ICSV14

Cairns • Australia

9-12 July, 2007

14th International Congress on Sound & Vibration

ROLLING ELEMENT BEARING FAULT DIAGNOSIS

BASED ON GENETIC ALGORITHMS

Yongxiang Zhang[1], R. B. Randall[2]

[1]School of Mechanical and Manufacturing Engineering,
Huazhong University of Science and Technology，Hubei, Wuhan, 430074, China
[2]School of Mechanical and Manufacturing Engineering,
University of New South Wales, Sydney, NSW 2052, Australia
zyx6302@public.wh.hb.cn

Abstract

The rolling element bearing is a key part in many mechanical facilities and the diagnosis of its faults is very important in the field of machinery health monitoring. Currently the resonant demodulation technique (envelope analysis) has been widely exploited in practice. However, much practical diagnostic equipment for carrying out the analysis gives little flexibility to change the analysis parameters for different working conditions, such as variation in rotating speed, and different fault types. Because the signals from a flawed bearing have features of non-stationarity, wide frequency range and weak strength, it can be very difficult to obtain the best analysis parameters for diagnosis. However, the kurtosis of the vibration signals of a bearing is different from normal to bad condition, and is robust in varying conditions. Secondly, as genetic algorithms have a strong ability for optimization, the authors present a model and algorithm to design the parameters for optimal resonance demodulation using kurtosis as a criterion. The feasibility and effectiveness of the proposed method are demonstrated by experiment and give better results than the classical method of arbitrarily choosing a resonance to demodulate. The method gives more flexibility in choosing optimal parameters than another optimization procedure based on the fast kurtogram.

1. INTRODUCTION

Rolling element bearings are the heart of almost every rotating machine, including planes flying in the sky and trains running on track. Bearing failures can sometimes cause both personal damage and economic loss, if the fault cannot be detected and diagnosed well in advance. Therefore they have received a lot of attention in the field of vibration analysis as they represent an area where much can be gained from the early detection of faults. R. F. Burchill et al. [1] presented the method of resonance demodulation to diagnose the fault of rolling element bearings in the 1970’s, and the SPM Company later developed an instrument to detect rolling element bearing faults based on measurement of the resonant responses of an accelerometer excited by the faults**.** In the first case, there was a general problem to find the optimum resonance to demodulate, to give the best separation of the bearing fault signal from background noise. In the second case, the resonance frequency used as a carrier was that of the accelerometer itself (approx. 33 kHz), and was not always optimal [2]. With large machines and large faults, the resonances excited are often lower in frequency than 30 kHz. With evolution of the technologies of the sensor, signal processing and engineering measurement, many approaches have been developed in recent years, such as wavelet analysis, EMD (empirical mode decomposition), cepstral analysis and so on [3-9]. Although these methods have added to the development of the condition monitoring of rolling element bearings, much practical diagnostic equipment for carrying out the analysis still gives little flexibility to change the analysis parameters for different working conditions. In many cases such as complicated working conditions and variation in the rotational speed, the diagnosis sensitivity is poor. Also back in the 1970's, kurtosis was suggested as a means of detecting the transient pulses arising from gear and bearing faults, but once again little guidance was given for choosing an appropriate frequency range for filtering bearing signals to maximize the kurtosis [10]. Currently a number of researchers have testified that the spectral kurtosis of the vibration of rolling element bearings can better characterize the transients which arise from the faults in a bearing and have developed some approaches to diagnose the fault of a bearing at the same time as indicating the optimum frequency bands for demodulation [11-14]. Since genetic algorithms have adaptive features and a strong ability to optimize parameters, this paper presents a new approach to detect the faults in rolling element bearings based on a genetic algorithm and the kurtosis as a criterion.

2. THE CHARACTERISTICS OF ROLLING BEARING SIGNALS

The key to fault diagnosis of a rolling element bearing is to capture the special symptoms arising from their faults. Acceleration signals measured on the casing consist of two parts: $y(t) = x(t) + n(t)$, where $x(t)$ is the defect-induced impulse responses and $n(t)$ is the background noise, including vibration signals generated by other components, such as rotor unbalance and gear meshing. Because of the structure and mode of operation of rolling element bearings, $x(t)$ has distinct features as follows

(1) **Wide frequency:** As a rolling element strikes a localized defect in a very short period, it excites a wide range of resonances of the structure of the bearing system. For initial local defects, almost all frequencies are excited equally, but with extended spalls whose surfaces tend to become smoother with wear, not all frequency ranges are equally excited, and it becomes more important to find the resonances actually excited in each case.

