Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Sound and Vibration 289 (2006) 1066–1090
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics

Hai Qiu[a,*], Jay Lee[a], Jing Lin[b], Gang Yu[c]

[a]*Center for Intelligent Maintenance Systems, University of Cincinnati, OH 45221, USA*
[b]*Institute of Acoustics, Chinese Academy of Sciences, 17 Zhongguancun Street, Haidian, Beijing 100080, China*
[c]*Department of Mechanical and Industrial Engineering, Northeastern University, 60 Huntington Ave 334SN, Boston, MA 02115, USA*

Received 26 March 2004; accepted 10 March 2005
Available online 31 May 2005

Abstract

De-noising and extraction of the weak signature are crucial to fault prognostics in which case features are often very weak and masked by noise. The wavelet transform has been widely used in signal de-noising due to its extraordinary time-frequency representation capability. In this paper, the performance of wavelet decomposition-based de-noising and wavelet filter-based de-noising methods are compared based on signals from mechanical defects. The comparison result reveals that wavelet filter is more suitable and reliable to detect a weak signature of mechanical impulse-like defect signals, whereas the wavelet decomposition de-noising method can achieve satisfactory results on smooth signal detection. In order to select optimal parameters for the wavelet filter, a two-step optimization process is proposed. Minimal Shannon entropy is used to optimize the Morlet wavelet shape factor. A periodicity detection method based on singular value decomposition (SVD) is used to choose the appropriate scale for the wavelet transform. The signal de-noising results from both simulated signals and experimental data are presented and both support the proposed method.

*Corresponding author. Tel.: +1 414 229 3106; fax: +1 414 229 3107.
E-mail addresses: haiqiu@uwm.edu (H. Qiu), jaylee@uwm.edu (J. Lee), jinglin@mail.ioc.ac.cn (J. Lin).

0022-460X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2005.03.007

1. Introduction

Rolling element bearing are of paramount importance to almost all forms of rotating machinery and are among the most common machine elements. Bearing failure is one of the foremost causes of breakdowns in rotating machinery and such failure can be catastrophic, resulting in costly downtime. In order to prevent these kinds of failures from happening, various bearing condition monitoring techniques have been developed. Among them, vibration analysis has been used extensively due to its intrinsic advantage of revealing bearing failure [1,12].

After nearly four decades of study of the bearing failure mechanism [2–4], the theoretical background of bearing failure modes has been covered quite comprehensively. The signature of a damaged bearing consists of exponentially decaying ringing that occurs periodically at the characteristic frequency [3]. The vibration signal of a defective bearing usually considers being amplitude modulated at the characteristic defect frequency. Matching the measured vibration spectrum with the defect characteristic frequency enables defect detection and enables diagnosis of the defective area.

As for the vibration signal of rolling element bearing, signal modulation effect and noise are two major barriers in incipient defect detection. Because of the amplitude-modulated effect, the spectrum of defect signals consists of a harmonic series of frequency components present at the bearing defect frequency with the highest amplitude around the resonance frequency [5]. To overcome this barrier, an effective signal demodulation technique should be developed. Most of the time vibration signals are collected with a vibration sensor installed on the bearing housing. The sensors are subject to collecting vibration responses from other mechanical components and noise sources. The inherent deficiency of the measuring mechanism introduces a great amount of noise to the signal. The signature of a defective bearing is spread across a wide frequency band and hence can easily become masked by noise and low frequency effects [6]. The weak signature, at the early stage of defect development, is even more difficult to detect. A signal enhancing method is needed to provide more evident information for incipient defect detection of rolling element bearings.

To solve the problem of modulation, a large variety of signal demodulation methods have been proposed. Eshleman [7] proposed a hardware-based signal envelope technique in which signals are passed through a capacitor to produce the demodulated time waveform. Fast Fourier Transform (FFT)-based Hilbert transform is the traditional method for deriving the signal envelope and has been widely used in roller bearing diagnostics [8]. More recently, the wavelet transform has been used for signal demodulation [5,6,9] and optimal band-pass filter design [10]. In summary, Hilbert transform and wavelet transform offer promising techniques for signal demodulation. However, those methods do not successfully address how to enhance the weak signature from a noisy signal and how to detect early stage defects.

The problem of signal de-noising has a strong connection to roller element bearing prognostics. De-noising and extraction of the weak signature are crucial to fault prognostics in which case features are often very weak and masked by noise. Prognostics is achieved by detecting the defect at its initial stage and alerting maintenance personnel before it develops into a catastrophic failure. The standard approach for extracting signals from a noisy background is to design an appropriate filter, which removes the noise components and at the same time, lets the desired signal go through unchanged. Based on noise type and application, different filters can be

designed to conduct the de-noising [11]. However, for a situation where the noise type and frequency range are unknown, traditional filter design could become a very challenging task. Therefore, research is focused on finding alternative methods. The wavelet transform has been widely used in signal de-noising due to its extraordinary time-frequency representation capability [13], which is discussed in detail later in this paper. Traditionally, most of the signal de-noising approaches are dealing with detecting smooth curves from the noisy raw signals. However, the vibration signal from faulty mechanical components, such as gears and bearings, are more like impulses, instead of smooth and continuous curves. This unique feature constrains the application of conventional signal de-noising method. A de-noising method based on Morlet wavelet analysis is proposed and applied to the feature extraction of gearbox vibration signals [21]. This method seeks optimal wavelet filters that only yield the largest kurtosis value for the transformed signal, whereas the periodicity of the signal is not addressed.

In this paper, the performance of wavelet decomposition-based de-noising and wavelet filter-based de-noising methods are compared based on signals from mechanical defects. The comparison results reveal that the wavelet filter is more suitable and reliable to detect weak and impulse-like signatures of mechanical defect signals, whereas the wavelet decomposition de-noising method can achieve satisfactory results on smooth signal detection. In order to select the optimal parameters for the wavelet filter, a two-step optimization process is proposed. Minimal Shannon entropy is used as a criterion to optimize the shape factor of a Morlet wavelet. A periodicity detection method based on Singular Value Decomposition (SVD) is used to choose the appropriate scale for the wavelet transform. The signal de-noising results from both simulated signals and the experimental data are presented and both support the proposed method.

The remaining sections of this paper are organized as follows: In Section 2, the concept of a wavelet transform is reviewed. In Section 3, the wavelet decomposition-based de-noising method is discussed in detail. A comparison study is presented using two sets of simulated signals. The results suggest that for impulse-like signals, wavelet decomposition-based de-noising method is not able to achieve a satisfactory level of performance. In Section 4, the Morlet wavelet filter and its underlying capability of detecting weak impulse signatures from a noisy background is discussed and demonstrated using simulated signals. In Section 5, the proposed method is validated using the data collected from bearing run-to-failure tests. The result demonstrates that by designing an optimal wavelet filter bearing defects can be detected at an early stage of development and therefore bearing prognostics is possible.

2. Wavelet transform

The wavelet is obtained from a single function $\psi_{(a,b)}(t)$ by translation and dilation:

$$\psi_{(a,b)}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \tag{1}$$

where a is the so-called scaling parameter, b is the time localization parameter and $\psi(t)$ is called the "mother wavelet". The parameters of translation $b \in R$ and dilation $a > 0$, may be continuous or discrete.

The wavelet transform of a finite energy signal $x(t)$ with the analyzing wavelet $\psi(t)$ is the convolution of $x(t)$ with a scaled and conjugated wavelet:

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t-b}{a}\right) \mathrm{d}t, \tag{2}$$

where $\psi^*(t)$ stands for the complex conjugation of $\psi(t)$.

The wavelet transform $W(a,b)$ can be considered as functions of translation b with each scale a. Eq. (2) indicates that the wavelet analysis is a time-frequency analysis, or a time-scaled analysis. Different from the Short Time Fourier Transform (SFFT), the wavelet transform can be used for multi-scale analysis of a signal through dilation and translation so it can extract time-frequency features of a signal effectively.

Wavelet transform is also reversible, which provides the possibility to reconstruct the original signal. A classical inversion formula is

$$x(t) = C_{\psi}^{-1} \iint W(a,b)\psi_{(a,b)}(t) \frac{\mathrm{d}a}{a^2} \mathrm{d}b, \tag{3}$$

where

$$C_{\psi} = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} \mathrm{d}\omega < \infty, \tag{4}$$

$$\hat{\psi}(\omega) = \int \psi(t) \exp(-\mathrm{j}\omega t)\, \mathrm{d}t. \tag{5}$$

3. Wavelet decomposition-based de-noising

The underlying model for the noisy signal is basically of the following form:

$$x(n) = s(n) + \sigma w(n), \quad n = 0, 1 \ldots, N-1. \tag{6}$$

In this simplest model, $w(n)$ is standard Gaussian white noise, independent and identically distributed (i.i.d.), denoted by $w(n) \overset{\text{i.i.d.}}{\sim} N(0,1)$, and σ is the noise level. The objective of de-noising is to suppress the noise part of the signal $x(n)$ and to recover $s(n)$. Theoretically, this is accomplished by reconstructing the signal from the noisy data such that the mean squared error between $s(n)$ and the reconstructed signal is minimized. From a statistical viewpoint, the model is a regression model over time and the method can be viewed as a nonparametric estimation of the function $s(n)$ using an orthogonal basis.

Wavelet de-noising is based on the principle of multi-resolution analysis [13]. By multi-level wavelet decomposition the discrete detail coefficient and approximation coefficient can be obtained. Grossmann [14] proved that the variance and amplitude of the details of white noise at various levels decreases regularly as the level increases, whereas the amplitude and variance of the wavelet transform of the available signal are not related to the change of scale. According to this property, noise can be weakened or even removed by adjusting the wavelet coefficients properly.

