

Application of improved wavelet total variation denoising for rolling bearing incipient fault diagnosis

W Zhang¹ and M P Jia^{1,2}

¹School of Mechanical Engineering, Southeast University, Nanjing 211189, China

E-mail: mpjia@seu.edu.cn

Abstract. When incipient fault appear in the rolling bearing, the fault feature is too small and easily submerged in the strong background noise. In this paper, wavelet total variation denoising based on kurtosis (Kurt-WATV) is studied, which can extract the incipient fault feature of the rolling bearing more effectively. The proposed algorithm contains main steps: a) establish a sparse diagnosis model, b) represent periodic impulses based on the redundant wavelet dictionary, c) solve the joint optimization problem by alternating direction method of multipliers (ADMM), d) obtain the reconstructed signal using kurtosis value as criterion and then select optimal wavelet subbands. This paper uses overcomplete rational-dilation wavelet transform (ORDWT) as a dictionary, and adjusts the control parameters to achieve the concentration in the time-frequency plane. Incipient fault of rolling bearing is used as an example, and the result shows that the effectiveness and superiority of the proposed Kurt-WATV bearing fault diagnosis algorithm.

1. Introduction

Rolling bearing is one of the most significant part of the mechanical transmission system. Incipient fault diagnosis is conducive to the formulation of reasonable maintenance plan, which can realize the maximization of production efficiency and change the disadvantages of the traditional Run-to-Breakdown maintenance way. When rolling bearing is in the incipient stage of failure, the fault feature is relatively weak. Hence, it is difficult to extract fault information due to the influence of vibration and noise of signal transmission path and other components.

In recent years, domestic and foreign scholars have done a great deal of researches on the incipient fault extraction in rolling bearings. Although the signal processing methods [1-5] show outstanding performance in extracting incipient fault feature, the methods themselves have some shortcomings. Wavelet transform [1], empirical mode decomposition [2,3] and some other methods [4,5] are all based on the construction of the best band-pass filter to obtain fault information, but in-band noise can not be removed, which affects the extraction of faults related information from the strong background noise.

Sparse representation has become a hot topic in the field of signal processing in recent years. The purpose of signal sparse representation is to express a signal with as few atoms as possible in a given overcomplete dictionary, to obtain a more concise representation way [6]. Liu *et al* [7] proposed a time-frequency atomic matching pursuit algorithm to analyze vibration signals and extract rolling bearing fault characteristics. Shi *et al* [8] applied an adaptive time-frequency decomposition method based on matching pursuit and Gaussian line-frequency dictionaries to analyze transient vibration signals. Yang *et al* [9] combined the basic pursuit and wavelet packet dictionary to extract bearing



fault features. They point out that basic pursuit can sparsely represent the rolling bearing fault features at fine resolution and enhance signal-to-noise ratio. Feng and Chu [10] applied several typical methods of sparse signal decomposition, including frame methods, best orthogonal basis, matching pursuit and basis pursuit, to extract the embedded meaningful information. Qin *et al* [11] combined basis pursuit with various transformation bases to decompose different components of mixture signals, such as impacts, harmonics and modulation components, which can effectively diagnose rotor and bearing faults. Feng and Liang [12] used iterative threshold method to suppress background noise and extracted the fault feature of planetary gearbox.

Wavelet thresholding often introduces artifacts such as spurious noise spikes and pseudo-Gibbs oscillations [13]. Total variation (TV) term is added into the optimization objective function to alleviate the spurious noise spikes and pseudo-Gibbs artifacts [14]. The existing sparse representation method based on wavelet dictionary can not adjust the time-frequency resolution. Therefore, the frequency resolution will decrease as the frequency increases, and the appropriate time-frequency division can not be selected according to the signal characteristics. The proposed algorithm utilizes ORDWT dictionary to achieve a better time-frequency distributions. In addition, the kurtosis is introduced to determine the optimal wavelet subbands, and the denoised results show that the impact features are more obvious. Lastly, experimental results indicate that the proposed algorithm can be effectively used to incipient rolling bearings fault diagnosis.

