Machinery vibration signal denoising based on learned dictionary and sparse representation

Liang Guo, Hongli Gao, Jun Li, Haifeng Huang and Xiaochen Zhang

School of Mechanical Engineering, Southwest Jiaotong University, ChengDu 610031 China

E-mail: guoliang2248@my.swjtu.edu.cn

Abstract. Mechanical vibration signal denoising has been an import problem for machine damage assessment and health monitoring. Wavelet transfer and sparse reconstruction are the powerful and practical methods. However, those methods are based on the fixed basis functions or atoms. In this paper, a novel method is presented. The atoms used to represent signals are learned from the raw signal. And in order to satisfy the requirements of real-time signal processing, an online dictionary learning algorithm is adopted. Orthogonal matching pursuit is applied to extract the most pursuit column in the dictionary. At last, denoised signal is calculated with the sparse vector and learned dictionary. A simulation signal and real bearing fault signal are utilized to evaluate the improved performance of the proposed method through the comparison with kinds of denoising algorithms. Then Its computing efficiency is demonstrated by an illustrative runtime example. The results show that the proposed method outperforms current algorithms with efficiency calculation.

1. Introduction

To make machines run in good operating condition and reduce the maintenance costs, health conditions needs to be monitored. Vibration signals collected from a machine contain rich information[1]. Hence, By inspecting the physical characteristics of the vibration signals, one is able to detect the presence of a fault in an operation machine, to localise the position of damaged parts, etc.

However, noise is the major barrier in defect detection for the mechanical vibration signal[2]. At present, in the field of eliminating noise, many technologies have been developed such as Moving Average algorithm[3], Fast Fourier Transform(FFT)-based Hilbert transform[4] and so on. Since the beginning of the 90's there has been considerably interest in the use of wavelet transforms for the removal of noise from signals. Mallat et.zz[5] describes the singularity of signal and noise based on multiscale, then raise the denoising algorithm based on maximum the norm of wavelet transform. Donoho develops many correlative algorithms based on wavelet transform. in[6], RiskShrink with the minimax threshold and VisuShrink with the universal threhold are introduced with solid mathematicians properties for signal denoising. At the same time, methods based on redundant representation and sparsity has drawn a lot of research attentions. Sparsity of the unitary wavelet coefficients was considered, leading to the celebrated shrinkage algorithm[7]. Wei et al.[8] combines the Morlet wavelet transform and sparse code shrinkage (SCS) to enhance the bearing vibration impulsive features and suppress residual noise. Those methods seek the maximum correlation between the fixed wavelet basic and raw signal.

In recent years, dictionary learning has emerged in signal filter and compression, due to its capability in adaptive for the raw signal. K-SVD method is proposed in [9] that generalize the K-Means clustering process. Based on the redundant dictionary trained by K-SVD, Chen et al.[10] extracts the impulse components of vibration signal and reduces the noise. And comparing with the mature shrinkage/thresholding algorithm, performances based on dictionary learning is much better. Because of the calculation of large matrix, the K-SVD algorithm has the problem of low efficiency in computing.

In this paper, a vibration signal denoising based on dictionary learning and sparse approximation (ODL-SAD) is proposed. The K-SVD algorithm and modified Online dictionary learning(ODL) algorithm are compared based on simulation and experiment mechanical vibration signal. The comparison results reveal that the modified ODL algorithm is more suitable to real calculation. Orthogonal matching pursuit is applied to sparse represent. At last, the denoised siganl can be got by reconstructing. The signal denoising results of different methods are shown and those support the proposed algorithm.

2. The adaptively vibration signal denoising

The noise reducing technique based on ODL-SAD is mathematical described as:

$$\begin{aligned} \overline{x} = \min_{x} \|x\|_0 \; s.t. \; \|y - \overline{D}x\|_2^2 \leq \gamma \\ \overline{y} = \overline{D}\overline{x}. \end{aligned} \tag{1}$$

where $\|.\|_0$ is the zero norm. $\overline{D}$ is a redundant dictionary which can be found at the $Section$2.1. $\overline{x}$ is the sparse coding of vector x. γ is the noise level. $\overline{y}$ is the reconstruction of raw signal y.

