Extraction method of composite fault features of gear transmission system based on demodulation of NMD and Teager energy operators

Jingyue Wang1,2,3, Yuefu Liu1, Haotian Wang4 and Jiaqiang E3

Abstract
In order to effectively identify and extract the composite fault characteristics of the gear transmission system, a composite fault diagnosis method combining nonlinear mode decomposition (NMD) and Teager energy operator demodulation is proposed. Because the envelope demodulation of Hilbert transform has the disadvantages of large amount of calculation and end effect, it uses Teager energy operator to solve the problem of large amount of calculation, and NMD solves the problem that the fault signal features is not easy to extract under the mode aliasing. First, the NMD method is used to decompose the fault simulation signal, and the nonlinear modal component with practical physical significance is obtained. Secondly, the Teager energy operator demodulation is carried out for the nonlinear modal components, and the demodulation results are analyzed to verify the feasibility of the method. Then, the method is applied to the composite fault diagnosis of gear pitting wear in gear transmission system, and the characteristic frequency obtained from the test data is compared with the calculated characteristic frequency. The comparative analysis shows that the method can separate the fault characteristic frequencies of large and small gears. The comparative analysis with EMD and EEMD methods in simulation signal analysis and experimental research shows that this method is superior.

Keywords
Nonlinear mode decomposition, composite fault of gear transmission system, feature extraction, Teager energy operator demodulation

Date received: 25 April 2020; accepted: 25 November 2020

Introduction
Gearbox is an integral part of most mechanical products. In order to prevent accidental failure and possible life loss, the fault diagnosis of gearbox has received great attention.1,2 The vibration signal of gearbox is usually complex, non-stationary and nonlinear.3 When a composite fault occurs, multiple faults will interfere with each other and be coupled with each other.4 In recent years, time-frequency analysis methods have been widely used in composite fault diagnosis. Wang et al.5 proposed a composite fault diagnosis method for Gearbox Based on improved integrated local mean decomposition. Through the analysis of simulation and experimental signals, it can be seen that the method effectively overcomes the mode mixing phenomenon and reduces the reconstruction error. At the same time, the method avoids the appearance of pseudo components and reduces the calculation amount. It is verified that the improved integrated local mean decomposition method is an effective method to extract composite fault features. Isham et al.6 proposed the VMDEA and ELM combined fan gearbox fault diagnosis method. This study provides a good solution for determining optimized VMD parameters to decompose vibration signals, and also provides a more effective fault diagnosis method for wind turbine gearboxes. Li et al.7 proposed a gearbox composite fault diagnosis method based on SWD-AVDIF. By analyzing the simulation signals and experimental signals of the gearbox

1School of Automobile and Transportation, Shenyang Ligong University, Shenyang, China
2State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing, China
3State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha, China
4School of Automation, Shenyang Aerospace University, Shenyang, China

Corresponding author:
Jingyue Wang, School of Automobile and Transportation, Shenyang Ligong University, Shenyang 110159, China.
Email: sylgwjy@126.com