2. Kurtosis-base sparse model and algorithmic framework

2.1. Sparse fault diagnosis model

When the rolling bearings appear local damage, the relative motion between components produce periodic impact, so the fault signal model of rolling bearing contains periodic impact feature. In the incipient stage of the rolling bearings failure, the amplitude of impact component is not only small, but also the impact features are easily drowned in the strong noise environment. Usually collected vibration signal contains periodic impact components, mechanical frequency components and noise components. Rolling bearing fault signal model can be written as follows:

$$y = x + f + v \quad (1)$$

where x is periodic impacts, f is mechanical frequency, v is noise. The total variation of signal x is defined as $TV(x) = \|Dx\|_1$, where D is first-order difference matrix,

$$D = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \dots & \dots & \\ & & & -1 & 1 \end{bmatrix} \quad (2)$$

In order to extract x from the observed signal y , we construct the following optimization problem by using the sparse priori knowledge of the periodic impact component under the wavelet dictionary representation:

$$\hat{w} = \arg \min_w \left\{ F(w) = \frac{1}{2} \|Wy - w\|_2^2 + \sum_j \lambda_j \varphi(w_j) + \beta \|DW^T w\|_1 \right\} \quad (3)$$

The first term in equation (3) is the fidelity term, which means signal reconstruction error. The last two terms are penalty function. The function $\varphi(w_j)$ can be expressed in the form of l_1 norm. The estimate of the signal x is then given by the inverse wavelet transform of \hat{w} : $\hat{x} = W^T \hat{w}$. The regularization parameters are $\lambda_j > 0$ and $\beta > 0$.

2.2. Overcomplete rational dilation wavelet dictionary

The selection and construction of the basis functions in redundant dictionaries is one of the important factors that determine the effectiveness of signal sparse representation. When the basis function is as close as possible to the internal structure of the signal, it is possible to obtain the sparse representation of the signal through a sparse solution algorithm, so as to reveal the effective components in the signal. For a simple signal, pre-constructed dictionaries can describe their sparsity more effectively. Continuous wavelet, discrete wavelet, complex wavelet, lifting wavelet, wavelet packet and multiwavelet transform are widely used in fault diagnosis [15]. As a new branch of wavelet series, ORDWT have attracted more attention recently in the field of fault diagnosis [16]. Compared with the traditional binary wavelet, ORDWT is a new type of wavelet tight frame, which has better translational invariance, adjustable time-frequency distribution and flexible oscillations of flexible wavelet atoms. ORDWT is composed of an iterative two-channel filter, as shown in figure 1. It can be seen that the upsampling is done only in the low-pass path. The dilation factor of ORDWT is q/p , where p and q are satisfied $1 < q/p < 2$, and p and q are coprime.

The frequency response of filters $H(w)$ and $G(w)$ is:

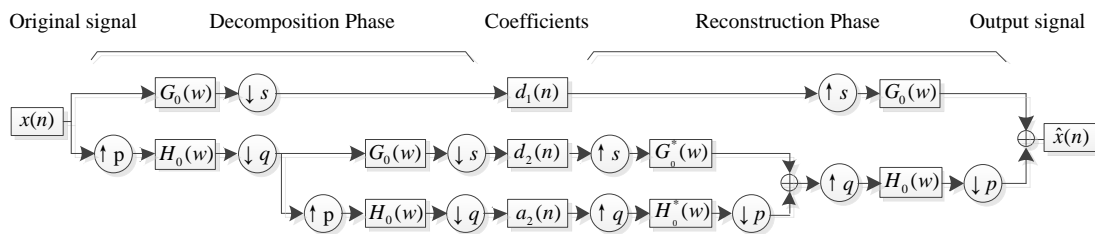


Figure 1. Filter-banks for ORDWT.

