Fault Detection Enhancement in Rolling Element Bearings Using the Minimum Entropy Deconvolution

Tomasz BARSZCZ(1), Nader SAVALHI(2)

(1) AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Krakow, Poland; e-mail: tbarszcz@agh.edu.pl
(2) Mechanical Engineering Department, Prince Mohammad Bin Fahd University (PMU)
AlKhobar 31952, Saudi Arabia; e-mail: nadersaw@hotmail.com

(received January 21, 2011; accepted February 10, 2012)

Minimum Entropy Deconvolution (MED) has been recently introduced to the machine condition monitoring field to enhance fault detection in rolling element bearings and gears. MED proved to be an excellent aid to the extraction of these impulses and diagnosing their origin, i.e. the defective component of the bearing. In this paper, MED is revisited and re-introduced with further insights into its application to fault detection and diagnosis in rolling element bearings. The MED parameter selection as well as its combination with pre-whitening is discussed. Two main cases are presented to illustrate the benefits of the MED technique. The first one was taken from a fan bladed test rig. The second case was taken from a wind turbine with an inner race fault. The usage of the MED technique has shown a strong enhancement for both fault detection and diagnosis. The paper contributes to the knowledge of fault detection of rolling element bearings through providing an insight into the usage of MED in rolling element bearings diagnostic. This provides a guide for the user to select optimum parameters for the MED filter and illustrates these on new interesting cases both from a lab environment and an actual case.

Keywords: rolling bearing, fault detection, Minimum Entropy Deconvolution (MED), wind turbine.

1. Introduction and background

1.1. Rolling element bearings

Rolling element bearings (REBs) are components, which transfer the load through elements in rolling contact. There are several types of these bearings depending on the rolling element type (ball, roller, spherical etc.); all of them share the same principle of operation. Figure 1 presents the layout of a rolling bearing under a unidirectional vertical load. The REB consists of: inner race, outer race, balls (or in general, rolling elements) and a cage, which holds the rolling elements in a given relative position.

Rolling element bearings are key components in modern machinery. Detection of their faults is very important, as it prevents any further deterioration of other components which may lead to a catastrophic failure. The deterioration of a faulty part (e.g. an outer race, inner race or a rolling element) can affect other supported components (e.g. gears, fans, etc.) and may result in their failure. The cost of a new bearing is much lower than that of a gearbox that has the faulty bearing went unnoticed.

Figure 2 presents a typical layout of a wind turbine drivetrain with marked locations of rolling bearings, while Fig. 3 shows the gearbox and the main bearing of
A 1.5 MW turbine. The typical wind turbine drivetrain (as shown in Fig. 2) consists of a main shaft, planetary gearbox, a two stage parallel gearbox and a generator. Depending on the location in a wind turbine drivetrain, the replacement of a bearing can cost between 2500 to 32000 EUR, while the replacement of a gearbox may cost between 75000 to 240000 EUR (Gajetzki, 2006). These operations depend very much on the accessibility to the wind turbine, which in turn depends on weather conditions, especially wind speed. This aspect is even more important for the offshore wind parks. Bearing spills, subject to the machine speed and load, usually propagate slowly, thus giving the analyst enough time for monitoring and maintenance scheduling before any catastrophic failure. Therefore, a huge body of research in the area of bearing diagnostics concentrated on the early detection of the bearing faults to enable providing enough lead time for maintenance purposes (Randall, Antoni, 2011). The knowledge about the technical status of the REB and its fault development and propagation are being employed to develop a reliable prediction of the remaining useful life of the rolling element bearings what is known as bearing prognostics (Heng et al., 2009). This is becoming an important aspect of the new trend in monitoring the health of rotating machines. Cempel proposed a set of methods for machinery components lifetime prediction and calculation of limit values (Cempel, 2008).

