



Lan Chen, Xiangfeng Zhang, Lizhong Wang \*, Kaihua Li and Yang Feng 回

School of Mechanical Engineering, Xinjiang University, Urumqi 830049, China; 107556521223@stu.xju.edu.cn (L.C.); xjuzxf@xju.edu.cn (X.Z.); 107552204213@stu.xju.edu.cn (K.L.); fengy@stu.xju.edu.cn (Y.F.)

\* Correspondence: wanglz@mail.xjtu.edu.cn

Abstract: A gearbox compound fault intelligent diagnosis method based on the period group sparse model is proposed for the problem that the fault features are coupled with each other and the fault components are superimposed on each other and difficult to be separated in the gearbox compound fault signal. Firstly, a binary sequence is constructed to embed the fault pulse period as a priori knowledge into the group sparse model to decouple and separate the composite fault signal while maintaining the amplitude and sparsity of the extracted features. Secondly, the wavelet packet energy features of the decoupled signals are extracted to improve the data quality while enhancing the characterization ability of the dictionary in the classification model. Finally, the wavelet packet energy features are imported into the sparse dictionary classification model, and the fault diagnosis is completed by outputting the fault categories using the self-driven characteristics of the data. The results show that the fault identification accuracy using the proposed method is 97%. In addition, the experimental validation under different states and working conditions with different rotational speeds allows the superiority and effectiveness of the algorithm in this paper to be tested and has the feasibility of a practical application in engineering.

Keywords: compound fault; gearbox; sparse representation; feature classification



app14104294

Lacidogna

(cc

4.0/).

check for

4294. https://doi.org/10.3390/

Academic Editor: Giuseppe

Received: 3 April 2024

Revised: 14 May 2024

Accepted: 14 May 2024

Published: 18 May 2024

(†)

L.; Li, K.; Feng, Y. Intelligent Diagnosis of Compound Faults of 1. Introduction Gearboxes Based on Periodical Group Sparse Model. Appl. Sci. 2024, 14,

With the increasing demand for energy in social development, wind turbines, gas turbines and other high-end equipment will be the main body of new energy supply under the carbon neutral strategy. As a key component of energy conversion, gearboxes are susceptible to irregular alternating loads and impact loads during operation, thus generating failures [1]. Timely and reliable condition monitoring and fault diagnosis technology is important in reducing operation and maintenance costs, improving production management capability, and providing "foresight" predictive maintenance solutions. Traditional time-frequency analysis methods, such as wavelet transform [2,3], empirical modal decomposition [4] and variational modal decomposition [5], have achieved satisfactory results in single-fault diagnosis. In the actual operating conditions, a single fault will have a chain reaction with each component of the transmission chain, and multiple faults will occur in successive cascades to form a compound fault. At the same time, various components in the signal are affected by the background noise, so that the vibration signal has strong Copyright: © 2024 by the authors. interference, nonlinearity, weakened fault characteristics, coupling modulation characteris-Licensee MDPI, Basel, Switzerland. tics and other characteristics, resulting in the weakening of the signal fault characteristics. Therefore, it is difficult to separate the composite fault components using traditional signal processing methods.

> Recently, sparse representation has attracted attention as a novel signal processing method in many fields, such as target detection [6], face recognition [7], and speech signal processing [8]. Sparse representation is essentially the use of a small number of atoms to

Attribution (CC BY) license (https://

creativecommons.org/licenses/by/



describe the internal structure and feature information of a signal without preprocessing operations such as denoising, the optimal selection of fault features and the stripping of working condition information, which enables a flexible, concise and adaptive representation of the signal. The focus of sparse representation is on the construction of the dictionary. The parametric dictionary consists of predefined basis functions to match the periodic impulse features in the fault signal. Sun et al. [9] designed a parametric dictionary that can highly match the bearing fault features and improve the iterative stopping criterion of the orthogonal matching tracking algorithm to successfully diagnose the rolling bearing fault signal. Zhang et al. [10] constructed an adaptive adjustment Gabor dictionary with the residual signal, which solved the shortcomings of the traditional method with large calculation and improved the signal sparsity while ensuring the accuracy of the fault features. Xia et al. [11] optimized the quality factor in the resonance-based sparse decomposition (RSSD) using the squirrel algorithm, and the obtained optimal resonance components effectively improved the separation accuracy of the bearing fault features. However, due to the variability and unpredictability of the pulse components in the actual vibration signal, the performance of the parametric dictionary is limited by the atomic structure [12], and the atoms obtained using explicit mathematical formulas are no longer adaptively adjusted to the changing characteristics of the signal. For this reason, learning dictionaries such as the optimal direction method and K-singular value decomposition (K-SVD) provide new inspiring ideas for fault diagnosis. Wang et al. [13] improved the objective function and constraints in the K-SVD model to construct a dictionary matching fault features without considering sparsity, and achieved a feature enhancement of weak bearing faults. Li et al. [14] used VMD to select the optimal components to form the training set, and then used K-SVD for dictionary learning. Compared with the parametric dictionary, the method applied to pipeline leakage vibration signals shows better compressive reconfigurability and sparsity. When extracting fault features using sparse representation, most studies use 10 or 11 norm and its deformers, which suffer from algorithmic degradation leading to amplitude underestimation when approximating the sparsity of the characterized pulse. For this reason, it has been proposed to embed the structural information of the signal into SR and exploit the sparsity within and across groups (SWAG) of the features [15] to improve the extraction accuracy of fault features. Ding et al. [16] constructed a periodic convolutional sparse representation using a learned dictionary and a Fourier parameter dictionary to extract fault classification and harmonic components so that the sparse coefficients have periodicity and group sparsity. The method performs well in extracting bearing fault features. Huang et al. [17] combined a sparse fidelity term and a generalized minimum-maximum concave penalty term in the sparse representation model to effectively prevent the signal amplitude from being underestimated and improve the reconstruction accuracy of gear and bearing fault features. Zhao et al. [18] introduced an enhanced sparse group lasso to improve the SWAG property, and then used an optimized minimization algorithm (Majorize-Minimization, MM) to iteratively solve the model to effectively extract the fault features of the inner and outer rings of the bearings while maintaining the magnitude and sparsity of the reconstructed signals.

