INTELLIGENT FAULT IDENTIFICATION OF PLANET BEARINGS USING DISCRIMINATIVE DICTIONARY LEARNING BASED SPARSE REPRESENTATION CLASSIFICATION FRAMEWORK

Yun Kong*, Tianyang Wang, Fulei Chu
Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China
kongy15@mails.tsinghua.edu.cn

Abstract: Planet bearing fault identification is an attractive but challenging task in numerous engineering applications, such as wind turbine and helicopter transmission systems. However, traditional fault characteristic frequency identification and impulsive feature extraction based diagnosis strategies are not sufficient to resolve the problem of planet bearing fault detection, due to complex physical configurations and modulation characteristics in planetary gearboxes. In this paper, a novel discriminative dictionary learning based sparse representation classification (SRC) framework is proposed for intelligent planet bearing fault identification. Within our approach, the optimization objective for discriminative dictionary learning introduces a label consistent constraint called 'discriminative sparse code error' and incorporates it with the reconstruction error and classification error to bridge the gap between the classical dictionary learning and classifier training. Therefore, not only the reconstructive and discriminative dictionary for signal sparse representation but also an optimal universal multiclass classifier for classification tasks could be simultaneously learnt in the proposed framework. The optimization formulation could be efficiently solved using the well-known K-SVD dictionary learning algorithm. The effectiveness of the proposed framework has been validated using experimental planet bearing vibration signals. Comparative results demonstrate that our framework outperforms the state-of-the-art K-SVD based SRC method in terms of classification accuracy for intelligent planet bearing fault identification.

Key words: Planet bearing; Fault identification; Discriminative dictionary learning; Sparse representation classification; K-SVD; Orthogonal matching pursuit.

1. INTRODUCTION

Planetary gearbox possesses attractive advantages such as large transmission ratio and excellent load-bearing capacity in a relatively compact structure, which make it widely applied in many engineering applications such as wind turbine and helicopter transmissions. Failures of planetary gearbox will not only reduce the reliability of engineering system but also lead to great operation and maintenance (O&M) costs. Vibration-based fault detection technique has been proven as one of the most effective techniques for condition monitoring and O&M costs saving for planetary gearbox [1]. However, vibration signals of planetary gearbox are more complicated than the parallel gearbox due to the complex physical configuration and kinematic mechanism. As illustrated in Figure 1(a), planetary gearbox comprises four main components, including ring gear, sun gear, planet carrier and multiple planet gears. Concerning the kinematic mechanism of planetary gearbox, sun gear rotates around its fixed axis while multiple planet gears not only rotate around their own centers but also revolve around the sun gear, as shown in Figure 1(b). As a result, the vibration transmission paths between meshing points of ring-planet or sun-planet gear pairs and the fixed sensor are periodically varying due to the revolving nature of planet gears [2]. Therefore, vibration-based fault diagnosis of planetary gearbox is challenging due to the complex physical configurations and revolving planet gears inducing modulation characteristics [3].

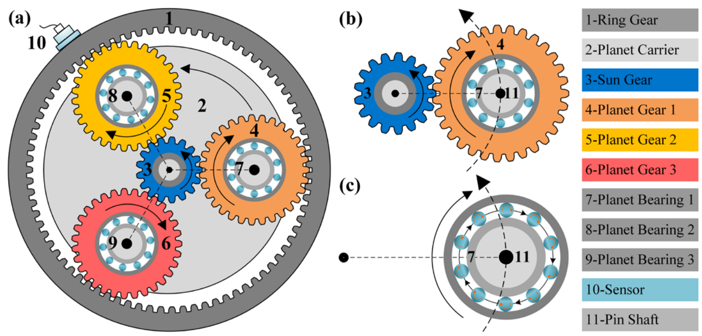

Figure 1: Physical Configurations of Planetary Gearbox and Kinematic Mechanism of Planet Bearings.

To address above challenges, many signal processing approaches have been developed for planetary gearbox fault diagnosis, such as statistical analysis [4], spectral kurtosis [5], time synchronous averaging [6], wavelet transform [7-8], demodulation analysis [9-10] and time frequency representation method [11]. Within these approaches, periodicity or time-varying characteristic frequency of fault-related features are analyzed to isolate the fault location, because the localized defect during meshing process will produce local anomalies repetitively and each gear has fault characteristic frequency proportional to the rotational frequency. To a certain extent, these contributions have successfully provided feasible means to address planetary gearbox fault detection. However, these reported literatures have been restricted to planetary gear fault detection and rarely devoted to identifying the planet bearing fault.

Planet bearing, as the most intricate component in term of kinematic mechanism, is also the most challenging component for fault identification in planetary gearbox. As illustrated in Figure1(c), the outer ring of planet bearing, which is fixed to the planet gear, not only rotates around its own center but also revolves around the sun gear. On the other hand, the inner ring stays relatively stationary to the planet carrier and only revolves around the sun gear. Therefore, vibration signature of planet bearings is complicated and significantly different from the common fixed-axis bearings due to the complex and time-varying vibration transmission paths. To understand the dynamic behavior of planetary gearbox containing localized planet bearing defects, Jain et al. first developed an analytical model to predict vibration signatures of the faulty planetary gearbox with localized planet bearing defects [12] and further predicted the fault signatures of planet bearing defects by identifying different sources of modulation sidebands [13]. Inspired by these thoughts, Feng et al. [14] developed analytical vibration signal models for defective planet bearings. Moreover, amplitude and frequency demodulation analyses were applied to identify the fault characteristic frequencies in demodulation spectrums of defective planet bearing vibration signals [15]. To locate the resonance frequency bands induced by planet bearing defects, Wang et al. proposed meshing frequency modulation index-based kurtogram [16] for filter design to detect planet bearing defects by identifying fault characteristic frequencies in envelope spectrums. These contributions have provided deep insights into understanding the vibration behavior of planetary gearbox in presence of defective planet bearings.

However, these reported literatures are only to reveal the frequency features for faulty planetary gearbox with planet bearing defects. These frequency features are too weak to identify due to the following reasons. First, the performance of the frequency feature based fault detection algorithms deteriorates easily due to planet bearing manufacturing error, misalignment and low- torque operating condition [17]. Second, vibration signatures of planet bearing are overwhelmed by gear meshing vibrations and other unwanted background noises [8, 16]. Third, the planet bearing defect-induced resonance frequency bands are very hard to determine for fault feature extraction due to heavy interferences from planetary gear meshing vibrations. Thus, these issues greatly challenge the fault characteristic frequency identification based diagnosis strategies.