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
composite fault, the results showed that the method could effectively separate the composite fault signals and strengthen the fault characteristics. Yang et al.\(^8\) put forward a gear fault diagnosis method combining EEMD and Choi Williams. Through the analysis of simulation and experimental signals, the adaptability and effectiveness of this method in gearbox fault diagnosis were verified. The traditional time-domain or frequency-domain analysis method is only suitable for linear stationary signals. The vibration signals of composite faults are mostly non-linear and non-stationary signals. Using only time-domain or frequency-domain analysis methods often fails to effectively extract fault features.\(^9\) Therefore, a time-frequency analysis method should be used in feature extraction of composite faults. Windowed Fourier transform (WFT) and wavelet transform (WT) are more commonly used time-frequency analysis methods in signal processing. Both are suitable for processing nonlinear, non-stationary signals.\(^10,11\) The difference between them is that they analyze the form of the components that appear in the signal. WFT has linear frequency resolution, that is, it is distinguished according to the frequency difference of each component. WT has logarithmic frequency resolution and is distinguished according to the component frequency ratio. According to Heisenberg’s uncertainty principle, for any kind of time-frequency transformation, the frequency resolution and time resolution cannot reach the optimum at the same time. The time resolution of the WT will increase as the frequency increases. The time resolution of the WFT is the same at all frequencies. So we merge the two. Using WT to get high temporal resolution, Use WFT for high frequency resolution. Huang first proposed a new time-frequency analysis method—empirical mode decomposition (EMD). It can decompose complex signals into finite intrinsic mode functions (IMF), which are suitable for nonlinear and non-stationary signals.\(^12\) Since this method came out, it is widely used in fault diagnosis of mechanical equipment.\(^13,14\) Wu and Huang\(^15\) also proposed an EMD method, adding white noise to the signal and then performing EMD decomposition. However, the signals decomposed by EMD and EEMD are prone to produce pseudo components. The non-sinusoidal complex components generated by the same excitation source will be decomposed into multiple simple components, which will increase the number of components. In addition, they are sensitive to noise and their robustness is not ideal. In view of their disadvantages, Iatsenko et al.\(^16\) proposed a new time-frequency analysis method—nonlinear mode decomposition (NMD), which can decompose the given signal into oscillation with practical physical significance and decompose it into nonlinear mode component (NM). In the process of decomposition, the advantages of WT and WFT are integrated, the accuracy of extraction and reconstruction of main signals is improved, and the noise robustness is very strong. At present, the envelope demodulation method used in many domestic and foreign literatures of fault diagnosis is Hilbert transform envelope demodulation.\(^17,18\) But there are also endpoint effects in Hilbert transform, which will affect the whole signal analysis results.\(^19\) Hilbert has a lot of calculations, the calculation of the Teager energy operator is relatively small. The calculation amount of Teager energy operator demodulation is mainly 4N times real number multiplication, equivalent to N times complex multiplication; The calculation amount of Teager energy operator demodulation is about 1/logN of Hilbert demodulation method.\(^20\) When \(N = 1024\), it is 1/10. Teager energy operator can enhance the transient characteristics of the signal and is suitable for detecting the impact component in the signal. This method has high time resolution, and it has good adaptive ability to the instantaneous change of signal. It has low computational complexity and high efficiency.\(^21\)

In this paper, a composite fault feature extraction method combining nonlinear mode decomposition (NMD) and Teager energy operator is proposed. The composite fault signal is decomposed by NMD, and then the obtained nonlinear modal component (NM) is demodulated by Teager energy operator, and the fault type is determined by comparing the obtained fault characteristic frequency with the calculated fault characteristic frequency.

The second part of this paper introduces the basic theory of the proposed method. The third part analyzes the simulation signal. The fourth part compares and analyzes the effectiveness of the three methods for composite fault feature extraction. The fifth part expounds the conclusion of this paper.

**The proposed method**

**Basic theory of NMD**

Under a certain excitation, nonlinear system will produce response, which can be regarded as NM component. It is assumed that the signal consists of a series of NM components and noise.

\[
s(t) = \sum c_i(t) + \eta(t) \quad (1)
\]

Among them, \(\eta(t)\) is noise, and \(c_i(t)\) is NM component. \(c_i(t)\) can be expressed as a series of FM and AM harmonics:

\[
c(t) = A(t)\varphi(t) = \sum A_h \cos \varphi_h \quad (2)
\]

\[
r_h = A_h \cos \varphi_h \quad (3)
\]

Where \(\varphi(t)\) period is 2 \(\pi\), \(r_h(t)\) is the h-th harmonic of \(c_i(t)\). \(r_1(t)\) is the main harmonic. The rest are sub-harmonics (\(h = 2, 3, 4 \ldots\)).