The general de-noising procedure involves three steps. The basic version of the procedure is as follows:

1. *Signal decomposition*. Choose a wavelet basis, and choose a level N. Compute the wavelet decomposition of the signal at level N.
2. *Threshold detail coefficients*. For each level from 1 to N, select a threshold and apply soft thresholding to the detail coefficients.
3. *Signal reconstruction.* Compute wavelet reconstruction using the original approximation coefficients of level N and the modified detail coefficients of levels from 1 to N.

Generally speaking, this method performs very well on Gaussian noise and can almost achieve optimal noise reduction while preserving the signal. However, there are still three issues which attract intensive research efforts.

The first is how to select an optimum wavelet for a particular kind of signal. Basically, the wavelet decomposition would achieve a better result if the wavelet basis is "similar" to the signal under analysis. Currently there are still no common guidelines for how to select the optimum wavelet basis, or how to select the corresponding shape parameter and scale level for a particular application.

The second issue is related to threshold selection and how to perform thresholding. Despite a large variety of threshold selection strategies proposed in recent literatures [16–19], threshold selection for a specific application where prior knowledge about the data is limited is still an open issue.

The third issue is about the sparseness of wavelet coefficients, which will be addressed in this paper. The wavelet decomposition-based de-noising method relies on the basic idea that the energy of a signal will often be concentrated in a few coefficients in the wavelet domain. Therefore, the nonlinear thresholding function will tend to retain a few larger coefficients representing the signal and at the same time tend to reduce the noise coefficients to zero. It works well if signal $s(n)$ is a smooth curve with little to no abrupt changes. However, if the signal $s(n)$ consists of a lot of impulse components, which is often true in machinery diagnostic applications, a sparse wavelet representation is very difficult to obtain. This adds great difficulty to the wavelet de-noising process.

To better understand this limitation, consider the wavelet decomposition results of two simulated signals, where signal (a) is the smooth function (sinusoidal) with additive white noise, and signal (b) is a series of impulses with additive white noise (Figs. 1 and 2).

Multiple-level wavelet decomposition was conducted on both simulated signals. As shown in Fig. 3, if the signal was decomposed into N scale, one approximation coefficient and N detail coefficients can be obtained.

The fourth-order Symlet wavelet was adopted as the mother wavelet and both signals were decomposed into five scales. All six of the coefficients of the series were plotted into one chart to check the data sparseness.

Fig. 4 shows that the pure sinusoidal signal has much better wavelet presentation in terms of sparseness. The detail coefficients at scale 1, 2, 3 and 4 are all approaching zero value. Conversely for the impulse series signal, none of the coefficients are near zero. This also implies that impulse signals have less capability of being compressed or de-noised.

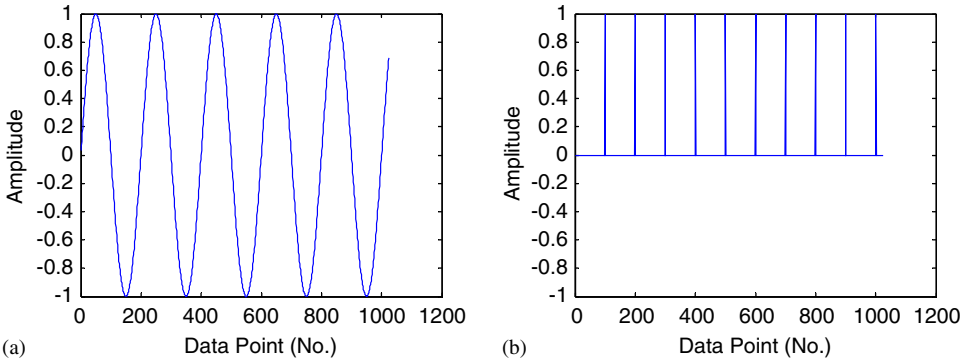

Fig. 1. Two simulated signals: (a) pure sinusoidal wave and (b) impulse series.

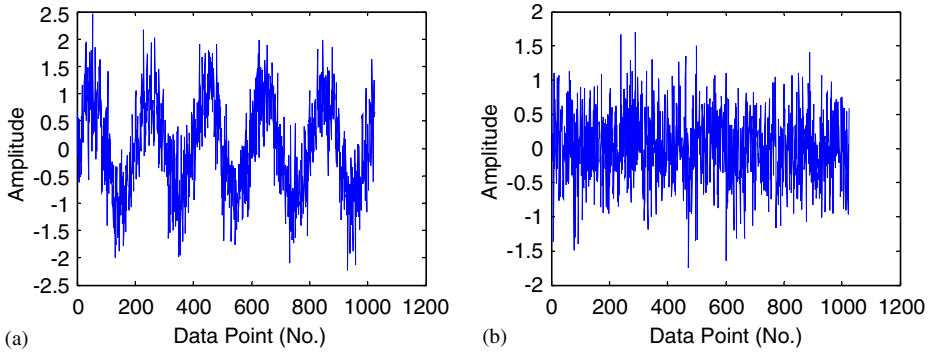

Fig. 2. Simulated signals with the addition of white noise.

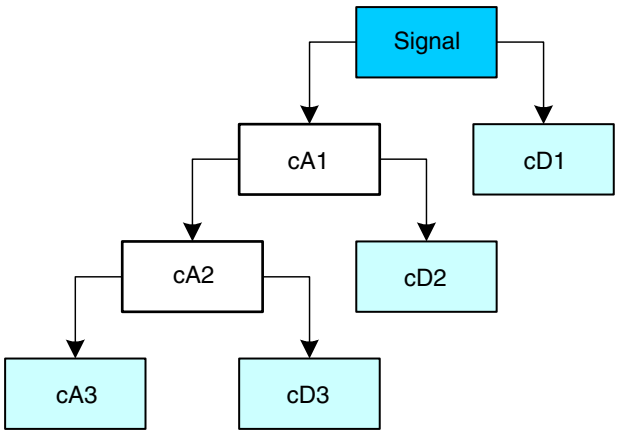

Fig. 3. Scheme of multiple level wavelet decomposition.

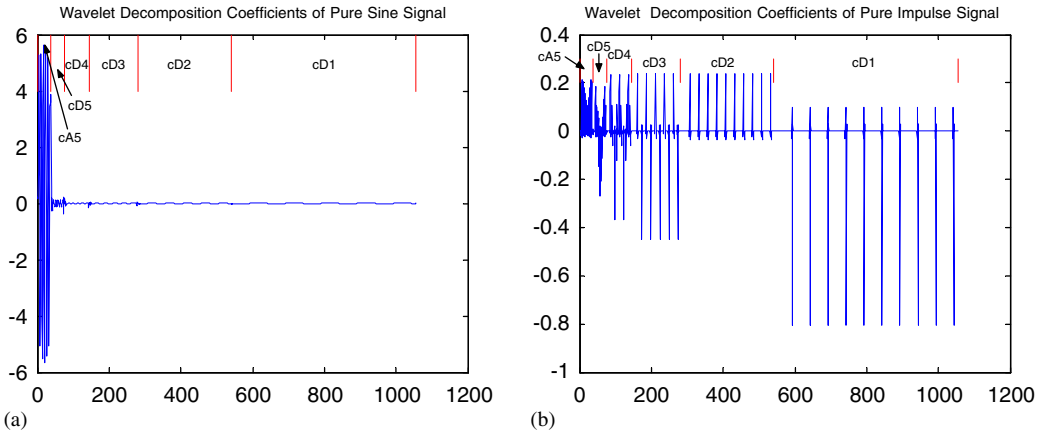

Fig. 4. Wavelet decomposition coefficients of two simulated signals: (a) pure sinusoidal wave and (b) impulse series.

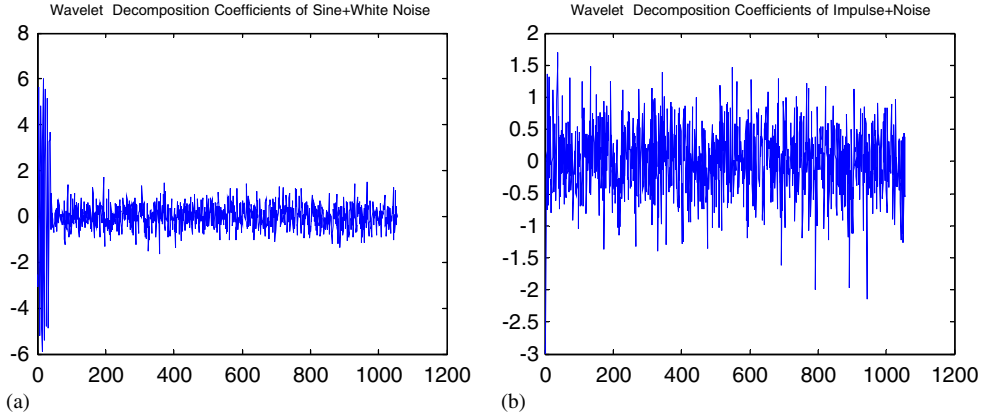

Fig. 5. Wavelet decomposition coefficients of two simulated signals with additive white noise: (a) pure sinusoidal signal and (b) impulse series.

The situation is much clearer by comparing the wavelet decomposition coefficients of the signals with added white noise as shown in Fig. 5. The coefficients curve shape in Fig. 5(a) still preserves the majority of information in Fig. 4(a), while the coefficient curve shape in Fig. 5(b) is greatly distorted from Fig. 4(b). Simply speaking, the de-noising can be treated as a process of recovering Fig. 4 from Fig. 5 by a particular thresholding strategy. Intuitively, to come out with a reasonable threshold to recover Fig. 4(a) from 5(a) is a relatively easy task. But it will be much more difficult to recover Fig. 4(b) from Fig. 5(b), even if a sophisticated thresholding strategy is applied.

The decomposition coefficients after thresholding are shown in Fig. 6. The threshold selection rule is based on Stein's Unbiased Estimate of Risk soft principle [19]. Fig. 7 shows the recovered signals after the thresholding procedure. Signal (a) is recovered quite well while signal (b) is distorted significantly.