Alternatively, classification-based intelligent identification approaches, which are free of detecting the weak frequency features induced by planet bearing defects, are more attractive and promising for planet bearing fault identification. Recently, dictionary learning based sparse representation classification (SRC) techniques have attracted extensive attentions in a wide range of academic communities such as image classification [18], face recognition [19] and computer vision [20]. The classical dictionary learning framework attempts to adaptively learn an overcomplete dictionary $\boldsymbol{D}\in\boldsymbol{R}^{n\times K}$ containing K signal-atoms of columns $\{\mathbf{d}_j\}^K_{j=1}$ so that the input signal $\boldsymbol{Y}\in\boldsymbol{R}^{n\times N}$ could be well approximated using sparse linear combinations of these atoms $\boldsymbol{Y}\approx\boldsymbol{DX}$ in terms of minimal reconstruction error $\|\boldsymbol{Y}-\boldsymbol{DX}\|_F^2$ [21]. Inspired by the merits of dictionary learning, the dictionary learning based SRC approaches have been developed for image classification [22-23] and fault identification [24]. Within the classical dictionary learning based SRC approach in [22], the main idea is to learn sub-dictionaries first for sparse representation of training signals from each category independently and then achieve the recognition tasks based on the corresponding reconstruction errors. Inspired by this idea, Zhao et al. [24] proposed the sparse representation classification for planet bearing fault identification, which achieved the recognition according to the minimal reconstruction error of testing signals with respect to these learned sub-dictionaries directly. However, these classical dictionary learning based SRC approaches only exploit the reconstruction power (best sparse representation of training signals for minimal reconstruction error) but neglect the discriminative power of the learned dictionary for recognition tasks. In contrast, a discriminative dictionary was achieved for pattern recognition by iteratively updating dictionary atoms based on the results of a linear predictive classifier [23]. In addition, the approach in [25] first learned a representative dictionary and then implemented the classifier training for recognition tasks. Nevertheless, these dictionary learning based SRC approaches treat the dictionary learning and classifier training as two separate processes, which might make the learned dictionary suboptimal for classification tasks. Therefore, learning a discriminative dictionary by identifying the discriminability of sparse codes and training the classifier model simultaneously in a supervised manner, could be promising to achieve better classification performance, which stimulated the idea of the discriminative dictionary learning based sparse representation classification framework in this paper.

In summary, to avoid the dilemmas of the frequency feature based diagnosis strategies for planet bearing fault identification and address the challenges for classical dictionary learning based SRC approaches, we propose a novel discriminative dictionary learning based sparse representation classification (DDL-SRC) framework for intelligent fault identification of planet bearings. The key features and contributions of this paper are presented as follows:

- To exploit class labels of training signals for supervised dictionary learning, the label information are associated with each dictionary atom to enforce the discriminability of sparse codes during dictionary learning process. Mathematically, we introduce a discriminative sparse code error into the optimization objective in the DDL-SRC framework.
- In addition to the reconstruction error for representative dictionary purpose, the discriminative sparse code error for discriminative dictionary purpose and the classification error for classifier training are jointly incorporated into the mixed optimization objective. Therefore, the proposed DDL-SRC framework could learn not only a discriminative dictionary but also an optimal multiclass classifier simultaneously to achieve better classification performance.
- The optimization problem in the proposed DDL-SRC framework could be efficiently solved with complexity bounded by the K-means Singular value decomposition (K-SVD) algorithm.
- In contrast to other existing pattern recognition methods, the proposed DDL-SRC framework is free of feature design and selection, which achieves the intelligent classification tasks via discriminative dictionary learning and classifier training directly from raw training signals.
- The DDL-SRC framework has been extended to achieve the intelligent fault identification of planet bearings and outperforms the state-of-the-art SRC method.

The rest of this paper is organized as follows: Section 2 presents the proposed DDL-SRC framework for intelligent fault identification, which involves discriminative dictionary learning, classification approach and intelligent fault identification for mechanical components. Section 3 validates the effectiveness of the DDL-SRC framework for intelligent fault identification of planet bearings using experimental data. Finally, conclusions are summarized in Section 4.

2. Discriminative dictionary learning based sparse representation classification framework for intelligent fault identification

In order to enhance the discriminative power of dictionary learning for classification tasks, learning a discriminative dictionary along with an optimal multiclass classifier simultaneously and leveraging the discriminability of sparse codes is preferred. To this end, the discriminative dictionary learning based sparse representation (DDL-SRC) framework is proposed for intelligent fault identification, which involves three procedures:

discriminative dictionary learning, classification approach and intelligent fault identification for mechanical components.

2.1 Discriminative Dictionary Learning

The discriminative dictionary learning aims to learn a reconstructive and discriminative dictionary and an optimal linear classifier simultaneously, which could bridge the gap between classical dictionary learning and classifier training in traditional SRC methods.

2.1.1 Optimization objective considering discriminative sparse codes and classifier training

In the discriminative dictionary learning, we aim to unify the dictionary learning and classifier training processes into one mixed optimization objective. In addition, the performance of linear classifier depends on the discriminability of sparse codes. As such, we prefer the supervised learning manner to leverage the label information of training signals for learning a reconstructive and discriminative dictionary. To this end, dictionary atoms should be chosen so that they represent a subset of the training signals ideally from one single class, and hence each dictionary atom will be associated with a particular label in our approach (see Figure2). Ideally, the discriminative sparse codes will be obtained over a discriminative dictionary, if we establish the explicit correspondences between dictionary atoms and the label information in a supervised fashion. The sparse representation using ideal discriminative sparse codes for optimal classification [27] is illustrated in Figure2. The ideal sparse codes $\boldsymbol{Q}$ are considered discriminative and optimal for classification tasks if $\boldsymbol{Q} = [q_1, q_2,\ldots,q_N] \in \boldsymbol{R}^{K\times N}$ where q_i is of the form of $[0,...,1, 1, 1,..,0]^{\mathrm{T}} \in \boldsymbol{R}^K$. Taking the case in Figure2 as an example, training signal matrix $\boldsymbol{Y} = [\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{Y}_3]$ contains training signals from three classes, where $\boldsymbol{Y}_1$ contains three samples y_1, y_2, y_3, $\boldsymbol{Y}_2$ contains four samples y_4, y_5, y_6, y_7, and $\boldsymbol{Y}_3$ contains two samples y_8, y_9. Dictionary $\boldsymbol{D}$ contains three sub-dictionaries for each class and each sub-dictionary $\boldsymbol{D}_l$ has three atoms. In this case, the resulting discriminative sparse codes are illustrated in Figure 2.