In summary, the vibration signal-based gearbox fault diagnosis technology is becoming increasingly mature, but the following problems still exist:

- (1) Most of the existing research focuses on single faults or localized compound faults, such as gear local faults, and bearing inner and outer ring compound faults. There is less research on composite faults formed in different positions of bearings and gears, etc.
- (2) Multiple faults in the transmission path due to mutual interference between fault features and energy loss and other effects of modulation coupling, decoupling separation that lead to an underestimation of the amplitude, thus leading to the signal reconstruction effect.
- (3) The traditional fault diagnosis using the fault feature frequency detection method; the detection performance is easily deteriorated due to factors such as part manufacturing

errors and shaft misalignment. The engagement vibration and background noise can submerge or interrupt the impulse features of the fault.

According to the above analysis, this paper proposes an intelligent diagnosis method for compound faults based on the periodic group sparse model. Different from the existing diagnostic methods, firstly, a binary periodic sequence is constructed to provide periodic prior knowledge for the fault signal, and a sparse model for handling compound faults is constructed by combining with a group sparse model with overlapping properties. Then, the MM algorithm is used to decouple and separate the fault signals; compared with other conventional methods, the proposed algorithm shows a higher discrimination and robustness in practical applications. Finally, the energy features of wavelet packets with the decoupled different fault states are extracted and imported into the sparse dictionary classification model as the data input to achieve fault feature classification and identification with the self-driven property of data. So far, the intelligent diagnosis of compound faults in gearboxes is realized. The effectiveness of the proposed method is verified with the analysis of the simulation signal and the experimental signal; the overall technology roadmap is shown in Figure 1.



Figure 1. General technology roadmap.

The arrangement of the paper is as follows: Section 1 is the introduction of the article. Section 2 introduces the theoretical basis of performance of different sparse models, LC-KSVD, and Wavelet packet energy features extraction. Section 3 introduces the fault diagnosis method proposed in this paper. Section 4 introduces the simulated signal analysis and results. Section 5 is the conclusion and future work of this paper.

## 2. Materials and Methods

## 2.1. Theoretical Methods

2.1.1. Performance Differences and Sparse Effect of Different Models

In actual working conditions, there is a force transmission between several components in the transmission chain of machinery and equipment. The sparse model using only the 11 norm cannot better capture the clustering trend among the sparse coefficients. This leads to its inability to effectively extract multiple fault feature information, while the 11 parametric number also has the problem that the signal amplitude is underestimated, which will reduce the reconstruction accuracy of the signal. The performance of different sparse models and their sparse effects are different, and the expressions of the related sparse models are shown below for the Lasso model, group Lasso model and group sparse model, respectively.

$$\hat{\psi}_{LASSO} = \underset{\psi}{\operatorname{argmin}} \|y - X\psi\|_2^2 + \lambda_1 \|\psi\|_1 \tag{1}$$

$$\underset{\omega_{1},\omega_{2}}{\operatorname{argmin}} \|y - \Phi_{1}\omega_{1} - \Phi_{2}\omega_{2}\|_{2}^{2} + \sum_{j=1}^{j_{1}+1} \lambda_{1,j} \|\omega_{1,j}\|_{1} + \sum_{j=1}^{j_{2}+1} \lambda_{2,j} \|\omega_{2,j}\|_{1}$$
(2)

$$\hat{\psi}_{Group-LASSO} = \underset{\psi}{\operatorname{argmin}} \|y - X\psi\|_2^2 + \lambda_2 \|\psi\|_2$$
(3)

$$\hat{\psi}_{SGL} = \underset{\psi}{\operatorname{argmin}} \|y - X\psi\|_{2}^{2} + \lambda_{1} \|\psi\|_{1} + \lambda_{2} \|\psi\|_{2}$$
(4)

where  $X \in \mathbb{R}^{N \times Q}$  is the dictionary matrix,  $y \in \mathbb{R}^N$  is the test signal, and  $\psi \in \mathbb{R}^Q$  is the corresponding sparse coefficient vector. $\lambda_1$  and  $\lambda_2$  are the adjustment parameters. The sparse effect of each model is shown in Figure 2. From the figure, it can be seen that the Lasso model, as a typical representative of the parametric-based sparse model, is based on the principle of optimizing the selection of each group of original signals so that each group of data maintains good sparsity. The RSSD objective function shown in Equation (2) is a typical Lasso model. The Group-Lasso model mainly groups the original signals and eliminates the relevant intra-group data, so that each group of data between the signals maintains good sparsity, but the processed data still have more redundant information. The group sparse model further optimizes the redundant data within groups by combining the advantages of both models, which can reduce the influence of redundant information on the signal. It is a good balance of the inter-group sparsity and intra-group sparsity of signal. Therefore, this paper selects the group sparsity-based model to deal with compound faults.