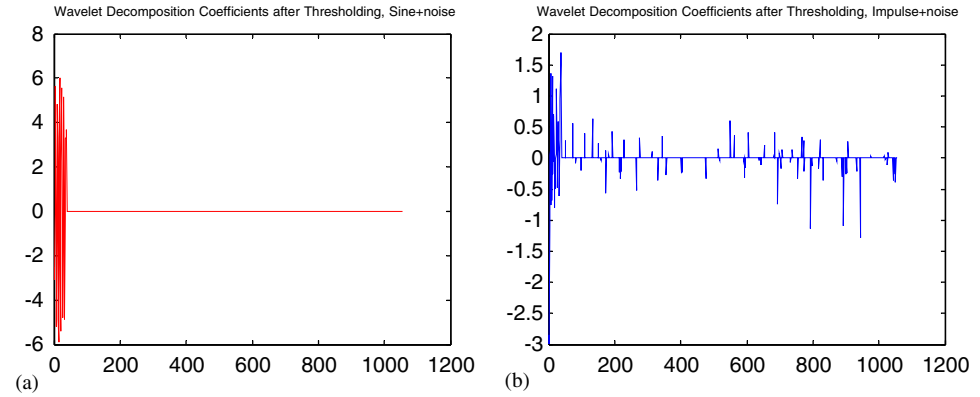

Fig. 6. Wavelet decomposition coefficients after de-noising by thresholding (a) pure sinusoidal signal with white noise and (b) impulse series with white noise.

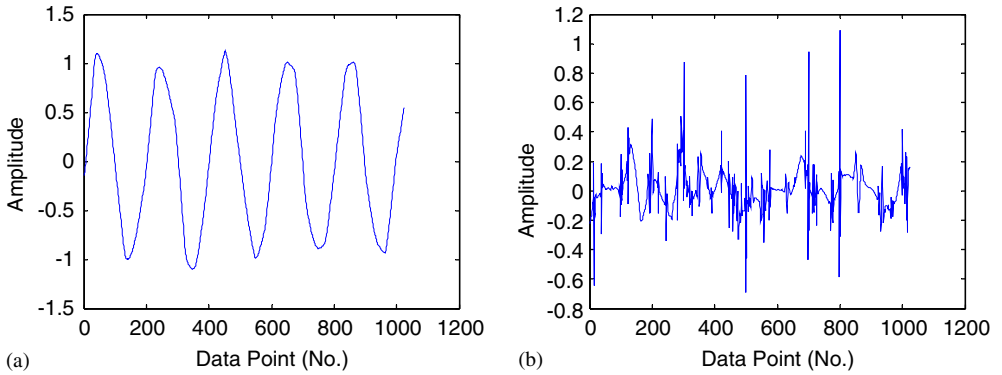

Fig. 7. The recovered signals obtained by wavelet de-noising (a) pure sinusoidal signal and (b) impulse series.

In summary, an orthogonal wavelet transform can compress the "energy" of the signal in a relatively small number of *big* coefficients, while the energy of the white noise will be dispersed throughout the transform with relatively *small* coefficients. The de-noising effect is impacted by the relative energy level of signal coefficients and white noise coefficients. Since only a small number of large coefficients can characterize the original signal (a), the pure sinusoidal signal, the classical wavelet de-noising method performs very well. However, it is much more challenging to de-noise the impulse series signal where wavelet coefficients are not sparse.

4. Optimal wavelet filter

4.1. The principle of the wavelet filter

Another method to extract useful information from a noisy signal is the wavelet filter. An important property of the Fourier transform is that convolution in one domain corresponds to

multiplication in the other domain. So Eq. (2) can take the following alternative form:

$$W(a,b) = \sqrt{a}F^{-1}\{X(f)\psi^*(af)\}, \tag{7}$$

where $X(f)$ and $\psi(f)$ are the Fourier transforms of $x(t)$ and $\psi(t)$, respectively, and F^{-1} denotes the inverse Fourier transform. Eq. (7) shows that the wavelet transform can also be considered as a special filtering operation. The frequency segmentation is obtained by dilating the analysis wavelet. In other words, the convolution process in the wavelet transform is simply a filtering operation if the daughter wavelet is treated as a filter kernel. The frequency response of the wavelet filter varies as the basis wavelet shape and scale changes, thus low-pass, high-pass, band-pass or even multiple-band pass filters can be built by reconstructing the wavelet coefficients at selected scales.

Another feature of Eq. (2) is that $W(a,b)$ gives the information of $x(t)$ at different levels of resolution and also measures the similarity between the signal $x(t)$ and the wavelet function. This implies that a wavelet can be used for feature discovery if the wavelet used is similar to the feature components hidden in the signal. To some extent, this convolution process of the daughter wavelet and the analyzed signal is similar to another classical concept of signal processing: *matched filtering,* which is originally derived from the correlation process.

In the application of machinery prognostics, what attracts attention in the original noisy signal is the periodicity and relative energy level of the impulse components, which are indicators of impacts due to cracks, spalling, or corrosion, etc. Therefore, the objective of weak signal detection is to *detect* the target signal, instead of *recreating* the signal. Specifically, in roller bearing prognostics, we will attempt to detect the hidden periodic impulses.

Fig. 8 shows a signal measured on a bearing with an outer race fault. The data was collected on a machinery fault simulator manufactured by SpectraQuest, Inc. The rotation speed was 50 Hz and the ball pass frequency on the outer race (BPFO) [3] is 152.6 Hz. That means the characteristic impulse period should be 1/BPFO, or equal to 0.0066 s. Since this measurement was conducted on a mature faulty bearing in a laboratory with the noise sources well under control, the signal clearly shows the characteristic impulse with the period as expected.

Enlarging Fig. 8 and only observing one impulse, the resonance excited by the impact is visibly demonstrated as shown in Fig. 9(a). The speed of impulse decay, the number of rebounds, and the

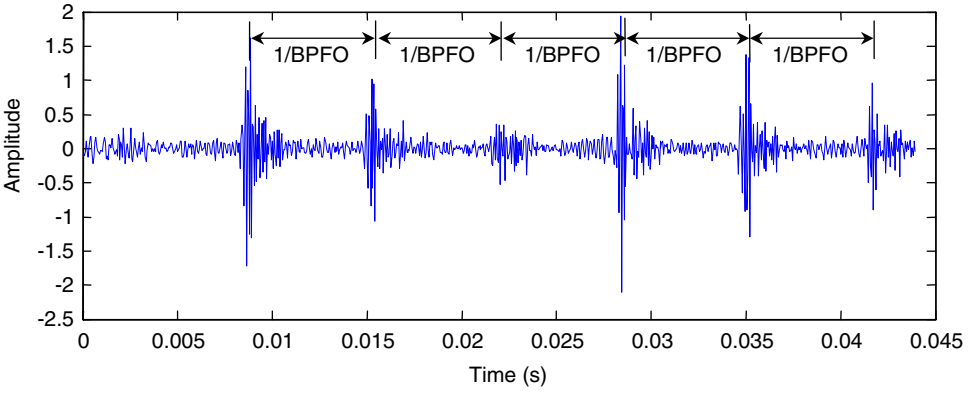

Fig. 8. Vibration signal of roller bearing with outer race fault.

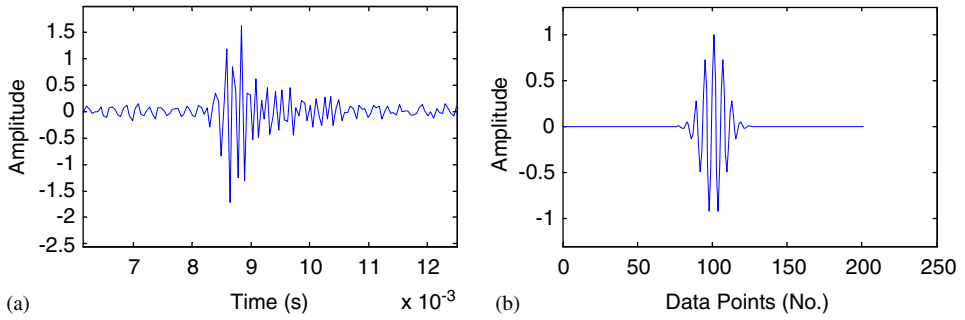

Fig. 9. Comparison of (a) mechanical impulse and (b) Morlet wavelet.

amplitude of impulse all depend on the system damping coefficient, resonance frequency, and signal transmission path. In addition, the signal can vary significantly with application-specific variables, including sensor location, sensor installation mechanism, sensor characteristics, etc. Therefore, the task of wavelet filter construction becomes identification of a wavelet that can be adjusted to have a "similar" curve shape with the bearing defect signature.

4.2. Introduction to the Morlet wavelet filter

The Morlet wavelet was defined as [20]

$$\psi(\omega) = \exp(-2\pi^2(v - v_0)^2), \tag{8}$$

which is a complex wavelet that can be decomposed into two parts, one for the real part, and the other for the imaginary part, as

$$\psi_r(t) = \frac{1}{\sqrt{2\pi}}\exp(-\beta^2 t^2/2)\cos(2\pi v_0 t), \tag{9}$$

$$\psi_i(t) = \frac{1}{\sqrt{2\pi}}\exp(-\beta^2 t^2/2)\sin(2\pi v_0 t), \tag{10}$$

where v_0 is a constant, and β is the shape parameter, which balances the time resolution and the frequency resolution of the Morlet wavelet.

Generally only the real part of the Morlet wavelet is used. The real part of the Morlet wavelet is a cosine signal that decays exponentially on both the left and right sides, and its function shape is very similar to an impulse. This similarity makes the Morlet wavelet very attractive and widely applied in mechanical fault diagnostic applications.

