Rolling bearing fault diagnosis based on GMCP sparse enhanced signal decomposition and TFM

Lingjun Xiao, Guozi Fu

Key Laboratory of Metallurgical Equipment and Control Technology, Wuhan University of Science and Technology, Wuhan, 430081; wanliguang@hbnu.edu.cn

Abstract. This paper presents a rolling bearing fault diagnosis method based on GMCP sparse enhancement signal decomposition and TFM. First, the algorithm uses the non-convex penalty function that not only increases the sparsity but also maintains the convexity of the cost function, that is, the characteristics of the generalized maximum minimum concave penalty (GMCP) to obtain a more accurate estimate. Then the time-frequency manifold (TFM) is used to process the signal. The advantage of TFM is its powerful noise reduction performance to reduce the noise of the signal. Finally, the fault frequency is extracted to realize the fault diagnosis of the rolling bearing. Through the analysis of experimental data, the excellent performance and application value of the proposed method in the diagnosis of rolling bearings are confirmed.

1. Introduction

In the field of rotating fault diagnosis, the sparse decomposition method is a signal processing method developed in recent years. It uses the sparsity built on an over-complete dictionary to denoise the original signal, such as matching pursuit (MP), k-means singular value decomposition (k-SVD) and orthogonal matching pursuit (OMP).

In recent years, a new method in the field of signal processing—the sparse auxiliary signal decomposition method based on morphological component analysis (MCA) theory has been proposed, which provides a new direction for signal decomposition. Cai et al. applied MCA to the fault diagnosis of gears and extracted a clear fault frequency [1]. In this study, the L1 norm is used for the specification of the MCA problem, because the L1 norm most effectively causes sparsity in the convex penalty [2]. However, L1 norm regularization usually suppresses the high-amplitude components of sparse coefficients, and suppresses characteristic frequency signals while eliminating noise and interference signals [3]. This will affect the feature frequency extraction and cause fault diagnosis to fail. In recent years, many researches on enhancing sparsity through penalized designs have made progress. In order to enhance the estimation accuracy of high amplitude components, non-convex sparsity penalty is necessary. However, usually the non-convex penalty cannot minimize the convexity of the cost function, and the cost function is often non-convex, and at the same time, irrelevant suboptimal local minimums can be obtained.

In 2014, Professor Selesnick designed a non-convex penalty function that not only increases the sparsity but also maintains the convexity of the cost function, that is, the Generalized Maximum Mini Concave Penalty (GMCP) [4]. Professor Selesnick proved that maintaining the minimization of the convexity of the cost function can be achieved through GMCP. The advantage of GMCP is that it can significantly improve the sparsity and it can estimate the characteristic signal more accurately. Wang S
et al. used the GMCP algorithm for the fault diagnosis of rolling bearings and realized effective
diagnosis of bearing faults [5].

Time-frequency domain signal processing methods have also been studied by many scholars, such
as short-time Fourier transform (STFT), Wigner-Ville distribution, wavelet transform (DWT), etc. In
2000, scholars such as H. S. Seung founded the theory of manifold learning, which found a new way
for time-frequency analysis. Time-frequency manifold (TFM) developed rapidly in subsequent
research [6]. The principle of TFM is to combine TFD with a manifold learning algorithm to achieve
the purpose of noise reduction on the target system in the time-frequency domain. The advantage of
TFM is that it has good denoising performance in the time-frequency domain.

The advantage of GMCP is that it can not only significantly improve the sparsity, but also can
estimate the characteristic signal more accurately, however, its noise reduction effect in the
frequency band is limited. The advantage of TFM is the powerful noise reduction performance, so this
article combines the characteristics of GMCP to better retain the impact components while reducing
noise and the powerful noise reduction performance of TFM for rolling bearing signal processing.
This method effectively extracts the characteristic frequencies, and the effectiveness and
competitiveness of the proposed method are verified by comparison with band-pass filtering, MP, and
wavelet analysis.