2.1.2. Label Consistency Dictionary Classification Algorithm

Since the traditional dictionary learning algorithm does not contain discriminative information, i.e., it cannot find an explicit correspondence between dictionary atoms and fault categories labels. To solve this problem, Jiang et al. [7] proposed a label consistent KSVD(LC-KSVD) dictionary learning algorithm. The algorithm adds label consistency discriminative sparse coding terms and optimal classification error terms to the objective function, so that dictionary atoms and fault categories are associated. The dictionary is

forced to make the sparse coefficients distinctive during the learning process through the discriminative sparse encoding matrix, i.e., the same class of signals produces similar sparse encoding. The dictionary reconfigurability and discriminability are maintained while integrating the advantages of traditional dictionaries. This algorithm has been widely used in face recognition and other fields.

The objective function of LC-KSVD as follows:

$$\langle D, A, W, X \rangle = \underset{D, A, W, X}{\operatorname{argmin}} \|Y - DX\|_{2}^{2} + \alpha \|Q - AX\|_{2}^{2} + \beta \|H - WX\|_{2}^{2}$$

$$s.t. \forall i, \|s_{i}\|_{0} \leq T$$

$$(5)$$

where  $||Y - DX||_F^2$  is the sparse fidelity term, i.e., denotes the reconstruction error. Where  $Y = [y_1, \dots, y_N] \in \mathbb{R}^{n \times N}$  is the set of N and n-dimensional training signals an  $D = [d_1, \dots, d_K] \in \mathbb{R}^{n \times K}$  is an overcomplete dictionary composed of K atoms.  $X = [x_1, \dots, x_N] \in \mathbb{R}^{K \times N}$  is the coefficient of the training signal after sparse decomposition.  $||Q - AX||_F^2$  denotes the discriminative sparse coding error, i.e., the label agreement term, which aims to constrain samples of the same class to produce similar sparse coding, and  $A \in \mathbb{R}^{K \times K}$  denotes the linear transformation matrix, with  $Q = [q_1, \dots, q_N] \in \mathbb{R}^{K \times N}$  denoting the discriminative sparse coding.  $||H - WX||_F^2$  denotes the classification error term of the classifier model, which helps to obtain  $H \in \mathbb{R}^{L \times N}$ , which denotes the category label matrix of the training samples and  $W \in \mathbb{R}^{L \times K}$  denotes the linear classifier parameters.  $\alpha$  and  $\beta$  are the weighting coefficients between the control correlation terms.

As shown in Figure 3, discriminative sparse coding is used to obtain the best classification results. Assume that the training sample matrix contains three classes of training signals  $Y = [Y_1, Y_2, Y_3]$ , where  $Y_1$  contains two samples  $y_1$  and  $y_2$ .  $Y_2$  contains three samples  $y_3$ ,  $y_4$ ,  $y_5$ ,  $Y_3$  contains four samples  $y_6$ ,  $y_7$ ,  $y_8$ ,  $y_9$ . The dictionary D contains three sub-dictionaries of the three categories, and each sub-dictionary  $D_i$  (i = 1, 2, 3) has three atoms, thus obtaining the ideal discriminative sparse coding matrix.



Figure 3. Using discriminative sparse coding to obtain the best classification results.

To facilitate the solution, Equation (5) is rewritten in the following form:

$$\langle D, A, W, X \rangle = \underset{D, A, W, X}{\operatorname{argmin}} \left\| \begin{pmatrix} Y \\ \sqrt{\alpha}Q \\ \sqrt{\beta}H \end{pmatrix} - \begin{pmatrix} D \\ \sqrt{\alpha}A \\ \sqrt{\beta}W \end{pmatrix} X \right\|_{F}^{2}$$

$$s.t.\forall i, \|x_{i}\|_{0} \leq T$$
(6)

$$\widetilde{Y}_{new} = \left(Y^T, \sqrt{\alpha}Q^T, \sqrt{\beta}H^T\right)^T \tag{7}$$

$$\widetilde{D}_{new} = \left(D^T, \sqrt{\alpha}A^T, \sqrt{\beta}W^T\right)^T$$
(8)

The merged atoms in the new dictionary are normalized by parametrization, and then the optimization problem is equated to the following form:

$$\left\langle \widetilde{D}_{new}, X \right\rangle = \underset{\widetilde{D}, X}{\operatorname{argmin}} \left\| \widetilde{Y} - \widetilde{D}_{new} X \right\|_{F}^{2} \quad s.t. \forall i, \|x_{i}\|_{0} \leq T$$
(9)

This type of model is equivalent to the optimization problem of the K-SVD algorithm, therefore, four variables, i.e., dictionary D, linear transformation matrix A, linear classifier W, and sparse coefficients X, are solved simultaneously using the K-SVD algorithm to obtain the global optimal solution. Then, the sparsity coefficient of the test sample is solved with the following equation.

$$x' = \operatorname{argmin} \|y - Dx\|_2^2 \quad \text{s.t.} \ \forall i, \|x_i\|_0 \le 0 \tag{10}$$

Finally, the linear classifier is used to obtain the class label vector l of the test samples, and the maximum value of the index corresponding to the selected l can distinguish the different fault classes in the test samples.

$$l = W'x' = [l_1, \dots, l_L]^T \ label(y_i) = \operatorname*{argmax}_{i} abs(l_j) \tag{11}$$

## 2.1.3. Wavelet Packet Energy Features Extraction

The decoupled and separated fault signal are both sparse, i.e., a large number of data sampling points in the signal are zero, and there is a presence of certain interference components in the decoupled single fault signal. Therefore, the direct segmentation of the fault signal as a dictionary and for learning will inevitably lead to the degradation of the dictionary quality and consume more computation time. For this reason, wavelet packet decomposition is firstly used to improve the data quality and enhance the variability of the energy distribution of different health states, which provides prerequisites for subsequent fault classification and identification.