A daughter Morlet wavelet is obtained by time translation and scale dilation from the mother wavelet,

$$\psi_{a,b}(t) = \psi\left(\frac{t-b}{a}\right) = \mathrm{e}^{-\frac{\beta^2(t-b)^2}{2a^2}}\cos\left[\frac{\pi(t-b)}{a}\right], \tag{11}$$

where a is the scale parameter for dilation and b is for time translation. By carefully choosing parameters a and β, a daughter Morlet wavelet that closely matches the shape of mechanical

impulse can be constructed, as shown in Fig. 9(b). And if the wavelet transform (or wavelet filtering) is conducted based on this *filter kernel*, it should be able to detect the "similar" components from the noisy signal.

In the next section, a two-step approach is proposed to find the appropriate parameters β and a that can construct an optimal wavelet filter.

4.3. Optimal selection of shape factor β based on Shannon entropy

The *sparseness* of wavelet coefficients is often used as the rule for evaluating the efficiency of wavelet transforms. The wavelet corresponding to the fewest and dominant wavelet transformation coefficients of a signal is ideal. An optimal wavelet transformation should be able to *condense* the signal into several large coefficients. The simplest definition of sparseness states that in a sparse matrix or vectors, most of the elements are zero. The sparseness of a series can be evaluated by various criteria.

The simplest way to measure the sparseness is the L_0 norm

$$L_0 = \sum_i v_i; \quad v_i \in \{0, 1\}$$

$$\{x_i > \text{Threshold} \rightarrow v_i = 1; \;\; x_i < \text{Threshold} \rightarrow v_i = 0\}.$$

If $L_0 = 0$ the vector x is completely sparse (i.e. contains only zeros). Quite obviously, L_0 norm is not very practical for measuring sparseness of noisy data. Adding a very small measurement noise makes completely sparse data completely non-sparse. Therefore, a variety of sparseness measurement criteria are proposed by researchers [24], such as L_p norm, Tanh-Function, kurtosis, etc. Among them, Shannon entropy is one of the well-adopted sparseness criterion.

Shannon entropy was first introduced by Shannon in connection with communication theory in 1948 [25,26]. Shannon entropy is defined as

$$H(p) = -\sum_{i=1}^{n} p_i * \log p_i, \qquad \sum_{i=1}^{n} p_i = 1,$$

where p_i is the probability of observing the ith possible value of random variable $X \in [x_1, x_2, \ldots, x_n]$.

Shannon entropy is the central role of information theory sometimes referred as measure of uncertainty. The entropy of a random variable is defined in terms of its probability distribution and can be shown to be a good measure of randomness and sparseness. Shannon Entropy, thus can be used to evaluate the sparcity of wavelet coefficients [15]. Wavelet transform coefficients with minimal Shannon entropy can be treated as the sparsest result. Therefore, the corresponding shape factor β can be adopted as the optimal result.

4.4. Optimal selection of scale a based on singular value decomposition

After the shape factor β is determined by the minimal Shannon entropy criterion, the next step is to choose the appropriate wavelet transformation scale a, in other words, the frequency range of the wavelet filter so that the periodic pattern of the noisy signal can be detected.

Since the objective of de-noising is to identify the weak periodic components from the noisy signal, the periodicity of wavelet coefficients can be used as the criterion for selecting the optimal scale a. The scale a that can reveal the strongest periodicity from the wavelet coefficients will be selected as the optimal wavelet transform scale.

Conventionally, signal periodicity detection can be conduced by spectral analysis methods such as Fourier analysis, power spectral density, periodogram, etc. Nevertheless traditional Fourier-based methods assumed that the signal can be decomposed into multiple components, where only the sinusoidal pattern is permissible for each component [27]. Therefore, when dealing with vibration signals of mechanical faults, whose patterns are more impulse-like and their amplitudes vary with time, Fourier-based method cannot reveal its periodicity explicitly. In addition, the weak domination of impulse series compared with the background noise adds another constraint to conventional methods.

Those constraints lead to the development of a new periodicity detection method. Singular Value Decomposition (SVD) can be applied to detect the periodicity of the time series [22]. It is much more powerful and sensitive in terms of information content and robustness than the presently available tools based on Fourier decomposition [27]. The SVD of an $m \times n$ matrix D is defined as the decomposition [23]

$$D = UEV^{\mathrm{T}}, \tag{12}$$

where U is $m \times m$ square matrix and V is $n \times n$ square matrix with orthogonal columns so that

$$U^{\mathrm{T}}U = I, \qquad V^{\mathrm{T}}V = I. \tag{13}$$

Additionally E is an $m \times n$ diagonal matrix, $E = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_{\mathrm{p}})$, with $p = \min(m, n)$, and the diagonal elements $[\sigma_1, \sigma_2, \ldots, \sigma_{\mathrm{p}}]$ of matrix E are the singular values of matrix D and non-negative numbers $[\sigma_1, \sigma_2, \ldots, \sigma_{\mathrm{p}}]$ are conventionally arranged as $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n \geqslant 0$. The power of the SVD becomes apparent as its connections with other fundamental topics of linear algebra are explored. For example, if D has rank r and $r > 0$, then D has exactly r strictly positive singular values, so that $\sigma_r > 0$ and $\sigma_{r+1} = \cdots = \sigma_{\mathrm{p}} = 0$. If D has full rank, all its singular values are nonzero.

Consider a periodic signal $X = [x_1, \ldots, x_l]$ with a period of length n. A matrix $\mathbf{X}$ can be formed by partitioning the series into periods and placing each period as a row of $\mathbf{X}$, as

$$\mathbf{X} = \begin{pmatrix} x(1) & \ldots & x(n) \\ x(n+1) & \ldots & x(2n) \\ \vdots & \ddots & \vdots \\ x((m-1)n+1) & \ldots & x(mn) \end{pmatrix}. \tag{14}$$

The matrix X has m repeated rows and is of rank 1. Therefore, it should have only 1 non-zero singular value σ_1 and $m - 1$ zero singular values.

Now consider the case of a periodic waveform with time-varying amplitude plus noise. Assuming the period length of the time series $X = [x_1, \ldots, x_l]$ is still n, a different size matrix $\mathbf{X}(\mathrm{round}(l/i), i)$, $2 \leqslant i \leqslant l/2$ can be formed by dividing the time series into segments with different lengths i. The matrix X may now be full rank due to the noise, but σ_1 would be very large

compared to the rest of singular values when $i = n$. Hence, the ratio

$$\delta_i = \left(\frac{\sigma_1}{\sigma_2}\right)^2 \tag{15}$$

will exhibit its maximal value at $i = n$. Then δ_i can be used to estimate the periodicity of the signal.

Connecting the two optimization processes, a two-step optimal parameter selection algorithm can be designed as depicted in Fig. 10.

An experiment based on simulated data is conducted to compare the periodicity detection capability of Fourier-transform-based method and SVD-based method. Fig. 11(a) shows a simulated impulse signal consisting of 10 impulses with period of 100 data points. Assuming that the data sampling rate is 1000 Hz, the impulse period is 0.1 s. Fig. 11(b) shows the simulated signal with additive white noise. Obviously, the additive white noise significantly weakens the periodic characteristic of the simulated signal. Fig. 11(c) shows the Fast Fourier Transform (FFT) spectrum of the simulated signal. The expected periodic component, which should be represented by a harmonic at 10 Hz (0.1 s) in the FFT spectrum, is heavily masked by the background noise and higher frequency harmonics. On the other hand, the periodicity of the simulated signal is clearly revealed by the SVD-based method as shown in Fig. 11(d). A peak at 0.1 s in Fig. 11(d) clearly indicates the existence of the periodic component.

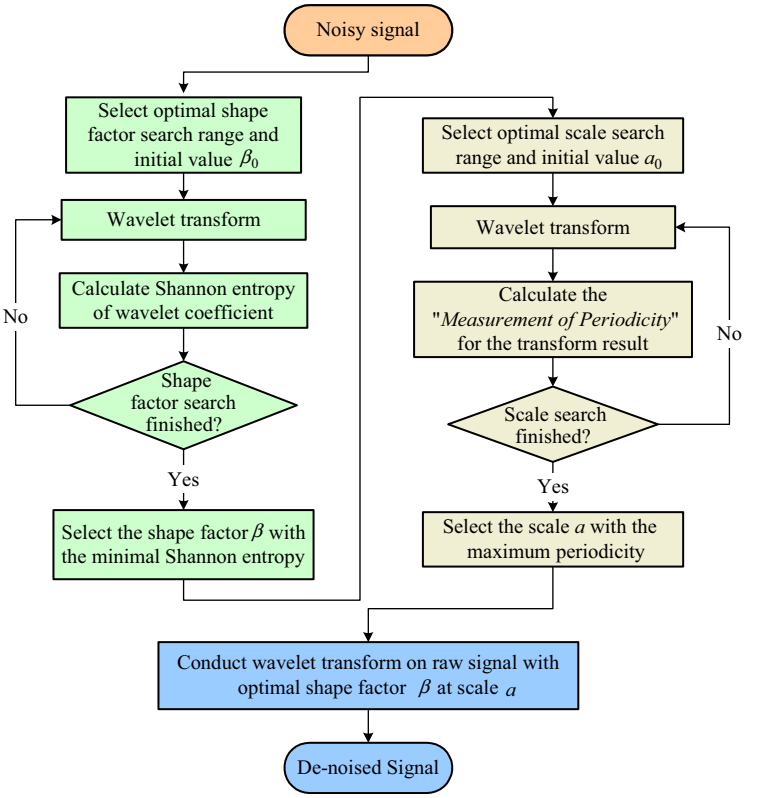

Fig. 10. Flowchart for selecting optimal shape factor and wavelet transform scale.

4.5. Comparison study on simulated signals

In order to compare the performance of Wavelet decomposition-based de-noising method and the wavelet filter, both methods are applied to the simulated data shown in Fig. 11(b). For the wavelet de-noising method, different thresholds are applied as well.