$$H(w) = \begin{cases} \sqrt{pq}, & w \in \left[0, \left(1 - \frac{1}{s}\right)\frac{\pi}{p}\right) \\ \sqrt{pq}\theta_H\left(\frac{w-a}{b}\right), & w \in \left[\left(1 - \frac{1}{s}\right)\frac{\pi}{p}, \frac{\pi}{q}\right) \\ 0, & w \in \left[\frac{\pi}{q}, \pi\right] \end{cases} \quad (4)$$

$$G(w) = \begin{cases} 0, & w \in \left[0, \frac{s-1}{s}\pi\right) \\ \sqrt{s}\theta_G\left(\frac{w-pa}{pb}\right), & w \in \left[\frac{s-1}{s}\pi, \frac{p}{q}\pi\right) \\ \sqrt{s}, & w \in \left[\frac{p}{q}\pi, \pi\right] \end{cases} \quad (5)$$

where $a = \left(1 - \frac{1}{s}\right)\frac{\pi}{p}$, $b = \frac{1}{q} - \left(1 - \frac{1}{s}\right)\frac{1}{p}$, $\theta_H = 1/2(1 + \cos w)\sqrt{2 - \cos w}$, $w \in [0, \pi]$; $\theta_G = 1/2(1 - \cos w)\sqrt{2 + \cos w}$, $w \in [0, \pi]$.

The variables p , q , s are the controlling parameters of ORDWT bases. The function $\theta_H(w)$ and $\theta_G(w)$, originating from Daubechies' orthonormal wavelet with 2 vanishing moments, are used to construct the transition bands of $H(w)$ and $G(w)$. According to equations(4) and (5), the frequency response curve of the filter can be obtained, as shown in figure 2. A complete description of ORDWT can be found in [17].

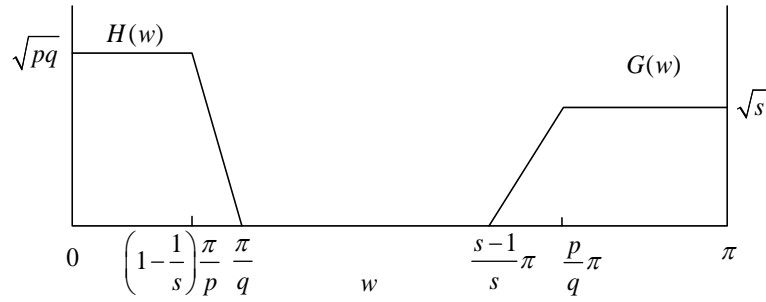


Figure 2. Frequency responses of the filter.

2.3. Optimization algorithm

ADMM is suitable for solving distributed convex optimization problems [18]. In order to solve the WATV denoising problem in equation (3), ADMM is used to solve equation (3):

$$\arg \min_{u,w} g_1(w) + g_2(u) \quad \text{subject to } u = w \quad (6)$$

where $g_1(w) = \frac{1}{2} \|Wy - w\|_2^2 + \sum_j \lambda \varphi(w_j)$ and $g_2(u) = \beta \|DW^T w\|_1$ are strictly convex. The augmented Lagrangian is given by

$$L(w, u, \mu) = g_1(w) + g_2(u) + \frac{\mu}{2} \|u - w - d\|_2^2 \quad (7)$$

ADMM transforms the joint optimization problem (7) of w and u into two simple sub-problems to solve. First, w and u are alternately minimized, and thus have the following iterative scheme:

$$w = \arg \min_w \{ g_1(w) + \frac{\mu}{2} \|u - d - w\|_2^2 \} \quad (8a)$$

$$u = \arg \min_u \{ g_2(u) + \frac{\mu}{2} \|u - d - w\|_2^2 \} \quad (8b)$$

$$d = d - (u - w) \quad (8c)$$

Combining the quadratic terms, equation (8a) may be written as follows:

$$w = \arg \min_w \sum_j \left\{ \frac{1}{2} (p_j - w_j)^2 + \lambda_j \varphi(w_j) \right\} \quad (9)$$

where $p = (Wy + \mu(u - d)) / (\mu + 1)$. It can be further expressed as:

$$w_j = \theta(p_j; \lambda_j) \quad (10)$$

where θ is the threshold function, which can be defined as $\theta(y; \lambda) = \arg \min_{x \in \mathbb{R}} \left\{ \frac{1}{2} (y - x)^2 + \lambda \varphi(x) \right\}$.