2. The basic principle of the proposed method

2.1. Generalized minimax concave penalty (GMCP)
The GMCP algorithm maintains the convexity of its sparse regularization linear least
squares problem through non-convex regularization of the cost function. GMCP is obtained through improvement and
promotion based on the multivariate study of Huber function.
First, for a given matrix \( B \in \mathbb{R}^{M \times N} \), the generalized Huber function is obtained as \( S: \mathbb{R}^N \to \mathbb{R} \)

\[
S(x; \lambda, B) = \min_v \left\{ \lambda \| v \|_1 + \frac{1}{2} \| B(x - v) \|_2^2 \right\}
\]

Function \( s \) is to find the optimal solution of L1 norm regularization. The GMCP function \( \psi: \mathbb{R}^N \to \mathbb{R} \) is defined as follows:

\[
\psi(x) = \lambda \| v \|_1 - S(x; \lambda, B) = \lambda \| v \|_1 - \min_v \left\{ \lambda \| v \|_1 + \frac{1}{2} \| B(x - v) \|_2^2 \right\}
\]

As shown in Fig. 1 (a), it is the generalized Huber function given \( \lambda = 1 \), and Fig. 1 (b) is its GMCP.
It can be observed from Fig. 1 (b) that GMCP itself is nonconvex, and GMCP also has the property of
regularization, that is, the large value is punished no less than the small value.

![Figure 1. Generalized Huber function and GMCP of generalized Huber function](image-url)
The MCA sparse assist method can be used to decompose vibration signals. In order to accurately decompose the components in the composite signal, a GMCP sparse enhancement signal decomposition method is proposed. The signal decomposition based on GMCP is defined as:

$$\psi(x_1, x_2; \lambda_1, \lambda_2, B_1, B_2) = \lambda_1 \| x_1 \|_1 + \lambda_2 \| x_2 \|_1 - \min_{v_1, v_2} \left\{ \lambda_1 \| v_1 \|_1 + \lambda_2 \| v_2 \|_1 + \frac{1}{2} \| B_1 (x_1 - v_1) + B_2 (x_2 - v_2) \|_2^2 \right\}$$

(3)

The essence of GMCP’s generalized maximum-minimum concave penalty algorithm in formula (3) is to optimize the regularization of L1 norm. It does not have a direct calculation formula. Here, it uses the iterative algorithm of double nested loops to estimate: The outer loop iteratively solves the optimal solution $x_{opt}$ and the inner loop to solve the optimal solution $v_{opt}$. Then there are:

$$x_{opt} = \arg \min_{x} \min_{v} \left\{ F_s(x, v) = \frac{1}{2} \| y - Ax \|_2^2 + \| \lambda \odot x \|_1 - \| \lambda \odot v \|_1 - \frac{\gamma}{2} \| A(x - v) \|_2^2 \right\}$$

(4)

See [4] for detailed derivation.

For the equipment vibration signal, the inherent component and the periodic transient component can be decomposed by $\hat{x}_1 = Ax_{opt}^1$ and $\hat{x}_2 = Ax_{opt}^2$ respectively. Among them, the periodic transient component, as the research object of fault diagnosis, is an important basis for judging faults.

2.2. TFM algorithm

TFM is a nonlinear manifold structure inherent in the time-frequency domain, which can be learned by phase space reconstruction (PSR) and manifold learning. The realized TFM is in the form of time-frequency representation, by maintaining the phase of the original signal, and using the inverse transformation of the STFT to achieve synthesis.

TFM theory is very complicated, please refer to the literature [7-8] for its detailed algorithm. Here only introduce the method of learning and synthesizing TFM and the main steps of TFM to reduce the noise of vibration signal $x$.