Fig. 12(a–d) show the de-noised signal by applying soft Stein's Unbiased Risk Estimation (SURE), soft heuristic SURE, soft minimaxi, and soft universal threshold, respectively [28,29]. SURE threshold is based on a quadratic loss function. An estimate of the risk is given for a particular threshold value t. Minimizing the risks in t gives a selection of the threshold value. Heuristic SURE is a heuristic variant of the SURE threshold. Minimaxi uses a fixed threshold chosen to yield minimax performance for mean square error against an ideal procedure. The minimax principle is used in statistics in order to design estimators. Since the de-noised signal can be assimilated to the estimator of the unknown regression function, the minimax estimator is the one that realizes the minimum of the maximum mean square error obtained for the worst function

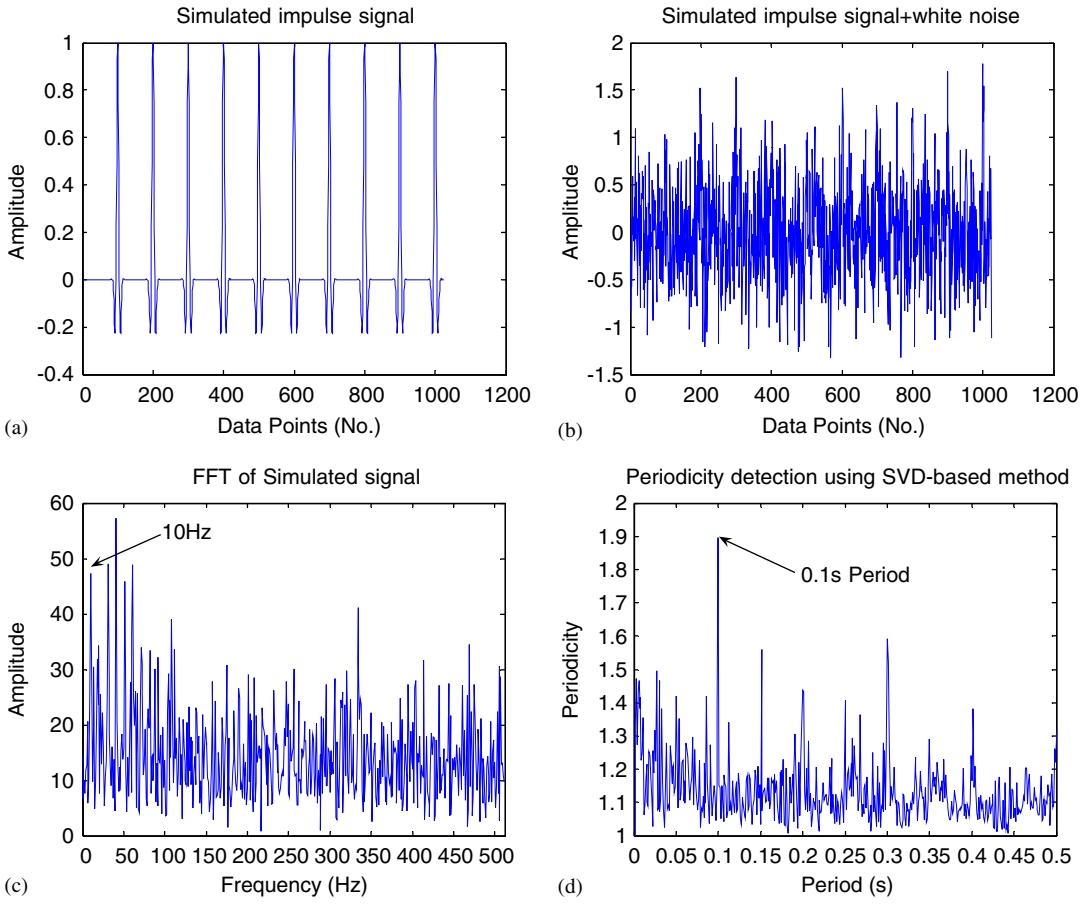

Fig. 11. (a) Simulated impulse signal, (b) simulated impulse signal with white noise, (c) fast Fourier transform spectrum of simulated signal and (d) periodicity detection using SVD-based method.

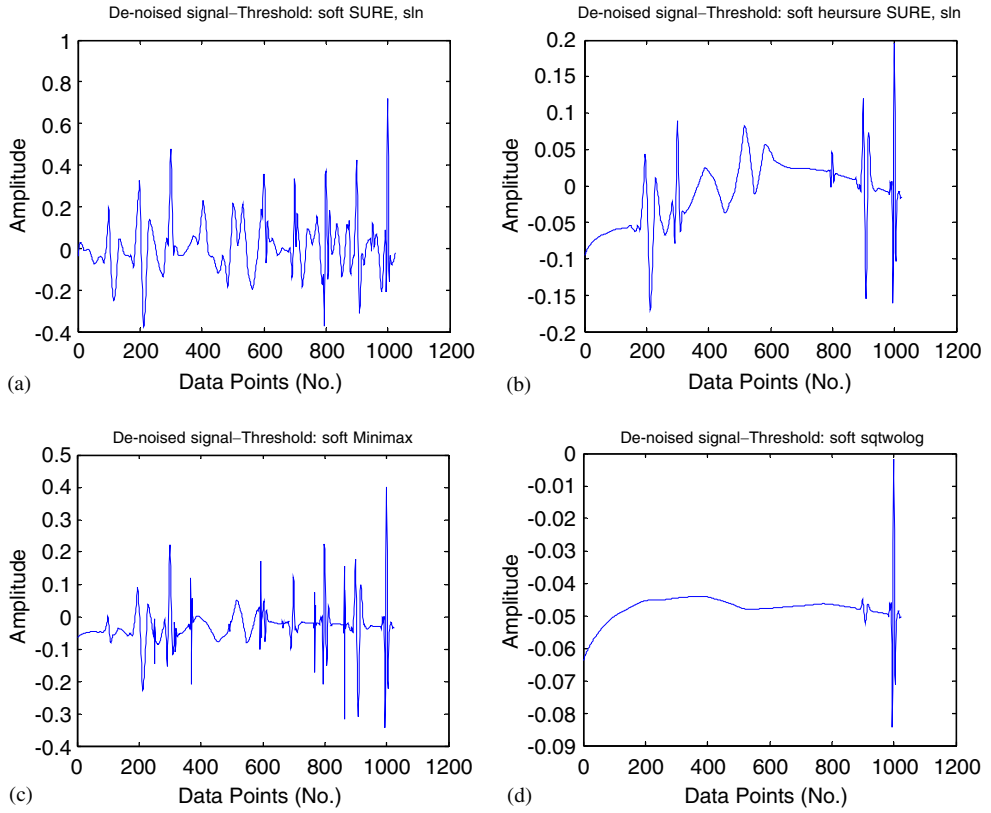

Fig. 12. De-noised signal by wavelet decomposition with different thresholds: (a) soft Stein's unbiased risk estimation (SURE), (b) soft heuristic SURE, (c) soft minimaxi and (d) soft universal threshold.

in a given set. Universal threshold is defined by the equation

$$t = \sqrt{2\log(n)},$$

where n is the signal length and s is the noise standard deviation.

Observing the de-noising results after appliying four different thresholding strategies, only the de-noised signal in Fig. 12(a) recovers the original signal partially. But the periodic feature is not as noticeable as it is in the original signal. In addition, there are other factors influencing the effectiveness of de-noising, such as the wavelet decomposition level and threshold rescaling method selection, which make the de-noising problem even more intricate. Since there are no explicit guidelines for how to tune the existing parameters, most of the time de-noising becomes a trial-and-error process.

For the comparison study, the wavelet filter method proposed in this paper is applied to the same set of simulated data. To find an optimal wavelet filter that can discover the periodic impulses from the noisy raw signal, the first step is to search for the optimal shape factor β. Increasing β from 0.1 to 20 and calculating the entropy of the corresponding coefficients, the optimal shape factor β leading to the minimal Shannon entropy relationship can be obtained.

As depicted in Fig. 13, the entropy exhibits its minimal value at $\beta = 0.5$. Therefore, $\beta = 0.54$ is selected as the optimal shape factor.

After the shape factor $\beta = 0.54$ is selected, a recursive route is carried out to find the optimal scale that can uncover the strongest periodicity from the wavelet transform results. The searching ranges for period and scale are set as [2,300] and [1,30], respectively. The measurement of periodicity δ_i is calculated and presented as a three-dimensional surface in Fig. 14. Strong

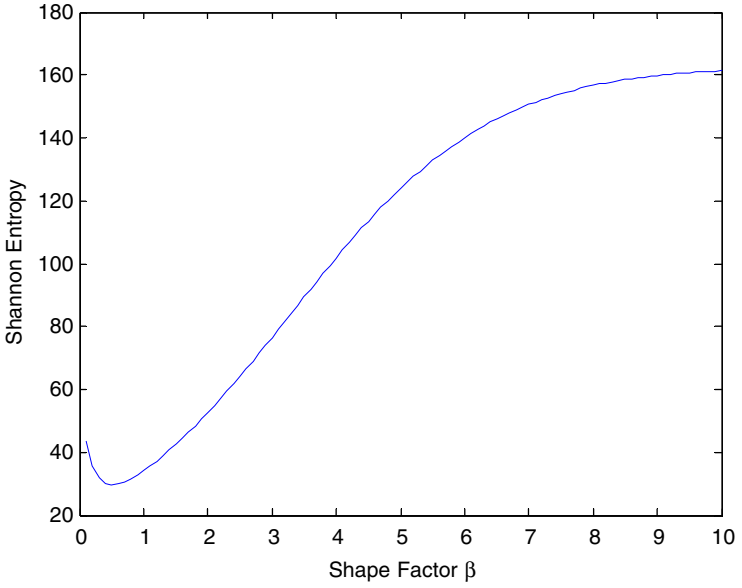

Fig. 13. The relationship between Shannon entropy of the wavelet transform coefficient and shape factor β.