2.1. Online Dictionary learning

A dictionary $D \in R^{N\times K}$ is a collection of K atoms that are column vectors of length N. Now there are two different methods of constructing dictionary. The first one is solid dictionary, such as Fourier Transformation matrix[11], Wavelet Transformation matrix[12] and so on. The other way of constructing dictionary is dictionary learning. Dictionary learning aims at finding the adaptive dictionary that yields sparse representation for the training signals. Let X be the sparse signal, corresponding to the sparse representation of training signals Y. Mathematically, the dictionary learning can be expressed as followed:

$$\left\{\overline{D}, \overline{X}\right\} = \arg\min_{D,X} \sum_{l=1}^{L} \|x_l\|_2 + \lambda \|y_l - Dx_l\|^2, \;\; \|d_i\| = 1 \tag{2}$$

where $\overline{D}$ and $\overline{X}$ are the optimized dictionary and sparse vectors. $\|d_i\|$ is the column vector of $\overline{D}$.

The classical dictionary learning algorithm optimizes the cost funciton:

$$D_n = \arg\min_{D} \sum_{i=1}^{I} \|y_i - Dx_i\|_2^2 + \lambda \|x_i\|_p \tag{3}$$

where D_n indicates the n_{th} interated dictionary D. The sparse coefficient X_{n+1} can be calculated through the dictionary D_n:

$$X_{n+1} = \arg\min_{X} \|x_i\|_1 s.t. \|y_i - Dx_i\|_2^2 \leq \gamma \tag{4}$$

where X_{n+1} indicate the $n+1_{th}$ sparse coefficients X. γ is the threshold value. In this stage, Orthogonal matching pursuit (OMP) is often applied because of its rate of convergence .

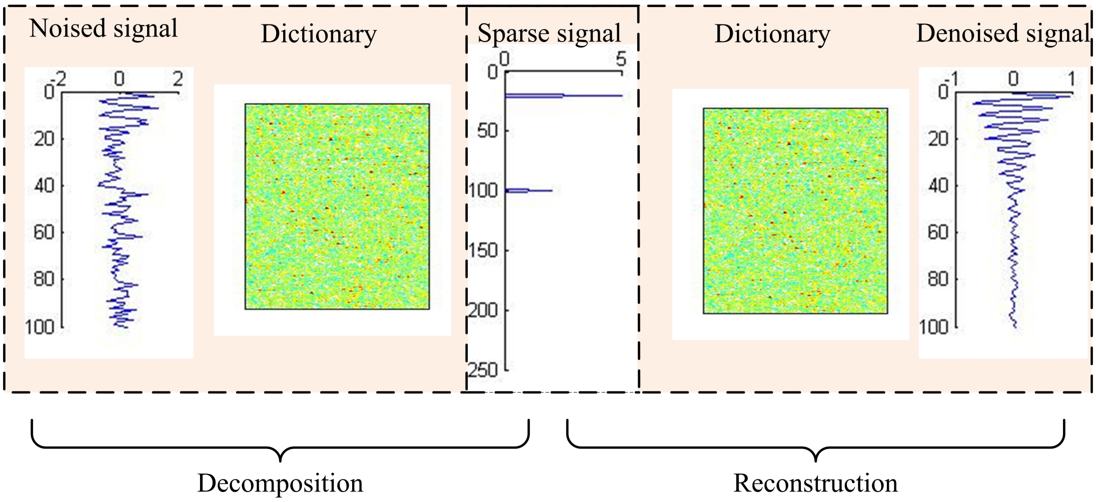


Figure 1: Two-stage dictionary sparse decomposition and reconstruction

In this paper, we use online dictionary learning to optimazation D. In online dictionary learning, Let $A_n \leftarrow A_{n-1} + x_n x_n^T, B_n \leftarrow B_{n-1} + y_n x_n^T$. With the matrix characteristics: $\|A\|_F = tr(A^H A)$, $tr(AB) = tr(AB)$ and $tr(A) = tr(A^T)$, where tr is matrix trace. Then we can derive $Eq.(3)$ to:

