


Article

# Optimized Dynamic Mode Decomposition via Non-Convex Regularization and Multiscale Permutation Entropy

Zhang Dang <sup>1,2,3</sup> , Yong Lv <sup>1,2,\*</sup>, Yourong Li <sup>1,2</sup> and Cancan Yi <sup>1,2</sup>

<sup>1</sup> Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China; dangzhang@wust.edu.cn (Z.D.); liyourong@wust.edu.cn (Y.L.); meycancan@wust.edu.cn (C.Y.)

<sup>2</sup> Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan 430081, China

<sup>3</sup> National Demonstration Center for Experimental Mechanical Education, Wuhan University of Science and Technology, Wuhan 430081, China

\* Correspondence: lvyong@wust.edu.cn; Tel.: +86-027-6886-2857; Fax: +86-027-6886-2212

Received: 22 January 2018; Accepted: 23 February 2018; Published: 27 February 2018

**Abstract:** Dynamic mode decomposition (DMD) is essentially a hybrid algorithm based on mode decomposition and singular value decomposition, and it inevitably inherits the drawbacks of these two algorithms, including the selection strategy of truncated rank order and wanted mode components. A novel denoising and feature extraction algorithm for multi-component coupled noisy mechanical signals is proposed based on the standard DMD algorithm, which provides a new method solving the two intractable problems above. Firstly, a sparse optimization method of non-convex penalty function is adopted to determine the optimal dimensionality reduction space in the process of DMD, obtaining a series of optimal DMD modes. Then, multiscale permutation entropy calculation is performed to calculate the complexity of each DMD mode. Modes corresponding to the noise components are discarded by threshold technology, and we reconstruct the modes whose entropies are smaller than a threshold to recover the signal. By applying the algorithm to rolling bearing simulation signals and comparing with the result of wavelet transform, the effectiveness of the proposed method can be verified. Finally, the proposed method is applied to the experimental rolling bearing signals. Results demonstrated that the proposed approach has a good application prospect in noise reduction and fault feature extraction.

**Keywords:** dynamic mode decomposition; sparse optimization; non-convex regularization; multiscale permutation entropy; feature extraction

---

## 1. Introduction

Generally, amplitude-modulated and/or frequency-modulated (AM-FM) multi-component signals collected by sensors are interfered with by background noise, resulting lower signal-to-noise ratio (SNR) [1]. The reason for the low SNR signals obtained by a determinate acquisition system are usually as follows: firstly, when a rolling bearing is in the infantile fault period, the fault feature is inconspicuous and buried in the background noise. Secondly, signal energy as well as incipient failure features will attenuate because of the long distance between the fault source and the sensor, which usually is located on the bearing pedestal. Moreover, the characteristic signal will also be attenuated when there is a long transmission path from the sensor to the acquisition device [2]. Thirdly, the signals collected by sensors contain strong background noise. The main transmission system of a machine is a complex system composed of multiple components such as motors, couplings, multiple

reduction gearbox. The vibration of all components will be collected by the sensor. Noisy signals make it difficult to identify the fault features in signal processing, therefore, it is necessary to perform denoising and characterize fault feature accurately on the low SNR vibration signal.

In the last few decades, many theoretical denoising methods have been proposed and are widely used in medical signaling, image processing and mechanical engineering signal processing [3–5]. Donolo and Johnstone [6], taking the advantage of the spatial diversity of wavelet basis functions, proposed a denoising method with a threshold based on wavelet transform. Their methods concentrate the energy of the ideal signal, benefiting from the effective wavelet, guaranteeing a certain de-noising effect. However, the selection of threshold functions is the main problem with wavelet de-noising algorithms, which usually generate a deviation between the calculated result and the real one. Empirical mode decomposition (EMD) [7] can be treated as a typical adaptive signal decomposition method of frequency-selection and filtering. The signal is decomposed into a series of intrinsic mode functions (IMFs) arranged from high frequency to low frequency. The source signal can be processed by abandoning high-frequency components for the consideration that noise components are generally distributed in the high-frequency domain. However, as EMD has the defects of modal aliasing, the ideal signal and noise coexist in the selected IMFs. This would result in the loss of useful information. Ensemble empirical mode decomposition (EEMD) [8] was developed to solve the mode-mixing problem of EMD by adding noise to the source signal. It is more accurate and effective for data decomposition than EMD, but it prompts new trouble by being time consuming. Variable mode decomposition (VMD) [9,10] is a signal decomposition method based on Wiener filtering, one dimensional Hilbert transform and heterodyne demodulation analysis, but the decomposition results are restricted by the selection of penalty parameters and the number of components. Singular value decomposition (SVD) [11], as a data-processing method without prior knowledge, has been successfully applied to signal denoising and proved to be effective. However, there exist two pivotal problems: one is that how to determine the effective rank order after decomposition, and the other one is that how to achieve the reconstruction matrix. Meanwhile, the noise reduction effect of SVD is not prominent when analyzing the low SNR signal.

Dynamic mode decomposition (DMD) [12] is an equation-free Koopman frequency analysis technique based on the theory of SVD and mode decomposition. It can extract the spatiotemporal coherent characteristics by decomposing a dynamic model into a series of single-frequency non-orthogonal modes based on its inherent space-time. The reconstructed signal can characterize the useful information of the original signal perfectly by superposing the filtered modes according to time coefficients. The algorithm is suitable for extracting dynamic information of nonlinear systems [13]. The instantaneous components and high-frequency noise components can be effectively removed, due to the dynamic delay processing in the algorithm [12]. The DMD algorithm has been well applied in the fluid mechanics community. Rowley [12] applied DMD in the jet cross flow field. The results show that DMD theory can extract the shedding frequency while the linear stability analysis method cannot. In addition, DMD can separate the shear layer and wall vortices, whereas the proper orthogonal decomposition (POD) algorithm is ineffective. Schmidt [13] applied DMD to the flow field in a variety of environmental conditions, indicating that the DMD theory is effective for feature extraction with numerical simulation and measured velocity field data. Moreover, DMD can separate video frames into a background and multiple foregrounds successfully when applied to video processing [14]. DMD can also be used for biometrics to detect fraudulent samples [15]. In the field of robotics and neuroscience, DMD has been used to estimate the perturbation of human-robot interactions [16], and the coherent modes in large-scale neural recording have been extracted [17]. Nowadays, theoretical research with DMD mainly focuses on two aspects. In the first aspect, a variety of improved methods are proposed based on the standard DMD to solve the problem that DMD cannot accurately extract the dynamic information due to the influence of signal noise. These strategies range from noise-corrected dynamic mode decomposition with control [18], total least-squares DMD [19], compressed sensing DMD [20] to sparsity-promoting DMD [21]. The improved DMD algorithm can extract fluid dynamic information

more accurately in varying degrees. In the second aspect, in view of the large data processing problems, researchers aim at reducing the computational complexity and the program storage space in either pre-processing or post-processing, using strategies including random DMD [22], random low-rank DMD [23], and tensor-based DMD [24]. The standard DMD and its improved algorithm are more suitable in analyzing periodic or quasi-periodic signals, but it is prone to generate some false mode components. Therefore, the theory of fault vibration signal processing and fault mode recognition with DMD needs further study.

Essentially, DMD is a hybrid algorithm based on mode decomposition and SVD, and it inevitably inherits the drawbacks of these two algorithms, including the selection strategy of truncated rank order and DMD modes. The number of truncation rank for the similar matrix needs to be artificially and blindly selected in advance; whether the number value is appropriate or not is of crucial importance for accurately representing the dynamic information of reconstructed signal. The measured vibration signal with the characteristic of non-linear and non-stationary is coupled with multi-components and contain strong environmental noise. To achieve the purpose of fault feature extraction with DMD, it requires separation of the fault feature components effectively from the DMD modes. It is necessary to establish an adaptive optimization algorithm based on DMD to achieve an accurate truncation of rank and screen the single-frequency non-orthogonal modes effectively.

In this paper, an optimized DMD algorithm is proposed for extracting the fault features drowned in the multi-component coupled and noisy mechanical signal. The idea of sparse optimization is adopted to select the optimal rank for the solution of truncated rank order selection problem. By introducing the constraint of non-convex regulation to the rank function, the similar matrix has low rank attribute, making the ideal components (signals) more centralized. As for the fact that the noise's multiscale permutation entropy (MPE) is larger than the ideal component's, the low rank modes are selected corresponding to faults and inherent features by a threshold, solving the difficulty in selecting useful DMD modes. Then the signal is reconstructed by the low rank DMD modes, and Fourier transform (FT) is conducted to achieve the purpose of denoising and fault diagnosis of the noisy mechanical signal.

The structure of this paper is as follows. Section 2 introduces the basic theory, including the standard algorithm of DMD, the sparse optimization algorithm with non-convex optimization regulation, and MPE. In Section 3, the proposed algorithm is applied to the normal and faulty bearing noisy simulation signals to verify its effectiveness. Analysis results of two experimental bearing fault signals are described in Section 4. Conclusions are summarized in Section 5.

## 2. Methodology

In this part, the standard DMD algorithm is introduced based on time series in fault diagnosis. To solve the problem of selecting the optimal rank, a technique based on sparse optimization is proposed. The details are described as in Section 2.2. Then, we introduce the basic algorithm of MPE, and show the basic flow chart in the form of illustration.

### 2.1. DMD with Time Series Signal

Supposing that a signal  $S$  is acquired by numerical simulation or a sensor with equal intervals between two successive sampling points,  $S = (x_1, x_2, \dots, x_i, \dots, x_N)$ ,  $x_i \in R$ , while  $\Delta t = x_{i+1} - x_i$ . Hankel matrix  $X$  can be obtained by a sliding window over the corresponding vector  $S$ .