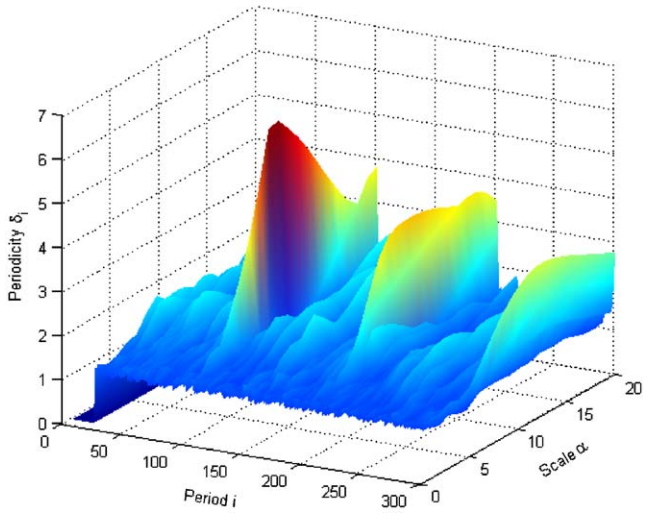

Fig. 14. Periodicity evaluation with different scale and different period.

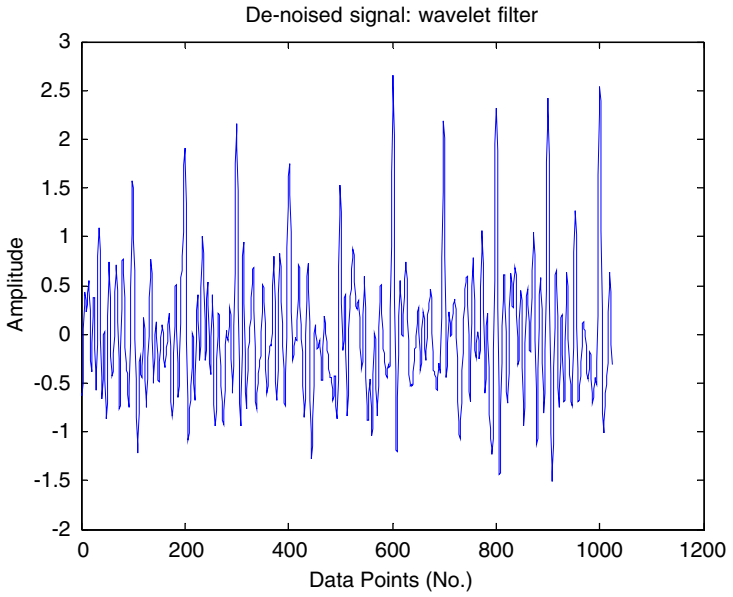

Fig. 15. De-noised signal by wavelet filter with optimal parameter $\beta = 5.4$ and $\alpha = 10$.

periodicity is discovered when the period i equals 100, 200, and 300. In addition, the periodicity reaches its maximum value when the scale $\alpha = 10$. Thus, we can conclude that conducting the wavelet transform at scale 10 can best reveal the periodicity of the signal.

Fig. 15 shows the de-noised signal by applying the Morlet wavelet filter with optimal shape parameter $\beta = 0.54$ and scale parameter $\alpha = 10$. The simulated impulses are conspicuously represented. Even though the noise level is still high, the periodic character of the simulated signal, which is the most important feature for fault diagnostics, is recovered significantly.

5. Experimental verification

5.1. Experimental setup

Most bearing diagnostics research involves studying the defective bearings recovered from the field, where the bearings exhibit mature faults, or from simulated or "seeded" damage. Simulated damages are typically induced by scratching or drilling the surface, introducing debris into the lubricant, or machining with an electrical discharge. Experiments using defective bearings have less capability to discover natural defect propagation in the early stages. In order to validate the wavelet filter methodology and truly reflect the real defect propagation processes, bearing run-to-failure tests were performed under normal load conditions on a specially designed test rig.

The bearing test rig hosts four test bearings on one shaft. The shaft is driven by an AC motor and coupled by rub belts. The rotation speed was kept constant at 2000 rpm. A radial load of 6000 lbs. is added to the shaft and bearing by a spring mechanism. All the bearings are force lubricated. An oil circulation system regulates the flow and the temperature of the lubricant.

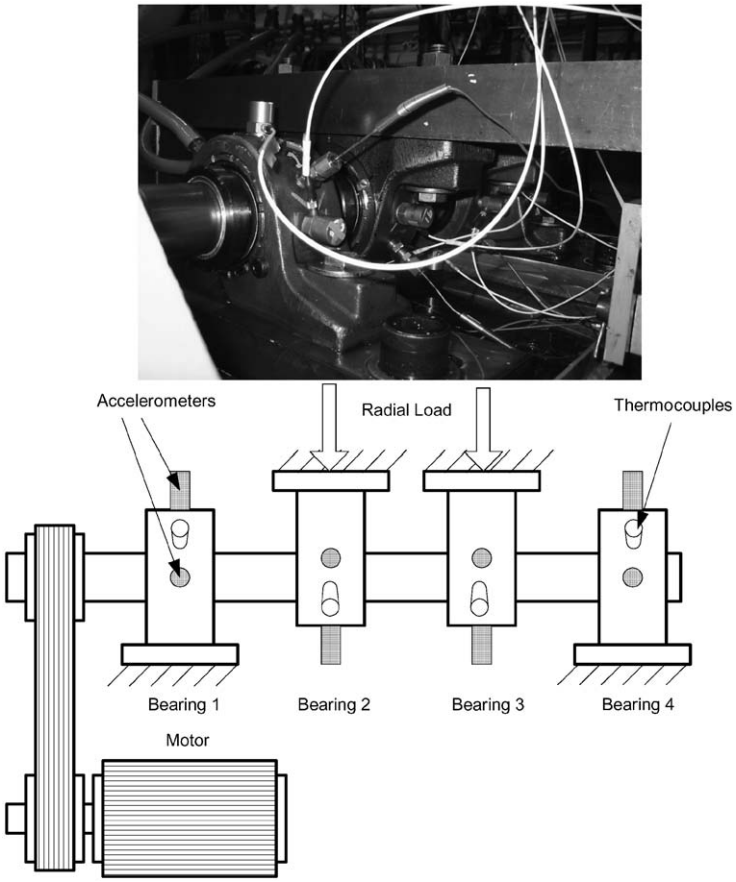

Fig. 16. Bearing test rig.

A magnetic plug installed in the oil feedback pipe collects debris from the oil as evidence of bearing degradation. The test will stop when the accumulated debris adhered to the magnetic plug exceeds a certain level and causes an electrical switch to close.

Four Rexnord ZA-2115 double row bearings were installed on one shaft as shown in Fig. 16. The bearings have 16 rollers in each row, a pitch diameter of 2.815 in., roller diameter of 0.331 in., and a tapered contact angle of 15.17°. A PCB 353B33 High Sensitivity Quartz ICP® Accelerometer was installed on each bearing housing. Four thermocouples were attached to the outer race of each bearing to record bearing temperature for monitoring the lubrication purposes. Vibration data was collected every 20 minutes by a National Instruments DAQCard-6062E data acquisition card. The data sampling rate is 20 kHz and the data length is 20480 points. Data collection is conducted by a National Instruments LabVIEW program.

5.2. Experimental results analysis

The test was carried out for 35 days until a significant amount of metal debris was found on the magnetic plug of the test bearing.

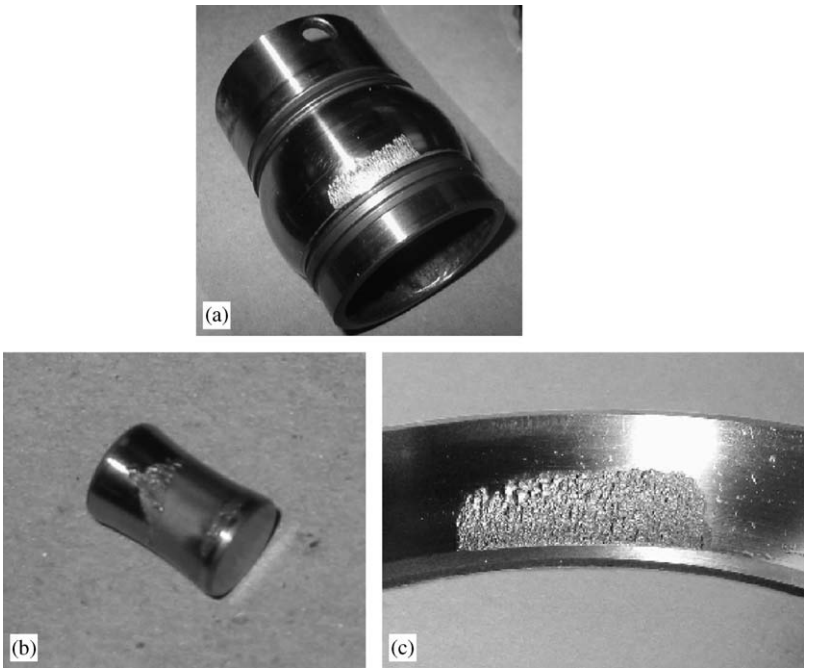

Fig. 17. Picture of bearing components after test: (a) inner race defect in bearing 3, (b) roller element defect in bearing 4 and (c) outer race defect in bearing 4.

An inner race defect was discovered in test bearing 3, and a roller element defect and outer race defect were found in test bearing 4 as shown in Fig. 17. Fig. 18 depicts the time domain features, root mean square (rms), and kurtosis for the entire life cycle for bearings 3 and 4 respectively.

Fig. 18(a) reveals that for the inner race defect, the change of rms can be divided into two stages. In the first stage, during the first 30 days of operation, no underlying trend was observed. After the test had been carried out for 30 days (approximately 86.4 million cycles), the rms started to increase and the rate of change also increased significantly.