$$D_n = \underset{D \in C}{argmin}(\frac{1}{2}Tr(D^T D A_n) - Tr(D^T B_n)) \tag{5}$$

where C is the criteria that dictionary need to meet. Then take a derivate with Eq (6) to D:

$$\frac{1}{2}D(A_n I + (A_n I)^T) - B_n = 0 \tag{6}$$

where I is unit matrix. Then the Newton iteration method is applied to Eq (7).Let $A = [a_1, ..., a_k] \in R^{k\times k} = \sum_{i=1}^{t} x_i x_i^T$, $B = [b_1, ..., b_k] \in R^{m\times k} = \sum_{i=1}^{t} y_i x_i^T$. A parameter u can be calculated by:

$$u_j = \frac{1}{A_{jj}}(b_j - Da_j) + d_j \tag{7}$$

where Ajj is the mian value of diagonal matrix. At last, the value of D is limited by:

$$d_j = \frac{1}{max(\|u_j\|_2, 1)}u_j. \tag{8}$$

where d_j is the $j-th$ column of dictionary D. And at every iteration, the algorithm uses the value of D_{n-1} as a warm restart for cumputing D_n that speeds up convergence.

2.2. Denoising based on learned dictioanry and sparse approximate

Online learning algorithm is adopted to form the adaptive dictionary. Coefficients of the most correlation between signals and atoms are calculated at each iteration of orthogonal matching pursuit(OMP). Once the termination condition of OMP is met, the sparsity vector is obtained. As shown in $Fig.1$, reconstruction is the dual process of decomposition. Only the atoms from the certain non-zeros in sparse signal are reconstructed, the noise components are removed. The above algorithm is called Online dictionary learning and sparse approximation denoising(ODL-SAD). The detailed steps of ODL-SAD are as follows:

(1): Prepare the training sets. Truncate signals Y into segments $[y_1, ..., y_n]$.
(2): Set dictionary learning parameters. Initial dictionary D_0, iteration number K.
(3): Update the dictionary by Eqs.(7) and (8) to form the signal adaptive dictionary $\overline{D}$.
(4): Set parameters of OMP. Matrix $\overline{D}$, vector Y and the error threshold γ.
(5): Obtain the sparse signal $\overline{X}$ through OMP by Eq.(1).
(6): Compute the reconstruction signal $\overline{Y}$
(7): Calculate the envelop spectrum F

3. Simulation

In this section, a simulation experiment is designed to illustrate the denoising ability of ODL-SAD. Generally, when rotating machinery with localized faults is in operation, a serial of impulse signals will be generated. And those signal are always embeded in heavy background noise[13]. The simulation signal is designed as follows:

$$x(t) = \sum_{i=0}^{M} A_i e^{-\beta_i(t/f_s - i/f_m)} sin(2\pi f_i(t/f_s - i/f_m) + \tau_i) + v(t) \tag{9}$$

where A_i is the amplitude, set to 1. The damping factor β_i is equalled to 500. The sampling frequency f_s is 20000Hz. f_m is the impulse frequency (100Hz). f_i is the oscillating frequency, set to 2000Hz, τ_i is the phase of the carrier wave setted random. The simulation signal waveform can be found in $Fig\ 2(a)$, Then the Gaussian white noise $v(t)$ is added to obtain a noise-contaminated signal with a signal-to-noise ratio -2dB. The noise signal is shown in $Fig\ 2(a)$. As you can see, the signal feature is almost swallowed by noise signal.

