A Novel Fault Diagnosis Method for Rolling Bearing Based on Improved Sparse Regularization via Convex Optimization

Dongjie Zhong,1,2 Cancan Yi,1,2,3 Han Xiao,1,2 Houzhuang Zhang,1,2 and Anding Wu4

1Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China
2Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan 430081, China
3Engineering Research Center for Metallurgical Automation and Measurement Technology of Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China
4Wenzhou Special Equipment Inspection and Research Institute, Wenzhou 325007, China

Correspondence should be addressed to Cancan Yi; meyicancan@wust.edu.cn

Received 12 January 2018; Revised 9 April 2018; Accepted 18 April 2018; Published 3 July 2018

Academic Editor: Minvydas Ragulskis

Copyright © 2018 Dongjie Zhong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Structural health monitoring and fault state identification of key components, such as rolling bearing, located in the mechanical main drive system, have a vital significance. The acquired fault signal of rolling bearing always presents the obvious nonlinear and nonstationary characteristics. Moreover, the concerned features are submerged in strong background noise. To handle this difficulty, a novel fault signal denoising scheme based on improved sparse regularization via convex optimization is proposed to extract the fault feature of rolling bearing. In this paper, the generalized minimax-concave (GMC) penalty is first researched to promote the sparsity of signal, which is based on traditional L1-norm and Huber function. It is designed to estimate the sparse solutions more accurately and maintain the convexity of the cost function. Then, the GMC penalty is extended to 1-D first-order total variation (TV) as nonseparability and nonconvex regularizer. Thus, a convex optimization problem, which involves a quadratic data fidelity term and a convex regularization term, is developed in this paper. To accelerate the convergence of the algorithm, it is solved by forward-backward (FB) iterative algorithm and thus the denoised signal can be obtained. In order to demonstrate its performance, the proposed method is illustrated for numerical simulation signal and applied in the feature extraction of the measured rolling bearing vibration signal.

1. Introduction

In the field of prognostics and health management (PHM) to mechanical equipment, the actual collected vibration signal contains wealthy information about operating status [1–3]. It is worthy to note that the faulty characteristic information can always be reflected from the measured signal. Rolling bearing is the key component in the main transmission system, and its operating performance is directly related to the status of the entire plant. When some faults occur in them, it will inevitably affect the work safety and production efficiency [4, 5]. Therefore, analyzing the possible fault characteristics of the rolling bearings has an important practical significance [6–8]. Modern signal processing methods provide the main technical means of rolling bearing fault diagnosis.

Generally, the acquired signal is coupled by useful information and the strong noisy component, which has the typical nonlinear and nonstationary characteristics. Hence, the main task of mechanical equipment structural health monitoring is to effectively remove the noise component and improve the signal-to-noise ratio [9, 10]. Traditional signal processing method is based on inner product operator, which is built on the analyzed signal and basis function. The most representative ones are short-time Fourier transform (STFT) [11] and wavelet transform (WT) [12, 13]. Since the size of the analyzed window is fixed, STFT lacks sufficient capacity to deal with the complex nonstationary signals. The
performance of WT depends on the selection of wavelet basis function and the decomposition level. Then, local mean decomposition (LMD) algorithm is proposed as an adaptive time-frequency analysis method [14]. However, it still has the problem of mode aliasing. Variational mode decomposition (VMD) [15, 16] is proposed based on Wiener filtering, one-dimensional Hilbert transform, and heterodyne demodulation analysis. It is still affected by the selection of penalty parameter and the number of signal components. Presently, synchrosqueezing transform (ST) has attracted much attention due to its properties in time-frequency reassignment [17, 18]. Nevertheless, its performance is unsatisfied since the cross-term interference and poor scale separation.