That is, $\theta(y; \lambda, a) = 0$ for all $|y| < \lambda$.

According to the proximity operators, equation (8b) can be transformed into the problem of TV denoising [19]. Hence, (8b) can be implemented as

$$u = v + W(\text{tvd}(W^T v, \beta/\mu) - W^T v) \quad (11)$$

where $tvd(W^T v, \beta/\mu) = \arg \min_x \left\{ \frac{\beta}{\mu} \|Dx\|_1 + \frac{1}{2} \|x - W^T v\|_2^2 \right\}$.

Therefore, solving problem equation (3) by equations (8a)-(8c), consists of wavelet threshold denoising equation (9) and the TV denoising equation (11), and see [20] for detail.

2.4. Kurt-WATV for incipient fault feature extraction

Kurtosis is a time statistics that reflects the distribution characteristics of vibration signals. It is very sensitive to impact feature, especially for incipient fault diagnosis. Kurtosis is defined as:

$$Kurt = \frac{E\{(x - \bar{x})^4\}}{\sigma^4} \quad (12)$$

where \bar{x} and σ are the mean and standard variance of signal x , respectively. $E\{\cdot\}$ is mathematical expectation. Therefore, the kurtosis is used to select the optimal wavelet subbands. Table 1 shows the proposed Kurt-WATV algorithm framework.

Table 1. Kurt-WATV algorithm framework.

1. Input:	$y, \lambda_j, \beta, \mu, p, q, s$
2. Construct a dictionary	Based on given p, q, s constructed wavelet dictionary W
3. Initialization	$u = Wy, d = 0$
4. Repeat	$p = (Wy + \mu(u - d))/(\mu + 1)$ $w_j = \theta(p_j; \lambda_j)$ $v = d + w$ $u = v + W(tvd(W^T v, \beta/\mu) - W^T v)$ $d = d - (u - w)$
Until convergence	
5. Select the optimal wavelet subbands	$w_{opt} = \max\{kurt[w_1(n)], kurt[w_2(n)], \dots, kurt[w_{J+1}(n)]\}$
6. Reconstruct signal	$x = W^T w_{opt}$

3. Experiment

In order to further demonstrate the performance of proposed Kurt-WATV algorithm, rolling bearing with incipient fault are used to diagnosis. The collected signal is provided by the Center of Intelligent Maintenance Systems. Figure 3 shows an experimental rig that comprises a motor and a shaft with four Rexnord ZA-2115 double-row bearings. The bearing parameters are shown in table 2. The motor speed is 2,000 r/min, and the sampling rate is 20 kHz. More information can be seen in [21] According to the structural parameters of rolling bearing (table 2), the fault feature frequency of outer race is equal to 236.4 Hz. The collected vibration signal from bearing 1 with outer fault is used to analyze.

Table 2. Parameters of the rolling bearing.