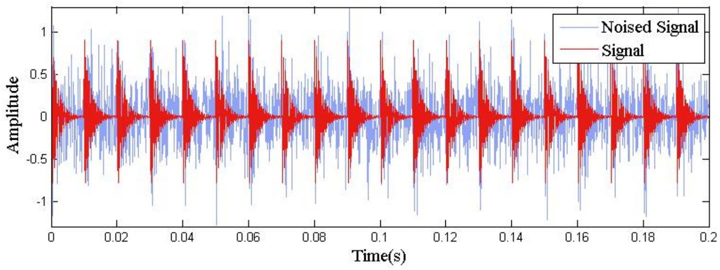
Table 1: Parameters of On-line dictionary learning denoising

Segment size	Dictionary size	Sparse coding	Initial dictionary
100	100*200	OMP	DCT

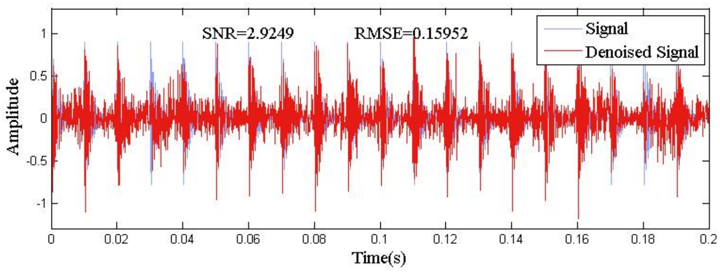
ODL-SAD is applied to analyse the noise-contained simulation signal. The parameters such as the Segment size, Dictionary size and so on can be found in $Table\ 1$. In order to match the atom length with impulse duration and identify all the impulse components, this paper chooses the segment size with 100. The question of dictionary size is an open class, we just simply choose a value we find empirically to perform well as $K/N = 2$[9]. Over-complete DCT is chosen as the initial fixed dictionary. Its update iterations is 20. With above parameters, the resulting signal is depicted in $Fig\ 2(d)$. The noise is almost wiped off while the waveform stay the same.

For comparison, the wavelet denoising and K-SVD dictioanry learning denoising are employed to analyse the same signal in $Fig\ 2(a)$. The K-SVD dictionary learning method has the same parameters as the proposed method, the only difference is the K-SVD instead of On-line learning algorithm. The filtered signal shows in $Fig\ 2(b)$. The signal shape is almost but large of noise is still kept. The wavelet method uses the minimax threshold selection rules and the Morlet wavelet is selected as the mother wavelet. The impulse component is clear obtained in $Fig\ 2(c)$, but attenuation component of signal is significant decreased. To make accurate comparison, Room mean suqare deviation ($RMSE$) and signal to noise (SNR) are adopted. Those results show that ODL-SAD has the highest $SNR = 3.3009$ and the lowest $RMSE = 0.15276$.

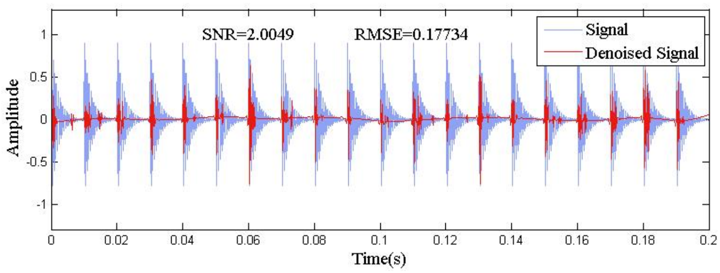
For future declare the ability of ODL to real-time computation, we compare the training time between ODL and K-SVD. $Fig\ 3$ shows test performance when training on the datasets of



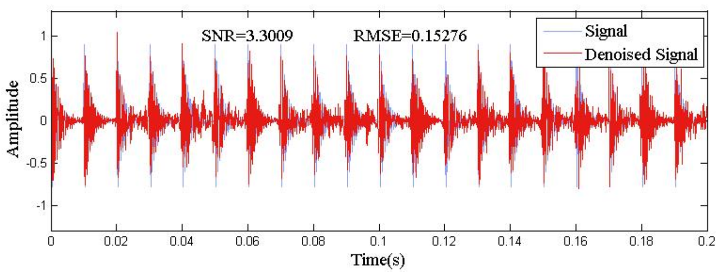
(a) The raw signal and noised signal



(b) The K-SVD dictioanry learning denoising



(c) The Wavelet thresholding denoising



(d) The On-line dictionary learning denoising

Figure 2: The results obtained by different kinds of denoising algorithms

$2 \sim 22$. As the datasets increase, the training time of K-SVD goes up linearly. However, ODL remains about the same.