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ x_2 & x_3 & \cdots & x_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ x_i & x_{i+1} & \cdots & x_{n+i-1} \\ \cdots & \cdots & \cdots & \cdots \\ x_m & x_{m+1} & \cdots & x_N \end{bmatrix} = [X_1, X_2, \cdots, X_n], \mathbf{X} \in \mathbb{R}^{m \times n} \tag{1}$$

where  $X_i = (x_i, x_{i+1} \dots x_{s+i-1})^T, m + n - 1 = N$ . Decomposing the Hankel matrix into two shift-stack matrices as in the form of Equation (2).

$$\mathbf{X}_t = [X_1, X_2, \cdots, X_{n-1}]; \mathbf{X}_{t+1} = [X_2, X_3, \cdots, X_n]; \mathbf{X}_t, \mathbf{X}_{t+1} \in \mathbb{R}^{m \times (n-1)} \tag{2}$$

Assume that the two continuous sequence matrices satisfy a mapping relation with optimal linear operator. DMD algorithm utilizes a low rank expression of matrix  $\mathbf{A}$  to capture the potential dynamic characteristics.

$$\mathbf{X}_{t+1} = \mathbf{A}\mathbf{X}_t \tag{3}$$

$\mathbf{A}$  can be also regarded as an operator to approximate the dynamic characteristics when the data are produced in nonlinear systems [13]. Matrix  $\mathbf{A}$  usually contains many complex entries if  $S$  is presented in high-dimensional space. The DMD algorithm finds an optimal similarity matrix  $\mathbf{F} \in \mathbb{C}^{r \times r}$  by projecting the matrix  $\mathbf{A}$  with its  $r$ -th order eigenvectors according to POD.

$$\mathbf{A} = \mathbf{U}\mathbf{F}\mathbf{U}^* \tag{4}$$

where  $\mathbf{U}^*$  is the complex conjugate transpose of POD mode  $\mathbf{U}$ , and  $\mathbf{U}$  is obtained by SVD of matrix  $\mathbf{X}_t$ .

$$\mathbf{X}_t \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^* \tag{5}$$

where  $\mathbf{U} \in \mathbb{C}^{m \times r}, \mathbf{V} \in \mathbb{R}^{r \times (n-1)}$  are known as the left and right eigenvectors respectively,  $\mathbf{U}^*\mathbf{U} = \mathbf{I}, \mathbf{V}^*\mathbf{V} = \mathbf{I}$ , The symbol  $*$  denotes the complex conjugate transpose.  $\mathbf{\Sigma}_r \in \mathbb{R}^{r \times r}$  is a diagonal matrix containing  $r$  non-zero eigenvalues  $\{\sigma_1, \sigma_2, \cdots, \sigma_r\}$  on its diagonal.

The matrix  $\mathbf{F}$  is determined by minimizing the Frobenius norm of the error between  $\mathbf{X}_{t+1}$  and  $\mathbf{A}\mathbf{X}_t$ .

$$\mathbf{\Psi} = \underset{\mathbf{F}}{\operatorname{argmin}} \|\mathbf{X}_{t+1} - \mathbf{U}\mathbf{F}\mathbf{\Sigma}\mathbf{V}^*\|_F^2 \tag{6}$$

Optimal solution of the above formula can be got directly.

$$\mathbf{F}_{DMD} = \mathbf{U}^*\mathbf{X}_{t+1}\mathbf{V}\mathbf{\Sigma}^{-1} \tag{7}$$

Equation (7) was proposed by Schmid [13], while the literature also introduced the application of DMD algorithm. Then a matrix  $\mathbf{F}_{DMD} \in \mathbb{R}^{r \times r}$  with lower dimension is obtained. After the similar transformation above,  $\mathbf{F}_{DMD}$  and  $\mathbf{A}$  are almost equivalent with each other. Hence, the eigenvectors and eigenvalues of  $\mathbf{A}$  are replaced by the eigenvectors and eigenvalues of  $\mathbf{F}_{DMD}$ .

Matrix  $\mathbf{F}_{DMD}$  optimizes the low-dimensional approximation of the internal data mapping matrix  $\mathbf{A}$  by the POD subspace modes of  $\mathbf{X}_t$ . The dynamic information of the  $r$ -dimensional subspace is expressed as:

$$\mathbf{X}_{t+1} = \mathbf{F}_{DMD}\mathbf{X}_t \tag{8}$$

Perform eigenvalue decomposition on the similarity matrix  $F_{DMD}$ :

$$F_{DMD} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^{-1} \tag{9}$$

where  $\mathbf{W} = [\omega_1, \omega_2, \dots, \omega_r] \in \mathbf{R}^{r \times r}$  are the eigenvectors of similarity matrix  $F_{DMD}$ , and  $\mathbf{\Lambda} = \text{diag}([\lambda_1, \lambda_2, \dots, \lambda_r]) \in \mathbf{R}^{r \times r}$  is a diagonal matrix containing the corresponding complex eigenvalue  $\lambda_i$ .

The evolution of the signal characteristics can be characterized by similarity matrix. In addition, the  $i$ -th eigenvector of original operator is presented by relevant feature vector of similarity matrix.

$$\phi_i = \mathbf{X}_{t+1} \mathbf{V}_i \mathbf{\Sigma}_i^{-1} \mathbf{W}_i \tag{10}$$

Equation (10), the project DMD approximate solution, is often called standard DMD mode. By firstly rewriting for convenience  $\varpi_i = \ln(\lambda_i) / \Delta t$ , the approximate solution of reconstruction matrix of DMD is then given by:

$$\mathbf{X}_{DMD} = \sum_{i=1}^r \phi_i \exp(\varpi_i \Delta t) b_i = \mathbf{\Phi} \exp(\mathbf{\Omega} t) b, \mathbf{X}_{DMD} \in \mathbf{R}^{m \times n} \tag{11}$$

where  $\mathbf{\Phi}$  is a matrix consisted of DMD modes  $\phi_i$ ,  $\mathbf{\Omega} = \text{diag}(\lambda_i)$  is a diagonal matrix whose entries are continuous-time eigenvalues  $\varpi_i$  of the similarity matrix  $F_{DMD}$ ,  $b$  is a vector containing the initial amplitude of each mode,  $b = \mathbf{\Phi}^\Gamma \mathbf{x}_t$ ,  $\Gamma$  denotes the Moore-Penrose pseudoinverse.

### 2.2. Sparse Optimized DMD via Non-Convex Regularization

As described in Section 2.1,  $r$ -th non-zero eigenvalues of the similarity matrix  $F_{DMD}$  are retained to reconstruct the analytical signal, and  $r$  is the truncation rank order for the SVD of  $\mathbf{X}_t$ . In the process of standard DMD, the rank of  $r$  needs to be manually preset. It plays a crucial role in accurately expressing the dynamic features of the original signal. If  $r$  is too large, it may reconstruct some unnecessary sparse components, such as noise, which cannot play a good denoising effect. Moreover, it increases the storage space and decreases the calculating efficiency. On the contrary, if  $r$  is too small, it will inevitably filter out useful low-rank components that symbolize the system dynamic information, failing to extract the signal physical characteristics. The difficulty lies in the determination of optimized truncation rank  $r$ .

Recently, many sparse optimization algorithms have been proposed in the fields of signal and image processing [16], compressive sensing [17], machine learning [25], and data mining [26] for consideration that the optimization variables have some sparse structures. Sparse optimization makes it possible to reconstruct high-dimensional signals and extract potential information from a small amount of data. Meanwhile, sparse optimization can greatly accelerate the calculation speed in large-scale optimization problems. Based on DMD theory, this paper introduces a parametric sparse optimization framework for the optimal low-rank approximation of  $\mathbf{X}_t$ .

If the shift-stack Hankel matrix  $\mathbf{X}_t$  take the form of

$$\mathbf{X}_t = \tilde{\mathbf{X}}_t + \mathbf{N}_t \tag{12}$$

where  $\tilde{\mathbf{X}}_t$  is the ideal clean matrix containing all the dynamical characteristics of the vibration system.  $\mathbf{N}_t$  is a random noisy matrix. A typical sparse optimization framework with minimal low rank matrix approximation is defined as follows.

$$\mathbf{X}^* = \underset{F}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|_F^2 + \lambda \Pi \right\} \tag{13}$$

where  $\lambda > 0$  is a regularization coefficient,  $\Pi$  is a sparse regularization.

The above minimum low-rank approximation problem is transformed into a standard problem of minimizing nuclear penalty when  $\Pi = \sum_{i=1}^k |\sigma_i(\mathbf{X}_t)|$ . Theoretically, the nuclear norm optimization framework is convex, and it can be directly solved by taking traditional SVD on the shift-stack Hankel matrix  $\mathbf{X}_t$ . However, the solution of convex norm-based nuclear norm optimization often leads to underestimate the rank of the original matrix [27]. Selesnick [28] proposed a sparse optimization method based on non-convex regularization to denoise a noisy signal. The calculation process can guarantee the strict convexity of Equation (13) and generates a unique solution. Inspired by the idea of [28,29], this paper adopts a sparse framework of non-convex optimization to obtain the best truncated rank of the shift-stack matrix. Non-convex optimization can approximate the non-zero singular value of the calculation matrix more accurately than nuclear norm optimization framework. Moreover, the non-convex method is more robust to strong background noise than nuclear norm optimization.