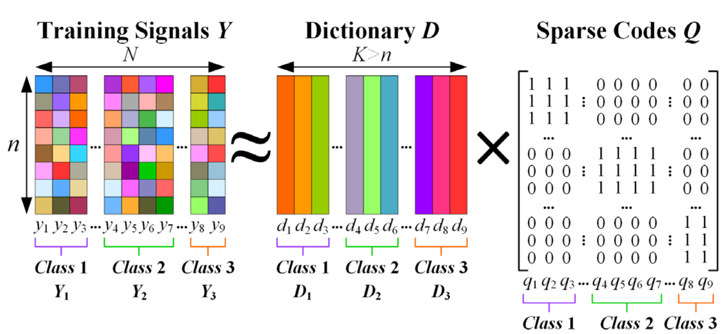

Figure 2: Sparse Representation Using Ideal Discriminative Sparse Codes for Optimal Classification.

Suppose a set of N n-dimension input signals $\boldsymbol{Y} = [y_1, y_2, \ldots, y_N] \in \boldsymbol{R}^{n\times N}$, the optimization objective for discriminative dictionary learning could be defined as follows:

$$\left\langle \boldsymbol{D},\boldsymbol{W},\boldsymbol{A},\boldsymbol{X}\right\rangle = \arg\min_{\boldsymbol{D},\boldsymbol{W},\boldsymbol{A},\boldsymbol{X}} \left\|\boldsymbol{Y}-\boldsymbol{D}\boldsymbol{X}\right\|_F^2 + \alpha\left\|\boldsymbol{Q}-\boldsymbol{A}\boldsymbol{X}\right\|_F^2 + \beta\left\|\boldsymbol{H}-\boldsymbol{W}\boldsymbol{X}\right\|_F^2 \quad s.t.\ \forall i, \left\|x_i\right\|_0 \le T , \tag{1}$$

where $\boldsymbol{X} = [x_1, x_2, \ldots, x_N] \in \boldsymbol{R}^{K\times N}$ are the sparse codes of the input signals $\boldsymbol{Y}$ with respect to dictionary $\boldsymbol{D} = [d_1, d_2, \ldots, d_K] \in \boldsymbol{R}^{n\times K}$. $\boldsymbol{A} \in \boldsymbol{R}^{K\times K}$ represents a linear transformation, which is designed to transform the obtained sparse codes $\boldsymbol{X}$ to be most discriminative in sparse feature space. $\boldsymbol{W} \in \boldsymbol{R}^{L\times K}$ denotes the linear classifier model parameters and $\boldsymbol{H} = [h_1, h_2, \ldots, h_N] \in \boldsymbol{R}^{L\times N}$ associates the label h_i with each training signal y_i. The first term in Equation (1) represents the reconstruction error. The second term in Equation (1) represents the discriminative sparse codes error, which enforces the sparse codes $\boldsymbol{X}$ to approximate the ideal discriminative sparse codes $\boldsymbol{Q}$. The third term in Equation (1) represents the classification error, which encourages achieving a relatively optimal classifier model parameters $\boldsymbol{W}$ for recognition tasks. Regularization parameters α and β control the relative contributions of the corresponding terms, respectively.

As a result, the above optimization objective for discriminative dictionary learning not only considers the reconstruction error, but also incorporates the discriminative sparse codes error and the classification error terms by leveraging the supervised information (label information $\boldsymbol{H}$ and discriminative sparse codes $\boldsymbol{Q}$). Therefore, this optimization objective is promising to learn a discriminative dictionary $\boldsymbol{D}$ and an optimal linear classifier jointly for classification tasks.

2.1.2 Numerical optimization algorithm

In this subsection, we will demonstrate that the proposed discriminative dictionary learning could be cast as a standard dictionary learning problem and solved using the well-known K-SVD algorithm. The proposed optimization objective in Equation (1) could be rewritten as follows:

$$\left\langle \boldsymbol{D},\boldsymbol{W},\boldsymbol{A},\boldsymbol{X}\right\rangle = \arg\min_{\boldsymbol{D},\boldsymbol{W},\boldsymbol{A},\boldsymbol{X}} \left\| \begin{pmatrix} \boldsymbol{Y} \\ \sqrt{\alpha}\boldsymbol{Q} \\ \sqrt{\beta}\boldsymbol{H} \end{pmatrix} - \begin{pmatrix} \boldsymbol{D} \\ \sqrt{\alpha}\boldsymbol{A} \\ \sqrt{\beta}\boldsymbol{W} \end{pmatrix} \boldsymbol{X} \right\|_F^2 ,\ s.t.\ \forall i, \left\|x_i\right\|_0 \le T , \tag{2}$$

Through defining the generalized training signal matrix $\boldsymbol{Y}_{\text{new}} = (\boldsymbol{Y}^T, \sqrt{\alpha}\boldsymbol{Q}^T, \sqrt{\beta}\boldsymbol{H}^T)^T$ and dictionary matrix $\boldsymbol{D}_{\text{new}} = (\boldsymbol{D}^T, \sqrt{\alpha}\boldsymbol{A}^T, \sqrt{\beta}\boldsymbol{W}^T)^T$, the optimization objective can be reformulated as follows,

$$\left\langle \boldsymbol{D}_{\text{new}}, \boldsymbol{X}\right\rangle = \arg\min_{\boldsymbol{D}_{\text{new}}, \boldsymbol{X}} \left\|\boldsymbol{Y}_{\text{new}} - \boldsymbol{D}_{\text{new}}\boldsymbol{X}\right\|_F^2 ,\ s.t.\ \forall i, \left\|x_i\right\|_0 \le T , \tag{3}$$

Consequently, this equivalent optimization is exactly the standard dictionary learning problem and can be efficiently solved by K-SVD algorithm.

Following the well-known K-SVD algorithm, we implement sparse coding and dictionary updating procedures alternately. In our approach, we utilize the OMP algorithm to accomplish the sparse coding. Within the dictionary updating procedure, we update one dictionary atom d_k^{new} and the associated nonzero sparse codes at a time for best reducing the optimization objective using SVD [18]. Let d_k^{new} be the k-th atom in dictionary $\boldsymbol{D}_{\text{new}}$ and x_R^k denotes the k-th row in sparse codes $\boldsymbol{X}$. In addition, we define $\boldsymbol{E}_k = (\boldsymbol{Y}_{\text{new}} - \sum_{j \neq k} d_j^{\text{new}} x_R^j)$, $\tilde{x}_R^k$ denotes the results of discarding the zero entries in x_R^k and $\tilde{\boldsymbol{E}}_k$ denotes the resulting $\boldsymbol{E}_k$ by discarding the corresponding columns. Hence, dictionary updating can be achieved by solving the following problem,

$$\left\langle d_k^{\text{new}}, \tilde{x}_R^k \right\rangle = \arg \min_{d_k^{\text{new}}, x_R^k} \left\| \tilde{\boldsymbol{E}}_k - d_k^{\text{new}} \tilde{x}_R^k \right\|_F^2, \tag{4}$$

The problem in Equation (4) could be solved by SVD for $\tilde{\boldsymbol{E}}_k$, namely, $\tilde{\boldsymbol{E}}_k = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}$. Then, the updated dictionary atom d_k^{new} and the associated nonzero sparse code $\tilde{x}_R^k$ are computed as follows,

$$d_k^{\text{new}} = \boldsymbol{U}(:,1), \tilde{x}_R^k = \Sigma(1,1)\boldsymbol{V}(:,1). \tag{5}$$

Finally, nonzero entries in x_R^k are correspondingly replaced by $\tilde{x}_R^k$. The resulting discriminative dictionary learning algorithm is detailed in Algorithm 1, which learns $\boldsymbol{D}$, $\boldsymbol{A}$ and $\boldsymbol{W}$ simultaneously, and thus could avoid the local minima problem [23] in traditional SRC methods.