For bearing 4, which exhibited mixed roller element and outer race defects, the trend of rms increased to a certain level, then decreased and rose again. The fluctuating trend can be explained by the nature of the propagating process of the damage. When the surface defect just initiated, small spalling or cracks were formed and were later smoothed by the continuous rolling contact. As the damage spread over a broader area, the vibration level rises again. The discovery in [30] also verified this "healing" phenomenon. The fluctuating trend of kurtosis features in both bearings 3 and 4 can also be explained by this theory.

The time domain feature also shows that most of the bearing fatigue time is consumed during the period of material accumulative damage, while the period of crack propagation and development is relatively short. This means that if the traditional threshold-based condition monitoring approach is used, the response time available for the maintenance crew to respond prior to catastrophic failure after a defect is detected in such bearings is very short. A prognostic approach that can detect the defect at the early stage is demanded so that enough buffer time is

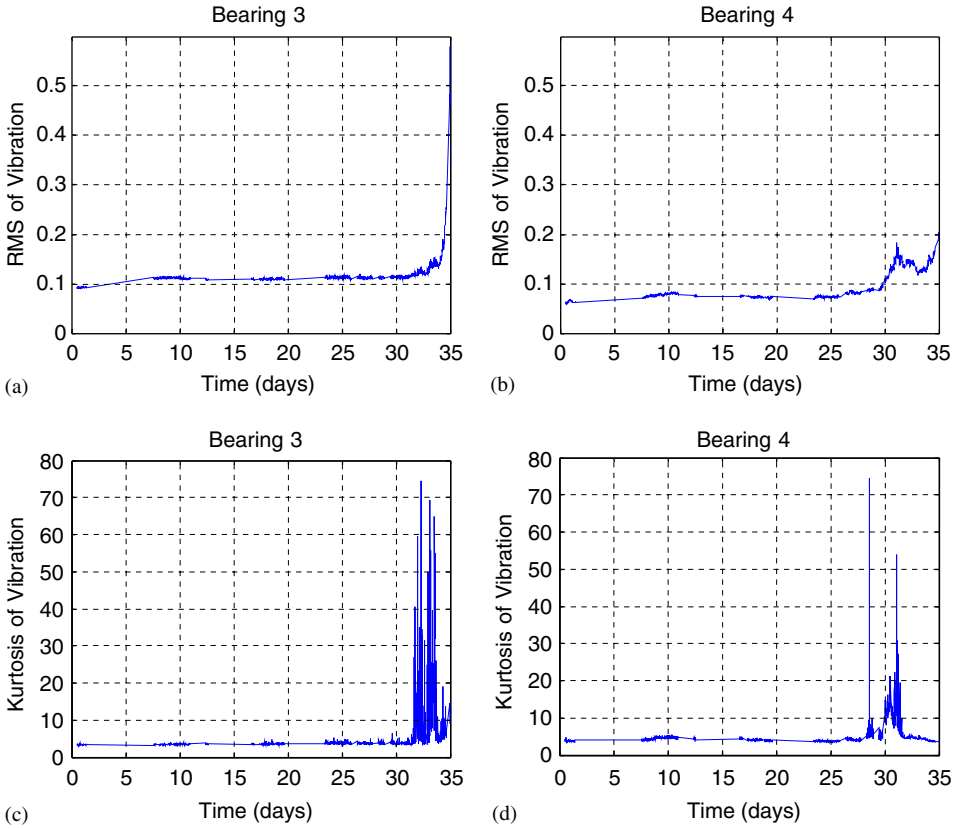

Fig. 18. Time feature (a) rms of bearing 3, (b) rms of bearing 4, (c) Kurtosis of bearing 3 and (d) Kurtosis of bearing 4 for the whole life cycle.

available for maintenance and logistical scheduling. This requirement is extremely important for some mission-critical situations, such as power plants and continuous production lines.

Fig. 19 presents the vibration waveform collected from bearing 4 at the last stage of the bearing test. The signal exhibits strong impulse periodicity because of the impacts generated by a mature outer race defect. The band pass frequency of the outer race (BPFO) is 236.4 Hz, so we expect to see a duration between the two conjoined impulses of $1/236.4 = 0.0042$ s. The vibration waveform clearly verifies the calculation.

However, when examining historical data and observing the vibration signal three days before the bearing failed, there is no sign of periodic impulse as shown in Fig. 20. The periodic impulse feature is totally masked by the noise.

The de-noising method proposed in this paper is used to enhance the signal shown above. At first, the optimal wavelet shape factor $\beta = 1.3$ is found by the minimal entropy method. Then scale a from 1 to 5 was scanned to find the optimal scale that can reveal the signal periodicity most clearly. Fig. 21 demonstrates the periodicity measurement when the scales for the wavelet filter are chosen as [1,5].

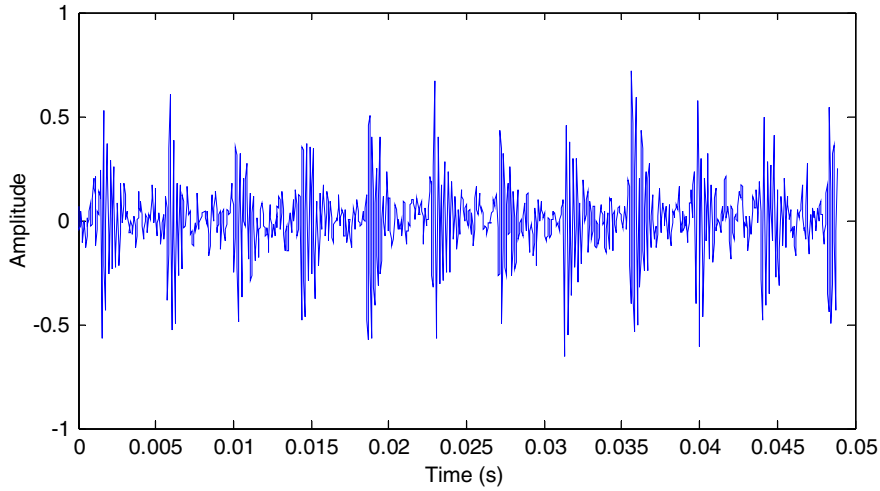

Fig. 19. The vibration signal waveform of a faulty bearing.

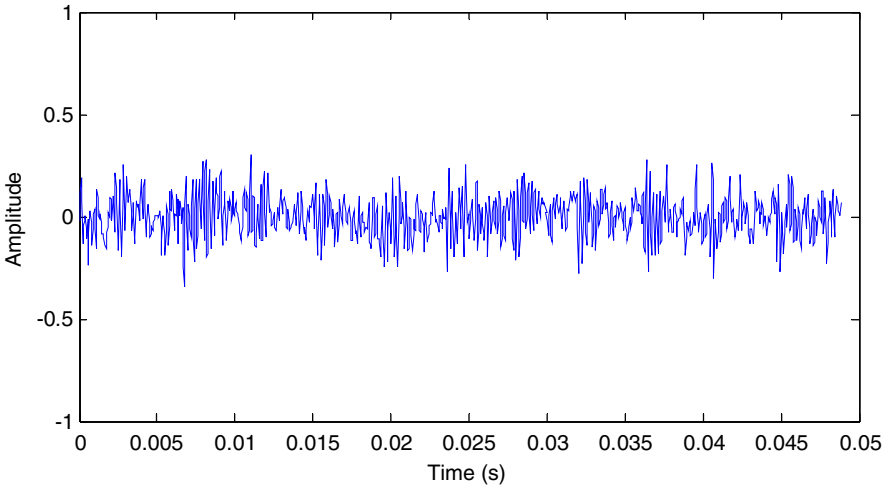

Fig. 20. The vibration waveform with early stage defect.

Notably, the selection of search range for optimal scale a is very important. It should be determined by checking the corresponding wavelet filter band pass width, otherwise a misleading result is very possible.

The periodicity of the de-noised signal reaches its maximum value using scale $a = 2.6$. Fig. 22 further illustrates that when scale $a = 2.6$ the periodicity measurement exhibits its maximum value at period $n = 89$ data points. Given the data sampling rate of 20 kHz and signal length of 20480 data points, period $n = 89$ actually means frequency $20480/89 = 230$ Hz, which is very close to the BPFO 236.4 Hz.

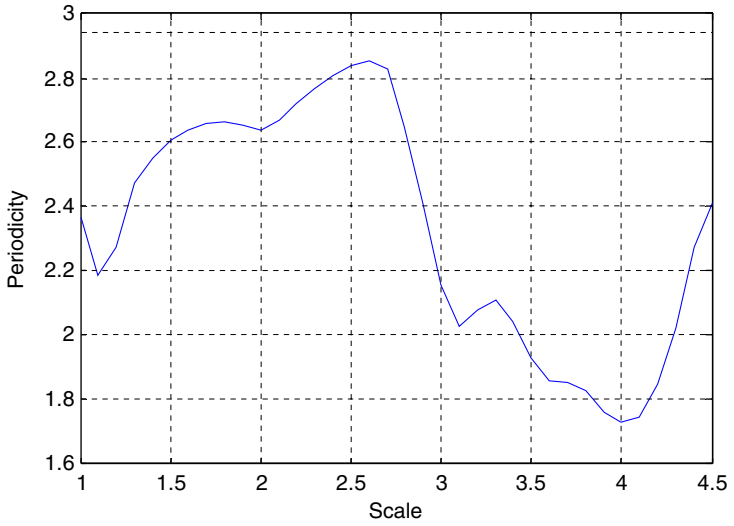

Fig. 21. The relationship of wavelet filter scale and periodicity of the de-noised signal.