(2) **Small energy:** The energy created by the defect is very small, and typically occupies less than one thousandth of the total signal energy. A band has to be found where the bearing signal dominates over other components.

(3) **Nonstationary signal:** Incipient bearing faults produce a series of repetitive short transient forces, which in turn excite structural resonances. Hence a reasonably versatile model for $x(t)$ is the generalized shot noise process [11]: $x(t) = \sum_k X_k h(t - \tau_k)$ where $h(t)$ is the impulse response resulting from a single impact and where $\{X_k\}$ and $\{\tau_k\}, k \in Z$ are sequences of random variables which account for possibly random amplitudes and random occurrences of the impacts, respectively. The stochasticity of the occurrences $\{\tau_k\}$ is caused by the random slips of the rolling elements, and the spacing between the pulses is a random

variable that varies by approximately 1% from the kinematic spacing that would occur in the case of no slip. In [6] it is shown that because there is no memory of previous slips, the resulting signals are not exactly cyclostationary, but can be termed “pseudo-cyclostationary” and usefully treated as though cyclostationary. The amplitudes $\{X_k\}$ represent the time-varying amplitude-modulation of the impacts, and typically have a deterministic component caused by periodic passage of the fault through the load zone and periodic variation of the signal transmission path, and a stochastic component caused by random variations in load and bearing component geometry. Because pulses are only generated by positive force between the bearing components, the amplitudes $\{X_k\}$ are always non-negative.

Because the signals generated by a defective rolling element bearing have the characteristics mentioned above, it is difficult to recognize their faults through simple frequency analysis. An early developed and widely used technique to tackle this issue is so-called envelope analysis. The key to this method is to choose a good band-pass filter. However, for fixed parameters of the band-pass filter (such as central frequency, bandwidth) which normally can’t be changed with the bearing working condition, the filter choice may not always be sensitive enough for some working conditions and/or fault types. A typical example is where a bearing runs over a wide range of speed. When it runs at very low speed, the defects may excite relatively low frequency resonances, while when it works at high speed, higher resonance frequencies may be excited. Thus, for a band-pass filter with constant parameters, it can’t meet the demands to detect the fault precisely for a bearing working over a wide speed range or varing conditions. The kurtosis of the vibration signal of a rolling element bearing can characterize the transients which result from the bearing fault, if they can be separated by an optimal filter from the background noise of harmless vibrations. Hence we decide to utilize a genetic algorithm to optimize the parameters of the band-pass filter using a cost function based on kurtosis.

3. THE ENVELOPE ANALYSIS BASED ON GENETIC ALGORITHM

The key to successful envelope analysis is to design a band-pass filter dynamically, whose filter parameters can be varied with the bearing working condition. For a band-pass filter, the six parameters to be determined are that it attenuates no more than *Rp* dB in the pass-band, and has