Pitch diameter d/mm	Ball diameter e/mm	Ball number Z	Contact angle $\alpha / ^\circ$
71.5	8.4	16	15.17

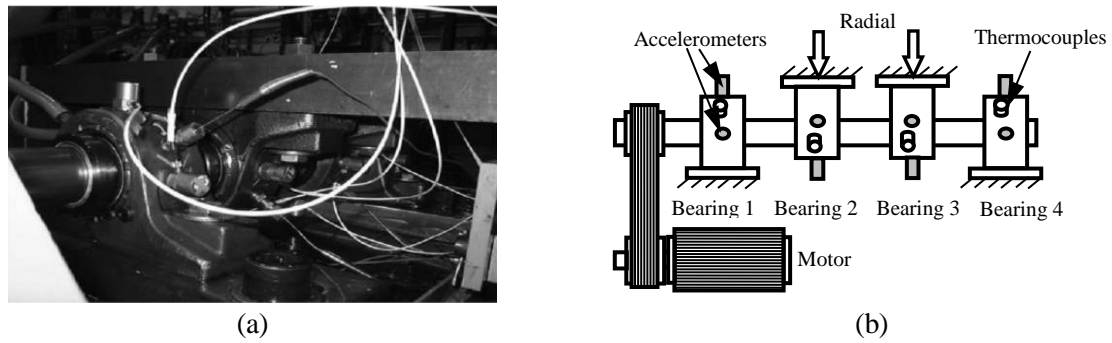


Figure 3. Experimental rig: (a) Bearing test rig; (b) Schematic presentation of sensor placement.

Figure 4 shows the raw vibration signal at the incipient stage of failure, figure 4(a) shows the time-domain waveform, in which the impulses are submerged in strong background noise. Figure 4(b) shows the envelop spectrum, in which the fault feature frequency is not clear. So we need to further processing of the raw vibration signal.

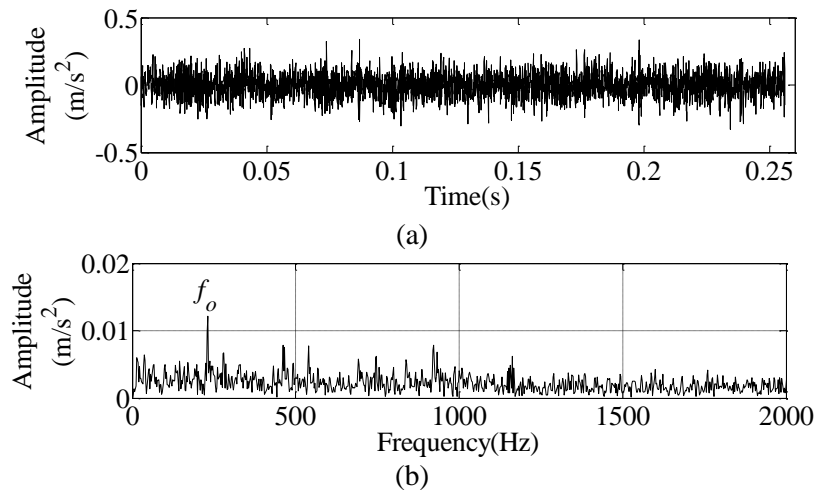


Figure 4. Experimental signal: (a) time-domain waveform; (b) envelop spectrum.

The proposed Kurt-WATV algorithm is employed enhance fault feature and weaken heavy noise. The range and constraint condition of the parameter space are set as follows:

$$T_1 = \{(p, q, s, J) \mid p, q, s, J \in \mathbb{Z}^+, 2 \leq p \leq 9, q = p + 1, 1 \leq s \leq 3, J = 20\} \quad (13)$$

The kurtosis of each wavelet subband is calculated and the result is shown in figure 5. The controlling parameters and wavelet subbands are chosen according to the kurtosis maximization principle. The optimal parameters are $p=6$, $q=7$, $s=2$, and the w_{16} wavelet subband is chosen to reconstructed signal, which reveal the hidden transient fault impulses. Figures 6(a) and 6(b) display the time-domain waveform and envelop spectrum using the proposed algorithm respectively. From figure 6(a), we can see the obvious periodic impacts. In the envelope spectrum, it can be seen that the fault feature frequency is attenuated while the other non-fault feature frequencies are weakened. Therefore, the proposed algorithm can successfully extract weak periodic impulses from strong background noise.