1.2. Diagnostics of rolling element bearings

Rolling element bearings fail in a number of different ways (spalling, brinelling, etc.) and as the result of a number of factors (fatigue, oil contaminations, overload, etc.). Fatigue results in the appearance of spills on the inner race, outer race or rolling elements. If one of the races has a spill, it will almost periodically impact with rolling elements. The fault signature is represented by successive impulses with a repetition rate depending on the faulty component, geometric dimensions and the rotational speed. The period between impacts is different for all the listed elements and depends on the geometry of the bearing, the rotational speed and the load angle. For a fixed outer race bearing, the theoretical ball bass frequencies (defect frequencies) are given in Table 1.

The impulses are usually described as pseudo cyclostationary to reflect the fact that the spacing between these impulses is not constant and usually varies between 1–2% as the result of slippage resulting from the variation of the load angle of the rolling element. This phenomenon can be used to distinguish between signals generated by gears and those generated by bearings. The immediate consequence of this 1–2% variation (jitter) is the smearing of the spacing between the repetition frequencies in the high frequency region. This makes the detection of the fault extremely difficult using the spectrum of the raw signal, especially that the generated impulses are usually rather
Table 1. Characteristic frequencies of fundamental REB faults.

<table>
<thead>
<tr>
<th>Fault description</th>
<th>Characteristic frequency</th>
<th>Type of fault</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ball Pass Frequency of the Outer race (BPFO)</td>
<td>$f_r \frac{N_r}{2} \left(1 - \frac{D_b \cos \phi}{D_p}\right)$</td>
<td>Outer race</td>
</tr>
<tr>
<td>Ball Pass Frequency of the Inner race (BPI)</td>
<td>$f_r \frac{N_r}{2} \left(1 + \frac{D_b \cos \phi}{D_p}\right)$</td>
<td>Inner race</td>
</tr>
<tr>
<td>Ball Spin Frequency (BSF)</td>
<td>$f_r \frac{D_p}{D_b} \left(1 - \frac{D_b^2 \cos^2 \phi}{D_p^2}\right)$</td>
<td>Rolling element</td>
</tr>
<tr>
<td>Fundamental Train Frequency (FTF)</td>
<td>$f \frac{1}{2} \left(1 - \frac{D_b \cos \phi}{D_p}\right)$</td>
<td>Cage</td>
</tr>
</tbody>
</table>

where: $f_r$ – rotational speed of the shaft, $D_b$ – rolling element diameter, $D_p$ – pitch diameter, $N_r$ – number of the rolling elements (for a single row), $\phi$ – load angle (contact angle from radial).

weak and are masked by other signal components like gearmesh, low shaft harmonics, noise etc. The usage of the high frequency resonance technique (HFRT) (DARLOW et al., 1974), which is also widely known as envelope analysis, and the usage of cyclostationarity analysis (ANTONI, 2007; 2009; RANDALL et al., 2001) provide the base for fault detection and diagnosis in rolling element bearings. As has been shown by many authors (e.g. KLEIN, 2003; HO, RANDALL, 2000), the envelope spectrum is a very efficient diagnostic tool for the aforementioned faults, as the information about the fault is extracted from the spacing between impulses but not by the excited frequencies. The process of obtaining the envelope spectrum is often referred to as the signal demodulation. There are several methods to properly select the frequency band for the performing of the demodulation. This is the so-called Optimum Frequency Band (OFB) selection problem. An informative source of rolling element bearing diagnostics can be recalled in RANDALL and ANTONI (2011).

Another important feature of REBs induced impacts is the fact that if the bearing load is constant and the outer race stationary, then the BPFO impulses will be similar to each other because conditions for the passage of each rolling element are uniform, i.e. no modulation occurs. This is not the case for the inner race and the rolling element faults as they experience varying load when moving in and out of the load zone, thus causing a modulation at the shaft speed for the inner race case and at the cage speed for the rolling element fault case. An inner race case fault signature is presented in Figs. 4a and 4b for both the time domain signature and the envelope spectrum, respectively. In Fig. 4b, the harmonics of the BPFI as well as sidebands at the shafts speed are observed. A similar pattern will be generated in the case of the rolling element fault, but in that case the impacts would be modulated by the cage speed (or fundamental train frequency – FTF).

