandwidth 250 Hz, highest kurtosis 7.7, is associated with the impulsive noise. The spectrum of the signal's squared envelope associated with this frequency band is plotted in Fig. 3d but does not provide any valuable diagnostic information. Fig. 4a shows the filtered signal in which an impulsive content exists but is certainly independent from the specified bearing fault.

The result of the Autogram is shown in Fig. 3c and the maximum value of the newly defined kurtosis (2) is 3.2, assigned to node (5,4) with center frequency of 468.75 Hz and bandwidth of 187.5 Hz. In comparison to the Fast Kurtogram, frequency band associated to the impulsive noise has low kurtosis value which indicates elimination of the impulsive noise after performing AC. The spectrum of the squared envelope signal for this node (frequency band) is displayed in Fig. 3e. and the harmonics of ball (roller) spin frequency (BSF) can be detected. Moreover, the filtered signal for this case is illustrated in Fig. 4b and, in contrast to the signal filtered by the Fast Kurtogram selected node, a series of transients is present which, based on its spectrum, is indeed related to the bearing fault. Furthermore, it is worth noting that presence of another resonance frequency can be identified in the right side of the Autogram. In this branch, node (5,21) with center frequency of 3843.75 Hz has the highest kurtosis and the spectrum of this node, which is not depicted here, gives an additional successful diagnosis.

As a consequence, these results show the improved capability of the Autogram over Fast Kurtogram in dealing with signals with impulsive noise.

4 Conclusion

This paper proposes a new enhanced Fast Kurtogram method to find the proper frequency band of demodulation for bearing faults diagnosis. This method is an attempt to overcome the major drawback of the conventional Fast Kurtogram, i.e. its vulnerable to impulsive noise which makes it more sensible to individual impulses than to series of transients. Based on the proposed method, first undecimated WPT (MODWPT) is employed to decompose the original signal at different levels, each divided in frequency bands called nodes. Then the unbiased autocorrelation is calculated for the squared envelope of all nodes. The kurtosis of the ACs is computed and afterward presented in form of a two dimensional colormap, called Autogram. At this point the node with the highest kurtosis value is chosen for further frequency analysis. The proposed method has been tested on experimental data and compared with literature results so to assess its performances in REBs diagnosis. Two data records from the Case Western Reserve University database where selected, in which the Fast Kurtogram was not able to find the optimal demodulation band for the second case. In general, the results of the proposed method are positive, and bearing characteristic frequencies even from signals with non-Gaussian background noise could be extracted.

References

[1] J. Antoni, The spectral kurtosis: a useful tool for characterising non-stationary signals, Mechanical Systems and Signal Processing 20 (2) (2006) 282–307.

[2] J. Antoni, Fast computation of the kurtogram for the detection of transient faults, Mechanical Systems and Signal Processing 21 (1) (2007) 108–124.

[3] Y. Lei, J. Lin, Z. He, Y. Zi, Application of an improved kurtogram method for fault diagnosis of rolling element bearings, Mechanical Systems and Signal Processing 25 (5) (2011) 1738–1749.

[4] T. Barszcz, A. JabŁoński, A novel method for the optimal band selection for vibration signal demodulation and comparison with the kurtogram, Mechanical Systems and Signal Processing 25 (1) (2011) 431–451.

[5] J. Antoni, Fast kurtogram.
URL https://mathworks.com/matlabcentral/fileexchange/48912-fast-kurtogram

[6] W. A. Smith, R. B. Randall, Rolling element bearing diagnostics using the case western reserve university data: A benchmark study, Mechanical Systems and Signal Processing 64 (2015) 100–131.

[7] A. T. Walden, Wavelet analysis of discrete time series, in: European Congress of Mathematics, Springer, 2001, pp. 627–641.

[8] J. Antoni, R. Randall, Differential diagnosis of gear and bearing faults, TRANSACTIONS-AMERICAN SOCIETY OF MECHANICAL ENGINEERS JOURNAL OF VIBRATION AND ACOUSTICS 124 (2) (2002) 165–171.

[9] J. Antoni, Cyclic spectral analysis in practice, Mechanical Systems and Signal Processing 21 (2) (2007) 597–630.

[10] Case western reserve university bearing data center website.
URL http://csegroups.case.edu/bearingdatacenter/home