In order to further verify the superiority of the proposed algorithm, wavelet thresholding methods are used to compare with the proposed algorithm in experimental case. Figure 7(a) displays the time-domain waveform of hard-thresholding denoising, which indicates that the periodic impacts are not obvious. Figure 7(b) shows the envelop spectrum of denoised signal. From the results, we can see the

amplitude of the fault feature frequency is not clear. There are some noise spikes in hard-thresholding denoising because the wavelet coefficients exceed the threshold. Figure 8 shows the time-domain waveform and envelop spectrum obtained after soft-thresholding denoising. Although the prominent peaks in figure 8(b) correspond to the feature frequency, the spectrum is again clouded with many unwanted small peaks due to background noise. The proposed algorithm in this paper is improved base on the method mentioned in [14], the denoised result obtained via the method in [14] is presented in figure 9. In the envelope spectrum, figure 9(b), the fault feature frequency of the periodic impact can not be obtained. We compare the result of the proposed algorithm with the result obtained using the method in [14]. The results demonstrated that Kurt-WATV is superior to the method proposed in [14]. It is mainly because the wavelet dictionary in [14] is a undecimated wavelet, and the time-frequency resolution is not adjustable. In this paper, time-frequency distribution characteristics can be varied by adjusting a set of controlling parameters. In addition, the optimal wavelet subband is selected using kurtosis index, which is equivalent to further performing band-pass filtering.

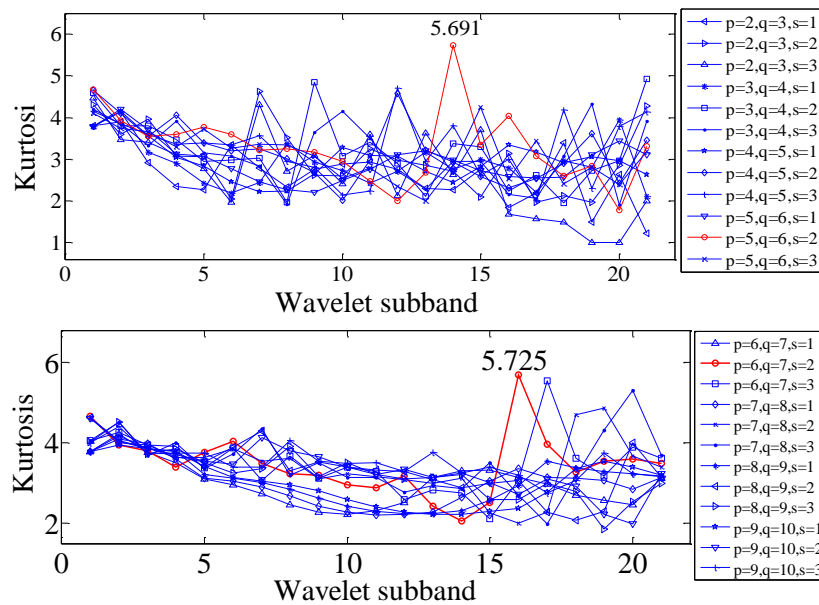


Figure 5. Kurtosis of the wavelet subbands by using parameter space T_1 .

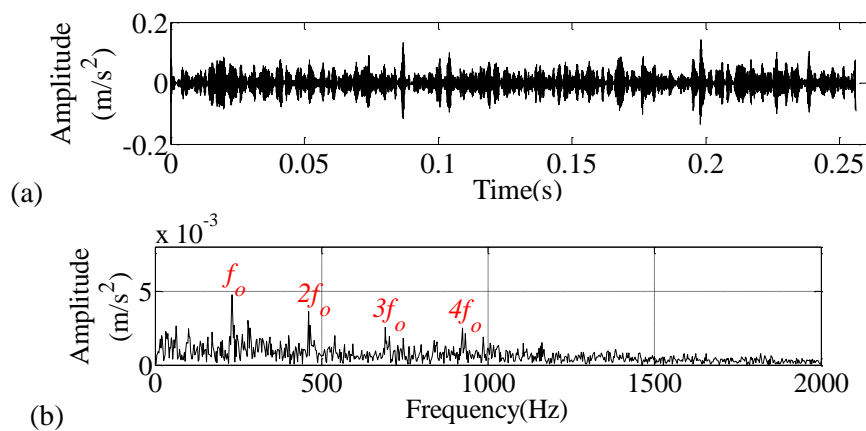


Figure 6. The denoised results using the proposed algorithm: (a) time-domain waveform and (b) envelope spectrum.