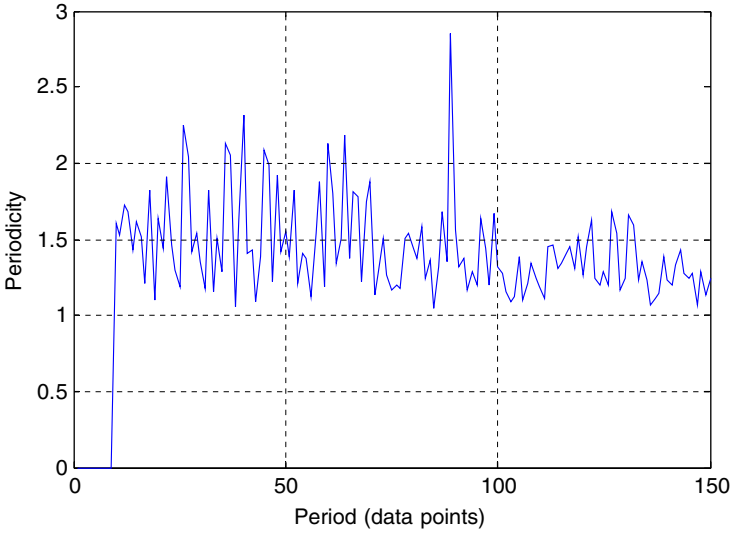

Fig. 22. Periodicity measurement when scale $\alpha = 2.6$.

Applying the wavelet filter with carefully selected shape factor $\beta = 1.3$ and scale $\alpha = 2.6$, the de-noised signal can be obtained as shown in Fig. 23. The periodic impulse feature is clearly discovered. The period of the impulse is 230 Hz, which is a strong evidence of bearing outer race degradation. Comparing Figs. 23 and 20, the wavelet filter-based de-noising method successfully enhanced the signal feature and provided potent proof for prognostic decision-making.

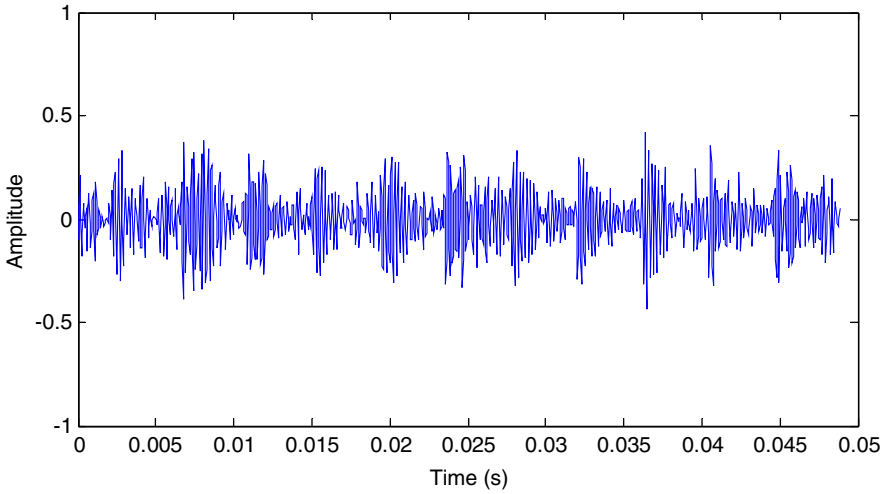

Fig. 23. The enhanced signal of early stage defect.

6. Conclusion and discussion

De-noising and extraction of the weak signature from the noisy signal are crucial to fault prognostics, in which case features are often very weak and masked by the background noise. Prognostics is achieved by detecting the defect at its initial stage and alerting the operator or maintenance personnel before the defect develops into a catastrophic failure.

The performance of traditional wavelet decomposition-based de-noising methods is greatly impacted by relative energy levels of signal coefficients and white noise coefficients. When dealing with smooth signals, satisfactory results can generally be achieved by manipulating the threshold. The underlying reason is because with smooth signals, a small number of large coefficients can characterize the original signal. However, it is much more challenging to de-noise impulse series signals where wavelet coefficients are not so concentrated.

The Morlet wavelet filter-based de-noising method is based on the idea of detecting the "similar" impulse components from the noisy signal by designing a daughter Morlet wavelet with specific shape factor β at certain scale a. This method is well suited for detecting the weak signature from a defective bearing signal where defect features are impulse-like. By applying the minimal Shannon entropy criterion, an optimal wavelet shape factor β with optimal time-frequency resolution capability can be obtained. The optimal scale a can be determined by Singular Value Decomposition (SVD)-based periodicity evaluation of wavelet transform results based on the assumption that the undetected signature is periodic.

The experimental results verify the effectiveness of the proposed method. The weak periodic impulse signature is successfully revealed and enhanced. Detection of the degradation signature at its early stage gives more time for maintenance reaction and business decision-making and also provides proof for prognostics. However, prognostics can only be accomplished by combining the knowledge from the physical degradation model and tendency forecasting analysis, which continues to make prognostics an attractive research topic.

Acknowledgement

This research was funded by the National Science Foundation Industry/University Cooperative Research Center (NSF I/UCRC) on Intelligent Maintenance Systems (IMS) under Grant No. 0117518. Rexnord Technical Services, an IMS sponsor, provided the testing facility and run-to-failure expertise for this research.

Reference

[1] P.D. McFadden, J.D. Smith, Vibration monitoring of rolling element bearings by high frequency resonance technique—a review, *Tribology International* 17 (1984) 3–10.
[2] O.G. Gustaffson, T. Tallian, Detection of damage of assembled rolling bearings, *ASME Transaction* 5 (1962) 197–209.
[3] P.D. McFadden, J.D. Smith, Model for the vibration produced by a signal point defect in a rolling element bearing, *Journal of Sound and Vibration* 96 (1) (1984) 69–82.
[4] D. Ho, R.B. Randall, Optimization of bearing diagnostic techniques using simulated and actual bearing fault signals, *Mechanical Systems and Signal Processing* 14 (5) (2000) 763–788.
[5] C.T. Yiakopoulos, I.A. Antoniadis, Wavelet based demodulation of vibration signals generated by defects in rolling element bearings, *Shock and Vibration* 9 (2002) 293–306.
[6] N.G. Nikolaou, I.A. Antoniadis, Rolling element bearing fault diagnosis using wavelet packets, *NDT&E International* 35 (2002) 197–205.
[7] R.L. Eshleman, Comments on rolling element bearing analysis, *Vibrations* 13 (2) (1997) 11–17.
[8] R.B. Randall, *Frequency analysis*, 3rd ed, Bruel&Kjaer, 1987.
[9] Yuh-Tay Sheen, Chun-Kai Hung, Construction a wavelet-based envelop function for vibration signal analysis, *Mechanical Systems and Signal Processing* 18 (2004) 119–126.
[10] J. Altmann, J. Mathew, Multiple band-pass autoregressive demodulation for rolling element bearing fault diagnosis, *Mechanical Systems and Signal Processing* 15 (2001) 963–977.
[11] S. Sophocles J. Orfanidis, *Introduction to Signal Processing*, Prentice Hall, 1996.
[12] J. Shiroishi, Y. Li, et al., Bearing condition diagnostics via vibration and acoustic emission measurement, *Mechanical Systems and signal processing* 15 (1997) 693–705.
[13] D.L. Donoho, I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, *Journal of American Statistical Assoiciation* 90 (1995) 1200–1224.
[14] A. Grossmann, Wavelet transform and edge detection, in: S. Albeverio, et al. (Eds.), *Stochastic Processes in Physics and Engineering*, D. Reidel Publishing Company, 1988, pp. 149–157.
[15] Jin Lin, Liangsheng Qu, Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis, *Journal of Sound and Vibration* 234 (1) (2000) 135–148.
[16] David L. Donoho, De-Noising by Soft-Thresholding, *IEEE Transactions on Information Theory* 41 (3) (1995) 613–627.
[17] S.G. Chang, Yu. Bin, M. Vetterli, Adaptive wavelet thresholding for image denoising and compression, *IEEE Transactions on Image Processing* 9 (9) (2000) 1532–1546.
[18] M. Hansen, Yu. Bin, Wavelet thresholding via MDL for natural images, *IEEE Transactions on Information Theory* 46 (5) (2000) 1778–1788.
[19] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, *Biometrika* 81 (1994) 425–455.
[20] P. Goupillaud, A. Grossmann, J. Morlet, Cycle-octave and related transforms in seismic signal analysis, *Geoexploration* 23 (1984–1985) 85–102.
[21] Jin Lin, Gearbox fault diagnosis using adaptive wavelet filter, *Mechanical Systems and Signal Processing* 17 (6) (2003) 1259–1269.
[22] P.P. Kanjilal, S. Palit, On multiple pattern extraction using singular value decomposition, *IEEE Transactions on Signal Processing* 43 (6) (1995) 1536–1540.

[23] G.H. Golub, C.F. Van Load, *Matrix Computations*, The John Hopkins Univ. Press, 1989.
[24] J. Juha Karvanen, A. Adrzej Cichocki, Measuring Sparseness of Noisy Signals, Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), Nara, Japan, April 2003.
[25] C.E. Shannon, W. Weaver, *Mathematical Theory of Communication*, Univ. of Illinois Press, Urbana, 1949.
[26] J.N. Kapur, H.K. Kesavan, *Entropy optimization principles with applications*, Academic Press, San Diego, 1992.
[27] P.P. Kanjilal, J. Bhattacharya, G. Saha, Robust method for periodicity detection and characterization of irregular cyclical series in term of embedded periodic components, *Physical Review E* 59 (4) (1999) 4013–4025.
[28] D.L. Donoho, De-noising by soft-thresholding, *IEEE Transactions on Information Theory* 41 (3) (1995) 613–627.
[29] Michel Misiti, Yves Misiti, Georges Oppenheim, Jean-Michel Poggi, *Wavlet Toolbox User's Guide for Use with Matlab*, The Mathworks, Inc., 2004.
[30] T. Williams, X. Ribadeneira, Rolling element bearing diagnostics in run-to-failure lifetime testing, *Mechanical Systems and Signal Processing* 15 (5) (2001) 979–993.