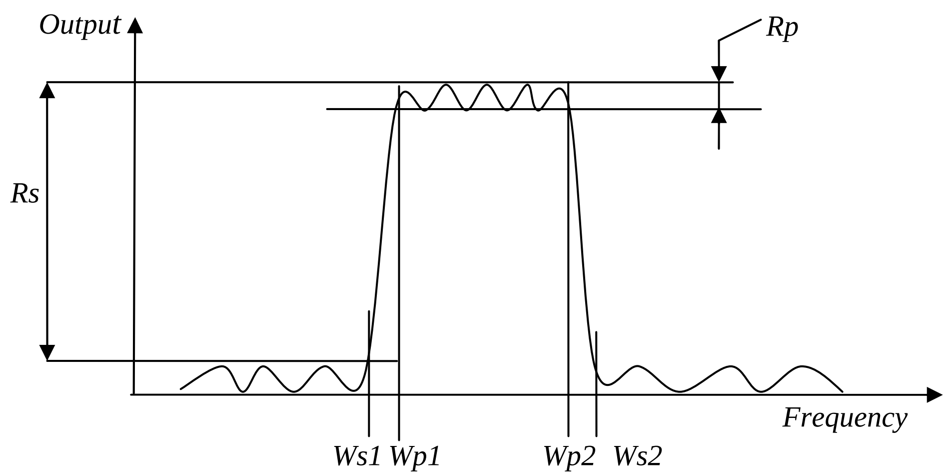

Fig 1 Band-pass Filter Sketch Map

at least *Rs* dB of attenuation in the stop-band, *Wp1*,*Wp2* and *Ws1*, *Ws2* are the pass-band and stop-band edge frequencies, shown in Fig 1. Previously these parameters were chosen according to practical experience; here we select them using a genetic algorithm, which has the ability to search the optimal solution over the whole zone and give fast convergence to solve the problem, in particular for nonlinear optimizations problems. The scheme for solution of the problem is shown in Fig 2.

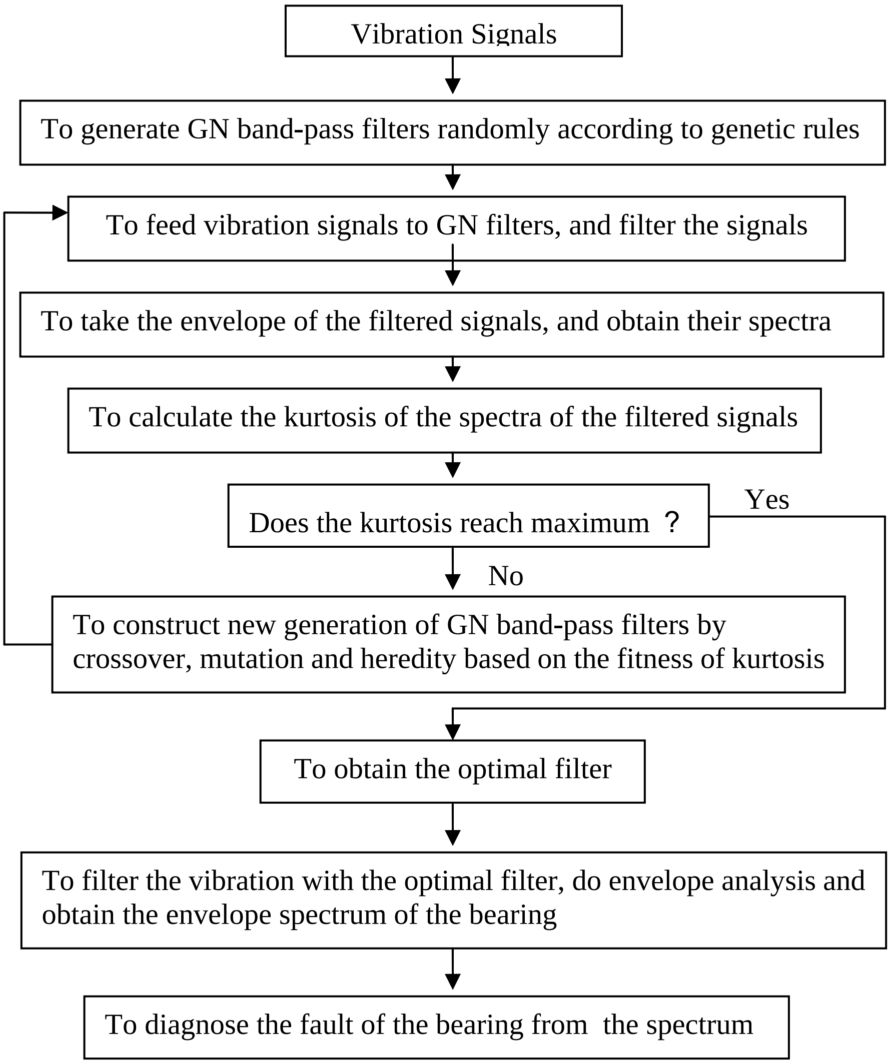

Fig 2 The flow chart of the genetic algorithm (GN = genotype population)