Let us define  $\Pi = \sum_{i=1}^k \phi(\sigma_i(X_t); \alpha)$ ,  $\phi: R \rightarrow R$  corresponding to the non-convex sparse regularization of sparse optimization. Non-convex penalty function must meet the following five properties:

- (1)  $\phi$  is symmetric and continuous in  $R$  space. i.e.,  $\phi(-x; a) = \phi(x; a)$ , which is continuous derivable in  $R$ ;
- (2)  $\phi'(x; a) > 0, \forall x > 0$ ;
- (3)  $\phi''(x; a) \leq 0, \forall x > 0$ ;
- (4)  $\phi'(0^+; a) = 1$ ;
- (5)  $\inf_{x \neq 0} \phi''(x; a) = \phi''(0^+; a) = -a$ ;

Chen [30] listed various penalty functions satisfying the above assumption in Table II. Here we employ the partly quadratic penalty function [27], defined as:

$$\phi(x; a) = \begin{cases} |x| - \frac{a}{2}x^2, & |x| < \frac{1}{a} \\ \frac{1}{2a}, & |x| \geq \frac{1}{a} \end{cases} \quad (14)$$

where  $\alpha$  is a non-convex penalty parameter and controls the degree of non-convexity of  $\phi$ , satisfying  $0 \leq \alpha \leq \frac{1}{\lambda}$ .

Finally, the sparse optimization objective function based on non-convex regularization is defined as Equation (15). The optimal truncation rank  $r$  of singular value which should be determined in the process of DMD algorithm is obtained by the non-convex optimization framework.

$$\mathbf{X}^* = \underset{F}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|_F^2 + \lambda \sum_{i=1}^k \phi(\sigma_i(\mathbf{X}_t); \alpha) \right\} \quad (15)$$

### 2.3. Optimized DMD Modes via MPE

The result of the DMD algorithm obtains a series of single-frequency modes that represent the dynamics of the original system. DMD modes are similar to the IMFs obtained by EMD, while each DMD mode corresponds to a single frequency. It can also be regarded as the vibration modes in mechanical vibration. All DMD modes express the inherent characteristics, fault information and noise background components (introduced by the shift-stack matrix  $\mathbf{X}_{t+1}$ ) of the original system information, making it difficult to diagnose mechanical fault features effectively. Approximate entropy (AE) [31] reflects the time series' complexity by calculating the approximate degree of two points in the given dimension and the probability of new mode when the dimension changes. AE has the characteristics of stable calculation results with short data sequence. Sample entropy [32,33] is an improved algorithm, and it performs better than AE to maintain the consistency of the calculation results. Permutation entropy (PE) [33] is an algorithm based on time series phase space reconstruction

and neighborhood value comparison, which can effectively describe the complexity of the system. PE has the advantages of simple algorithm, fast computing speed, and reliable calculation results [34]. MPE is an improved algorithm based on PE. In addition, it is especially suitable for nonlinear data. In this paper, MPE threshold is used to identify and remove the modes of noise components. The remaining low-rank components are reconstructed by DMD to achieve the purpose of noise reduction and fault feature extraction for noisy fault signals.

The basic idea is that it first gets the time series coarse-grained in time-scale, then calculates PE on the coarse-grained data. The process of MPE is briefly described here. For more detailed information about MPE, it can be referred to in [33,35].

First, a given time series  $\{x_k, k = 1, 2, \dots, N\}$  is coarse-grained.

$$x_j^s = \frac{1}{s} \sum_{i=(j-1)s+1}^{js} x_k \quad (1 \leq j \leq \lfloor \frac{N}{s} \rfloor) \quad (16)$$

where  $S$  is the scale factor,  $S = 1, 2, \dots$ , when  $S = 1$ , the coarse-grained sequence is the original sequence, and the result of MPE is PE.  $\lfloor \frac{N}{s} \rfloor$  means rounding  $\frac{N}{s}$ . If  $S \geq 2$ , the original sequence is coarse-grained with the length of  $\lfloor \frac{N}{s} \rfloor$ .

Secondly, the phase space reconstruction is carried on the coarse-grained time series.

$$x_i^d = [x_{(i)}, x_{(i+\tau)}, \dots, x_{(i+(m-1)\tau)}] \quad (17)$$

where  $i = 1, 2, \dots, N$ .  $d \geq 2$  is the embedding dimension;  $\tau$  is the delay time; Each row vector in the reconstructed phase space matrix is a reconstruction vector, and there are  $d$  reconstruction vectors in total.

Considering every two data in the form of ascending or descending order, the  $j$ -th reconstruction vector has  $d!$  possible ordinal patterns. Finally, the MPE value  $H$  is defined as:

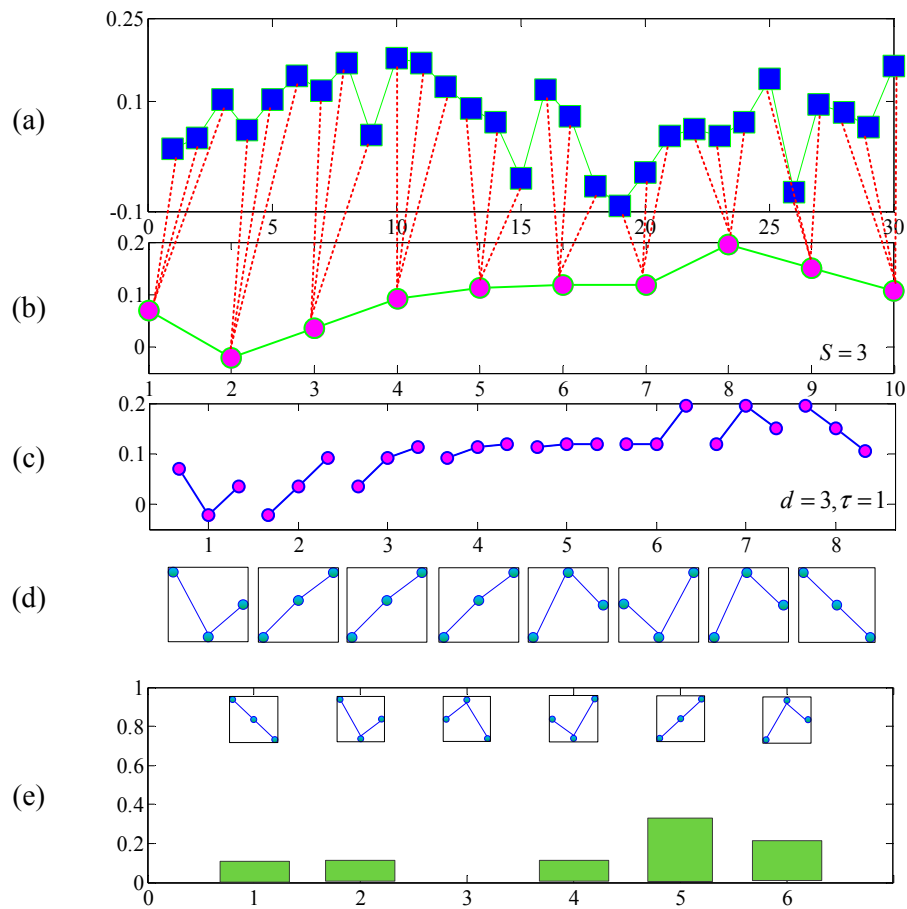
$$H(N, d, \tau) = -\frac{1}{\ln(d!)} \sum_{i=1}^{d!} \frac{p_i}{N} \ln\left(\frac{p_i}{N}\right) \quad (18)$$

where  $p_i$  represents the frequency of a particular sequence pattern in all reconstruction vectors.

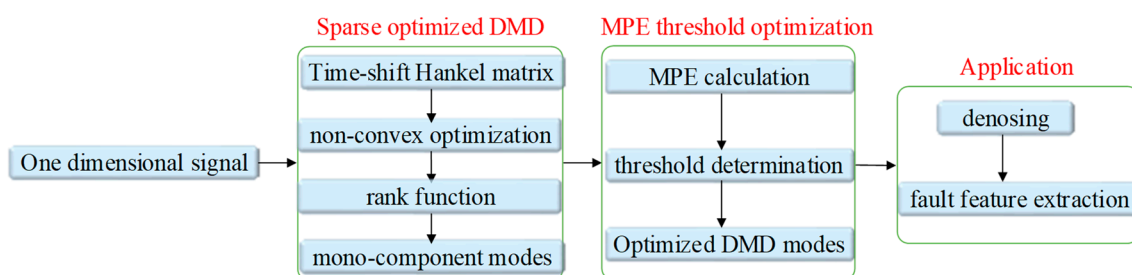
Figure 1 shows the calculation process of  $p_i$  with a few capacity sample data sequences. The regular parameter  $\ln(d!)$  ensures that the maximum value of  $H(N, d, \tau)$  is 1.

MPE is a measure of system complexity. The larger the entropy, the more complex the system components, and the closer the system is to the stochastic signal. Conversely, the smaller the entropy value, the simpler the system components appears. For the fixed-length analysis sequence, the result of MPE dependent on the scale factor  $s$  and the embedding dimension  $d$ . Generally, if  $d$  is too small, the reconstructed vector contains little information, the algorithm loses its validity and cannot detect the kinetic mutation of the sequence. If  $d$  is too large, the reconstruction of phase will homogenize the time series, which is not only time-consuming but also unable to reflect the small changes of the sequence. Bandt [33] and Zunino [34] suggested that  $d$  should be in the range 3–7. Time delay  $\tau$  has little effect on the MPE of time series [35,36], and in most literatures [36–38], authors employ  $\tau = 1$ . In this paper, we employ  $s = 5$ ,  $d = 6$ ,  $\tau = 1$  in the calculation of MPE. In [36,38,39], the length  $N$  of time series are suggested as  $N > 5d!$ , thus the sample points of the simulation signal should be larger than 3600.