Algorithm 1 Discriminative dictionary learning

Input: $\boldsymbol{Y}$, $\boldsymbol{Q}$, $\boldsymbol{H}$, K, T, α, β

Output: $\hat{\boldsymbol{D}}, \hat{\boldsymbol{A}}, \hat{\boldsymbol{W}}$

Initialize: compute $\boldsymbol{D}^{(0)}$, $\boldsymbol{A}^{(0)}$, $\boldsymbol{W}^{(0)}$ (See Sec. 2.1.3)

Initialize $\boldsymbol{Y}_{\text{new}} = (\boldsymbol{Y}^T, \sqrt{\alpha}\boldsymbol{Q}^T, \sqrt{\beta}\boldsymbol{H}^T)^T$, $\boldsymbol{D}_{\text{new}} = (\boldsymbol{D}^T, \sqrt{\alpha}\boldsymbol{A}^T, \sqrt{\beta}\boldsymbol{W}^T)^T$ and set $J = 1$.

while: until J reaches the preset maximum number of iterations $J_{\max}$.

- **Sparse coding:** compute sparse codes x_i while keeping $\boldsymbol{D}_{\text{new}}$ fixed.
  $x_i^* = \arg \min_{x_i} \left\| y_{\text{new}}^i - \boldsymbol{D}_{\text{new}} x_i \right\|_2^2 \quad s.t.\ \forall i, \left\| x_i \right\|_0 \leq T$
- **Dictionary update:**
  For $k = 1,\ldots,K$, repeat updating $\boldsymbol{D}_{\text{new}}$ along with $\boldsymbol{X}$ as follows,
  (1) Compute $\tilde{\boldsymbol{E}}_k$ and perform SVD operation $\tilde{\boldsymbol{E}}_k = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}$;
  (2) Update d_k^{new} and $\tilde{x}_R^k$ by $d_k^{\text{new}} = \boldsymbol{U}(:,1), \tilde{x}_R^k = \Sigma(1,1)\boldsymbol{V}(:,1)$.
- Set $J = J+1$.

End while

Obtain the desired $\hat{\boldsymbol{D}}, \hat{\boldsymbol{A}}, \hat{\boldsymbol{W}}$ from $\boldsymbol{D}_{\text{new}}$ for classification using Equation (10).

2.1.3 Implementation details

In this subsection, several algorithm implementation details are presented, including algorithm initializations and obtaining the desired dictionary and linear classifier from $\boldsymbol{D}_{\text{new}}$.

(1) Initialization

In the proposed discriminative dictionary learning algorithm, initializations for $\boldsymbol{D}^{(0)}$, $\boldsymbol{A}^{(0)}$ and $\boldsymbol{W}^{(0)}$ are required. As for $\boldsymbol{D}^{(0)}$, we could exploit the standard dictionary learning to obtain sub-dictionaries for each class using several times of K-SVD, and then concatenate these sub-dictionaries of each K-SVD to accomplish the initialization for $\boldsymbol{D}^{(0)}$. Meanwhile, the label information $\boldsymbol{H}$ and discriminative sparse code $\boldsymbol{Q}$ could be determined according to the class information. As for $\boldsymbol{A}^{(0)}$ and $\boldsymbol{W}^{(0)}$, the multivariate ridge regression model [28] with quadratic loss and L_2-norm regularization is employed,

$$\boldsymbol{A} = \arg\min_{\boldsymbol{A}} \|\boldsymbol{Q} - \boldsymbol{A}\boldsymbol{X}\|_F^2 + \lambda_1 \|\boldsymbol{A}\|_F^2 , \tag{6}$$

$$\boldsymbol{W} = \arg\min_{\boldsymbol{W}} \|\boldsymbol{H} - \boldsymbol{W}\boldsymbol{X}\|_F^2 + \lambda_2 \|\boldsymbol{W}\|_F^2 . \tag{7}$$

The multivariate ridge regression models yield the solutions for $\boldsymbol{A}^{(0)}$ and $\boldsymbol{W}^{(0)}$ as follows,

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} + \lambda_1 \boldsymbol{I})^{-1} , \tag{8}$$

$$\boldsymbol{W} = \boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} + \lambda_2 \boldsymbol{I})^{-1} . \tag{9}$$

Therefore, provided the initialized dictionary $\boldsymbol{D}^{(0)}$, we could compute sparse codes $\boldsymbol{X}^{(0)}$ of training signals $\boldsymbol{Y}$ using OMP algorithm. Then $\boldsymbol{A}^{(0)}$ and $\boldsymbol{W}^{(0)}$ can be computed by Equation (8) and (9).

(2) Obtaining the desired dictionary and linear classifier from $\boldsymbol{D}_{\text{new}}$

The solution $\boldsymbol{D}_{\text{new}}$ in Equation (3) could be efficiently achieved by K-SVD and the resulting $\boldsymbol{D}_{\text{new}}$ is L_2-norm normalized column-wise, i.e., $\forall k, \left\| \left(d_k^T, \sqrt{\alpha} a_k^T, \sqrt{\beta} w_k^T \right)^T \right\|_2 = 1$. However, the directly extracted dictionary $\boldsymbol{D} = [d_1, d_2, \ldots, d_K]$ from $\boldsymbol{D}_{\text{new}}$, will not satisfy the column normalization constraint for sparse representation. Therefore, proper transformation from $\boldsymbol{D}_{\text{new}}$ to the desired discriminative dictionary $\hat{\boldsymbol{D}}$ is required. This transformation can be achieved by normalizing the extracted dictionary $\boldsymbol{D}$ column-wise and scaling $\boldsymbol{A}$ and $\boldsymbol{W}$ correspondingly as follows,

$$\begin{aligned}
\hat{\boldsymbol{D}} &= \left[\frac{d_1}{\|d_1\|_2}, \frac{d_2}{\|d_2\|_2}, \ldots, \frac{d_K}{\|d_K\|_2} \right] \\
\hat{\boldsymbol{A}} &= \left[\frac{a_1}{\|d_1\|_2}, \frac{a_2}{\|d_2\|_2}, \ldots, \frac{a_K}{\|d_K\|_2} \right] , \\
\hat{\boldsymbol{W}} &= \left[\frac{w_1}{\|d_1\|_2}, \frac{w_2}{\|d_2\|_2}, \ldots, \frac{w_K}{\|d_K\|_2} \right]
\end{aligned} \tag{10}$$

Therefore, provided the learned dictionary $\boldsymbol{D}_{\text{new}}$, we could directly extract the dictionary $\boldsymbol{D}$, linear transformation $\boldsymbol{A} = [a_1, a_2, \ldots, a_K]$ and classifier model $\boldsymbol{W} = [w_1, w_2, \ldots, w_K]$ from $\boldsymbol{D}_{\text{new}}$.