![Fig. 4. Inner race fault: a) time domain signature, b) envelope spectrum.](image)

1.3. A model of the measured vibration signal with defective rolling element bearings

To illustrate the content of a measured vibration signal with a defective rolling element bearing, a simple model of the generation process is presented in Fig. 5. The symbol "**h**" represents the convolution of the combined vibration signal (deterministic signals, bearing a defective signal and noise) with the transfer path between the vibration source and the sensor location. In reality, the mechanism is far more complex as it involves a number of vibration sources which may be added or convolved in rather different forms.
entropy deconvolution technique is introduced and explained in Sec. 2. This section presents also the preprocessing proposed method, which also includes removing the deterministic components from the signal and prewhitening the signal spectrum prior to the application of the MED technique. Section 3 presents two main cases to illustrate the application of the MED to enhance the detecting and diagnosing faults in rolling element bearings. This describes the results of the MED application, first to data taken from a test rig and then to data measured on a real wind turbine with a REB fault. The MED technique was able to increase the kurtosis (which is a measure of the signal “peakedness”) from 4 to almost 22. The enhancement is also presented in both the time domain signal and in the envelope spectrum. Section 4 sets the summary and conclusion of this paper. One important aspect of this paper is the optimizing of the order of the pre-whitening filter and the length of the MED filter. Both the order of the pre-whitening method and the MED filter are optimized by maximizing the signal kurtosis since the analysis is aimed mainly at the increase of the “impulsiveness” of the signal in order to enhance the impulses generated by the impact of the defective rolling elements.

2. Minimum Entropy Deconvolution Technique

The Minimum Entropy Deconvolution technique (MED) is a type of the system identification method that was originally proposed by Wiggins (1978). Its main original use was to aid the extraction of reflectivity information in seismic data in order to identify and locate layers of subterranean minerals. MED has shown its effectiveness in deconvolving the impulse excitations from a mixture of response signals (Nandi et al., 1997; Boumahdi, Lacoume, 1995). It was used initially by Endo and Randall (2007) in the machine condition monitoring field in order to enhance the impulses arising from spalls and cracks in gears. It was then adopted by Sawalhi et al. (2007) to enhance the detection of spalls in rolling element bearings in high speed machines. The use of MED in the latest context improved the use of the Spectral Kurtosis (Antoni, Randall, 2006) as a surveillance technique. In this work, we show that MED has also a great potential to enhance the detection of bearing faults when they are masked by the effect of the transfer path. We also show its potential use for enhancing impulses generated by extended faults. These two main cases are presented and illustrated by processing signals from test rigs and actual machines.

Figure 6 illustrates the deconvolution process involved in the MED filtering when used to enhance the detection of bearing faults. In order to gain the full benefit from using the MED technique for rolling element bearings, it is recommended that the signal

For a clear diagnosis of the bearing fault a number of techniques have been proposed to separate deterministic components from the bearing component. Such techniques as the discrete random separation (DRS) (Antoni, Randall, 2004), self adaptive noise cancellation (SANC) (Ho, Randall, 2000) and the time synchronous averaging (TSA) (Braun, 2010), which benefits from the slippage phenomena, has been proposed with good results. This technique was applied to wind turbines diagnostics by Barszcz (2009). A number of papers proposed methods to improve the signal to noise ratio of the REB fault component by selecting a frequency range in which the energy of the signal component is relatively stronger. Different criteria for the so-called optimum frequency band (OFB) were proposed. A very successful approach (kurtogram), based on maximizing the kurtosis of the band filtered signal, was proposed by Antoni and Randall (2006). Recently, Barszcz and Jabłoński (2011) proposed a criterion of the kurtosis of the envelope spectrum (protrugram), which is more robust to random impulsive impacts. Another problem of detection of small impulses induced by REB faults is the transfer path between the faulty component (e.g. inner race) and the vibration sensor. Impacts, which are initially sharp after travelling the distance between the bearing and the sensor, may be very distorted. The method was originally proposed by Sawalhi et al. (2007) with the application to a test rig. This paper provides a means for removing the effect of the transfer path (h) through inverse filtration. The aim is to design an inverse filter to remove/minimize the effect of the transfer path filter. The base to optimize this filter is to minimize the entropy of the signal or, in other words, to maximize the impulsiveness (kurtosis) of the signal. This filter will then be used to deconvolve (as opposed to convolve) the signal, thus recovering the defective bearing fault signal (impulses) in a rather clear way. The filter used to do this is referred to as minimum entropy deconvolution (MED).