4. Experiment

To demonstrate the effectiveness of ODL-SAD in actual signal detection, an experiment of detecting the bearing outer race defect is considered.

Bearing run-to-failure tests were performed under normal load conditions on a specially designed test rig. The test rig is shown in $Fig.$ 4. This test rig hosts four test bearings on

Journal of Physics: Conference Series **628** (2015) 012124 doi:10.1088/1742-6596/628/1/012124

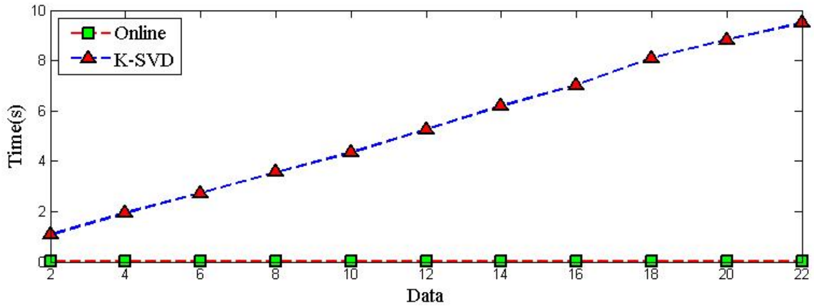


Figure 3: The runtime of K-SVD and Online dictionary learning

one shaft driven by an AC motor and coupled by rub belts. The rotation speed is kept 2000rpm. A radial load of 6000lbs is added to the shaft and bearing by a spring mechanism. All the bearings are forced lubricated.

Four Rexnord ZA-2115 double row bearings were installed on one shaft as shown in $Fig.$ 4. The test bearings have 16 rollers in each row, a pitch diameter of 2.815in, roller diameter of 0.331in, and a tapered contact angle of 15.17°. Vibration data was collected every 20 minutes by a National Instruments DAQCard-6062E data acquisition card. The data sampling rate is 20kHz and the data length is 20480 points. Depending on paper[14], the outer-race fault frequency f_0 of bearing is:

$$f_0 = \frac{n}{2}(1 - \frac{d}{D_p}\cos\theta)f_r \tag{10}$$

where n is number of balls, d is the diameter of the rolling element, D_p is the groove section size, θ is the contact angle, f_r is the shafe frequency.

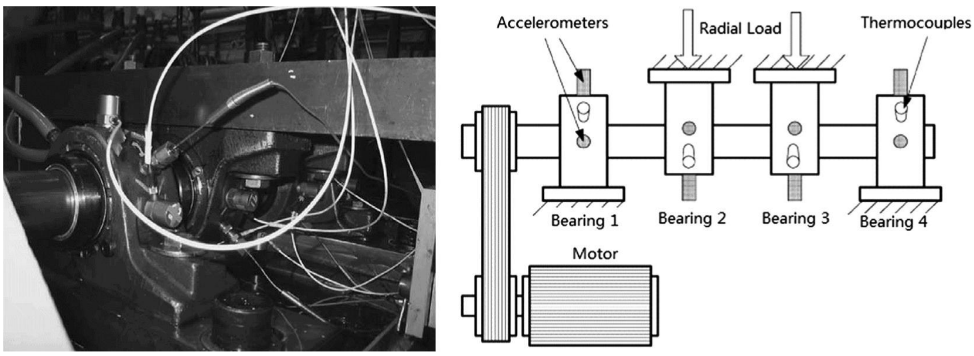
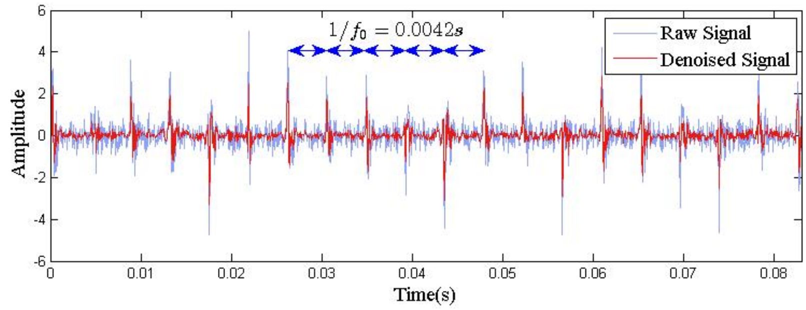