Essentially, the signal denoising can be achieved by the calculation of a sparse approximate solution to the measured vibration signal. Thus, the novel denoising method based on convex optimization and sparsity has publicly employed in signal processing and image enhancement [19, 20]. It has been successfully applied in the field of mechanical fault diagnosis [21], spectral data processing, and baseline correction [22]. Total variation (TV) is the main content of convex optimization algorithm, which involves a quadratic data fidelity term and a convex regularization term [23]. It is developed on sparse signal models. Based on that, the first-order TV [24] and higher degree TV (HDTV) [25] regularization has been researched. However, the experimental analysis demonstrated that it may work when the signal is piecewise constant and it often produces undesirable staircase artifacts. Subsequently, wavelet total variation denoising [26] is proposed to improve the denoising performance, while the estimation of noise variance and penalty function selection is an inevitable question. Using either (1) $L_1$-norm regularization and convex optimization or (2) nonconvex regularization and nonconvex optimization, the calculation of a sparse approximate solution to a linear system of equations is often performed. However, it tends to provide solutions that deviate the real values. Afterwards, a new penalty is suggested, which is a multivariate generalization of the minimax-concave (MC) penalty and is defined as generalized MC (GMC) penalty [27, 28]. GMC penalty simultaneously involves the generalized Huber function and the regular $L_1$-norm regularization. In other words, it is an innovative nonseparability and nonconvex penalty function, which is used for achieving the sparse enhancement and signal smoothing. A nonseparable penalty has more superiority in meeting the requirement of preserving the convexity of the objective function [29]. The published result fully indicates that the GMC penalty is obviously superior to the common penalty function in noise artifacts removing for time-frequency analysis [27].

For early and weak fault detection, it has the feature of latent and dynamic response as the multifactor coupling and complex transmission path [30]. Since the fault is in its early state, the energy generated during the operation is low, and the signal that can be received by the sensor after attenuation is extremely weak [31]. Therefore, early failure signal is easily submerged by background noise, and effective extraction of these characteristics has been a difficult problem. Inspired by the idea of jointing the nonconvex penalty and convex optimization algorithm, the traditional $L_1$-norm regularization term is replaced by GMC penalty in the TV denoising scheme in this paper so as to effectively realize fault state identification. In order to verify the rationality and feasibility of the proposed method, it is used to analyze the numerical simulation and the actual fault vibration signal of rolling bearings in the bearing test rig. The result demonstrated that the proposed method has obvious advantages over traditional methods such as WT, TV, and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [32].

The rest of the paper is organized as follows. In Section 2, the basic ideas of GMC penalty and the researched convex optimization denoising algorithm are introduced. The simulation signal analysis is described in Section 3. The fault-bearing data from bearing test rig is analyzed in Section 4. Section 5 gives the final conclusions.

2. Theory Descriptions

2.1. Generalized Minimax-Concave (GMC) Penalty. Generally, the objective function is convex, and the optimization problem of constraint variable value in a convex set is called convex optimization problem. The penalty function is a measure of constraint violation, which makes the constraint zero when the constraint is satisfied. The one-dimension time series $x \in R^N$ can be simply expressed as $x = (x_1, x_2, \ldots, x_N)$. Then, the $L_1$-norm and $L_2$-norm is denoted as $\|x\|_1 = \sum_{n=1}^{N} |x_n|$ and $\|x\|_2 = (\sum_{n=1}^{N} |x_n|^2)^{1/2}$ respectively.

The Huber function $s(x)$ can be defined as:

$$s(x) = \begin{cases} \frac{1}{2} x^2, & |x| \leq 1 \\ |x| - \frac{1}{2}, & |x| \geq 1. \end{cases}$$  \hspace{1cm} (1)

The minimax-concave (MC) penalty $\phi(x)$ is defined as:

$$\phi(x) = \begin{cases} |x| - \frac{1}{2} x^2, & |x| \leq 1 \\ \frac{1}{2}, & |x| \geq 1. \end{cases}$$  \hspace{1cm} (2)

Figures 1(a) and 1(c) have illustrated the Huber function and MC penalty. Hence, the relationship between Huber function and MC penalty can be described as

$$\phi(x) = |x| - s(x).$$  \hspace{1cm} (3)

Defining the scaled versions of the Huber function and MC penalty is convenient. For $b \neq 0$, the scaled Huber function $s_b(x)$ is determined by

$$s_b(x) = \begin{cases} \frac{1}{2} b^2 x^2, & |x| \leq \frac{1}{b^2} \\ |x| - \frac{1}{2b^2}, & |x| \geq \frac{1}{b^2}. \end{cases}$$  \hspace{1cm} (4)
Similarly, the scaled MC penalty function \( \phi_b(x) \) is given by
\[
\phi_b(x) = \|x\|_1 - s_b(x).
\]  

Figures 1(b) and 1(d) demonstrate the scaled Huber function and scaled MC penalty, respectively. Let \( B \in \mathbb{R}^{M \times N} \) and we next define the generalized Huber function \( S_B(x) \) as
\[
S_B(x) = \inf_{v \in \mathbb{R}^N} \left\{ \|v\|_1 + \frac{1}{2} \|B(x - v)\|_2^2 \right\}. 
\]  