Besides, the desired $(\hat{\boldsymbol{D}}, \hat{\boldsymbol{A}}, \hat{\boldsymbol{W}})$ could be computed using the learned $(\boldsymbol{D}, \boldsymbol{A}, \boldsymbol{W})$ and Equation (10).

2.2 Classification Approach

Another critical procedure for the proposed DDL-SRC framework is the classification approach for recognition tasks. The classification in our DDL-SRC framework is carried out by two steps, namely sparse coding of testing signals and label estimation using classifier $\hat{\boldsymbol{W}}$.

2.2.1 Sparse coding of testing signals

For test signal y_i, sparse coding could be achieved by solving the following problem,

$$\hat{x}_i^* = \arg\min_{x_i} \left\| y_i - \hat{\boldsymbol{D}}\hat{x}_i \right\|_2^2, \quad s.t. \forall i, \left\| \hat{x}_i \right\|_0 \leq T , \tag{11}$$

The sparse coding could be efficiently solved using OMP algorithm [26].

2.2.2 Label estimation using classifier model

Estimating the label information of test signal y_i in the proposed framework is carried out according to the optimized sparse codes $\hat{x}_i^*$ and the linear classifier model $\hat{\boldsymbol{W}}$. We simply employ the linear classifier to estimate the label vector $\boldsymbol{l} = \hat{\boldsymbol{W}}x_i^*$. The predicted label j^* of y_i is determined by the index corresponding to the largest absolute value in label vector as follows,

$$j^* = \arg\max_j \ \text{abs}(l_j), \ \boldsymbol{l} = \hat{\boldsymbol{W}}x_i^* = [l_1, ..., l_L]^T . \tag{12}$$

2.3 Intelligent Fault Identification for Mechanical Components

In this section, we aim to extend the DDL-SRC framework for intelligent fault identification. To this end, we first propose an overlapping segmentation strategy for vibration signals and then summarize the overall procedures of the DDL-SRC framework for intelligent fault identification.

2.3.1 Overlapping segmentation strategy for vibration signals

Proper segmentation strategy is crucial to extract robust two-dimensional features from one-dimensional vibration signals. Periodic self-similarity is a robust feature in vibration signals of rotating machinery under stationary operating conditions. Hence, it is favorable to develop a wise segmentation strategy, which could take the best of self-similarity information and avoid the boundary artifacts between adjacent segments as well. To this end, we introduce a wise segmentation operator $\boldsymbol{R}$ to partition the original vibration signals into a set of local segments. The segmentation operator $\boldsymbol{R}$ is parameterized with two critical parameters, namely the segmentation window size W and overlapping rate δ. The principle of segmentation operator is illustrated in Figure 3. Let $y \in \mathrm{R}^{m}$ denotes the original vibration signal, the segmentation operator $\boldsymbol{R}: R^{m} \mapsto R^{W \times N}$ maps original vibration signals y into a signal matrix $\boldsymbol{Y} \in \mathrm{R}^{W \times N}$ as follows,

$$\boldsymbol{Y} = \underbrace{\left[\boldsymbol{R}_1\ \boldsymbol{R}_2 \ldots \boldsymbol{R}_{N-1}\ \boldsymbol{R}_N\right]}_{\boldsymbol{R}} y = \boldsymbol{R}(y)\ . \tag{13}$$

where the operator $\boldsymbol{R}_i : R^{m} \mapsto R^{W}$ represents the procedure that first takes the i-th segment from the original vibration signal y and then transposes the obtained segment, thus $\boldsymbol{R}_i(y) = y_i^{\mathrm{T}} = \boldsymbol{Y}(:,i)$.

Taking into account the multi-state vibration signals with different health conditions of mechanical components, we could employ the overlapping segmentation operator $\boldsymbol{R}$ to transform original vibration signals of each state. Then, constructing the overall training/testing signal matrix is accomplished by concatenating all signal matrix for different states. This overall overlapping segmentation strategy for overall signal matrix construction is illustrated in Figure 3.

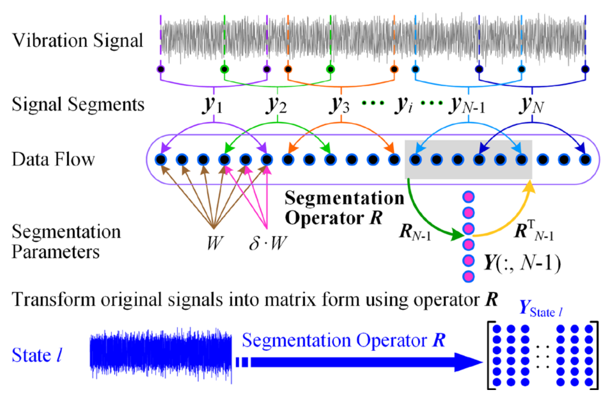

Figure 3: Overlapping Segmentation Strategy for Vibration Signals

As a result, the overlapping segmentation strategy could leverage the self-similarity feature maximally and construct the overall training/testing signal matrix, which serves as a premise for our proposed DDL-SRC framework for intelligent fault identification using vibration signals.

2.3.2 Overall procedures of DDL-SRC framework for intelligent fault identification

In this section, we summarize the overall algorithm procedures of the DDL-SRC framework for intelligent fault identification using vibration signals as follows.

Step 1 Acquire vibration signals of mechanical components under various health conditions and construct training/testing signal matrixes using the overall overlapping segmentation strategy.

Step 2 Implement the discriminative dictionary learning algorithm for learning the discriminative dictionary and an optimal multiclass linear classifier model simultaneously.
Step 3 Implement the sparse classification using the learned discriminative dictionary and predictive multiclass linear classifier model in step 2.
Step 4 Achieve intelligent fault identification of mechanical components using the results of sparse representation classification.

3. CASE STUDY: INTELLIGENT FAULT IDENTIFICATION FOR PLANET BEARINGS

In this section, experiments and vibration measurements are conducted on a planetary gearbox test bench, which are used to validate the effectiveness of the proposed DDL-SRC framework for intelligent fault identification of planet bearings. In addition, the state-of-the-art dictionary learning based sparse representation classification (DL-SRC) method in [24] is compared for demonstrating the superiority of the proposed framework.