The purpose of NMD decomposition is to decompose the NM signal. The characteristic parameters of
each harmonic \( r_j(t) \) are obtained. Amplitude is \( A_k(t) \), phase is \( \phi_k(t) \), frequency is \( v_k(t) = \frac{\partial \phi_k(t)}{\partial t} \).

To construct WT and WFT:

(1) Wavelet transform the original signal. Given the signal \( s(t) \), WT can be expressed as:

\[
w_r(\omega, t) = s^+(\mu)\psi^\prime \left( \frac{\omega(\mu - t)}{\omega_p} \right) \frac{\omega d\mu}{\omega_p} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega \tau} G(t) \overline{\psi^\prime \left( \frac{\omega(\tau)}{\omega_p} \right)} d\tau
\]

In the formula, \( \omega \) is the frequency, \( \mu \) and \( \varepsilon \) are the other forms of time \( r \) in the time and frequency domains respectively; \( \omega_p \) is the peak frequency of wavelet; \( \psi \) is a normal wavelet basis function; \( \overline{\psi} \) and \( \psi^\prime \) are the Fourier transform and complex conjugate function of \( \psi(t) \) respectively.

(2) Extraction time-frequency ridge \( \omega_p \)

A curve consisting of a series of ridge points is called a ridge curve. Represents the peak frequency of each moment. In the paper, the ridge extraction method proposed by Iatsenko et al.\(^{22} \) is adopted. The ridge \( r_1(t) \) of the main harmonic \( \omega_p(t) \) is a sequence of peaks that maximizes the function path. For wavelet transform (WT), the path function can be expressed as:

\[
S[\omega_p(t)] = \sum_{n=1}^{N} \left| \ln[G_c(\omega_p(t_n), t_n)] - \frac{1}{2} \left( \frac{\Delta \ln \omega_p(t_n)}{\text{std}[\Delta \ln \omega_p]} - \frac{\Delta \ln \omega_p}{\text{std}[\Delta \ln \omega_p]} \right)^2 \right|
\]

Where, \( G_c(\omega, t) \) is the formula represented by equation (4). \( \omega_p(t_n) \) is the expression of ridge \( \omega_p \) when \( t = t_n \); \( t_n \) is the \( n \)-th peak; \( \langle f(t) \rangle \) and \( \text{std}[f(t)] \) represent the average and standard deviation of \( f(t) \) respectively.

(3) Main harmonic reconstruction

After the ridgeline curve is extracted, the main harmonic \( r_1(t) \) needs to be reconstructed. There are two methods to reconstruct the main harmonic: the direct method and ridge method. Because the ridge method has good robustness. Therefore, we first use ridge method to reconstruct the main harmonic. For wavelet transform, the ridge method can be expressed as:

\[
v^r(t) = \omega_p(t) \exp(\delta \ln v_0(t)) \\
A^r(t) = \exp(i \varphi(t)) = \frac{\sum_{n=1}^{N} G_c(\omega_p(t_n), t_n)}{\varphi'(\omega_p(t_n))}
\]

(4) WT or WFT selection criteria

The paper combines the advantages of WT and WFT. It enables NMD to adaptively select the most suitable type of time-frequency transformation. The empirical formula can be expressed as:

\[
R = (1 + V[\nabla_t (\omega(t), v(t))]^{-1} + (1 + V[\nabla_t A(t), v(t)]^{-1}
\]

When \( R < 1 \), WFT is used, and WT is used otherwise.

(5) Determination of reconstruction method

Since there are two reconstruction methods, in order to make the reconstructed harmonics more accurate and reasonable. According to different time-frequency analysis methods and two reconstruction methods, the final reconstruction method is determined.