Ultimately, the flowchart of the algorithm based on DMD via non-convex regularization and MPE is shown in Figure 2.



**Figure 1.** An example for the schedule of calculating  $p_i$  with MPE. (a) noisy time series with length of  $N = 30$ ; (b) coarse-grained time series within non-overlapping time slice referred to as the scale factor  $s = 3$ ; (c) data segments with embedded dimension  $d = 3$  and time delay  $\tau = 1$ ; (d) ordinal patterns according to the data segments; (e) The frequency of each particular ordinal patterns of all the data segments.



**Figure 2.** A flow chart of the proposed algorithm.

### 3. Algorithm Application in Fault Bearing Simulation Signal

Here we adopt the dynamic model of bearing fault signal established by Randall [40]. The mathematical model of fault rolling bearing is described in Equation (19). According to the characteristics of bearing structure, instantaneous impact signal should be generated when bearing parts pass through the fault location. The impact motivates the bearing system oscillating by its natural frequency with high-frequency attenuation. Define  $T$  as a period of two intervals,  $S(t)$  as the natural frequency oscillation function, and  $A_k$  as the  $i$ -th amplitude of shock response. The interference of additive noise  $n_t$  (zero mean additive background noise) should be taken into consideration in the



simulation signal in view of poor working environment of rolling bearings. Therefore, the mathematical model of fault rolling bearing can be described as:

$$\begin{cases} f(t) = \sum_{k=1}^M A_k S(t - kT - \tau_k) + n(t) \\ A_k = a_k \cos(2\pi f_m t + \varphi_A) + c_A \\ S(t - kT - \tau_k) = \exp[-B(t - kT - \tau_k)] \cdot \sin[2\pi f_n(t - kT - \tau_k) + \gamma] \end{cases} \quad (19)$$

where  $a_k$  is the  $k$ -th impact energy;  $\gamma$  and  $\varphi_A$  are the initial phases;  $f_m$  is the modulation frequency;  $B$  is the attenuation coefficient depended on the bearing system;  $\tau_k$  is the time lag from its mean period due to the presence of tiny fluctuation, and  $c_A$  is a random constant.

Time domain simulation signals of rolling bearing with normal bearing and inner race, outer race and roller faults are respectively plotted in Figure 3. It should be pointed out that Gaussian white noise is added so as to SNR reach to 0.5 precisely. The parameters selection of Equation (19) is listed in Table 1, where  $f_r$  and  $f_c$  are denoted as rotational frequency and cage frequency.  $f_i, f_o, f_b$  respectively represent the failure frequency of the inner circle, the outer ring, and the rolling element. The sampling frequency is selected as 5 KHz, and the number of total sampling points is 8000.

Table 1. The parameters selection with simulation signals of rolling bearing.

$a_k$	$\gamma$	$\varphi_A$	$B$	$\tau_k$	$c_A$	$f_r(\text{Hz})$	$f_i(\text{Hz})$	$f_o(\text{Hz})$	$f_b(\text{Hz})$	$f_c(\text{Hz})$
3.5	0	0	750	0.01	1	25	180	130	80	12

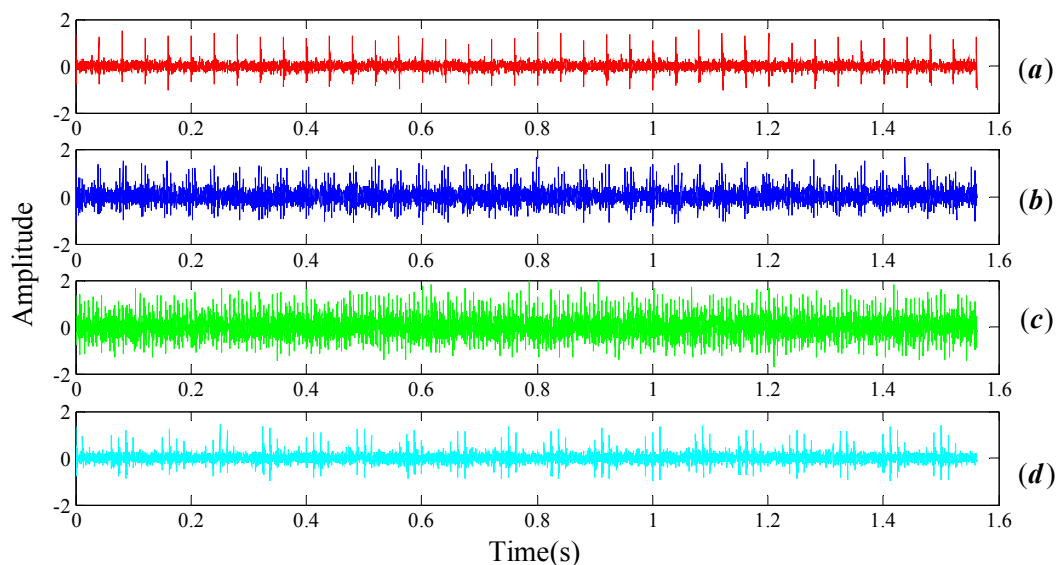
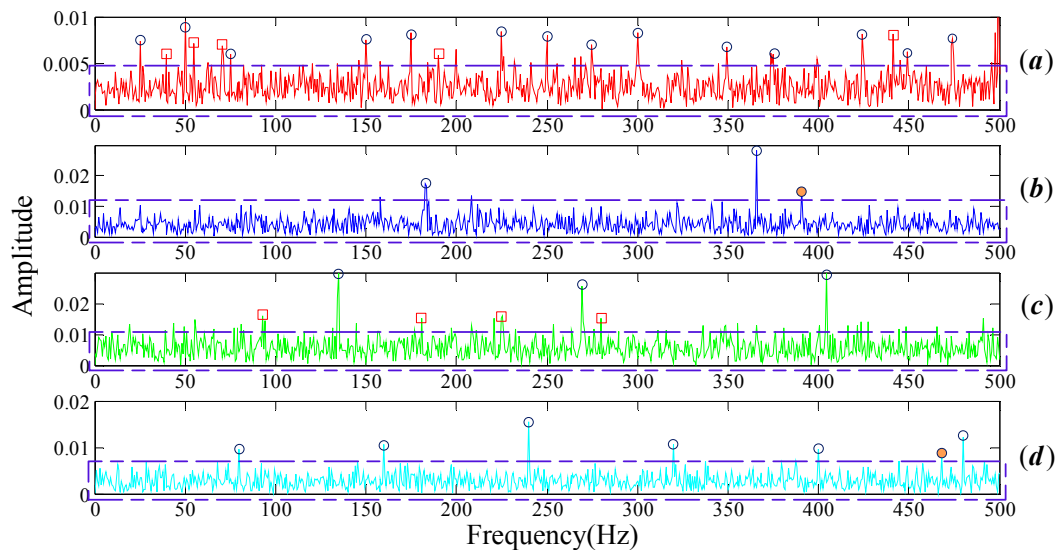


Figure 3. Time domain simulation signals of rolling bear including noise with SNR 0.5 (a) normal bearing; (b) inner fault; (c) outer fault; (d) cage fault.

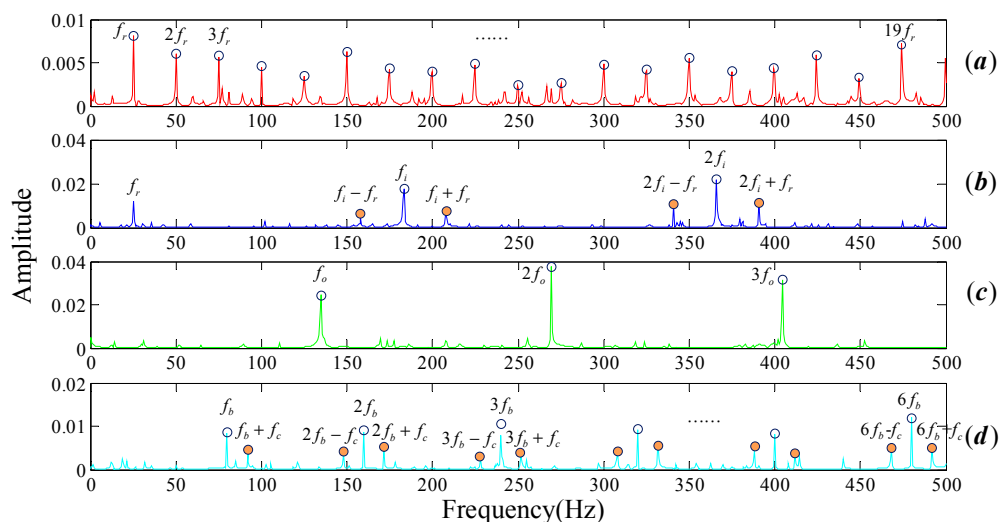
The FT spectrum analysis of the bearing noisy signal is shown in Figure 4. There are a large number of noisy spectral components, as shown by the blue dotted boxes in Figure 4a–d, which severely affect the feature identification of the fault. Rotational frequency and its resonance frequencies (as marked in blue circles in Figure 3a) should be easy to recognize in the spectrum domain when the bearing is under healthy conditions. However, numbers of noise frequencies revolve around the center frequencies, as marked in the brown dotted box in Figure 3a, resulting in the fact that we cannot interpret the signal properly. The same situation appears in the bearing inner race fault signal spectrum, as shown in Figure 3c. In general, the three common types of faults on the inner race, the outer race and the rolling elements can be modeled as  $f_m = f_r, f_m = 0, f_m = f_c$ , respectively. That is to say, rotational

frequency, 0, cage frequency should modulate their center frequency, respectively. In Figure 3b,d, the center frequency belongs to the inner race and balls can be identified, but it is difficult to identify their respective modulation frequency as they are submerged by the noise components.