Then, we propose a multivariate generalization of the MC penalty. The basic idea is to generalize (6) using the \( L_1 \)-norm and the generalized Huber function. Thus, we define the generalized MC (GMC) penalty function \( \psi_B(x) \) as follows:
\[
\psi_B(x) = \|x\|_1 - S_B(x). 
\]  

\subsection*{2.2. The Denoising Algorithm Based on Convex Optimization.}

Let \( y \in \mathbb{R}^M \) be the original observed signal, \( A \in \mathbb{R}^{M \times N} \) and \( \lambda > 0 \) is the regularization parameter. A commonly used approach to obtain an optimal sparse approximate solution is to minimize the following objective function [24]:
\[
Q(x) = \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1. 
\]  

Undoubtedly, this is a typical convex optimization problem of first-order TV denoising, which is comprised by a quadratic fidelity term and an \( L_1 \)-norm regularization term. To achieve the signal denoising and early fault feature extraction, a sparse regularization denoising algorithm based on convex optimization is proposed in this paper. Specifically speaking, the traditional \( L_1 \)-norm regularization item is replaced by the GMC penalty term and it is extended to the first-order TV denoising scheme. Thus, we redefine the objective function \( F : \mathbb{R}^N \rightarrow \mathbb{R} \) as follows:
\[
F(x) = \frac{1}{2} \|y - Ax\|_2^2 + \lambda \psi_B(Dx), 
\]  

where \( A \) is an oversampled inverse discrete Fourier transform, \( \psi_B : \mathbb{R}^N \rightarrow \mathbb{R} \) is the generalized MC (GMC) penalty defined by (7), \( \lambda \) is the regularization parameter, and the symbol of \( Dx \) stands for the TV operator. The bidirectional matrix \( D \in \mathbb{R}^{(n-1)\times n} \) is defined as
\[
D = \begin{bmatrix}
-1 & 1 \\
-1 & 1 \\
M & M \\
-1 & 1
\end{bmatrix}
\]  

The penalty \( \psi_B \) is parameterized by a matrix \( B \) and the convexity of \( F(x) \) depends on \( B \) being suitably prescribed. It also should be pointed out that the choice of \( B \) will depend on \( A \). If
\[
B^TB \leq \frac{1}{\lambda} A^TA,
\]  

then \( F(x) \) is a convex function. It is easy to satisfy the convexity condition. Given the matrix \( A \in \mathbb{R}^{M \times N} (N \geq M) \), we may simply set
\[
B = \sqrt{\frac{N}{\lambda}} A.
\]
where $M$ is the signal length and $N$ is the transform length. It should be pointed out that $0 \leq \gamma \leq 1$ always meets the convexity condition. Generally, we use a nominal range of $0.5 \leq \gamma \leq 0.8$ so as to obtain better performance. In order to minimize the cost function by using the approximate algorithms, we rewrite it as a saddle-point problem:

$$
(x^{opt}, v^{opt}) = \arg \min_{x \in \mathbb{R}^n} \max_{v \in \mathbb{R}^n} F(x, v),
$$

where $F(x, v) = 1/2\|y - Ax\|_2^2 + \lambda \|Dx\|_1 - \lambda \|v\|_1 - (\gamma/2)\|A(Dx - v)\|_2^2$ is the saddle function and $x^{opt}$ presents the denoised signal.

According to abovementioned, it is obvious that the selection of regularization parameter $\lambda$ also has a significant influence in denoising performance. Commonly, we chose $\lambda = 2$ to achieve the convergence of algorithm. Therefore, the forward-backward (FB) iterative algorithm can be used to solve the problem $F(x, v)$ of this kind of saddle point [33]. The resulting iterative threshold algorithm uses the soft-threshold function, which is defined as

$$
\text{soft}(y; \lambda) = \begin{cases} 
0 & |y| \leq \lambda \\
(y - \lambda \text{sign}(y)) & |y| \geq \lambda.
\end{cases}
$$

The flowchart of the presented method in this paper is plotted in Figure 2.

3. Simulation Signal Analysis

Without loss of generality, the numerical signal analysis is used to simulate the rolling bearing fault feature identification. It is composed by frequency modulation signal,
Figure 3: The synthetic signal of simulation.

Figure 4: The result provided by wavelet denoising algorithm.

Figure 5: The result provided by the typical total variation (TV) denoising.
harmonic signal, and strong background noise components.