(1) When WFT is the optimal time-frequency transform, the given signal \( G(s) \) is transformed by WFT:

\[
G_c(\omega, t) = \sum_{n=1}^{N} \left| \ln[G_c(\omega_p(t_n), t_n)] - \frac{1}{2} \left( \frac{\Delta \ln \omega_p(t_n)}{\text{std}[\Delta \ln \omega_p]} - \frac{\Delta \ln \omega_p}{\text{std}[\Delta \ln \omega_p]} \right)^2 \right|
\]

Where, \( g(t) \) is the Gaussian window function. The time-frequency ridge can be extracted according to equation (5). For WFT, according to the difference between the two, it is only necessary to change the logarithmic frequency scale \( \ln \omega_p \) in equation (5) to rule \( \omega_p \). Then the parameters of \( A^{(d, r)}(t), \varphi^{(d, r)}(t), v^{(d, r)}(t) \), respectively, using the direct method and ridge line method refactoring (\( d \) represent direct method, \( r \) represent direct method):

\[
\left\{
\begin{array}{l}
v^{(d)}(t) = \omega_p(t) + \delta \varphi(t) \\
A^{(d)}(t) \exp(i \varphi(t)) = \frac{2G_c(\omega_p(t), t)}{G_c(\omega_p(t), t) - (\omega(t) - (\omega_p))}
\end{array}
\right.
\]

\[
\left\{
\begin{array}{l}

A^{(d)}(t) \exp(i \varphi(t)) = C_g^{-1} \sum_{\omega_p(t)}^{g(t)} G_c(\omega, t) d\omega \\
C_g = \frac{1}{C_g^{-1} \sum_{\omega_p(t)}^{g(t)} G_c(\omega, t) d\omega} \frac{C_g^{-1} \sum_{\omega_p(t)}^{g(t)} G_c(\omega, t) d\omega}{A^{(d)}}
\end{array}
\right.
\]

\[
\left\{
\begin{array}{l}

v^{(d)}(t) = \text{Re} \left[ \sum_{\omega_p(t)}^{g(t)} G_c(\omega, t) d\omega - \omega \right] \\
\overline{\omega} = \frac{1}{C_g^{-1} \sum_{\omega_p(t)}^{g(t)} G_c(\omega, t) d\omega} \frac{C_g^{-1} \sum_{\omega_p(t)}^{g(t)} G_c(\omega, t) d\omega}{A^{(d)}}
\end{array}
\right.
\]
Where, $\delta$ is the discrete correction coefficient.

First, calculate the weighted average $\bar{A}^{(d,r)}$, $\bar{q}^{(d,r)}$, $\bar{v}^{(d,r)}$. The reconstruction method is determined according to the inconsistency of the reconstruction method. For harmonic parameters, the inconsistency can be expressed as:

$$e^{(d,r)}_{\phi} = k^{(d,r)}_{\phi} \sqrt{1 - \left| \exp(\phi(t)) - \phi(t) \right|^2}$$

$$e^{(d,r)}_{v} = k^{(d,r)}_{v} \sqrt{<\left[ \bar{v}^{(d,r)}(t) - \bar{v}^{(d,r)}(t) \right]^2>}$$

Where $k^{(d,r)}_{\phi,v}$ is the coefficient of adjustment performance, $k^{(d,r)}_{\phi,v} = [3.4.2], k^{(d,r)}_{\phi,v} = 1$.

(2) According to formula (11), if $e^{(d,r)}_{\phi} < e^{(d,r)}_{v}$, the direct method is used, otherwise the ridge method is used. According to the above, WT is set as the best, the direct method is used to reconstruct the harmonic parameters.

$$A^{(d)}(t) = \frac{C_{\phi}^{-1} \int \omega(t) w(t)}{\int w(t)}$$

$$\psi^{(d)}(t) = \frac{D_{\phi}^{-1} \sum_{t=0}^{t} \bar{W}(t) w(t)}{C_{\phi}^{-1} \sum_{t=0}^{t} \bar{W}(t) w(t)}$$

$$\gamma^{(d)}(t) = \frac{\gamma^{(d)}(t)}{\gamma^{(d)}(t)}$$

(6) Extraction of sub-harmonics

After obtaining the main harmonic $r_{1}(t)$, its instantaneous frequency is $v^{(1)}(t)$. For the sub-harmonic $r_{h}(t)$, its ridge line $\omega^{(h)}$ is the closest to the peak sequence of $h^{(1)}(t)$ in the direction of amplitude increase. After finding the sub-harmonic ridge line $\omega^{(h)}$, the direct method and the ridge method are used to reconstruct the sub-harmonics $r_{h}(t)$. Then, the optimal sub-harmonic parameter is determined by equation (11).