Wavelet packet decomposition adds the decomposition of the high frequency part of the signal to the wavelet transform. Through the multi-scale decomposition of frequency bands, the decomposition process is neither redundant nor loose, and at the same time more detailed and comprehensive, so that it has stronger transient information extraction ability. It has a better performance of time-frequency localization analysis and thus improves the time-frequency resolution.

Wavelet packet decomposition obtains the signal components of the corresponding sub-bands using the multi-level division and multi-scale decomposition of the signal frequency bands. The characteristic information contained in the original vibration data is dispersed into the sub-band signals, and according to the different energy distributions of the signals in different frequency intervals that will be healthy, the difference in the energy characteristic distribution can be used as an important criterion for fault classification and identification. The steps for wavelet packet energy feature extraction are as follows:

- (1) The fault signal is decomposed into  $2^n$  frequency bands,  $S_{i,j}$  denotes the decomposed signal of the *j*th node in the *i*-th layer. The wavelet packet coefficients of the node are  $x_{i,j}$ .
- (2) The corresponding low-frequency and high-frequency coefficients are obtained after multiple layer decomposition, and the signal is reconstructed, and the expression of the reconstructed signal is as follows:

$$S = S_{(i,0)} + S_{(i,1)} + \ldots + S_{(i,2^n - 1)}$$
(12)

(3) Calculate the energy of each sub-band signal.

$$E_{ij} = \int |S_{i,j}(t)|^2 dt = \sum_{k=1}^n x_{i,j}(k)$$
(13)

(4) Calculate the total energy and obtain the fault feature vectors with different energy percentages after normalization.

$$E = \sum_{j=0}^{2^{i}-1} E_{i,j} \tag{14}$$

$$T = \left[\frac{E_{i,0}}{E}, \frac{E_{i,1}}{E}, \dots, \frac{E_{i,2^n-1}}{E}\right]$$
(15)

#### 2.2. Proposed Fault Diagnosis Method

2.2.1. Compound Fault Diagnosis Method Based on Period Group Sparse Model

Suppose that the fault signal x is recovered from the noisy signal y, where x is with the group periodicity property. However, the large magnitude coefficients in the signal do not exist independently of each other, but are composed in group clusters. The l1 parametric and other sparse models cannot capture the clustering trend of the coefficients, so He et al. [19] proposed an overlapping group sparse model to recover the fault signal through constructing an objective function containing a nonconvex penalty function. However, in practical fault diagnosis, the pulse lengths of various types of fault information cannot be determined in advance, so the penalty function is improved by assuming that each group has the same size and maximum overlap [18] and defined as shown in Equation (16). Define the index sets  $\Omega$  and  $\psi$  as shown in Equation (17).

$$P(\mathbf{x}) = \lambda \sum_{i \in \Omega} \left( \left[ \sum_{j \in \psi} |x(i+j)|^2 \right]^{\frac{1}{2}} + \rho \sum_{j \in \psi} |x(i+j)| \right) \\ = \lambda \sum_{i \in \Omega} \left( \|\mathbf{x}_{i,K}\|_2 + \rho \|\mathbf{x}_{i,K}\|_1 \right)$$
(16)

$$\Omega = [1, \dots N] \quad \psi = [0, 1, \dots K - 1] \tag{17}$$

where  $\lambda$  and  $\rho$  are the regular term parameters. The length of the signal is N and each group has the same size K.i is the group index and j is the coefficient index.  $x_{i,K} = [x(i), \dots, x(i+K-1)] \in \mathbb{R}^K$  indicates that the size of the *i*th group is K.

To decouple multiple faults from a composite fault signal, the penalty term of the overlapping group sparse group model is modified by the feature frequencies of the faults. A binary sequence is used to make the penalty term concise and highly structured.

$$b = \underbrace{[\underbrace{1, 1, \dots, 1}_{N_1}, \underbrace{0, 0, \dots, 0}_{N_0}, \underbrace{1, 1, \dots, 1}_{N_1}, \dots, \underbrace{0, 0, \dots, 0}_{N_0}, \underbrace{1, 1, \dots, 1}_{N_1}]}_{K}$$

$$N_0 + N_1 \approx F_s / f_{c,} N_0 + N_1 = (K + N_0) / M$$
(19)

where 
$$N_1$$
 is the estimated length of the oscillation wave and  $N_0$  is the time interval between  
each pulse. *K* is the size of the group.  $F_s$  and  $f_c$  denote the sampling frequency and fault  
characteristic frequency, respectively. *M* is the number of  $N_1$  in each group. The three  
parameters satisfy the relation in Equation (19).

The penalty term and the corresponding term in the binary sequence are elementwise multiplied so that inter-group sparsity is maintained between different faults, while intra-group sparsity is formed by maintaining a certain degree of sparsity in the feature information of individual faults. Where the first part of the penalty term represents intergroup sparsity and the second part represents intra-group sparsity.  $\odot$  denotes the multiplication between corresponding elements in the matrix.

$$P(\mathbf{x}; \mathbf{b}) = \sum_{i \in \Omega} \phi(\|\mathbf{b} \odot \mathbf{x}_{i,K}\|_2) + \rho M N 1 \|\mathbf{x}\|_{1,w}$$
(20)

 $\phi$  is a non-convex penalty function, such as an absolute value function. The nature of the function is proved in the literature [20]. The convexity of the model is maintained while promoting sparsity. The expression of the group sparse model is formed as follows:

$$\min_{\mathbf{x}_{1},\dots,\mathbf{x}_{2}} \frac{1}{2} \left\| \mathbf{y} - \sum_{l=1}^{L} \mathbf{x}_{l} \right\|_{2}^{2} + \sum_{l=1}^{L} \lambda_{l} P(\mathbf{x}_{l}; \mathbf{b}_{l})$$
(21)