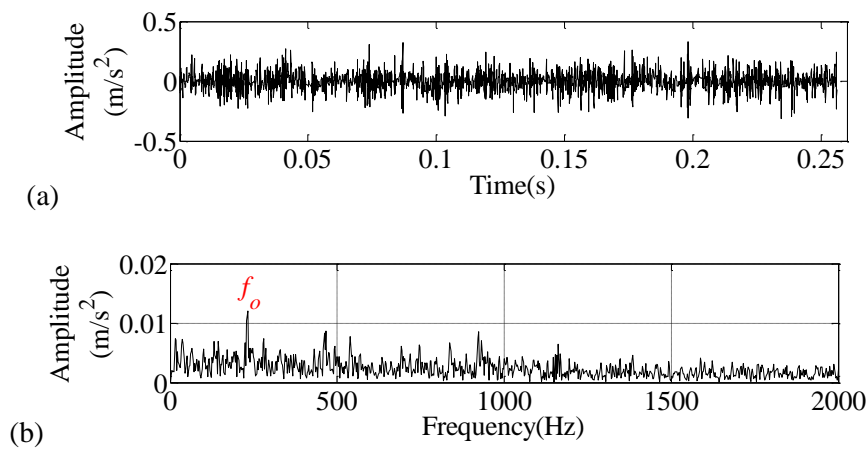


Figure 7. The denoised results obtained by hard-thresholding denoising: (a) time-domain waveform; (b) envelope spectrum

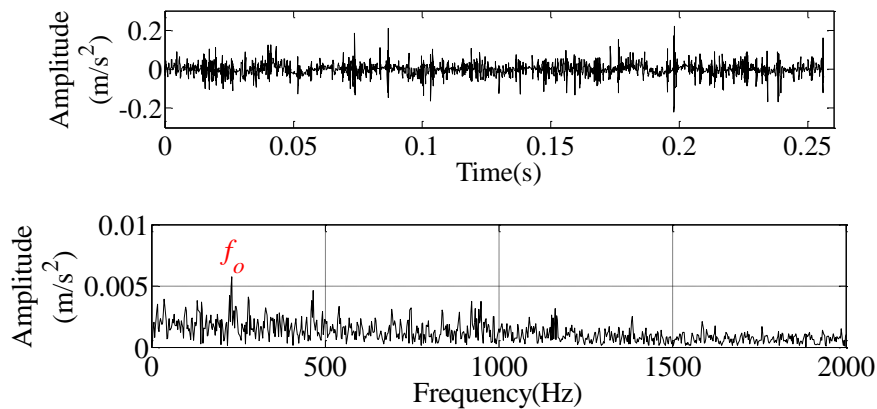


Figure 8. The denoised results obtained by soft-thresholding denoising: (a) time-domain waveform; (b) envelope spectrum