(7) Identification of true harmonics

The amplitude, phase, and frequency of true harmonics should be related to the main harmonic. After extracting the primary harmonic $r_{1}(t)$ and the secondary harmonic $r_{h}(t)$, we need to determine if they are true harmonics. By using the method of time shift instead of data, the independent zero hypothesis test of the main harmonic $r_{1}(t)$ and the sub-harmonic $r_{h}(t)$ is carried out, so as to identify the authenticity of the harmonic. The alternative parameter of the main harmonic $r_{1}(t)$ can be expressed as:

$$A^{(1)}(\tau) = A^{(1)}(\tau - T_{d})$$

$$\phi^{(1)}(\tau) = \phi^{(1)}(\tau - T_{d})$$

$$v^{(1)}(\tau) = v^{(1)}(\tau - T_{d})$$

$$\Delta T_{d} = \{t_{i = 1,3...N_{0}} / 2(2) \}$$

Where, $N$ is the sampling length, $t_{i}$ is the sampling frequency, $d$ is the number of time shift points of the substitute data and $M$ is the maximum time shift of the substitute data.

Alternative parameter process of subharmonic $r_{h}(t)$:

(1) Move its signal forward in the time domain by $T_{d}/2$; G($\omega$, $\tau + \Delta T_{d}/2$) or W($\omega$, $\tau + \Delta T_{d}/2$).

(2) Use step (6) to select the optimal substitute harmonic parameters $A^{(h)}_{\phi}$, $A^{(h)}_{v}$, $A^{(h)}_{\gamma}$.

(3) Introduce the statistical feature $q_{\gamma}(A^{(h)}_{\gamma})$ to measure the degree of consistency between the primary and secondary harmonics. Its dependence can be expressed as:

$$q_{\gamma}^{(h)} = \exp \left\{ - \sqrt{\frac{<A^{(h)}(i) - A^{(h)}(i)>_2^2>}{<A^{(h)}(i) - A^{(h)}(i)>^2}} \right\}$$

$$q_{\phi}^{(h)} = \exp \left\{ - \sqrt{\frac{<\phi^{(h)}(i) - \phi^{(h)}(i)>^2}{<\phi^{(h)}(i)>} \right\}$$

$$q_{v}^{(h)} = \exp \left\{ - \sqrt{\frac{<\gamma^{(h)}(i) - \gamma^{(h)}(i)>^2}{<\gamma^{(h)}(i)>} \right\}$$

(14) Consistency coefficient can be expressed as:

$$\rho^{(h)}(A^{(1)}_{\phi},A^{(1)}_{\gamma},A^{(1)}_{\gamma}) = (q_{\gamma}^{(h)})^{\alpha_{\gamma}}(q_{\phi}^{(h)})^{\alpha_{\phi}}(q_{v}^{(h)})^{\alpha_{v}}$$

The parameter $\rho^{(h)}$ is the weight of each consistency coefficient $q_{\gamma}^{(h)}, q_{\phi}^{(h)}, q_{v}^{(h)}$ is the consistency coefficient when $\Delta T_{0} = 0$.

(4) In order to reduce the error caused by harmonic interference, the previous true harmonic must be subtracted and then harmonic identification (it is also applicable to the main harmonic). When identifying the sub-harmonic $r_{h}(t)$, the main harmonic $r_{1}(t)$ should be subtracted.

(8) Noise detection.