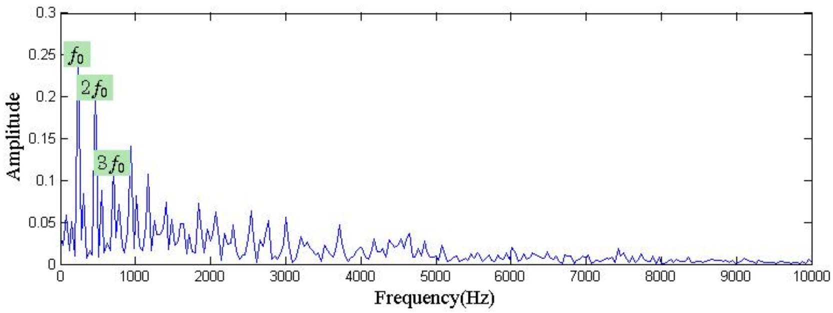
Figure 4: Bearing test rig and sensor placement illustration

Fig 5(a) presents the impulse periodicity vibration signal of bearing 1 with the outer race defect. By Eq (10), the out race band pass frequency is 263.4Hz. The waveform verifies the calculation. ODL-SAD is applied to enhance the signal shown above. The parameters are shown in $Table$ 1, while the segment size is 80 and the dictionary size is $80 * 160$. $Fig.5(a)$ shows that the noise component is eliminated. To further discover the frequency characteristic, envelop spectrum is illustrated in $Fig5(b)$ and $Fig5(c)$. In the envelop spectrum obtained by ODL-SAD presented in $Fig5(c)$, even the higher order harmonic ($e.g.$, $4f_0, 5f_0$) are evident.

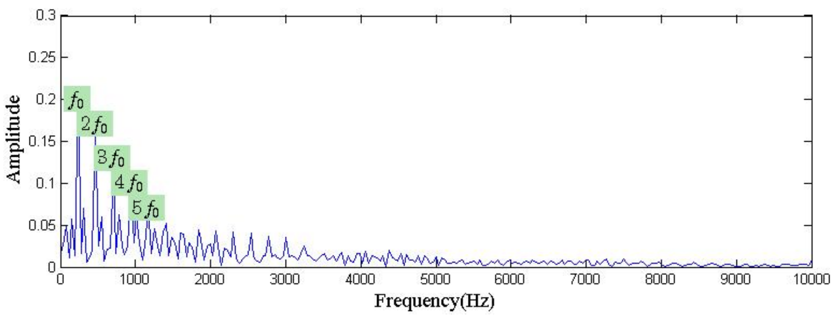
The computing efficiency of ODL and K-SVD based on actual signal is shown in Fig 6. That indicts the run times of ODL algorithm is irrelevant to the size of training datasets. Nevertheless, K-SVD is proportional to data volume.