The numerical simulation signal can be expressed as follows:

\[ s_1 = 0.5 \cos(2\pi f_2 t) + \sin(2\pi f_1 t) \],
\[ s_2 = 0.3 \sin(2\pi f_3 t) \],
\[ s_3 = 1.5 \cdot \text{randn}(1, 1024) \],
\[ y = s_1 + s_2 + s_3 \].

where \( s_3 \) is expressed a Gaussian white noise with a variance of 1.5, the composite signal \( y \) is a typical noisy multicomponent signal and the feature frequency is, respectively, set as \( f_1 = 30 \) Hz, \( f_2 = 200 \) Hz, and \( f_3 = 100 \) Hz.

Figure 3 shows the time and frequency responses of the synthetic signal \( y \) with strong noisy signal components. Figure 3(b) suggests that only the feature frequency \( f_2 \) and \( f_3 \) can be inspected, and the baseline of frequency spectrum has generated a drift. Unfortunately, the phenomenon of frequency modulation, namely, \( f_2 \pm f_1 \), is hardly identified since the interference of noise components. Nevertheless, frequency modulation is an important tool in fault state identification. Thus, we can make a conclusion that the traditional frequency spectrum analysis method should be improved.

The immediate idea is that an effective denoising algorithm is performed before frequency spectrum analysis.

Subsequently, the commonly used method wavelet denoising and total variation (TV) denoising algorithm have been employed to analyze the complex simulation signal. The wavelet denoising is performed by wavelet packet scheme. Moreover, the wavelet base function is determined as “db5,” and the decomposition level is selected as 5. Then, the signal reconstruction is achieved using the
coefficient of the node (3, 0). Figure 4 has plotted the result generated by wavelet analysis in time domain and frequency domain. Judging from Figure 4, we can make a conclusion that the interesting components have been removed, and WT denoising method has failed to feature extraction. Traditional TV denoising algorithm has also been regarded as comparative analysis method in this paper, and the result is shown in Figure 5. According to Figure 5, we observe that most concerned signal components have been deleted, and its performance is still unsatisfied.

Subsequently, CEEMDAN is employed to simulation signal analysis, and the result is drawn in Figure 6. Original signal is decomposed into nine intrinsic mode functions (IMF). For the parameter setting for CEEMDAN algorithm, the noise standard deviation is 0.2, and the number of realizations is 200. It can be seen from Figure 6 that only the feature frequency $f_1 = 30$ Hz can be found in IMF2 and IMF3. Motivated by the classical first-order total variation (TV) denoising scheme, a sparse regularization method based on convex optimization is presented in this paper. In order to verify the effectiveness of this method, the researched method based on optimization method has also been applied to it. The parameters of the proposed method are chosen as $\gamma = 0.8$ and $\lambda = 2$. Figure 7 shows the result obtained by the proposed method. It is obvious that the feature frequency $f_2$ and $f_3$ can be clearly inspected. Most importantly, the frequency modulation $f_2 \pm f_1$ can also be determined. Comparing Figure 3(b) with Figure 7, it is obvious that the baseline in frequency domain has been corrected. The experimental results completely demonstrate that this proposed algorithm outperforms the classical wavelet denoising and TV method.

The actual noise reduction effect is directly related to the intensity of noise components. Theoretically, the smaller noise intensity will lead to the better actual denoising performance. When the intensity of Gaussian noise varies from 0.5 to 1.5, the root mean square error (RMSE) between the original noisy signal and the denoised signal is listed in Table 1. The result illustrated that the proposed method has advantage against the strong noise, and its robustness has been proved. The algorithms are run on a computer with an Intel Core i3-4160 CPU and 8.0 GB RAM. The computational costs of four signal processing methods for the simulated signal are listed in Table 2. Since the iterated operation with 200 times, the computational efficiency of CEEMDAN is low. It demonstrates that the computational complexity of the proposed method is acceptable.

### 4. Experimental Analyses

Rolling bearing is an important component of rotating machinery. Its main function is to support the mechanical rotary body, reduce the friction coefficient of its motion, and guarantee its accuracy. Practically, the measured bearing failure signal is more complicated than the numerical simulation signal. Hence, the proposed method is also used for experimental signal analysis. The test rig was equipped with a NICE bearing with the outer race fault shown in Figure 8. The red circle indicates the location of the outer ring fault. The following parameters about the rolling bearing are listed in Table 3. The outer race fault conditions is expressed as follows: 300 lbs of load, input shaft rate $f_r = 25$ Hz, and sample rate of $f_s = 48828$ Hz. According to the theoretical calculation, the characteristic frequency of the outer ring fault is determined as $f_o = 80$ Hz.