The extracted true harmonics are added to obtain the NM component, which is then subtracted from the original signal. Until there is only noise in the remaining signal. The surrogate test method with Fourier transform is used to detect whether the residual signal contains noise. The statistical characteristic quantity $D$ of the alternative data is expressed as:

$$D(\alpha_{A},\alpha_{A}) = \alpha_{A}Q[A^{(\phi)}/\alpha_{A}Q[V^{(\phi)]}$$

$$Q[f(x)] = \int \frac{|f(x)|^{2}}{\int |f(x)|^{2} dx} \ln \int \frac{|f(x)|^{2}}{\int |f(x)|^{2} dx}$$

(16)
Where, \( Q[A(e)] \) and \( Q[v(e)] \) are the discrete spectral entropy of amplitude and frequency respectively. The inspection steps are as follows:

1. Calculate the statistical characteristic quantity \( D_0(\alpha_A, \alpha_v) \).
2. Create \( N_s \) Fourier transform alternatives, and calculate the corresponding \( D_s = 1, 2, \ldots, N_s \)(\( \alpha_A, \alpha_v \)).
3. If the confidence level of \( D_s \cdot D_0 \) is greater than or equal to the preset confidence level, the noise test zero hypothesis is rejected. \( D(1,1), D(0,1), \) and \( D(1,0) \) are used as discriminant statistics to test three different ones (\( \alpha_A, \alpha_v \)). If at least one of them negates the zero hypothesis, it can be considered that the signal is not noise, and the test should be continued. The NMD flowchart is as follows (Figure 1):

**Basic theory of Teager energy operator**

The Teager energy operator is a non-linear operator, and the energy operator method is developed on the signal form of equation (17):

\[
x(t) = a(t) \cos[\varphi(t)]
\]  

(17)

An energy operator is defined as follows:

\[
\Psi(x) = \dot{x}^2 - x\ddot{x}
\]  

(18)

\( \dot{x} \) and \( \ddot{x} \) are the first reciprocal and second derivative of \( x(t) \), respectively. There are:

\[
\Psi[x(t)] = [a(t) \cos[\varphi(t)]]^2 + a^2(t) \varphi''(t) \frac{\sin[2\varphi(t)]}{2} + \cos^2[\varphi(t)]\Psi[a(t)]
\]  

(19)
In general, the variation of the modulated signal is much slower than that of the carrier signal, so the amplitude-frequency value can be treated approximately as a constant. Thus there are:

$$C \frac{a(t)}{C_{138}}' = 0,$$

$$y_0(t)' = (20)$$

$$u_{xt}(\tau) = a^2(t) v_2(t) (21)$$

The instantaneous amplitude and phase of the modulated signal are:

$$|a(t)| = \Psi[x(t)] / \sqrt{\Psi[x'(t)]}$$

$$a(t) = \Psi[x(t)] / \sqrt{\Psi[x'(t)]}$$

$$\omega(t) = \sqrt{\Psi[x(t)] / \Psi[x'(t)]}$$

Therefore, this method can be used to demodulate and analyze the signal.

In this paper, NMD method and Teager energy operator demodulation method are combined to perform NMD decomposition for composite fault signals, and then Teager energy operator demodulation for decomposed signals. At the same time, it is compared with EMD and EEMD. The specific fault diagnosis process is shown in Figure 2.

**Simulation analysis**

This paper compares NMD with EMD and EEMD to verify the advantages of NMD. In the paper, the sum of two AM and FM signals and one random signal is selected for simulation analysis. The simulated signals and components are as follows:

$$x(t) = x_1(t) + x_2(t) + 0.1 \cdot randn$$

$$x_1(t) = (1.5 + 0.5 \sin(26\pi t)) \cdot \cos(600\pi t)$$

$$x_2(t) = (1 + \sin(20\pi t)) \cdot \cos(200\pi t)$$

(25)

Among them, randn is random noise of N [0,1], the sampling time is 1 s, and the sampling frequency is 1024 Hz. The simulation signal and each component is shown in the Figure 3:

Next, the NMD, EMD, and EEMD decompositions are performed on the simulation signals. Their decomposition results are shown in Figures 4 to 6 respectively.