The basic idea of the MM algorithm is to define an easily solvable auxiliary function to replace the optimization problem to be solved when the optimization problem is difficult to solve, and to indirectly achieve the solution of the original optimization problem by iteratively optimizing the auxiliary function several times, thus accelerating the convergence speed and reducing the time of the computation [21]. The optimized model is solved iteratively using the MM algorithm to obtain the decoupled sparse reconstructed signal.

$$\mathbf{x}_{l}^{(k+1)} = \mathbf{Q}_{l}^{-1} soft(p, \mu \lambda_{l} \rho M N_{1} \mathbf{w})$$
(22)

$$\mathbf{Q}_{l}^{-1} = \operatorname{diag}(1 + \mu\lambda_{l}r(\cdot;\mathbf{x}_{l}^{k})) \tag{23}$$

$$soft(\mathbf{x}, t) = sign(\mathbf{x})\max(|\mathbf{x}| - t, 0)$$
(24)

where *k* denotes the number of iterations and l denotes the number of faults.

# 2.2.2. Main Steps in the Proposed Method

In summary, the main steps of the method proposed in this paper are as follows:

- (1) Decoupling of compound fault signals using a periodic group sparse model.
- (2) Wavelet packet energy feature extraction: The single fault features decoupled in the first part are extracted using wavelet packet energy features to fill in the gaps in the reconstructed signal except for sparse non-zero terms, so as to improve the data quality and thus the characterization ability of the dictionary. At the same time, compared with the traditional signal directly used for segmentation, the dimensionality of the data after wavelet packet processing is reduced, which reduces the computational time and increases the differentiation between different health states.
- (3) Dictionary training and learning: Set the category labels of different health states, and import the wavelet packet energy features into the dictionary classification model for dictionary training. By discriminative sparse coding in the model, the dictionary and classifier are obtained by solving the model using algorithms such as K-SVD and OMP.
- (4) Fault classification and identification: The test samples are decomposed on the dictionary to obtain sparse representations of different health state signals, and after calculating the sparse coefficients and classifier products, the index term corresponding to the largest element of them is used for fault identification and classification.

# 3. Simulated Signal Analysis

The proposed method is compared and analyzed with the resonance decomposition approach based on the l1-parametric penalty term to verify the effectiveness of the proposed algorithm. In this section, two simulations are used to verify the analysis. First, a simulation signal containing two different fault components is constructed to show the intra-group sparsity of a single fault and the inter-group sparsity between different faults after decoupling the proposed model. The simulation signal with two fault components of 55 Hz and 30 Hz is shown in Figure 4. In Figure 4a, the apparent periodic component cannot be found. From Figure 4c, it can be seen that the 55 Hz fault feature and its multiplication are obvious, and the 30 Hz component has been drowned out by the noise.



**Figure 4.** Simulated signals of two different components: (**a**) time domain signals; (**b**) frequency domain signal; (**c**) envelope spectrum.

The decoupling separation is performed through the use of RSSD in reference [11] and the method proposed in this paper, and the processing results are shown in Figure 5. The resonance decomposition, as a typical sparse model based on the l1-parametric penalty term, can only keep good inter-group sparsity between single faults since there is no constraint effect of the l2 norm penalty term fails to produce good sparsity among different faults. From Figure 5a,c, it can be seen that the fault frequencies and their multiplicities are not obvious and are affected significantly from the disturbance components. Figure 5b,d show the performance difference between the sparse model based on the periodic group and the sparse model based on the l1 norm. Two different fault features are clearly extracted by the decomposition of the proposed method, and the two fault components maintain good inter-group sparsity, while the within-group sparsity between single fault components is good. In addition, the magnitude of the decoupled signal is not underestimated due to the weighting effect of the vanes in the penalty term, which gives the model good amplitude preservation and maintains a good reconstruction accuracy.



**Figure 5.** Extracted two fault characteristics: (**a**) the 55 Hz fault feature extracted using RSSD; (**b**) 55 Hz fault features extracted using the proposed method; (**c**) the 30 Hz fault feature extracted using RSSD; (**d**) the 30 Hz fault features extracted using the proposed method.

Subsequently, a compound fault simulation signal containing a bearing fault  $x_1(t)$ , a gear fault  $x_2(t)$  and a Gaussian white noise n(t) is constructed as follows:

$$x(t) = x_1(t) + x_2(t) + n(t)$$
(25)

$$x_1(t) = \sum_k a_k h_1(t - kT_1 - \tau_k) \ h_1(t) = e^{-2\xi \pi f_{n1}t} \sin\left(2\pi f_{n1}\sqrt{1 - \xi^2 t}\right)$$
(26)

$$x_{2}(t) = \sum_{m=0}^{M} A_{m} \left[ 1 + \sum_{k=0}^{K} A_{m,k} \cos(2\pi k f_{r} t + \alpha_{m,k}) \right] \cos(2\pi m f_{z} t + \beta_{m})$$
(27)

where, the pulse amplitude of the bearing fault  $a_k = 1$ , the bearing fault period  $T_1 = 1/70$ , resonance frequency  $f_{n1} = 2000$ , damping ratio  $\xi = 0.02$ ; the amplitude of the gear fault signal  $A_m = 0.6$ , the rotation frequency of the shaft where the gear is located  $f_r = 15$  Hz, the gear meshing frequency  $f_z = 200$  Hz, and the phase of the order meshing frequency  $\beta_m$  is 0. The *k*th amplitude  $A_{m,k}$  and phase  $\alpha_{m,k}$  of the *m* order meshing frequency is 0; add a noise with a variance of 0.5.