3.1 Experiment Description

The planetary gearbox experiment setup is illustrated in Figure 4(a), which consists of a drive motor, an encoder for rotational speed measurement, a planetary gearbox, a magnetic powder brake for applying load and the vibration measurement system. To simulate the local defects in planet bearings, minor damages are introduced on surface of different planet bearing components using electrical discharge machining technique. As a result, vibration measurements could be performed on the planetary gearbox with four different health conditions (namely, healthy, outer ring defect, inner ring defect and rolling element defect), as illustrated in Figure 4(b). Detailed physical parameters of the planetary gearbox are listed in Table 1. As for the experiment measurement, the sun gear rotates at a constant frequency of 24.97 Hz and vibration signals are sampled at a frequency of 20480 Hz.

Table 1: Physical Parameters of the Experimental Planetary Gearbox.

Gear	Tooth number	Planet bearing	Size (mm)
Carrier	--	Roller diameter	9
Sun gear	13	Diameter of pitch circle	36
Ring gear	92	Number of rollers	10
Planet gear	38(3)	Contact angle (°)	0

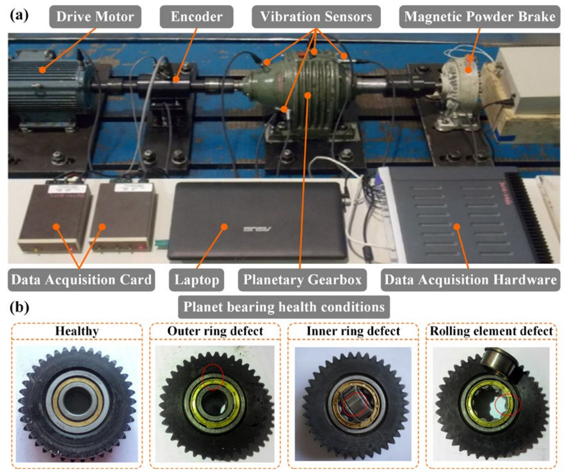

Figure 4: Planetary Gearbox Experiment Setup and Planet Bearings with Four Health Conditions.

3.2 Planet Bearing Fault Identification

Vibration signals of planetary gearbox with four different health conditions are measured in two different processes, the one of which with time duration of 60 seconds serve as the training signals and the other with time duration of 60 seconds serve as the testing signals for the proposed DDL-SRC framework. The time domain waveforms of vibration signals of planetary gearbox in presence of four different health conditions of planet bearings are illustrated in Figure 5. In order to reduce the computation burden of discriminative dictionary learning algorithm, all original vibration signals are downsampled at a sampling frequency of 1024 Hz.

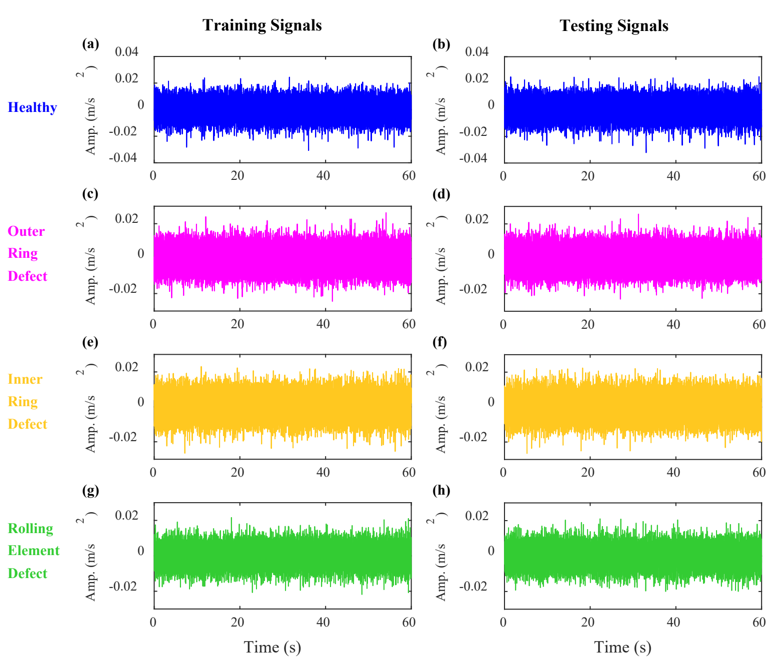

Figure 5: Vibration Signals of Planetary Gearbox with Four Health Conditions of Planet Bearings.

3.2.1 Intelligent identification of planet bearing defects

In this subsection, we evaluate the effectiveness and superiority of the proposed DDL-SRC framework over the DL-SRC method in [24] for intelligent fault identification of planet bearings.

In this experiment, dictionary size is set as 4350 such that an average of 1088 dictionary atoms corresponds to each health condition of planet bearing. The parameters for overlapping segmentation strategy are chosen that the segmentation window size W be 1800 and overlapping rate δ be 0.97. Hence, both the training and testing signal matrixes for each health condition consist of 1105 samples. Besides, the sparsity threshold T and maximum iteration number $J_{\max}$ are set as 4 and 35 for discriminative dictionary learning, respectively. The regularization parameters α=0.006 and β=0.001 are used in this case. As a result, the confusion matrix of the proposed DDL-SRC framework and classification performance including the comparison with the state-of-the-art DL-SRC method [24], are illustrated in Figure 6(a) and (b), respectively.

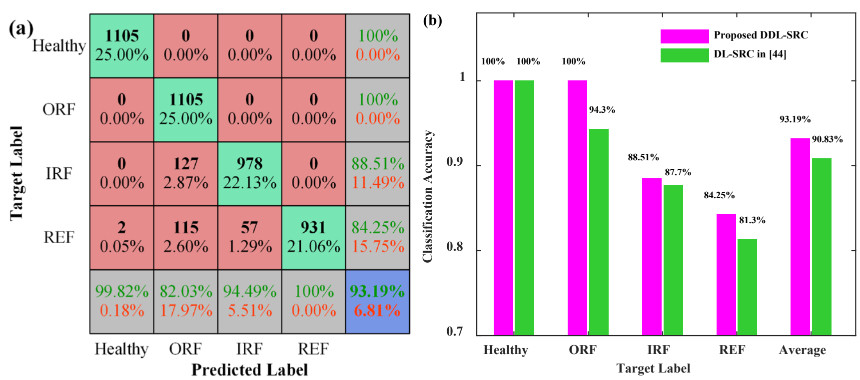

Figure 6: Confusion Matrix of The Proposed DDL-SRC Framework and Classification Performance for Intelligent Fault Identification of Planet Bearings.