1) Given a signal $x$, where $N$ data points are represented by $[x_1, x_2, ..., x_N]$, according to the following equation, the size of $m \times n$ (n=N-m+1) is calculated by PSR Data matrix $P$:

$$P_{(i,k)} = x_{k+(i-1)\tau}$$

(5)

Where $j \in [1, N-(m-1)\tau]$ (6)

2) Perform STFT on each row of matrix $P$ with the following formula to obtain $S_j(k, v)$, $j=1, 2, ..., m$, and calculate the corresponding amplitude and phase parts:

$$S_j(k, v) = \sum_{l=-\infty}^{\infty} P_j[l]w[k-l]e^{-\frac{j2\pi}{M}v}$$

(7)

In formula (7), k and v are the positions of the time axis and the frequency axis, respectively, M is the number of discrete frequency points in the STFT, (w(k)) is the short-term analysis window, and $P_j$ is the position of the matrix $P$ of length n j line.

3) Select the frequency band of interest (the main mode can be displayed) to get the time-frequency distribution (TFD) of m. In dynamics, these TFDs can be considered to be produced by m-dimensional manifolds polluted by noise.

4) Use the local tangent space alignment (LTSA) algorithm to calculate TFM, and select the first TFM label for analysis because it has the smallest manifold reconstruction error. The first method reveals the inherent time-frequency structure of the signal, and the influence of noise is small, so it has a good denoising effect.
5) Update the STFT result with the original phase and TFM feature as the new amplitude, and get \( S_j(k, v), j = 1, 2, ..., m \):

\[
S_k^j(v) = S_k(v) e^{\frac{2\pi}{\nu} k \nu}, \quad j = 1, 2, ..., m
\]

6) Through time-frequency synthesis, a new \( m \times n \) data matrix \( \hat{P} \) in phase space is obtained

\[
\hat{P}[k] = \frac{1}{M \omega[0]} \sum_{j=0}^{M-1} S_j(k, v) e^{\frac{2\pi}{\nu} k \nu}, \quad j = 1, 2, ..., m
\]

7) Finally, use the PSR synthesis method to reconstruct the denoising signal \( y(t) \)

\[
y_i = \sum_{q=\{I(j, k)\}} \frac{\hat{P}[q]}{C_i}, \quad i = 1, 2, ..., N; \quad j = 1, 2, ..., m
\]

Where \( \{I(j, k)\} \) is the subscript set of signal elements that meet the requirements of \( k + (j-1)r = i(k \in [1, N - (m-1)r]) \) and \( C_i \) is the number of elements in \( \{I(j, k)\} \).

TFM has the advantage of maintaining the inherent time-frequency structure. Due to the existence of nonlinear components, the signal \( y(t) \) processed by TFM will have an incorrect amplitude compared with the original signal \( x(t) \). LTSA is an optimal alignment algorithm for solving eigenvalue problems. Based on LTSA, the obtained first optimal TFM feature vector \( Te_1 \) should satisfy the following constraints:

\[
Te_1T_{e1}^T = 1
\]

Under this constraint, the TFM feature vector \( Te_1 \) can be uniquely determined. The resulting TFM \( Te_1 \) reveals the inherent time-frequency structure, but in order to meet the constraints, it must lose meaningful amplitude information. Therefore, we hope to combine TFM with sparse decomposition to solve this problem.

After TFM analysis, the obtained signal has good denoising and sparsity. However, due to the constraints of TFM learning, its amplitude information is invalid. Therefore, in this section we use sparse decomposition to extract the sparse component of the TFM result \( y(t) \). These sparse components can be regarded as sparse atoms of the original signal \( x(t) \) in mathematics to realize the idea of signal reconstruction.

3. Experiment analysis

The process of the method proposed in this article is as follows:

---

![Figure 2. Flow chart of the proposed method](image-url)
In order to evaluate the effectiveness of the sparse signal reconstruction method based on TFM in the enhancement of bearing fault features, the proposed method will be applied to analyze the measured data. Then, the application research of measured engineering data is carried out, and rolling bearing is selected as the research object here. In order to evaluate the performance of the proposed method, the GMCP, MP method, and DWT-based denoising method are selected to compare the results of the proposed method.