The simulation signals of the composite fault are shown in Figure 6. Figure 6a,b show the fault signals of the bearing and gear, respectively. Figure 5c shows the simulated signal of the composite fault containing noise, from which no significant periodic pulse is found. The envelope spectrum of Figure 6d reveals only the meshing frequency  $f_m$  and the rotational frequency  $f_r$ , with no bearing-related fault information.

The proposed method is used to decouple the composite fault signal. Where, the values of  $N_0$  and  $N_1$  are set to 4 as recommended in the reference [22]. The processing results are shown in Figure 7. As shown in the time domain plots in Figures 6c and 7a, the periodic fault pulse components of the bearings and gears are very obvious.

From the envelope spectrum in Figure 7b, it can be seen that the fault characteristic frequency  $f_o$  of the bearing and its multiple frequency  $2f_o$  are very obvious. In Figure 6d, the rotational frequency  $f_r$  and its multiples  $2f_r$  and  $3f_r$  are found to be significant. Meanwhile, the information of the fault characteristics centered on the meshing frequency  $f_m$  and with  $f_m + f_r$  and  $f_m - f_r$  as the side bands is clear, thus proving the existence of bearing fault characteristics and gear fault characteristics in the signal, and also verifying that the proposed model has good performance in intra-group sparsity for a single fault and intergroup sparsity between different faults, which shows the effectiveness of the algorithm.



**Figure 6.** Constructed simulation signal of bearing-gear: (**a**) bearing fault signal; (**b**) gear fault signal; (**c**) simulation signal of compound fault of bearing-gear including noise; (**d**) the envelope spectrum of the bearing-gear compound fault.



**Figure 7.** Decoupled two fault signals: (**a**) time domain of bearing fault signal; (**b**) envelope spectrum of bearing fault signal; (**c**) time domain of gear fault signal; (**d**) envelope spectrum of gear fault.

## 4. Experimental Verification and Analysis

4.1. PHM Data Challenge Fault Dataset

Gearbox failure data from the 2009 PHM Data Challenge dataset. These vibration data represent the operating condition of a typical industrial gearbox. The gears on the input and output shafts have 32 and 80 teeth, respectively. The two gears on the intermediate shaft have 96 and 48 teeth, respectively. The primary and secondary ratios are 3 and 1.67, respectively, and one sensor is installed on the input and one on the output side of the gearbox for data acquisition. The gearbox structure used for data collection is shown in Figure 8. In this paper, the performance of the proposed algorithm is tested by using the data set of spur gears with the output side sensor with a significant coupling modulation of the output side signal.



Figure 8. Internal structure of gearbox.

The fault data set contains two groups as shown in the green area (group A) and red area (group B) in Figure 8. Group A contains three fault types: input shaft unbalance, intermediate shaft rolling body fault (input side), and output shaft outer ring fault (input side). Group B contains six fault types: input shaft unbalance, input shaft inner ring fault (input side), intermediate shaft rolling body fault (input side), output shaft outer ring fault (input side) and output shaft so gear broken teeth. The bearing parameters are the following: the rolling body diameter is 7.94 mm, the bearing pitch diameter is 33.5 mm, and the rolling body number is 8. The sampling frequency is 66.67 KHz.

The input shaft speed of 1800 rpm/min in the data set of group A was selected and the fault frequencies were calculated to be 30 Hz, 19.92 Hz and 18.31 Hz. The composite fault signal is shown in Figure 9. The fault features are submerged except for the rotational frequency component.

The proposed algorithm is used to decouple the compound fault signals. The root mean square error (RMSE) is shown in Equation (8). The RMSE indicates the ability to maintain the amplitude of the features.  $\lambda$  and  $\rho$  balance the inter-group sparsity and intra-group sparsity, respectively. The RMSE of the two parameters at different ranges are shown in Figure 10, and  $\lambda = \rho = 1 \times 10^{-4}$  is chosen. Since the noise estimation intensity is generally consistent for both data sets, the same parameters are set for the following categories of operating conditions.

RMSE = 
$$\sqrt{\frac{1}{N} \left( \|\mathbf{x} - \mathbf{f}(\mathbf{x})\|_2^2 \right)}$$
 (28)

The decoupled vibration signals of different states are cut off into segments of 950 points each in length. Each segment is used as a sample. A total of 800 sets of data are selected from the group A fault data set for analysis, where 400 sets are randomly selected as the training data and the remaining 400 sets are used as the test data. The wavelet packet energy features of different states are extracted, and the dimension of sample data is changed from 950\*400 to 32\*400. A total of 1440 sets of data are selected from the failure data set B for analysis, where 720 sets are randomly selected as the training data and the remaining 720 sets are used as the test data. After extracting wavelet packet energy features for different health states, the sample data dimension is transformed from 950\*720 to 32\*720.

When using the proposed algorithm to identify and classify different health states of group A, each parameter is set: the number of atoms of the sub-dictionary corresponding to each category is 80 and the sparsity is 1. Then, the number of atoms of the dictionary is 320 and the sparsity is 4;  $\alpha = 0.001$  and  $\beta = 0.03$ . The fault classification recognition accuracy is shown in the confusion matrix of Figure 11a. The overall recognition accuracy is 97% and the health status recognition accuracy of each component is greater than 90%.