**Figure 4.** Frequency spectrum of rolling bear simulation signals with noisy (a) normal bearing; (b) inner fault; (c) outer fault; (d) cage fault.

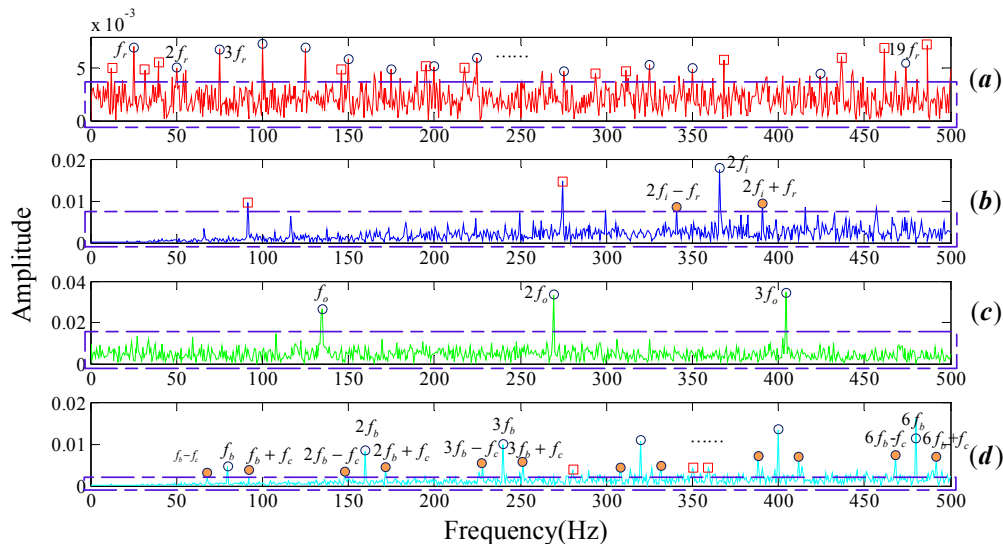
In this paper, the parameters  $\lambda$ ,  $a$  are selected as 1.2, 0.5 respectively in the process of sparse optimization using non-convex optimization regularization. The MPE threshold is selected as 0.7 (refer to the conclusion section). We recover the signals by DMD modes whose MPE values are less than the threshold. Spectrums were obtained by performing FT on the reconstruction signals, as shown in Figure 5.



**Figure 5.** Frequency spectrum of the noisy signals provided by the proposed algorithm. (a) Normal bearing; (b) inner fault; (c) outer fault; (d) cage fault.

Meanwhile, in order to illustrate the effectiveness of our method, multi-scale wavelet transform is applied on the four noisy signals by taking the wavelet function db1, WT results are shown in Figure 6. Though it is obvious that the harmonic frequencies of  $f_r$  and  $f_o$  can be identified in Figure 6a,c, there are still some frequencies attributed by the noise involving around the mid-frequencies, as can we see in

the brown dotted box. Meanwhile the spectra of the proposed algorithm are cleaner and more readable than the results of FT and WT. Not only can we pick out their center and modulation frequencies in Figure 5, but also there is few noise frequencies surrounding them. Our method has tremendous advantages in both de-noising and fault feature extraction compared with FT and WT.



**Figure 6.** Frequency spectrum of the noisy signals provided by WT. (a) Normal bearing; (b) inner fault; (c) outer fault; (d) cage fault.

#### 4. Algorithm Application in Experimental Signals of Rolling Bear

Signals of bearing systems collected by sensors from practice spots are much more complex than the simulation signals, which contain multi-component signals and background noises transmitted by other components. In order to verify the effectiveness of the proposed algorithm in signal de-noising and fault diagnosis, we apply the proposed algorithm to the actual bearing signals. We test the bearing signals from two experimental environments. First, the bearing signal is from the published bearing test data at the university of Cincinnati [41]; the second comes from our rotating machinery vibration analysis test rig.

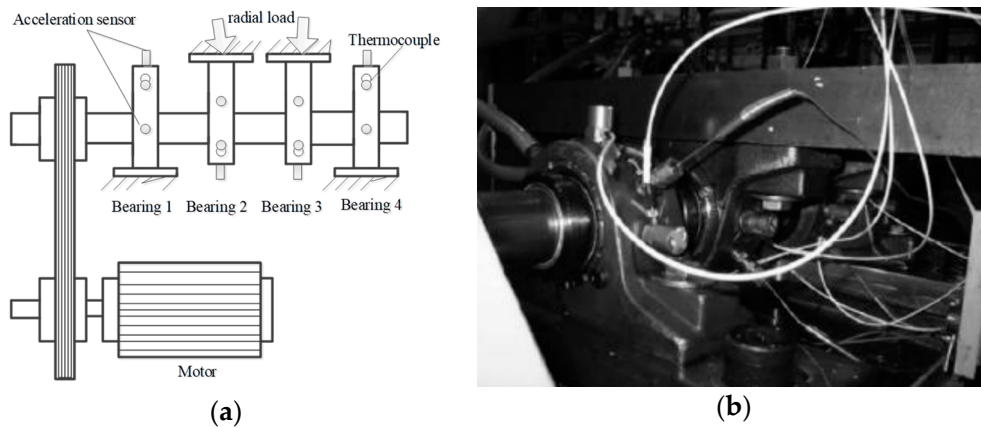
The Intelligent Maintenance Systems of Cincinnati is shown in Figure 7. Four Rexnord ZA-2115 double row bearings were installed on a shaft, which was driven by an AC motor at the speed of 2000 RPM via sever rub belts. Each row of the bearing has  $Z = 16$  roller elements. The circle diameter of the bearing bitch and roller elements are respectively  $D = 28.15$  mm,  $d = 3.31$  mm, and the bearing's contact angle is  $\alpha = 15.17^\circ$ . They utilized high sensitivity quartz ICP accelerometers (PCB 353B33), installed on the bearing housing, collecting the vibration signals and NI DAQ Card 6062E collecting the signals.

We take their experimental data of fourth channel in Set No. 3 to verify our method. The total number of the time series data is 20,480, and the sampling frequency is 20 KHz. Finally, the bearing failed because of outer race fault. The rotation frequency and outer ring fault frequency can be respectively calculated as  $f_r = 33.33$  Hz,  $f_o = 236.17$  Hz by formulas [42].

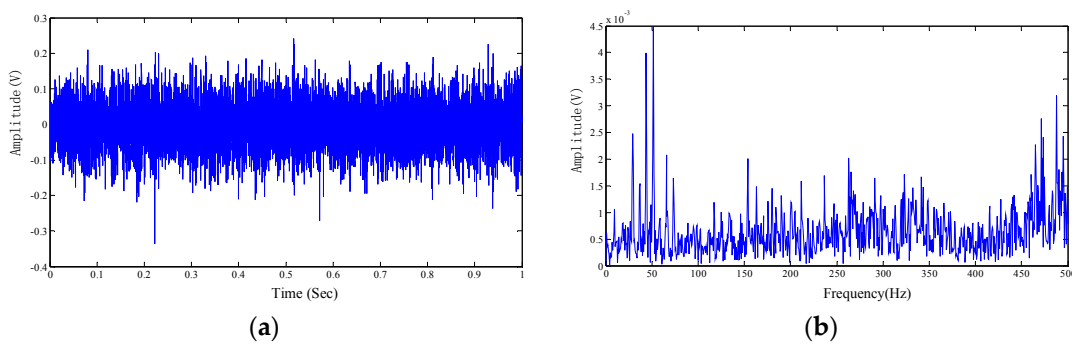
The time domain and the frequency domain of the selected signals are shown in Figure 8. Obviously, it is hard to identify whether there is a bearing fault from Figure 8b for such large amounts of interference frequency presented in the frequency spectrum.

Then we apply our algorithm on the chosen signal with chosen parameters  $\lambda = 1.2$ ,  $a = 0.5$ ,  $s = 5$ ,  $d = 6$ ,  $\tau = 1$ . The truncation rank is calculated as 484 by applying non-convex optimization regularization on the rank function. The same number of single-frequency DMD modes are obtained correspondingly. The MPE values corresponding to each DMD modes are shown

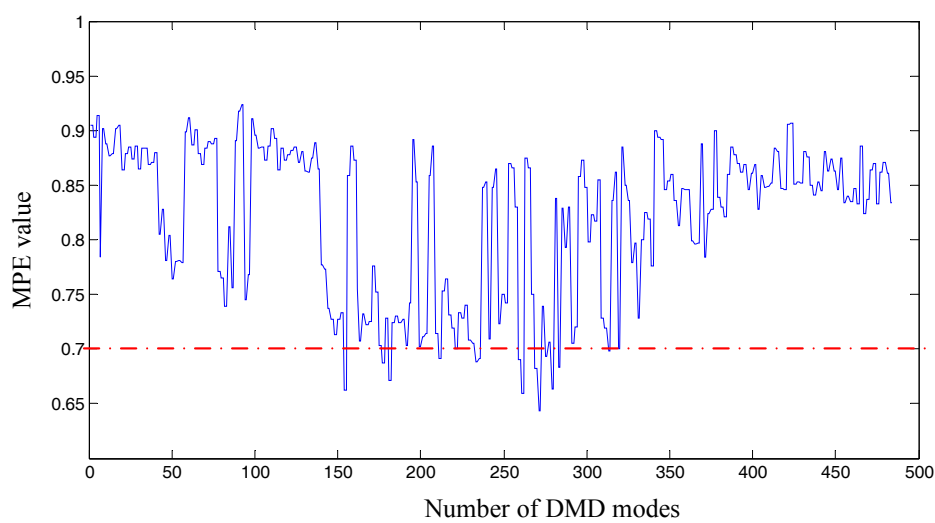
in Figure 9. We recover the wanted signals by DMD modes whose MPE values are less than 0.7. Spectrums are obtained by performing FT on the reconstruction signal, as shown in Figure 10.