![Measuring point](image)

1, 3-motor, 2-torque sensor and encoder

Figure 3. Schematic diagram and picture of bearing test bench

The experimental data is obtained from the bearing data published on the Internet by the Data Center of Western Reserve University. The photo of the experimental device is shown in Figure 3. The experimental data of the tested bearing at the transmission end was selected this time, and the fault damage points were artificially generated by electric sparks. The loss parts were in the rolling elements. The sampling frequency was 12 kHz, the rotation speed was 1796 r/min, and the rotation frequency $f_r=29.9$ Hz. The characteristic frequency calculation formula can be obtained: the characteristic frequency of the rolling element $f_b=141.1$ Hz.

The following uses GMCP sparse enhancement signal decomposition, DWT noise reduction, MP reconstruction noise reduction and GMCP combined with TFM methods to analyze the original signals of rolling element failures. The result is shown in Figure 4-8.

![Figure 4](image)

Figure 4. Time domain diagram, envelope spectrum diagram and HHT spectrum diagram of the original signal of bearing rolling element failure

![Figure 5](image)

Figure 5. Time domain time domain, frequency domain and HHT spectrum after GMCP sparse enhancement signal decomposition of the original signal of bearing rolling element fault

![Figure 6](image)

Figure 6. Time domain, frequency domain and HHT spectra of DWT noise reduction of the original signal of bearing rolling element fault
The results of the different methods shown in Figure 4 to Figure 8 clearly show that the time-domain signal of GMCP combined with TFM has shock components at regular intervals, and there is almost no noise between intervals, and the envelope spectrum is very clearly extracted to the rotational frequencies $f_r$ and $f_b$. And the $n$ multiplication of $f_b$, the frequency distribution in TFD is very concentrated, the bandwidth is very narrow, and the resolution is better. This shows that the frequency components are concentrated and there is almost no noise distribution, so it can be judged that the rolling elements of the bearing are malfunctioning. However, the performance of other methods in the time domain, envelope spectrum and TFD distribution is far from the proposed method in terms of noise reduction and feature extraction, and it is impossible to clearly and effectively extract the effective feature frequency and its multiplication. So it is impossible to determine the type of failure of the bearing. In summary, the effectiveness of the proposed method is proved.

4. Conclusion

Based on the idea of combining advantages, this paper proposes a method of combining GMCP sparse enhancement signal decomposition and TFM to realize the diagnosis of faulty bearings. The method combines the advantages of GMCP in significantly improving the sparsity of the signal without suppressing the characteristic frequency and the advantages of TFM with strong nonlinear noise reduction. The measured data proves the effectiveness of the proposed method.

References

[1] Cai G, Chen X, He Z. Sparsity-enabled signal decomposition using tunable Q-factor wavelet transform for fault feature extraction of gearbox[J]. Mechanical Systems and Signal Processing, 2013, 41(1-2): 34-53.
[2] Bruckstein A M, Donoho D L, Elad M. From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images[J]. SIAM Review, 2009, 51(1): 34-81.
[3] Selesnick I W, Bayram I. "Enhanced Sparsity by Non-Separable Regularization[J]. IEEE Transactions on Signal Processing, 2016,64(9): 2298-2313.
[4] Selesnick I W. Sparse regularization via convex analysis[J]. IEEE Transactions on Signal Processing, 2017, 65(17): 4481-4494.
[5] S.Wang, I. Selesnick, G. Cai, Y. Feng, X. Sui, X. Chen, Nonconvex sparse regularization and convex optimization for bearing fault diagnosis, IEEE Trans. Ind.Electron. 2018,65 (6) :7332-7342.
[6] Ding X, Li Q, Lin L, et al. Fast time-frequency manifold learning and its reconstruction for transient feature extraction in rotating machinery fault diagnosis[J]. Measurement,
2019,(141):380-395.

[7] Q. He, Y. Liu, Q. Long, and J. Wang, “Time-frequency manifold as a signature for machine health diagnosis,” IEEE Trans. Instrum. Meas., vol. 61, no. 5, pp. 1218-1230, May 2012.

[8] Q. He, X. Wang, and Q. Zhou, “Vibration sensor data denoising using a time-frequency manifold for machinery fault diagnosis,” Sensors, vol. 14, no. 1, pp. 382-402, 2014.