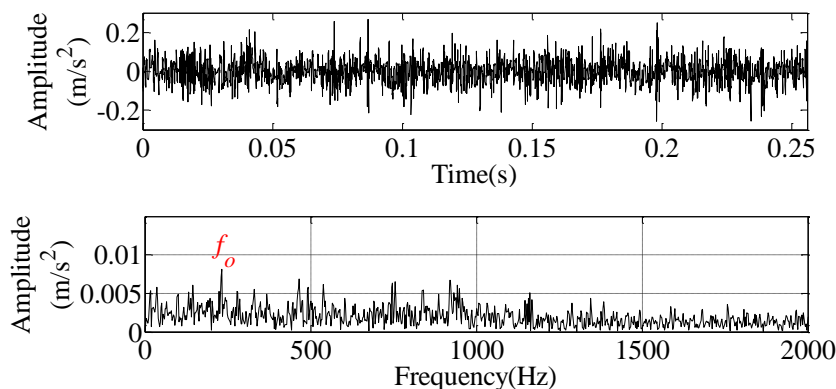


Figure 9. The denoised result obtained via the method in [14]: (a) time-domain waveform; (b) envelope spectrum

4. Conclusion

Aiming at the difficulty of extracting the incipient fault features of rolling bearing, a sparse optimization objective function is constructed. The regularization term of the objective function

contains a TV term, which can effectively avoid noise spikes and pseudo-Gibbs phenomenon. ADMM is used to solve the constructed optimization objective function, and kurtosis is used as the choice of the optimal wavelet subbands, which is equivalent to a band-pass filter to extract the resonance frequency band of related fault features. The experimental signal is used to verify the the proposed algorithm. The denoised results indicate that the proposed Kurt-WATV can effectively detect fault feature frequency of rolling bearing at the incipient fault stage. The proposed Kurt-WATV, however, is limited by the lack of theoretical support for determining the parameters λ and β . Hence, future study in this area is parameter optimization.

Acknowledgments

The authors very appreciate the support from the National Natural Science Foundation of China (Nos. 51675098).

References

- [1] Wang D, Guo W and Wang X J 2013 *Appl Soft Comput* **13** 4097-104
- [2] Wang H C, Chen J and Dong G M 2014 *Mech Syst Signal PR* **48** 103-19
- [3] Lv Y, Yuan R and Song G B 2016 *Mech Syst Signal PR* **81** 219-34
- [4] Park C S, Choi Y C and Kim Y H 2013 *Mech Syst Signal PR* **38** 534-48
- [5] Antoni J 2007 *Mech Syst Signal PR* **21** 108-24
- [6] Stéphane M 2009 *A Wavelet Tour of Signal Processing: The Sparse Way* 3rd Ed (Burlington, MA: Academic Press)
- [7] Liu B, Ling S F and Gribonval R 2002 *NDT&E INT* **35** 255-62
- [8] Shi D F, Tsung F and Unsworth P J 2004 *Mech Syst Signal PR* **18** 127-41
- [9] Yang H Y, Mathew J and Ma L 2005 *Mech Syst Signal PR* **19** 341-56
- [10] Feng Z P and Chu F L 2007 *J Sound Vib* **302** 138-51
- [11] Qin Y, Mao Y F and Tang B P 2013 *J Sound Vib* **332** 5217-35
- [12] Feng Z P and Liang M 2014 *J Sound Vib* **333** 5196-211
- [13] Coifman R R and Donoho D L 1995 *Wavelets & Statistics* **2** 125-50
- [14] Ding Y and Selesnick I W 2015 *IEEE Signal Proc Lett* **22** 1364-8
- [15] Yan R Q, Gao R X and Chen X F 2014 *Signal Process* **5** 1-15
- [16] Chen B Q, Zhang Z S, Sun C, Li B, Zi Y Y and He Z J 2012 *Mech Syst Signal PR* **33** 275-98
- [17] Ding Y and Selesnick I W 2015 *IEEE Signal Proc Let* **9** 1364-8
- [18] Boyd S, Parikh N, Chu E, Peleato B and Eckstein 2011 *J Found Trends in Mach Learn* **3** 1-122
- [19] Condat L 2013 *IEEE Signal Proc Lett* **11** 1054-7
- [20] Pustelnik N, Pesquet J C and Chaux C 2012 *IEEE Trans Signal Process* **2** 968-73
- [21] Qiu H, Lee J, Lin J and Yu G 2006 *J Sound Vib* **289** 1066-90