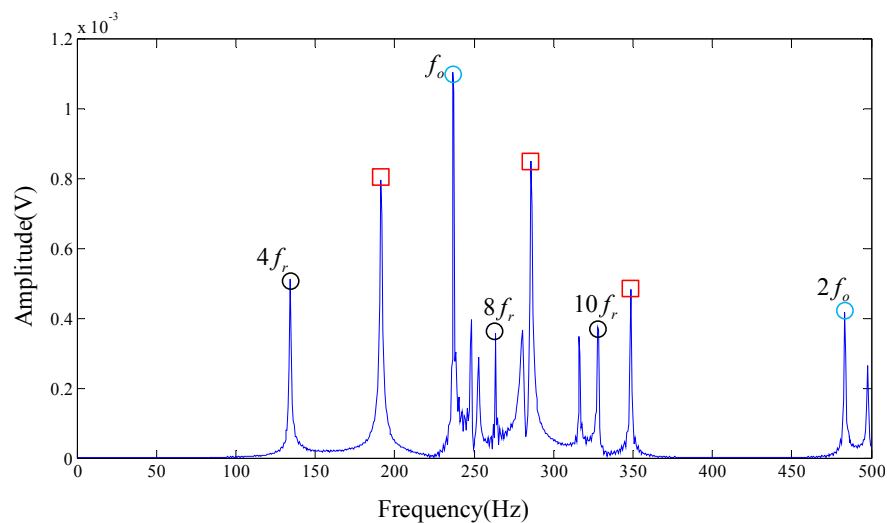
**Figure 7.** The Intelligent Maintenance Systems of Cincinnati: (a) the schematic diagram of the bearing test rig; (b) spot photo of the bearing test rig.



**Figure 8.** The outer race fault signal of the Intelligent Maintenance Systems of Cincinnati in time domain and frequency: (a) Time domain; and (b) Frequency domain of FT.



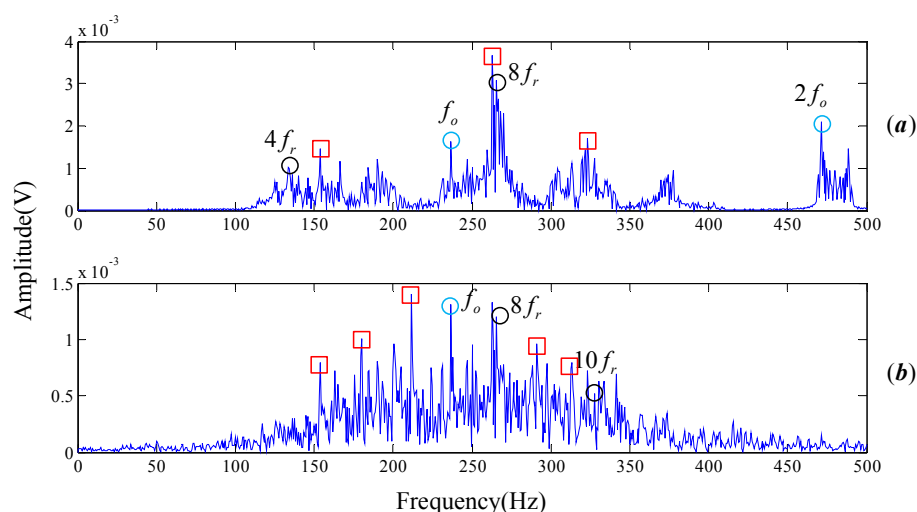
**Figure 9.** MPE values of DMD modes, the original signal is from Cincinnati.



**Figure 10.** The result provided by our algorithm, the original signal is from Cincinnati.

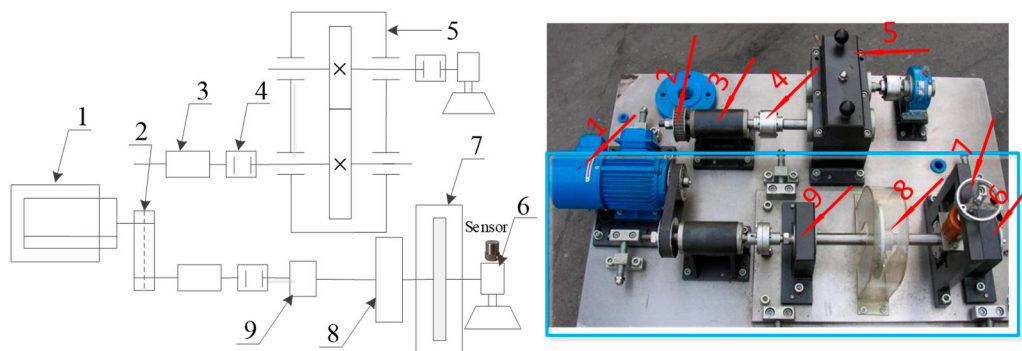
As shown in Figure 10 there are a few characteristic frequencies located in the frequency spectrogram provided by our algorithm. Outer race failure frequency  $f_o$  and its resonant frequency  $2f_o$  can be easily captured. Also, some resonant frequencies of the rotation frequency  $f_r$  are presented in the spectrum. However, regrettably, some of the frequencies we cannot account for are still in Figure 10, which are shown in the red box, revealing that the signal of the actual bearing system is complex. Yet we still got the failure frequency of the bearing test signal, which is in line with experimental results.

Wavelet packet decomposition and EEMD are applied on the experimental signal, as shown in Figure 11. Figure 11a is the spectrum of the recover signal transformed by wavelet packet, with db15 and the default threshold. The optimal amplitude of the noise is set as 0.2 and the ensemble size is 100 for the EEMD. Figure 11b shows the spectrum of IMF3 which has the largest correlation with the original signal. A large amount of unexplainable interference frequencies, visible in the spectrogram, seriously affect the recognition of the rotation frequency and outer ring fault frequency. Thus, the optimized DMD outperforms wavelet packet decomposition and EEMD in the mode decomposition of rolling bearing experimental signal.

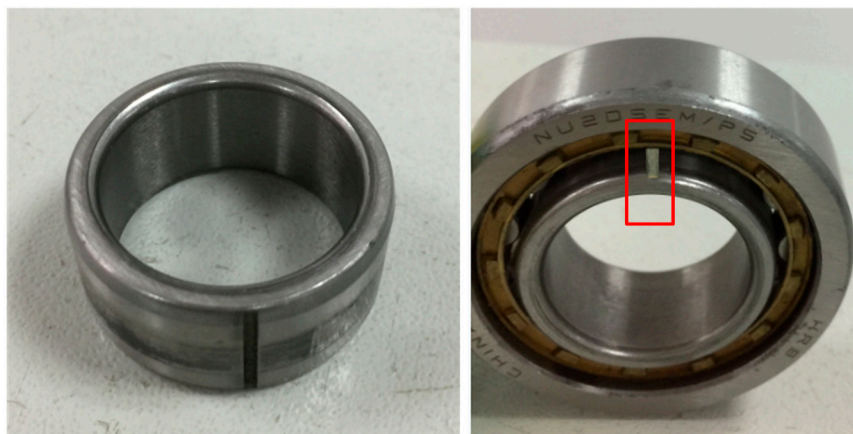


**Figure 11.** Frequency spectrogram provided by wavelet packet decomposition and EEMD. (a) Results of wavelet packet decomposition; (b) Frequency spectrogram of IMF3 decomposed by EEMD.

Next, we apply the algorithm on the experimental signal measured by our rotating machinery vibration analysis test rig. The physical structure of the experimental rig is shown in Figure 12. The whole device consists of a variable speed drive motor, shaft, gear box, deflection wheel, bearing and governor. A picture of the sensor in Figure 12 refers to the position where the single channel vibration acceleration sensor (PCB-352C33, NY, USA) is installed. A cylinder roller bearing NU205 with artificial defects on the outside surface of the inner race was mounted in the housing. The fault bearing physical diagram is shown in Figure 13. The number of rolling elements is  $Z = 13$ , and the cylinder roller diameter is  $d = 7.5$  mm, pitch diameter of the testing bearing is  $D = 39$  mm. We set the motor at the speed of 980 RPM. So, rotation frequency and outer ring fault frequency can be calculated as  $f_r = 16.3$  Hz,  $f_i = 126.3$  Hz, respectively [42].