As illustrated in Figure 6(a), the proposed DDL-SRC framework achieves no false prediction for healthy planet bearings and faulty planet bearings with outer ring defect. Additionally, the overall percentage of which the faulty planet bearing is wrongly predicted as the healthy one is as low as 0.045% (i.e., 2/(4*1105)). As a whole, the proposed DDL-SRC framework achieves an excellent average identification accuracy as higher as 93.19% for intelligent fault identification of planet bearings in this general case. In contrast, our proposed framework always achieve better classification accuracy than the state-of-the-art DL-SRC method [44], as illustrated in Figure6(b).

3.2.2 Parameter analysis

In this subsection, we discuss the effects of related algorithm parameters on classification performance of the DDL-SRC framework. As for the fault identification of planet bearings in this experiment, we evaluate several crucial algorithm parameters with cross-validation strategy. These parameters are divided into four different groups and each group of parameters will be discussed with other groups fixed. More specifically, we firstly analyze the effect of dictionary size K and with other parameters fixed. Then we will fix the dictionary size and check the performance of regularization parameters α and β. Next, the segmentation window size W and overlapping rate δ are discussed. Finally, the sparsity threshold T and maximum number of iterations J_{max} for discriminative dictionary learning are validated.

(1) *Dictionary Size K*. Dictionary size K determines the number of atoms in the learned dictionary. On the one hand, a higher dimension of feature subspace will contain more information so as to enhance the representational ability of dictionary. On the other hand, a higher dimension of dictionary will generally be more discriminative than a lower one for classification tasks [37]. To validate these two suggestions, dictionary size K is varied from 2400 to 4400 and the corresponding classification accuracies are illustrated in Figure 7(a). It can be concluded that increasing the dictionary size will lead to a higher classification performance and the highest classification accuracy can be achieved around the maximum considered dictionary size.

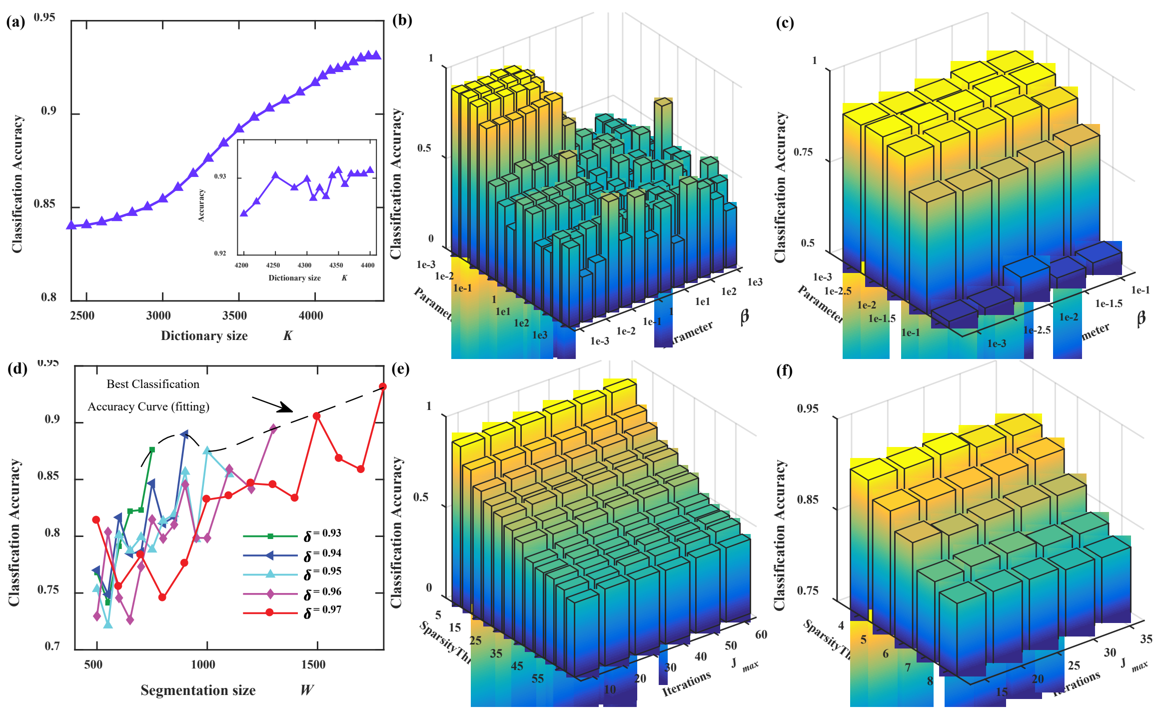

Figure 7: Parameter Analysis on Classification Performance. (a) Effect of Dictionary Size K; (B) Effect of Regularization Parameters α and β; (c) Enlarged Illustration of (b); (d) Effect of Segmentation Size W and Overlapping Rate δ; (e) Effect of Sparsity Threshold T and Maximum Number of Iterations J_{max}; (f) Enlarged Illustration of (e).

(2) *Regularization parameters α and β*. The regularization parameters determine the relative contributions of the discriminative sparse code error and classification error terms

in the optimization objective. A balanced pair of regularization parameters will lead to the desired dictionary with both reconstructive and discriminative power, and an optimal classifier model for classification tasks. To investigate their effects, both α and β are varied from 10^{-3} to 10^{3} exponentially and the corresponding results are illustrated in Figure 7(b). By checking these classification accuracies, we can get the conclusions. Firstly, regularization parameters with relatively small values could lead to higher classification performance than pairs of regularization parameters with greater values. Secondly, the regularization parameters with the order of magnitude of 10^{-3} achieve the best classification accuracy in our task, as shown in Figure 7(c).

(3) *Segmentation window size W and overlapping rate* δ. These two parameters influence the results of the overlapping segmentation strategy for vibration signals. In order to take the best of self-similarity information in vibration signals, we consider the cases with relatively large overlapping rate, namely δ=0.93:0.01:0.97, and varying segmentation sizes. The corresponding results are illustrated in Figure 7(d). By carefully checking these classification accuracies, we can get the following conclusions. Firstly, though there are few outliers like jump points, increasing the segmentation size generally could achieve a better classification performance. Secondly, the best classification accuracy using a higher overlapping rate is generally better than that using a lower overlapping rate, as illustrated by the 'best classification accuracy curve' in Figure 7(d).

(4) *Sparsity threshold T and maximum number of iterations* $J_{\max}$. Finally, the effects of sparsity threshold and maximum number of iterations on classification accuracy are studied and illustrated in Figure 7(e) and (f). These two parameters mainly affect the discriminative dictionnary learning algorithm. The following conclusions can be drawn from Figure 7(e) and (f). Firstly, smaller sparsity thresholds T could lead to better classification performance than greater sparsity thresholds. Secondly, the classification accuracy is merely promoted very little as the increase of maximum number of iterations $J_{\max}$. In our experiment, sufficiently better classification performance could be obtained with 35 iterations.