The time domain waveform and frequency spectrum analysis results of the collected vibration signals are plotted in Figure 9. From the Figure 9(a) of the time response about measured bearing fault signal, the impact characteristics are obvious. However, we cannot detect the characteristic frequency of the outer ring in Figure 9(b). The frequency spectrogram analysis result indicates that the components of measured faulty signal are complicated. As the interference of the strong noisy components, the fault feature frequency can hardly been identified. The ideal result is to remove unnecessary signal components by effective denoising algorithm, and to retain or amplify the related signal components.

Then, the common signal processing methods, such as wavelet analysis and TV denoising, have been applied to the real inspected signal. For the method of wavelet denoising, we perform the wavelet packet decomposition denoising scheme. It should be noted that the wavelet base function is determined as “db5,” and the decomposition level is selected as 3. The coefficient of the node (2, 0) is used to signal reconstruction. Figures 10(a) and 10(b) show the results obtained by wavelet denoising and TV denoising, respectively. There is no obvious feature frequency corresponding to outer ring
fault in Figure 10(a). Figure 10(b) suggested that the method of TV denoising has removed most of the useful signals, which is not suitable for vibration signal processing.

Similarly, CEEMDAN is also used to analyze the experimental data, and the corresponding result is plotted in Figure 11. Unfortunately, only the characteristic frequency of the outer ring fault $f_o$ and its double frequency $2f_o$ can be inspected in IMF2, while the unwanted signal components have not been removed and it interfere with the identification of fault features. Since CEEMDAN still lacks theoretical support and exits the problem of mode aliasing, it fails to analyze the complex nonstationary vibration signal.

Finally, the proposed method based on GMC penalty function is performed to validate the method and illustrate its superiority. The concerned parameter of the proposed algorithm is chosen as $\gamma = 0.5$ and $\lambda = 2$, respectively. The computed result is drawn in Figure 12. Obviously, feature frequency of outer ring fault $f_o$ and its multiple frequencies ($2f_o$, $3f_o$, $4f_o$, $5f_o$, and $6f_o$) had been both identified in Figure 12. Meanwhile, the rotational frequency $f_r$ can also be determined. The abovementioned characteristics fully indicate that the fault occurs in the outer ring, which is in accordance with the actual situation. Through the comparative analysis between Figures 10 and 12, the results show that the proposed method has obvious advantages in the fault state recognition of rolling bearings under the environmental of strong noise.

Theoretically, traditional signal processing methods such as fast Fourier transform and wavelet transform are based on the idea of matching the analyzed signal to the base function. Its performance largely depends on the signal structure and the base function selection. CEEMDAN is a typical signal adaptive decomposition algorithm, while it is restricted by endpoint effect and mode mixing. Distinguished it, convex
optimization denoising algorithm such as the proposed method is a typical iterative optimization process for objective functions. Using the nonconvex penalty function or cost function to realize the high efficiency, it is more suitable for analyzing the complex vibration signals.

### 5. Conclusions

For rolling bearing fault diagnosis, a novel method based on the improved sparse regularization via convex optimization is proposed in this paper. The main findings of this paper include (1) based on $L_1$-norm and the Huber function, a novel nonconvex penalty function has been researched, namely, GMC penalty. It is employed to achieve the sparse representation of the signal and estimate the sparse solutions more accurately. (2) The proposed method is firstly put forward by combining the GMC penalty and TV denoising scheme. It is designed to enhance the performance of noise reduction for the complex vibration signals. (3) Compared with the traditional methods such as WT and TV denoising, the proposed method has better performance, as demonstrated by both numerical and experimental studies.

### Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

### Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this article.

### Authors’ Contributions

Dongjie Zhong and Cancan Yi conceived and designed the experiments. Dongjie Zhong and Han Xiao performed the experiments. Dongjie Zhong and Cancan Yi analyzed the data. Houzhuang Zhang and Anding Wu contributed to materials/analysis tools development. Dongjie Zhong, Cancan Yi, and Han Xiao wrote the paper.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant nos. 51475339, 51605344, and 51375354, the Natural Science Foundation of Hubei Province under Grant no. 2016CFA042, the Applied Basic Research Programs of Wuhan Science and Technology Bureau under Grant no. 2017010201010115, the Engineering Research Center for Metallurgical Automation and Measurement Technology of Ministry of Education (Wuhan University of Science and Technology) under Grant no. MARC201701, Guangxi Key Laboratory of Optoelectronic Information Processing Open Foundation of China under Grant no. KFJJ2017-01, and Wenzhou Municipal Administration of Quality and Technology Supervision Research Project (201709).

### References