(a) The raw signal and denoised signal



(b) The envelop spectrum of raw signal



(c) The envelop spectrum of denoised signal

Figure 5: The results of ODL-SAD for bearing vibration signal

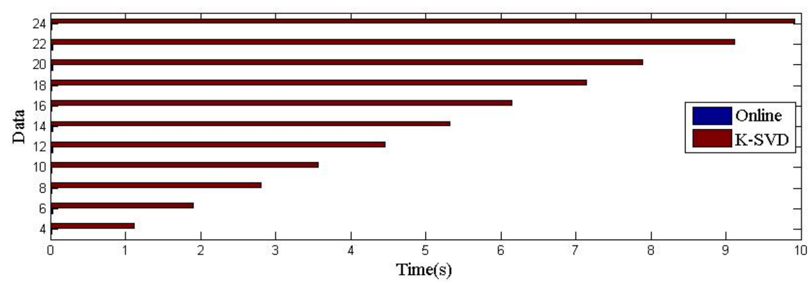


Figure 6: The runtime of K-SVD and Online dictionary learning for bearing vibration signal

5. Conclusion and future work

Signal denoising is crucial to machine health monitoring and damage assessment. The performance of traditional wavelet transform-based methods is greatly impacted by the mother wavelet. Dictionary learning is an alternative method to construct the transfer matrix. However, it is much more challenging to computing the large matrix. This paper presents a On-line dictionary learning and sparse approximation denoising algorithm(ODL-SAD). The transfer matrix of sparse signal is established by dictionary learning. The modified On-line algorithm runs the training processing real-time. In the simulation analysis experiment verification, the RMSE and SNR are compared between denoised signal and raw signal with different algorithms. Results show that the proposed algorithm capture the efficiency results with less deformation. And it is significantly faster than traditional dictionary learning algorithm. Multi-Scale dictionary learning for sparse and redundant signal representations is considering to improve the flexibility.

Acknowledgments
This research was funded by the National Science Foundation (NSF) under Grant No.51275426

References

[1] Y. Lei, J. Lin, Z. He, Y. Zi, Application of an improved kurtogram method for fault diagnosis of rolling element bearings, Mechanical Systems and Signal Processing 25 (5) (2011) 1738–1749.
[2] H. Qiu, J. Lee, J. Lin, G. Yu, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, Journal of sound and vibration 289 (4) (2006) 1066–1090.
[3] S. Jha, O. Singh, R. K. Sunkaria, Modified approach for ecg signal denoising based on empirical mode decomposition and moving average filter, International Journal of Medical Engineering and Informatics 6 (2) (2014) 165–188.
[4] R. B. Randall, Frequency analysis, Brül & Kjor, 1987.
[5] S. Mallat, W. L. Hwang, Singularity detection and processing with wavelets, Information Theory, IEEE Transactions on 38 (2) (1992) 617–643.
[6] D. L. Donoho, J. M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (3) (1994) 425–455.
[7] E. J. Candes, D. L. Donoho, Recovering edges in ill-posed inverse problems: Optimality of curvelet frames, Annals of statistics (2002) 784–842.
[8] W. He, Z.-N. Jiang, K. Feng, Bearing fault detection based on optimal wavelet filter and sparse code shrinkage, Measurement 42 (7) (2009) 1092–1102.
[9] M. Aharon, M. Elad, A. Bruckstein, -svd: An algorithm for designing overcomplete dictionaries for sparse representation, Signal Processing, IEEE Transactions on 54 (11) (2006) 4311–4322.
[10] J. Mairal, M. Elad, G. Sapiro, Sparse representation for color image restoration, Image Processing, IEEE Transactions on 17 (1) (2008) 53–69.
[11] M. Goodwin, Matching pursuit with damped sinusoids, in: Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on, Vol. 3, IEEE, 1997, pp. 2037–2040.
[12] P. Vera-Candeas, N. Ruiz-Reyes, M. Rosa-Zurera, D. Martinez-Munoz, F. Lopez-Ferreras, Transient modeling by matching pursuits with a wavelet dictionary for parametric audio coding, Signal Processing Letters, IEEE 11 (3) (2004) 349–352.
[13] X. Chen, Z. Du, J. Li, X. Li, H. Zhang, Compressed sensing based on dictionary learning for extracting impulse components, Signal Processing 96 (2014) 94–109.
[14] R. B. Randall, J. Antoni, Rolling element bearing diagnosticsa tutorial, Mechanical Systems and Signal Processing 25 (2) (2011) 485–520.