In summary, the proposed DDL-SRC framework is effective using proper algorithm parameters and outperforms the state-of-the-art DL-SRC method for fault identification of planet bearings, which indicates that it is promising to achieve intelligent machine fault diagnosis.

4. Conclusion

In this paper, a novel discriminative dictionary learning based sparse representation classification framework is proposed for intelligent fault identification of planet bearings. In contrast to traditional frequency feature based diagnostic strategies for planet bearing fault detction, the proposed framework could learn adaptive sparse features directly from raw vibration signals for reliable identification performance. The main conclusions are drawn as follows.

(1) The proposed framework bridges the gap between the reconstructive dictionary learning and classifier training in classical sparse representation classification methods, by introducing a discriminative sparse code error term and incorporating it with the

reconstruction error and classification error into one mixed optimization formulation.

(2) The proposed framework could learn a discriminative dictionary and an optimal multiclass linear classifier simultaneously for better classification performance over classical sparse representation classification methods. Additionally, the proposed discriminative dictionary learning problem could be efficiently solved by a simple extension of the K-SVD algorithm.

(3) In contrast to other conventional classifier based pattern recognition methods, the proposed framework is free of feature design and selection. It can adaptively learn adaptive features from raw vibration signals for intelligent fault identification.

(4) The proposed framework outperforms the existing state-of-the-art sparse representation classification method and achieves intelligent fault identification of planet bearings with the highest classification accuracy among all reported literatures.

References

[1] Z. Hameed, Y.S. Hong, Y.M. Cho, S.H. Ahn, C.K. Song, Condition monitoring and fault detection of wind turbines and related algorithms: A review, Renew. Sustain. Energy Rev. 13 (2009) 1–39.

[2] Z. Feng, M.J. Zuo, Vibration signal models for fault diagnosis of planetary gearboxes, J. Sound Vib. 331 (2012) 4919–4939.

[3] Y. Lei, J. Lin, M.J. Zuo, Z. He, Condition monitoring and fault diagnosis of planetary gearboxes: A review, Measurement 48 (2014) 292–305.

[4] W. Bartelmus, R. Zimroz, A new feature for monitoring the condition of gearboxes in non-stationary operating conditions, Mech. Syst. Signal Process. 23 (2009) 1528–1534.

[5] T. Barszcz, R.B. Randall, Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine, Mech. Syst. Signal Process. 23 (2009) 1352–1365.

[6] P.D. McFadden, A technique for calculating the time domain averages of the vibration of the individual planet gears and the sun gear in an epicyclic gearbox, J. Sound Vib. 144 (1991) 163–172.

[7] Y. Kong, T. Wang, Z. Li, F. Chu, Fault feature extraction of planet gear in wind turbine gearbox based on spectral kurtosis and time wavelet energy spectrum, Front. Mech. Eng. 12 (2017) 406–419.

[8] Y. Kong, T. Wang, F. Chu, Meshing frequency modulation assisted empirical wavelet transform for fault diagnosis of wind turbine planetary ring gear, Renew. Energy. 132 (2019) 1373–1388.

[9] Z. Feng, M. Liang, Y. Zhang, S. Hou, Fault diagnosis for wind turbine planetary gearboxes via demodulation analysis based on ensemble empirical mode decomposition and energy separation, Renew. Energy. 47 (2012) 112–126.

[10] Z. Feng, X. Lin, M.J. Zuo, Joint amplitude and frequency demodulation analysis based on intrinsic time-scale decomposition for planetary gearbox fault diagnosis, Mech. Syst. Signal Process. 72–73 (2016) 223–240.

[11] Z. Feng, M. Liang, Fault diagnosis of wind turbine planetary gearbox under nonstationary conditions via adaptive optimal kernel time-frequency analysis, Renew. Energy. 66 (2014) 468–477.

[12] S. Jain, H. Hunt, Vibration response of a wind turbine planetary gear set in the presence of a localized planet bearing defect, Proceedings of the ASME 2011 International

Mechanical Engineering Congress and Exposition (2011) 943–952.
[13] S. Jain, P.E. Whiteley, H. Hunt, Detection of planet bearing faults in wind turbine gearboxes, Proceedings of the International Conference on Noise and Vibration Engineering (2012) 4351–4362.
[14] Z. Feng, H. Ma, M.J. Zuo, Vibration signal models for fault diagnosis of planet bearings, J. Sound Vib. 370 (2016) 372–393.
[15] Z. Feng, H. Ma, M.J. Zuo, Amplitude and frequency demodulation analysis for fault diagnosis of planet bearings, J. Sound Vib. 382 (2016) 395–412.
[16] T. Wang, F. Chu, Z. Feng, Meshing frequency modulation (MFM) index-based kurtogram for planet bearing fault detection, J. Sound Vib. 432 (2018) 437–453.
[17] M. Inalpolat, A. Kahraman, A dynamic model to predict modulation sidebands of a planetary gear set having manufacturing errors, J. Sound Vib. 329 (2010) 371–393.
[18] Z. Jiang, Z. Lin, L.S. Davis, Label consistent K-SVD: Learning a discriminative dictionary for recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 2651–2664.
[19] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 1–18.
[20] X. Mei, H. Ling, Robust visual tracking and vehicle classification via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 2259–2272.
[21] M. Aharon, M. Elad, A. Bruckstein, K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation, IEEE Trans. Signal Process. 54 (2006) 4311–4322.
[22] J. Mairal, F. Bach, J. Ponce, G. Sapiro, A. Zisserman, Discriminative learned dictionaries for local image analysis, IEEE Conf. Comput. Vis. Pattern Recognition, CVPR (2008) 1–8.
[23] D.S. Pham, S. Venkatesh, Joint learning and dictionary construction for pattern recognition, IEEE Conf. Comput. Vis. Pattern Recognition, CVPR. (2008) 1–8.
[24] C. Zhao, Z. Feng, X. Wei, Y. Qin, Sparse classification based on dictionary learning for planet bearing fault identification, Expert Syst. Appl. 108 (2018) 233–245.
[25] K. Huang, S. Aviyente, Sparse representation for signal classification, Advances in Neural Information Processing Systems (2007) 609–616.
[26] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inf. Theory 53 (2007) 4655–4666.
[27] Y. Zhang, Z. Jiang, L.S. Davis, Learning structured low-rank representations for image classification, IEEE Conf. Comput. Vis. Pattern Recognition, CVPR. (2013) 676–683.
[28] G.H. Golub, P.C. Hansen, D.P. O'Leary, Tikhonov regularization and total least squares, SIAM J. Matrix Anal. Appl. 21 (1999) 185–194.