1.4. Paper layout and contribution

The paper is organized as follows: after this introductory and background section (Sec. 1), the minimum

![Fig. 5. Model of the generation of the vibration signal from a machine with gears and bearings.](image-url)
The criterion is: the maximum value of kurtosis. Thus, the optimization that recovers the output signal (of an inverse filter) with the general filter given by:

\[ y[n] = \sum_{l=1}^{L} f_k[l] \varepsilon_1[n - l]. \]  

Maximum criterion from (2) will occur when:

\[ \frac{\partial \{O(f[l])\}}{\partial f[l]} = 0. \]  

With the formula:

\[ \frac{\partial y[n]}{\partial f[l]} = \varepsilon[n - l], \]

Equation (5) can be presented as:

\[ b = Af, \]

where b is left side of Eq. (5), A is the second sum on the right side of (5) and f is the first sum on the right side of (5).

The proposed algorithm has been implemented in this study using the Objective Function Method (OFM) given in (LEE, NANDI, 2000). This method is an iterative optimization process, which is designed to maximize the kurtosis of the MED output (thus minimizing the entropy). The OFM achieves this by changing the values of the coefficients in the MED filter. The optimization process finishes when the values of the coefficients converge within the specified tolerance (Eqs. (7), (8)). The iteration finishes when the expected value of error ‘E(err) ≤ tolerance’.

If ‘E(err) > tolerance’, the filter coefficients \( f^{(0)} = f^{(1)} \) are updated and the process is iteratively repeated. A condition was also set to abort the iteration if the value of the E(err) was found to be diverging. The detailed steps of this implementation can be recalled from ENDÖ and RANDALL (2007).

\[ err = (f^{(1)} - \mu f^{(2)}) / \mu f^{(0)}, \]  

\[ \mu = \left( E(f^{(0)})^2 E(f^{(1)})^2 \right)^{1/2}. \]  

The next section shows the results of application of this algorithm to the test rig data and to the data from a real REB failure.
3. Case studies

A number of tests have been carried out to illustrate the benefits gained from using the MED technique to recover the impulses generated from the impacts of rolling elements with the fault region. The enhancement resides in increasing the impulsiveness of the signal and revealing the masked bearing fault signature. This has been verified by plotting the time domain signals and the envelope spectra during the different stages of processing.

The criteria for selecting the filter order for both the AR process as a pre-whitener and the length of the MED filter were based on maximizing the kurtosis of the output signal as the main aim is to enhance the impulsiveness of the residual and filtered signals.

3.1. UNSW fan test rig

3.1.1. Test rig and measurement description

Measurements were carried out on a bladed disk test rig at the Vibration and Acoustics Lab at UNSW. The test rig, which is shown in Fig. 7, was originally designed to develop models and techniques for monitoring the health of turbomachinery blades. It has 19 flat blades attached to a disk, which is mounted on a shaft. The shaft is supported by two self-aligning, double row ball bearings, which are mounted on sleeves and are contained within plummer blocks. The test rig is driven by a motor, which is coupled to the test rig via a 1:1 ratio multi-rib V belt (not shown in the figure). The motor speed can be controlled via a VVVF (Variable Voltage, Variable Frequency) drive. An accelerometer was placed midway between the casing and the defective bearing to mask the signature of the defect through the transfer function and the dominance of the blade pass harmonics. The signals were captured at a sampling frequency of 65.5 kHz for both inner race and outer race seeded faults (see Fig. 8). They are measured and analysed at a speed of 2400 rpm.