3.1 Expression of the parameters of a filter by the genotype

The parameters of a filter must be expressed by the genotype for a genetic algorithm. Here we propose a coding approach for the genotype based on binary bits. We use 49 binary bits to express a filter, where R_s occupies seven bits from the first to the seventh bit, R_p occupies two bits from the eighth to the ninth bit, W_{s1} , W_{p1}, W_{p2} , W_{s2} each occupy ten bits and the order arrangements are the same as R_s and R_p, the expression being shown as follows.

$$\underbrace{*******}_{R_s}\ \underbrace{**}_{R_p}\ \underbrace{**********}_{W_{s1}}\ \underbrace{**********}_{W_{p1}}\ \underbrace{**********}_{W_{p2}}\ \underbrace{**********}_{W_{s2}}$$

Where * stands for 0 or 1. The genotype has the relationships with R_s, R_p, W_{s1}, W_{p1} , W_{p2} and W_{s2} shown by formulas (1) to (6).

$$R_s = \sum_{i=1}^{7} chrom(i) \cdot 2^{i-1} \tag{1}$$

$$R_p = \sum_{i=1}^{2} chrom(i+7) \cdot 2^{i-1} \tag{2}$$

$$W_{s2} = \left[\sum_{i=1}^{10} chrom(i+39) \cdot 2^{i-1} \right] \cdot {}^{F_s}\!/_{2} \Big/ 1024 \tag{3}$$

$$W_{p2} = \left[\sum_{i=1}^{10} chrom(i+29) \cdot 2^{i-1} \right] \cdot W_{s2} / 1024 \tag{4}$$

$$W_{p1} = \left[\sum_{i=1}^{10} chrom(i+19) \cdot 2^{i-1} \right] \cdot W_{p2} / 1024 \tag{5}$$

$$W_{s1} = \left[\sum_{i=1}^{10} chrom(i+9) \cdot 2^{i-1} \right] \cdot W_{p1} / 1024 \tag{6}$$

In the formulas (1) to (6), Fs is sampling frequency, *chrom*(*i*) stands for the binary bit of the genotype. Hence, if the genotype of a filter has been determined, the parameters of a filter will be selected completely.

3.2 Genetic algorithm operation

After the formula of the filter has been expressed by genotype, the genetic algorithm can be carried out as follows.

(1) Initial group of genotype

For starting the genetic algorithm operation, the initial group of genotypes must be randomly generated. The size of the genotype population (*GN*) is decided by experience. From experiments, we find that $40 \le GN \le 100$ is suitable. In this paper *GN* is equal to 60.

(2) Fitness of genotype

The fitness of a filter is an indicator to distinguish the fault effectively for different bearing working conditions. The advantage of the kurtosis as fitness parameter is that it takes high values in the presence of the fault signal $x(t)$, whereas it is ideally zero when only background noise $n(t)$ is present. This is true when the impulse responses generated by the impacts are sufficiently well separated and when the signal-to-noise ratio is sufficiently high. As a matter of fact, background noise often embodies strong vibrations from several competing sources (e.g. harmonics of rotating parts, random impulses from friction and contact forces, flow noise, etc.) which span a large frequency range and seriously mask the signal of interest. As a result, the kurtosis is unable to capture the peakiness of the fault signal and hardly departs from a value of zero. In this situation, the kurtosis as a global indicator is not appropriate, and it is preferable to apply it locally in different frequency bands. This is exactly what the spectral kurtosis does. According to Ref.[14], if *y* is the squared envelope, then the spectral kurtosis can be calculated using equation (7), to give a single value for each filter.

$$kurtosis(y) = \frac{mean(y^2)}{(mean(y))^2} - 2 \tag{7}$$

The value 2 is subtracted here to obtain a value zero for the squared envelope of Gaussian noise. The best filter is chosen as the one that gives the highest value of kurtosis. This filter is retained and then re-used to obtain the squared envelope signal and the squared envelope spectrum.