It can be seen from the decomposition results that NMD only decomposes the simulated signal into two NM components. There are no redundant other components. The decomposed NM1 and NM2 correspond to the simulation signals X1 and X2 respectively, and are basically consistent. However, after EMD decomposition, we can actually get more components, six of which are IMF1–IMF6, among which IMF4–IMF6 is basically meaningless. After EEMD decomposition, more components can be obtained, the first six are IMF1–IMF6, among which IMF4–IMF6 is basically a pseudo component, which has no practical significance. Therefore, from the perspective of the decomposition
effect of simulation signals, the decomposition effect of NMD is indeed better than that of EMD and EEMD.

Next, the simulated components obtained from NMD, EMD and EEMD decomposition are demodulated by Teager energy operator respectively, obtain the NM simulation component demodulation spectrum (as shown in Figure 7), the IMF simulation component demodulation spectrum after EMD decomposition (as shown in Figure 8) and the IMF simulation component demodulation spectrum after EEMD decomposition (as shown in Figure 9). From the demodulation results of the NM simulation component and the demodulation results of the IMF simulation component decomposed by EMD and EEMD, it can be seen that the NM simulation component can be extracted to 1 and 2 times the characteristic frequency of the simulated signal fault, respectively. The IMF1 simulation component of EMD decomposition can be extracted to 1 frequency and 2 frequency, but the IMF2 simulation component of EMD decomposition can only be extracted to 1 frequency. The IMF1 simulation component of EEMD decomposition can also extract 1 and 2 times of frequency, but the IMF2 simulation component of EEMD decomposition can only extract 1 times of frequency. Although the demodulation results of EMD and EEMD simulation signal components can also get fault characteristic frequency, the number of peak points is less, and the phenomenon of mode aliasing occurs in the demodulation spectrum of EMD and EEMD simulation components. Therefore, the demodulation results show that the decomposition effect of NMD is indeed better than EMD and EEMD, and it can effectively suppress the phenomenon of modal aliasing, the decomposition and demodulation results of the simulated signal illustrate the advantages of the NMD method. Therefore, this paper uses the decomposition method of NMD.

**Experimental analysis of gearbox composite fault diagnosis**

NMD, EMD, and EEMD methods are compared and applied to the complex fault diagnosis of gear-gear transmission system respectively, so as to verify the
Figure 6. EEMD simulation signal decomposition results.

Figure 7. Demodulation spectrum of NM simulation component: (a) demodulation spectrum of NM1 simulation component and (b) demodulation spectrum of NM2 simulation component.

Figure 8. Demodulation spectrum of IMF simulation component decomposed by EMD: (a) demodulation spectrum of IMF1 simulation component decomposed by EMD and (b) demodulation spectrum of IMF2 simulation component decomposed by EMD.
effectiveness of this method in practical application. The equipment model used in the experiment is HFXC-I. The experimental equipment is shown in Figure 10. By manual setting, it is set up as a composite failure of pitting of big gear and wear of small gear. The acceleration sensor is used in the experiment, and the experiment is carried out under the load of 0.2 A, the sampling time is 2 S, the sampling frequency is 5120 Hz, and the measured speed is $n = 825$ r/min. The number of teeth of the small gear and the large gear is $Z_1 = 55$ and $Z_2 = 75$. The pinion shaft is the driving shaft. By calculation, the fault characteristic frequency of the large gear is $f_0 = 10.08$ Hz. The fault characteristic frequency of the pinion is $f_r = 13.75$ Hz. The calculation formula of fault characteristic frequency is shown in equations (26) and (27).

$$f_r = \frac{n}{60}$$  \hspace{1cm} (26)

$$f_0 = \frac{f_r \cdot Z_1}{Z_2}$$  \hspace{1cm} (27)