![Fig. 7. Schematic presentation of the UNSW bladed test rig.](image1)

![Fig. 8. Seeded faults: a) outer race, b) inner race.](image2)

3.1.2. Results for the outer race fault

Figure 9 shows the time domain signals during the different stages of processing. The envelope spectrum corresponding to each can be seen in Fig. 10. The raw time domain signal shown in Fig. 9a has a low kurtosis value (2.99) and shows no significant impact pattern that could be attributed to a bearing fault. When plotting the squared envelope spectrum for a frequency band from 1000 Hz to 25000 Hz (Band limits selected to reflect the frequency content of the signal), it is noticed that this spectrum is dominated by the shaft harmonics/sidebands. However, the ball pass frequency of the outer race (BPFO) and its harmonics are clearly present in the spectrum, but they have a lower amplitude.

In the second stage of processing (Fig. 9b), the signal 9a was order tracked with the aid of a once per revolution tachometer, and the synchronously averaged signal (containing the harmonics of the shaft speed) was removed. The plotted signal in 9b shows the residual, while 10b shows its corresponding spectrum. Note that the removal of the shaft harmonics had a very minimal effect on the raise of the kurtosis of the signal. It also had a minimum contribution to enhancing the harmonics of the BPFO. In fact the harmonics of the shaft speed are seen to remain in the envelope spectrum. This is believed to be due to...
modulating sidebands rather than true harmonics. The last two subplots in Figs. 9 and 10 show the results of pre-whitening and MED, respectively. Although their corresponding spectra show a clear extraction of the BPFO harmonics and a complete separation from the shaft harmonics and sidebands, the MED result is far superior and gives a very clear extraction and enhancement of the impacts. This is well reflected in a comparison of the kurtosis values.

The order of the AR model (4) used to pre-whiten the signal and the length of the MED filter (4096) has been selected by plotting the AR filter order and the length of the MED filter against the kurtosis values and selecting this one that maximized the kurtosis (high impulsiveness). The results are shown in Figs. 11a and 11b for the AR process and the MED filter, respectively. For pre-whitening purposes a low model order is enough to give a high kurtosis. For the MED filter, it is observed that the longer the filter the higher the kurtosis value (associated with a long tail). The variation between the kurtosis values above a filter length of 1024 samples is not dramatic (from 8 to 8.8), but it is a burden on the computational work. So a filter length between 1024 and 4096 samples would be suitable. Filters with length above 4096 samples will slightly increase the kurtosis, but will require a huge memory.

3.1.3. Results for the inner race fault

Figures 12, 13 and 14 present the results for the inner race fault in a similar way to what has been presented earlier in Figs. 9, 10 and 11 for the outer race fault. The most important issue that has to be no-
Fig. 12. Vibration signals (acceleration) for the case of an inner race fault: a) raw measured signal, b) after the removal of synchronous average, c) signal (b) pre-whitened (N=6), d) signal (c) after using the MED.

Fig. 13. Squared enveloped spectrum (band pass from 1000–25000 Hz) for the an inner race fault: a) processing the raw signal, b) residual, c) pre-whitened, d) processing the MED filtered signal.

Practice here is the effect of using the MED on enhancing not only the BPFI and its harmonics, but also their sidebands. This shows the effect of the ball when it enters and exits the load zone. The individual pulses are clearly seen in the MED filtered signal.

3.2. Wind turbine extended inner race fault

MED has been attempted on a signal taken from a wind turbine with extended inner race spalls. The turbine was of the GE 1.5sl type from one of the German wind parks. Figure 15 presents the layout of the wind turbine with location of sensors.

This is a 1500 kW turbine with the doubly fed generator and pitch control (HAU, 2006). The main rotor with three blades is supported by the main bearing and transmits the torque to the planetary gear. The planetary gear input is a plate, to which the main rotor is connected. The planetary gear has three planets,
with their shafts attached to the plate. The planets roll over the stationary ring and transmit the torque to the sun. The sun shaft is the output of the planetary gear. Further, the sun drives the two-stage parallel gear. The parallel gear has three shafts: the slow shaft connected to the sun shaft, the intermediate shaft and the fast shaft, which drives the generator. The doubly fed generator produces an AC current of slightly varying frequency (the nominal rotational speed of the generator shaft is 1800 rpm). The turbine has had a bearing fault on the generator shaft in its inner race as seen in Fig. 16.