When setting the same parameters as above to be applied to group B, there is a different health state recognition classification; the recognition accuracy is shown in Figure 11b. Due to the increase in health status categories, the classification is relatively difficult and the recognition accuracy decreases slightly. However, it basically meets the recognition requirements. The recognition accuracy of each state is mostly kept around 90%, and the overall recognition accuracy is 91%.



Figure 9. Compound fault signal: (a) time domain; (b) frequency domain; (c) envelope spectrum.



**Figure 10.** Root mean square error of  $\lambda$  and  $\rho$  at different values.



**Figure 11.** The confusion matrix obtained from the proposed method. (**a**) The accuracy of fault identification with four different states. (**b**) The accuracy of fault identification with six different states.

#### 4.2. Evaluate Model Performance

The parameters  $\alpha$  and  $\beta$  control the weights of the relevant terms in the model and have an important impact on the classification accuracy. Therefore, the effect of the two parameters on the recognition accuracy of the model at different values was tested and analyzed using the working condition of 30 Hz in the data set A. As can be seen in Figure 12, the overall recognition accuracy is maintained above 91% when the two parameters are varied from  $10^{-3}$  to  $10^3$ , showing the strong robustness and generalizability of the model.

To demonstrate the superiority of the proposed algorithm, it is compared with various fault identification methods. As can be seen from the recognition accuracies in Figure 13, LC-KSVD obtains a higher accuracy than other classification algorithms. The difference is that the traditional sparse representation classification (SRC) directly draws random samples from the training data to form a dictionary without learning the dictionary, which cannot accurately match the fault features leading to a larger classification error. The K-SVD algorithm learns the dictionary obtained from the training samples, but has no

discriminative capability. It fails to generate corresponding sparse codes for the vibration signals of the same fault class. The discriminative K-SVD (D-KSVD) algorithm adds the classification error term to the objective function, but the discriminative information is contained in the whole dictionary and classifier, and no explicit relationship can be found between the corresponding sub-dictionary atoms and the category labels. And LC-KSVD takes the above defects into account, so that the same category gets similar sparse encoding and further improves the accuracy of fault identification.



**Figure 12.** The accuracy of classification with different parameters  $\alpha$  and  $\beta$  of algorithm.



Figure 13. Performance comparison with different fault identification methods.

Finally, a group A data set with different operating conditions was selected for analysis. The experimental data are selected for analysis under light load and heavy load conditions with six working conditions at input shaft speeds of 2100 rpm/min, 2400 rpm/min and 3000 rpm/min, and the results are shown in Table 1. The results show that the algorithm of this paper is applied to the health status identification and classification under different working conditions, and the feature classification effect performs well.

Fault Categories	Rotational Speed: 2100 rpm/min		Rotational Speed: 2400 rpm/min		Rotational Speed: 3000 rpm/min	
	High Load	Low Load	High Load	Low Load	High Load	Low Load
Normal	97.0	97.0	97.0	98.0	96.0	98.0
Rolling element fault	91.0	84.0	89.0	87.0	93.0	85.0
Outer ring fault	94.0	93.0	90.0	93.0	97.0	88.0
Shaft imbalance	92.0	90.0	91.0	90.0	95.0	91.0
Overall Accuracy	93.5	91.0	91.8	92.0	95.3	90.5

Table 1. Fault classification identification performances under different working conditions.

## 5. Conclusions and Future Work

This paper proposes an intelligent diagnosis method for gearbox compound faults based on a periodic group sparse model and verifies its effectiveness through experiments. First, a group sparse model with overlapping characteristics is constructed by combining the respective advantages of different models. Then, a binary periodic sequence is constructed as prior knowledge of fault components, and the penalty term of the group sparse model is improved to achieve inter-group sparsity among different faults and intra-group sparsity of a single fault. Finally, combining the label-consistent sparse dictionary achieves fault classification and identification under different states.

In the experimental verification, simulation signals containing two different fault components and the PHM Data Challenge gearbox fault dataset were used. The experimental results show that the proposed method exhibits high accuracy and robustness under various conditions. Specifically:

- 1. Simulation Signal Experiment: In the simulation signal containing 55 Hz and 30 Hz fault components, the proposed method successfully extracted their respective fault features without underestimating the amplitude, maintaining high signal reconstruction accuracy.
- 2. PHM Dataset Experiment: Using the PHM Data Challenge gearbox fault dataset under different rotational speeds and load conditions, the proposed method achieved an overall accuracy of 97% in identifying compound faults. In the dataset containing six fault types, the overall recognition accuracy was 91%.

Compared to traditional methods such as Sparse Representation Classification (SRC) and K-SVD, the proposed method improved fault recognition accuracy by 6% and 4%, respectively.

Additionally, tests conducted in actual industrial environments demonstrated that the method can stably and effectively identify gearbox compound faults under various working conditions, further proving its feasibility in engineering applications.

Future research can further explore the following directions:

- 1. Model Optimization: Improving the penalty term and algorithms in the periodic group sparse model to enhance the accuracy and computational efficiency of fault feature extraction.
- 2. Multi-fault Type Recognition: Extending the method to handle more types of compound faults, especially those involving complex mechanical systems with multiple fault scenarios.
- 3. Real-time Application: Applying the proposed diagnostic method in actual industrial environments to verify its real-time performance and stability under different working conditions.

4. Data-driven Optimization: Utilizing big data and machine learning techniques to further optimize the fault diagnosis model, enhancing its adaptability and accuracy.

In summary, the proposed intelligent diagnosis method based on a periodic group sparse model shows promising applications in detecting compound faults in gearboxes and provides valuable references and directions for future research.