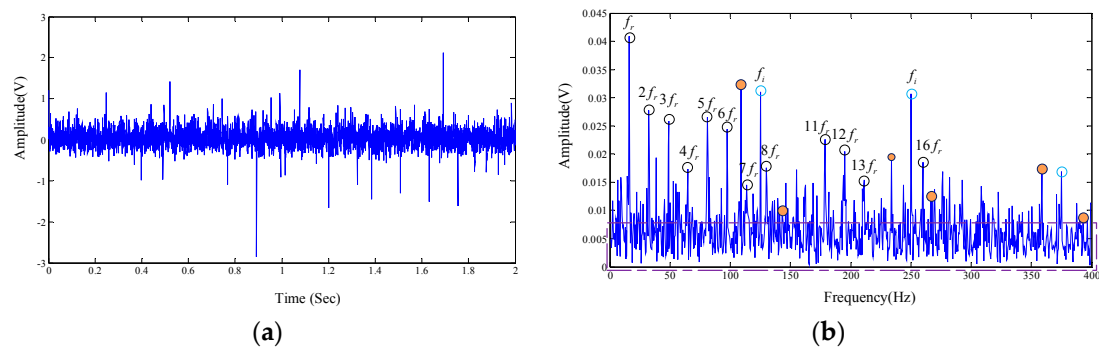
**Figure 12.** Rotating machinery vibration analysis test rig. (a) Structure diagram of test device; (b) physical object of the device; 1—motor, 2—belt pulley, 3—drive shaft support, 4—coupler, 5—gear box, 6—bearing pedestal, with replaceable rolling bearing NU205, 7—belt pulley, 8—radial force device, 9—bearing pedestal, with rolling bearing not replaced.



**Figure 13.** Physical diagram of the fault bearing.

A total number of 4000 sample points, shown in Figure 14a in time domain, are taken into analysis with the proposed algorithm. The frequency domain is shown in Figure 14b.

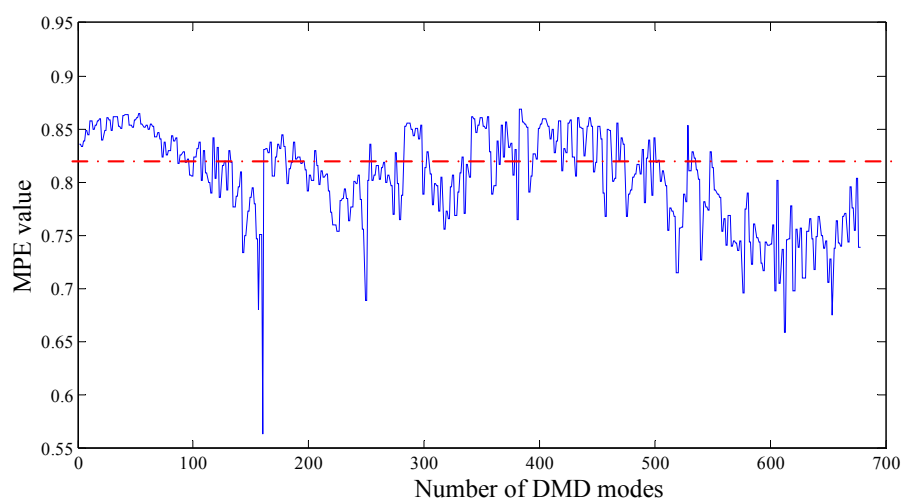
As is shown in Figure 14b, rotational frequency and its resonant frequencies are presented in the spectrum diagram. The center frequency of the inner race can also be easily found out, but the modulation components are not obvious. Meanwhile, there are large amounts of interference frequency presented in the frequency spectral. The spectrum of the test signal is not intuitionistic, which makes it difficult to diagnose the fault in view of the technicians.



**Figure 14.** The measured signal in time domain and frequency, the original signal is from our test rig. (a) Time domain; and (b) Frequency domain of FT.

Subsequently, the proposed algorithm is applied to the tested signal with the same parameters  $\lambda = 1.2$ ,  $a = 0.5$ ,  $s = 5$ ,  $d = 6$ ,  $\tau = 1$ . 682 single-frequency DMD modes are obtained by sparse optimization algorithm. Figure 15 shows the MPE values of each DMD modes. DMD modes whose MPE values are less than 0.82 are reconstructed to form a recovery signal. Spectrums are obtained by performing FT on the reconstruction signal, as shown in Figure 16.

Wavelet packet decomposition and EEMD are also employed on our experimental signal, as shown in Figure 17. The methods and parameters used in the analysis are consistent with the signal processing procedure of Cincinnati. Here the spectrum of IMF1 decomposed by EEMD is selected because it has the largest correlation with the original signal, as shown in Figure 17b. Obviously, Figure 16 is more readable than Figures 13b and 17. The rotational frequency  $f_r$  and its harmonics ( $2f_r, 3f_r, 4f_r, 5f_r, \dots$ ) are easily identified. Not only could the center frequency be pointed out clearly but also the modulation frequency of the inner face. That is to say, we are more confident to judge that the inner race of the testing rolling bearing has encountered failure, which is highly consistent with the actual situation. The basic reason the frequencies are easy to read is that there are a few interference frequencies nearby them. Therefore, the proposed algorithm of noise reduction and feature extraction is proved to be practical and to outperform other mode decompositions such as wavelet packet decomposition and EEMD.



**Figure 15.** MPE values of DMD modes, the original signal is from our test rig.

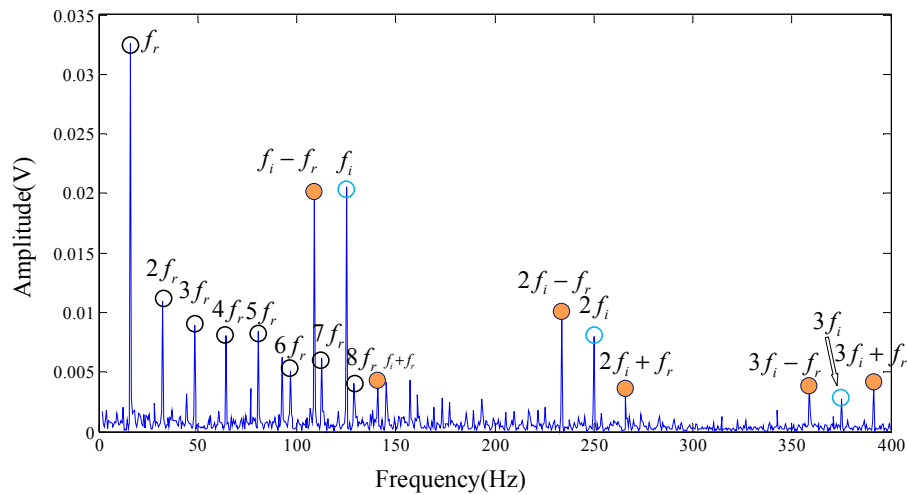


Figure 16. The result provided by the proposed algorithm, the original signal is from our test rig.

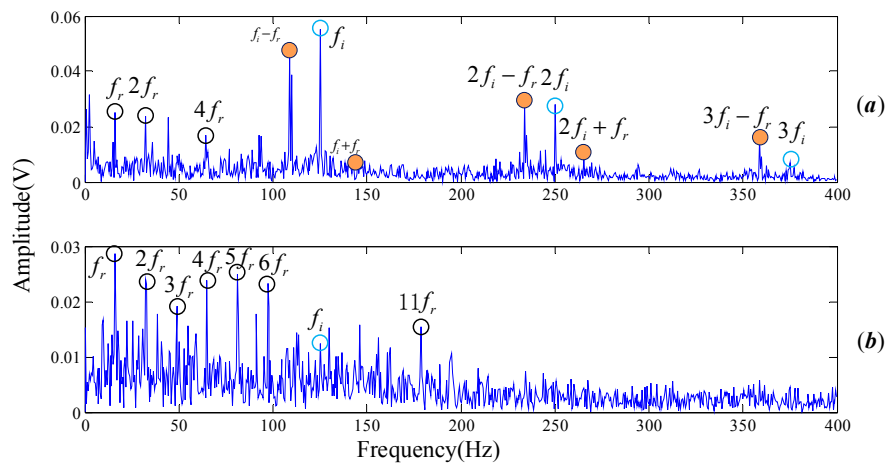


Figure 17. Frequency spectrogram provided by wavelet packet decomposition and EEMD. (a) Results of wavelet packet decomposition; (b) Frequency spectrogram of IMF1 decomposed by EEMD.

### 5. Conclusions

As a powerful tool for Koopman spectrum analysis technology with the characteristic of equation-free and data driven, DMD decomposes the time series into a series of single-frequency modes without any assumption. Theoretically, dynamic features of the original system including fault features and noise components exist in these DMD modes only if the truncation rank order  $r$  is appropriate. The truncated rank order needs to be preset artificially in the standard DMD algorithm, which is a crucial problem. Another puzzle with the application of standard DMD is of how to define a selection strategy to filter the undesired DMD modes in the process of signal reconstruction.

By proposing solutions for the selection strategy of truncated rank order and DMD modes, this paper puts forward a novel denoising and feature extraction algorithm for multi-component coupled noisy mechanical signals. Non-convex penalty regularization is adopted to the rank function of the shift-stack matrix, and a series of optimal DMD modes are obtained by this sparse optimization method. MPE is performed as a tool of optimal mode selection. After calculating the MPE of each DMD mode, signal is reconstructed by the modes whose entropies are smaller than a threshold. The proposed algorithm is successfully applied in rolling bearing simulation and experimental signals, demonstrating that it has a good application prospect in noise reduction and fault feature extraction.