![Fig. 16. Spalled inner race of a wind turbine.](image)

Since the rotational speed of the turbine was varying, there was a need to resample the signal to the order domain. Moreover, the changes of the rotational speed were occasionally significant and affected the vibration signal level. We have manually selected the data when the rotational speed was within a narrow range and its variation was less than 3%. In another case a technique for normalization of the signal level should be applied. A similar technique was proposed by Gibiec (2006).

The raw acceleration time domain signal, the results of the different processing stages, their corresponding envelope spectra and the selection criteria are shown in Figs. 17, 18 and 19, respectively.

As can be seen from Fig. 17, the application of the MED has significantly increased the kurtosis of the vibration signal. While the kurtosis of the raw signal was just 4.03, it reached 21.88 after application of the obtained inverse filter. These results can be also clearly observed in Fig. 18, which presents the envelope spectra of signals from Fig. 17. It is observed in the envelope spectra that MED not only causes an increased clarity of the BPFI harmonics, but also discloses the presence of strong modulation by the rotational speed of the shaft. The harmonic spacing in Fig. 18 equals 283.87 Hz, which was found equal to the repetition period of the BPFI (ball pass frequency inner ring). The sidebands were spaced by 30.11 Hz, which is the rotational speed of the generator shaft during the measurement session.

![Fig. 17. Vibration signals (acceleration) for the extended inner race fault: a) raw measured signal, b) after the removal of synchronous average, c) signal (b) pre-whitened, d) signal (c) after using the MED.](image)

![Fig. 18. Envelope spectra (band pass from 1000 to 10000 Hz): a) raw, b) residual after subtracting the synchronous average, c) signal (b) pre-whitened, d) the MED result.](image)
Finally, the dependency between the AR and MED algorithm parameters were plotted (see Fig. 19). The trend observed earlier in the experimental test rig (Figs. 11 and 14) can be clearly seen here. Again a low AR order model has been used (AR (1)) and a filter length of 4096 was used (although 1024 or 2048 would also give good enough results).

4. Summary and conclusions

This paper presents a further development of the minimum entropy deconvolution (MED) method to aid extracting faults in rolling element bearings. The MED technique was applied to signals with defective bearings taken from an experimental test rig and a wind turbine. Prior to applying the MED techniques on the signals, order tracking was used and a synchronous averaged signal was obtained. The synchronous averaged signal (containing deterministic components) was subtracted from the total signal to get a residual signal, which contains fault impulses. The residual signal was then pre-whitened to aid further the enhancement of the impulses by minimizing the variation between adjacent frequencies. The MED was then applied with the aim of removing the effect of the transfer path (deconvolution) and enhance the clarity of the impulses and then the detection and diagnoses of the bearing fault. It is shown that in both cases, the test rig and the wind turbine, MED significantly increase the peakedness of the vibration signals and the clarity of the impulses. This has been illustrated in both the time domain signals and further observed in the envelope spectra. In particular, the modulations at the shaft speed in the case of inner race faults were dramatically enhanced and observed with the introduction of the MED technique. The selection of the filter length for the MED and the model order for pre-whitening are based on maximizing the kurtosis of the signal, which in effect means more clarity in the impulses and a better detection and analyses of the fault. It is observed that for pre-whitening purposes a low model order is usually required to achieve a high kurtosis. For the MED filter, it is observed that the longer the filter the higher the kurtosis value (associated with a long tail). The variation between the kurtosis values above a filter length of 1024 samples is not dramatic, but burdens the computational work. So a filter length between 1024 and 4096 samples would be suitable. Filters with length above 4096 samples will slightly increase the kurtosis, but require a huge memory.

Acknowledgment

The authors would like to express their gratitude to the company SeaCom GmbH (Herne, Germany) for access to the wind turbine data.

References

9. BRAUN S. (2010), The synchronous (time domain) average revisited, The 24th International Conference on