Author Contributions: L.C.: Methodology, data curation, writing—original draft, software. X.Z.: Conceptualization, methodology, supervision. Y.F.: Language editing, methodology, review. L.W.: Language editing, review. K.L.: Investigation, supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by National Natural Science Foundation of China (52265016).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

**Conflicts of Interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Chen, X.F.; Guo, Y.J.; Xu, C.B.; Shang, H. Review of Fault Diagnosis and Health Monitoring for Wind Power Equipment. *China* Mech. Eng. 2020, 31, 175–189.
- 2. Beale, C.; Niezrecki, C.; Inalpolat, M. An adaptive wavelet packet denoising algorithm for enhanced active acoustic damage detection from wind turbine blades. *Mech. Syst. Signal Process.* **2020**, *142*, 106754. [CrossRef]
- 3. Pan, C.; Chen, X.; Cai, G. Identification of winding vibration characteristics of three-phase unbalanced transformer based on scale-energy ratio of wavelet packet. *J. Instrum.* **2020**, *41*, 129–137.
- 4. Wang, L.; Shao, Y. Fault feature extraction of rotating machinery using a reweighted complete ensemble empirical mode decomposition with adaptive noise and demodulation analysis. *Mech. Syst. Signal Process.* **2020**, *138*, 106545. [CrossRef]
- 5. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. IEEE Trans. Signal Process. 2014, 62, 531–544. [CrossRef]
- 6. Yang, Z.Z.; Fan, L.; Yang, Y.P. Improved low-rank and sparse decomposition with application to object detection. *Chin. J. Sci. Instrum.* **2019**, *40*, 198–206.
- 7. Jiang, Z.L.; Lin, Z.; Davis, L.S. Label Consistent K-SVD: Learning a Discriminative Dictionary for Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2651–2664. [CrossRef]
- 8. Ding, S.; Zhao, G.; Liberatore, C.; Gutierrez-Osuna, R. Learning Structured Sparse Representations for Voice Conversion. *IEEE-ACM Trans. Audio Speech Lang. Process.* 2020, *28*, 343–354. [CrossRef]
- Sun, R.B.; Yang, Z.B.; Zhai, Z.; Chen, X.F. Sparse representation based on parametric impulsive dictionary design for bearing fault diagnosis. *Mech. Syst. Signal Process.* 2019, 122, 737–753. [CrossRef]
- 10. Zhang, X.; Liu, Z.; Wang, L.; Zhang, J.; Han, W. Bearing fault diagnosis based on sparse representations using an improved OMP with adaptive Gabor sub-dictionaries. *Isa Transactions* **2020**, *106*, 355–366. [CrossRef]
- 11. Xia, J.; Jia, M.P. Rolling bearing fault diagnosis with a resonance-based sparse decomposition and squirrel optimization algorithm. *J. Vib. Shock* **2021**, *40*, 250–254.
- 12. Zhao, Z.; Qiao, B.; Wang, S.; Shen, Z.; Chen, X. A weighted multi-scale dictionary learning model and its applications on bearing fault diagnosis. *J. Sound Vib.* **2019**, *446*, 429–452. [CrossRef]
- 13. Wang, H.Q.; Ren, B.Y.; Song, L.Y.; Dong, F.; Wang, M. Sparse Representation Method Based on Termination Criteria Improved K-SVD Dictionary Learning for Feature Enhancement. *J. Mech. Eng.* **2019**, *55*, 35–43. [CrossRef]
- 14. Li, S.Y.; Mao, W.P.; Cheng, Z.H. Compressed sensing method for leakage vibration signal in water-supply pipelines based on VMD and K-SVD dictionary learning. *Chin. J. Sci. Instrum.* **2020**, *41*, 49–60.
- 15. Bayram, I.; Bulek, S. A Penalty Function Promoting Sparsity Within and Across Groups. *IEEE Trans. Signal Process.* **2017**, *65*, 4238–4251. [CrossRef]
- 16. Ding, C.C.; Zhao, M.; Lin, J. Sparse feature extraction based on periodical convolutional sparse representation for fault detection of rotating machinery. *Meas. Sci. Technol.* **2021**, *32*, 015008. [CrossRef]
- 17. Huang, W.; Song, Z.; Zhang, C.; Wang, J.; Shi, J.; Jiang, X.; Zhu, Z. Multi-source fidelity sparse representation via convex optimization for gearbox compound fault diagnosis. *J. Sound Vib.* **2021**, *496*, 115879. [CrossRef]
- Zhao, Z.; Wang, S.; Sun, C.; Yan, R.; Chen, X. Sparse Multiperiod Group Lasso for Bearing Multi-fault Diagnosis. *IEEE Trans. Instrum. Meas.* 2020, 69, 419–431. [CrossRef]
- He, W.; Ding, Y.; Zi, Y.; Selesnick, I.W. Sparsity-based algorithm for detecting faults in rotating machines. *Mech. Syst. Signal Process.* 2016, 72–73, 46–64. [CrossRef]

- 20. Chen, P.Y.; Selesnick, I.W. Group-Sparse Signal Denoising: Non-Convex Regularization, Convex Optimization. *IEEE Trans. Signal Process.* **2014**, *62*, 3464–3478. [CrossRef]
- 21. Sun, Y.; Babu, P.; Palomar, D.P. Majorization-Minimization Algorithms in Signal Processing, Communications, and Machine Learning. *IEEE Trans. Signal Process.* 2017, *65*, 794–816. [CrossRef]
- 22. Zhao, Z.B.; Wu, S.M.; Qiao, B.J.; Wang, S.B.; Chen, X.F. Enhanced Sparse Period-Group Lasso for Bearing Fault Diagnosis. *IEEE Trans. Ind. Electron.* 2019, *66*, 2143–2153. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.