DMD algorithm has been widely used in the field of fluid mechanics and video processing. This paper presents an optimized DMD algorithm and applies it to mechanical vibration signal. The parameters of non-convex penalty regularization and MPE in our algorithm are proved to be appropriate by repeated verification. The non-convex optimization method is used to extract the dynamic characteristics of the original system more comprehensively than selecting truncation rank with a “elbow” of the singular value. The dynamic characteristics, fault features and noise components of the original signal can be identified by using MPE method. However, for filtering the noise components, entropy threshold needs to be set by multiple attempts in our algorithm. For example, when dealing with the experimental signal of Cincinnati, we firstly set the threshold values of 0.7, 0.75, 0.8, etc., and then observed the effect of the recovery signal through the spectrum diagram respectively. We found that when the threshold value was 0.7, the fault characteristics were more obvious, and the interference frequencies were less. Therefore, 0.7 was selected as the criterion. Research on different vibration signals, various transmission systems and different degrees of fault damage to select a certain MPE threshold is our future work.

**Acknowledgments:** This work was supported by the National Natural Science Foundation of China under Grants No. 51475339, the Natural Science Foundation of Hubei province under Grant No. 2016CFA042. The Applied Basic Research Programs of Wuhan science and technology bureau under Grants No. 2017010201010115. The authors also appreciate the free download of the original bearing failure data and one photo picture provided by the intelligent maintenance system center of University of Cincinnati.

**Author Contributions:** Zhang dang was in charge of contributing the analysis method and the program codes, analyzing the data, and writing the paper. Yong Lv and Yourong Li were in charge of the supervision of this research and helping in revising. Cancan Yi was responsible for the experimental design and data analysis. All authors have read and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Golafshan, R.; Sanliturk, K.Y. SVD and Hankel matrix based de-noising approach for ball bearing fault detection and its assessment using artificial faults. *Mech. Syst. Signal Process.* **2016**, *70–71*, 36–50. [[CrossRef](#)]
2. Yi, C.; Lv, Y.; Xiao, H.; Huang, T.; You, G. Multisensor signal denoising based on matching synchrosqueezing wavelet transform for mechanical fault condition assessment. *Meas. Sci. Technol.* **2018**. [[CrossRef](#)]
3. AlMahamdy, M.; Riley, H.B. Performance study of different denoising methods for ECG signals. *Procedia Comput. Sci.* **2014**, *37*, 325–332. [[CrossRef](#)]
4. Schlüter, S.; Sheppard, A.; Brown, K.; Wildenschild, D. Image processing of multiphase images obtained via X-ray microtomography: A review. *Water Resour. Res.* **2014**, *50*, 3615–3639. [[CrossRef](#)]
5. Rai, A.; Upadhyay, S.H. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribol. Int.* **2016**, *96*, 289–306. [[CrossRef](#)]
6. Donoho, D.; Johnstone, I. Adapting to Unknown Smoothness via Wavelet Shrinkage. *Publ. Am. Stat. Assoc.* **1995**, *90*, 1200–1224. [[CrossRef](#)]
7. Lv, Y.; Yuan, R.; Song, G. Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing. *Mech. Syst. Signal Process.* **2016**, *81*, 219–234. [[CrossRef](#)]
8. Yang, Z.X.; Zhong, J.H. A hybrid EEMD-based SampEn and SVD for acoustic signal processing and fault diagnosis. *Entropy* **2016**, *18*, 112. [[CrossRef](#)]
9. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [[CrossRef](#)]
10. Wang, X.B.; Yang, Z.X.; Yan, X.A. Novel particle swarm optimization-based variational mode decomposition method for the fault diagnosis of complex rotating machinery. *IEEE/ASME Trans. Mechatron.* **2018**, *23*, 68–79. [[CrossRef](#)]
11. Lv, Y.; He, B.; Yi, C.; Dang, Z. A novel scheme on multi-channel mechanical fault signal diagnosis based on augmented quaternion singular spectrum analysis. *J. Vibroeng.* **2017**, *19*, 955–966.
12. Rowley, C.W.; Mezić, I.; Bagheri, S.; Schlatter, P.; Henningson, D. Spectral analysis of nonlinear flows. *J. Fluid Mech.* **2009**, *641*, 115–127. [[CrossRef](#)]

13. Schmid, P.J. Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **2010**, *656*, 5–28. [[CrossRef](#)]
14. Grosek, J.; Kutz, J.N. Dynamic mode decomposition for real-time background/foreground separation in video. *arXiv* **2014**, arXiv:1404.7592.
15. Tirunagari, S.; Poh, N.; Windridge, D.; Ho, A.T.S. Detection of Face Spoofing Using Visual Dynamics. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 762–777. [[CrossRef](#)]
16. Berger, E.; Sastuba, M.; Vogt, D.; Jung, B. Dynamic Mode Decomposition for perturbation estimation in human robot interaction. In Proceedings of the Ro-Man IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2016; pp. 593–600.
17. Brunton, B.W.; Johnson, L.A.; Ojemann, J.G.; Kutz, J.N. Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *J. Neurosci. Methods* **2015**, *258*, 1–15. [[CrossRef](#)] [[PubMed](#)]
18. Dawson, S.; Hemati, M.; Williams, M.; Rowley, C.W. Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Exp. Fluids* **2016**, *57*, 1–19. [[CrossRef](#)]
19. Hemati, M.S.; Rowley, C.W.; Deem, E.A.; Cattafesta, L.N. De-Biasing the Dynamic Mode Decomposition for Applied Koopman Spectral Analysis. *J. Nonlinear Sci.* **2015**, *25*, 1–40.
20. Brunton, S.L.; Proctor, J.L.; Kutz, J.N. Compressive sampling and dynamic mode decomposition. *arXiv* **2013**, arXiv:1312.5186.
21. Jovanovic, M.; Schmid, P.; Nichols, J.W. Sparsity-promoting Dynamic Mode Decomposition. *Phys. Fluids* **2013**, *26*, 561–571. [[CrossRef](#)]
22. Erichson, N.B.; Steven, L.B.; Kutz, J.N. Randomized Dynamic Mode Decomposition. *arXiv* **2017**, arXiv:1702.02912.
23. Erichson, N.B.; Donovan, C. Randomized low-rank Dynamic Mode Decomposition for motion detection. *Comput. Vis. Image Underst.* **2016**, *146*, 40–50. [[CrossRef](#)]
24. Klus, S.; Gelß, P.; Peitz, S.; Schütte, C. Tensor-based dynamic mode decomposition. *arXiv* **2016**, arXiv:1606.06625.
25. Zhao, X.; Li, W.; Zhou, L.; Song, G.; Ba, Q. Application of support vector machine for pattern classification of active thermometry-based pipeline scour monitoring. *Struct. Control Health Monit.* **2015**, *22*, 903–918. [[CrossRef](#)]
26. Metzler, C.A.; Maleki, A.; Baraniuk, R.G. From Denoising to Compressed Sensing. *IEEE Trans. Inf. Theory* **2016**, *62*, 5117–5144. [[CrossRef](#)]
27. Parekh, A.; Selesnick, I.W. Enhanced Low-Rank Matrix Approximation. *IEEE Signal Process. Lett.* **2015**, *23*, 493–497. [[CrossRef](#)]
28. Selesnick, I.W.; İlker, B. Sparse Signal Estimation by Maximally Sparse Convex Optimization. *IEEE Trans. Signal Process.* **2014**, *62*, 1078–1092. [[CrossRef](#)]
29. Chen, P.Y.; Selesnick, I.W. Translation-invariant shrinkage/thresholding of group sparse signals. *Signal Process.* **2013**, *94*, 476–489. [[CrossRef](#)]
30. Chen, P.Y.; Selesnick, I.W. Group-sparse signal denoising: Nonconvex regularization, convex optimization. *IEEE Trans. Signal Process.* **2014**, *62*, 3464–3478. [[CrossRef](#)]
31. Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [[CrossRef](#)] [[PubMed](#)]
32. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [[CrossRef](#)] [[PubMed](#)]
33. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [[CrossRef](#)] [[PubMed](#)]
34. Zunino, L.; Olivares, F.; Scholkmann, F.; Rosso, O.A. Permutation entropy based time series analysis: Equalities in the input signal can lead to false conclusions. *Phys. Lett. A* **2017**, *381*, 1883–1892. [[CrossRef](#)]
35. Li, Y.; Xu, M.; Wei, Y.; Huang, W. A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. *Measurement* **2016**, *77*, 80–94. [[CrossRef](#)]
36. An, X.; Pan, L. Bearing fault diagnosis of a wind turbine based on variational mode decomposition and permutation entropy. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2017**, *231*, 200–206. [[CrossRef](#)]

37. Wu, S.D.; Wu, P.H.; Wu, C.W.; Ding, J.J.; Wang, C.C. Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy* **2012**, *14*, 1343–1356. [[CrossRef](#)]
38. Riedl, M.; Müller, A.; Wessel, N. Practical considerations of permutation entropy. *Eur. Phys. J. Spec. Top.* **2013**, *222*, 249–262. [[CrossRef](#)]
39. Amigó, J.M.; Zambrano, S.; Sanjuán, M.A. Combinatorial detection of determinism in noisy time series. *Europhys. Lett.* **2008**, *83*, 60005. [[CrossRef](#)]
40. Randall, R.B.; Antoni, J.; Chobsaard, S. The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals. *Mech. Syst. Signal Process.* **2001**, *15*, 945–962. [[CrossRef](#)]
41. Intelligent Systems Division. Available online: <https://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/#bearing> (accessed on 27 August 2017).
42. Han, T.; Liu, X.; Tan, A. Fault Diagnosis of Rolling Element Bearings Based on Multiscale Dynamic Time Warping. *Measurement* **2017**, *95*, 355–366. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).