(3) Replicate

The partial children genotypes come from the replication of the parent genotype, (whose probability of fitness is higher the value *Ps* (replicative probability). The fitness probability λ_i of a signal from a filter is as follows

$$\lambda_j = {q_j}/{q_{\max}} \quad , \quad q_j = \frac{Kurtosis(j)}{\frac{1}{GN}\sum_{j=1}^{GN} Kurtosis(j)} \qquad (8)$$

Here $q_{\max} = \max(q_j),\ j = 1,2,\ldots GN$. If λ_j is larger than *Ps*, the genotype will be retained in the children genotypes.

(4) Crossover

A crossover operation is to choose a pair of parent genotypes randomly, and according to two random numbers *N1* and *N2* (1<*N2* < *N1*< 49) exchange segment bits of two genotypes from the *N1*-th bit to the *N2*-bit respectively. The crossover probability *Pc* is selected by experience, hence the children genotype is *GN* * *Pc*/2.

(5) Mutation

A mutation operation is to select *Pm***GN* parent genotypes according to mutation probability *Pm*, then change each bit from original “1” to “0” or original “0” to “1” to form children genotypes.

(6) Terminating conditions

The genetic algorithm operation is terminated when satisfying one of the following conditions:

(a) The maximum value of the fitness doesn’t change during certain generations.

(b) The fitness is larger than a given value. For example, if kurtosis(*y*) > 15, the fault of a bearing can be recognized by experience.

4. EXPERIMENTAL INVESTIGATION

In order to testify the feasibility and effectiveness, signals were used that had been collected by N. Sawalhi at The University of New South Wales. In the test rig, a single stage gearbox is driven primarily by a 3-phase electric motor, but with circulating power via a hydraulic motor/pump set. The input and output shafts of the gearbox are arranged in parallel and each shaft is supported by ball bearings (Koyo 1205), one of which is used for investigation. A rough fault was introduced into the inner race of the ball bearing. This was performed using electric spark erosion and generated a rough surface over half the inner race. The signals were collected at a speed of 600 rpm using an accelerometer positioned on the top of the gearbox casing near the bearing. It had previously been found that this extended rough fault was difficult to detect by conventional envelope analysis. A typical experimental result is shown in Fig 3. The defect frequency at which the ball passes the defect on the inner race (BPFI) can be estimated as 71.1 Hz at the speed 600 rpm and the pulse series is modulated by the shaft speed 10 Hz. According to the previous analysis, we designed a filter based on the genetic algorithm, for which the parameters of the genetic algorithm are: *GN*=60, *Ps* = 0.59, *Pc* = 0.4, *Pm* = 0.01. As explained above, when the kurtosis of the envelope signal reaches 12, the fault signal caused by a defective rolling element bearing can be separated from the background noise, so the target kurtosis is assumed to be 15. The parameters for the optimal filter were found to be: $R_s = 32.9$ dB, $R_p = 1$ dB, $W_{s1} = 0.431$ kHz, $W_{p1} = 14.21$ kHz, $W_{p2} = 21.37$ kHz, $W_{s2} = 26.88$ kHz. When the vibration signal is filtered by the optimal filter and then envelope analyzed, the envelope spectrum of the defect bearing is illustrated in Fig 3, and the kurtosis reaches 23.6. From Fig 3(b), it is verified that the spectrum of the original vibration signal cannot detect any bearing fault indication, while the spectrum from the envelope analysis of the optimal filter signal not only eliminates the strong gear vibration signal but also indicates the defect frequency 71 Hz (including its second and third harmonic frequencies 142Hz and 213Hz) surrounded by sidebands spaced at shaft speed 10 Hz. Therefore according to this information we can determine the fault and its location in the bearing, as well as the severity of the fault using the amplitude of the spectrum. On the other hand, using traditional envelope analysis with the parameters of $W_{p1} = 8.00$ kHz, $W_{p2} = 10.00$ kHz (chosen because there is a local resonance in the spectrum in this range) we see in Fig. 4 that we cannot get the desired result for the same signal, since the envelope spectrum is dominated by harmonics of shaft speed, in particular the gearmesh frequency (harmonic 32). We also used the fast kurtogram method [15] to process the same signal, and obtained a somewhat similar result, however, it should be realized that the fast kurtogram has a limited range of possible centre frequencies, in particular when the bandwidth is wide. The total frequency band is divided successively into 2, 4, 8….etc bands, which thus specifies the centre frequencies for each bandwidth, whereas using the GA method of this paper, the centre frequency,bandwidth and lower and upper edges of the filter are optimized independently

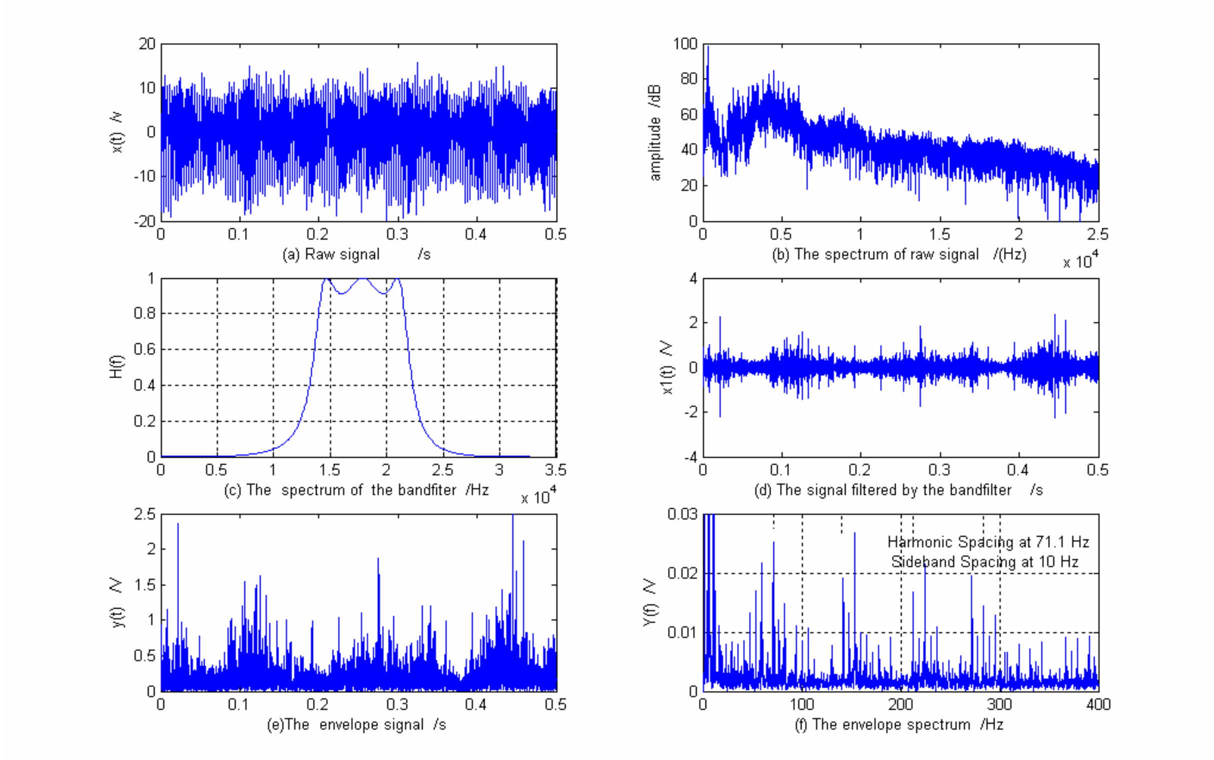

Fig 3 The vibration signals and their spectra for a defect bearing (inner defect)
(a) Original vibration signal (b) The amplitude spectrum of original signal
(c) The amplitude spectrum of the band-pass filter (d) The signal filtered by the BP filter (e) The envelope signal (f) The envelope spectrum

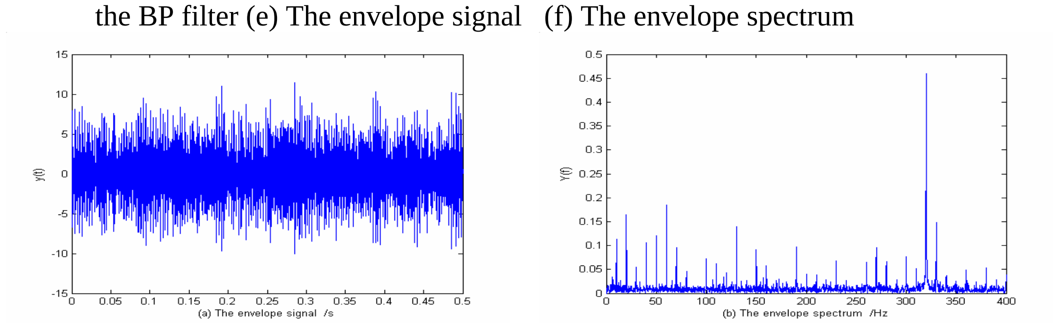

Fig 4 Traditional envelope analysis
(a) The signal filtered by the traditional BP filter (b) The traditional envelope spectrum

5. CONCLUSION

In this paper, the authors present an approach for diagnosis of rolling element bearings based on kurtosis of the bearing transient signal and a genetic algorithm to select an optimum bandpass filter. The advantages are firstly that because the spectral kurtosis of the bearing vibration signals robustly indicates the transients corresponding to the bearing fault, it is a good dynamic fault indicator under complicated bearing working conditions. Secondly the genetic algorithm based on spectral kurtosis has a strong ability to optimize the parameters for the envelope analysis very fast and with minimal constraints on centre frequency and bandwidth. The experimental investigation has testified the feasibility and effectiveness in comparison with traditional envelope analysis and even with an alternative optimization method based on the fast kurtogram.

6. ACKNOWLEDGEMENTS

The first author wishes to express his gratitude to The China Scholarship Council for financially supporting him as a visiting professor at The University of New South Wales.

REFERENCES

[1] R. F. Burchill, "Resonant structure techniques for bearing fault", Proceedings of Mechanical Failure Prevention Group, Novemember 1972, Gaithersburg, Md..

[2] H. Engja, J. Lippe, "Vibration analysis used for detection of roller bearing failure", *Norwegian Maritime Research* **3**, 23-33(1977).

[3] Huang Dishan, "A wavelet-based algorithm for the Hilbert transform", *Mechanical Systems and Signal Processing***10**(2),125-134(1996)

[4] P. Tse, Y.H. Peng, R. Yam, "Wavelet analysis and envelope detection for rolling element bearing fault diagnosis—their effectiveness and flexibility", *Journal of Vibration and Acoustics* **123** (3), 303–310(2001).

[5] J. Altmann, J. Mathew, "Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnosis", *Mechanical Systems and Signal Processing* **15** (5), 963–977(2001).

[6] J. Antoni, R.B. Randall, "Differential diagnosis of gear and bearing faults", *Journal of Vibration and Acoustics* **124** (2), 165–171(2002).

[7] Cheng Junsheng, Yu Dejie, Yang Yu , "A fault diagnosis approach for roller bearings based on EMD method and AR model" , *Mechanical Systems and Signal Processing* **20** , 350–362(2006)

[8] N.E. Huang, et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", *Proceedings of the Royal Society of London* A 454 (1998) 903–995.

[9] Yuh-Tay Sheen, Chun-Kai Hung, "Constructing a wavelet-based envelope function for vibration signal analysis", *Mechanical Systems and Signal Processing* **18**, 119–126(2004).

[10] D. Dyer, R. M. Stewart, "Detection of rolling element bearing damage by statistical vibration analysis", *Journal of Mechanical Design* **100,** 229-235(1978).

[11] Jérôme Antoni, "The spectral kurtosis: a useful tool for characterizing non-stationary signals", *Mechanical Systems and Signal Processing* **20**, 282–307(2006).

[12] Jérôme Antoni, R.B. Randall, "The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines", *Mechanical Systems and Signal Processing***20**, 308–331(2006).

[13] Jérôme Antoni, "Fast computation of the kurtogram for the detection of transient faults", *Mechanical Systems and Signal Processing***21**, 108–124(2007).

[14] N. Sawalhi, R. B. Randall, H. Endo, "The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis", *Mechanical Systems and Signal Processing,* In press (2007).