The original composite fault signals were decomposed by EMD, EEMD and NMD methods, and the first two signals obtained by the decomposition of EMD and EEMD were taken, namely, IMF1 and IMF2. The results are shown in Figures 11 and 12, NMD decomposition results in two NM components as shown in Figure 13. Then the signals IMF1 and IMF2 obtained from the decomposition of EMD and EEMD were demodulated by Teager energy operator respectively and the demodulation results of the EMD decomposed signal is shown in Figure 14, and the demodulation results of the EEMD decomposed signal is shown in Figure 15. Then the two NM components obtained by NMD decomposition are demodulated by Teager energy operator, and the demodulation results is shown in Figure 16.

As shown in Figure 14(a), the demodulation results of the IMF1 component after EMD decomposition only extracted the 1-fold, 2-fold, and 3-fold frequency of the characteristic frequency of the large gear. As shown in Figure 15(a), the demodulation spectrum of IMF1 component decomposed by EEMD extracts the 1-fold frequency, 2-fold frequency, and 3-fold frequency of large gear fault feature frequency, and the 2-fold frequency of small gear fault feature frequency. As shown in Figure 16(a), the NM1 demodulation spectrum only extracted the 1-fold, 2-fold, and 3-fold frequency of the characteristic frequency of the large gear. The demodulation results show that EMD and NMD

![Figure 10. Experimental equipment diagram.](image-url)
can extract the 1-fold, 2-fold, and 3-fold frequency of the fault feature frequency of large gear. Although EEMD can also extract the 1-fold, 2-fold, and 3-fold frequency of the fault feature frequency of large gear, it also extracts the 2-fold frequency of the fault feature frequency of small gear, which shows that EEMD cannot extract the fault feature frequency of large gear alone. As shown in Figure 14(b), the demodulation results of imf2 component after EMD decomposition extract the 1-fold and 2-fold frequency of pinion fault feature frequency, and extract the 1-fold, 2-fold, and 3-fold frequency of large gear fault feature frequency. As shown in Figure 15(b), the demodulation spectrum of IMF2 component after EEMD decomposition only extracted the 1-fold, 2-fold, and 3-fold the fault characteristic frequency of pinion. As shown in Figure 16(b), the NM2 demodulation spectrum only extracted the 1-fold, 2-fold, and 3-fold frequency of the characteristic frequency of the pinion. The demodulation results show that the EEMD and NMD methods can extract the characteristic frequency of the pinion. The effect is better than the EMD method.

Therefore, through the comparative analysis of the demodulation results, we can know that the EMD method can extract the fault feature frequency of the large gear separately, but it can’t extract the fault feature frequency of the small gear separately. The EEMD method can extract the fault characteristic frequency of
the small gear, but it cannot extract the fault characteristic frequency of the large gear. It shows that the EMD and EEMD methods cannot effectively extract the fault characteristic frequency of the large and small gears separately. NMD method can not only extract the fault feature frequency of small gear, but also the fault feature frequency of big gear. The superiority of NMD method is illustrated. The NMD method can separate the fault characteristic frequency of the gears separately. It shows that the NMD method is better than the EMD and EEMD methods. Therefore, NMD method is suitable for composite fault diagnosis of gear transmission system, and the effect is good.

Conclusion

The structure of gear transmission system is complex, and it is easy to be affected by noise and other external factors, so it is not easy to extract fault feature frequency and identify fault type. Therefore, this paper proposes a method of Composite Fault Feature Extraction Based on the combination of nonlinear mode decomposition (NMD) and Teager energy operator.

1. Through comparative analysis with EMD and EEMD methods, it can be seen that the composite fault feature extraction method combining the nonlinear mode decomposition (NMD) and
Teager energy operator proposed in this paper effectively suppresses the modal aliasing phenomenon.

2. Through the analysis of simulation signals and experimental data, we can see that the method proposed in this paper can effectively extract the composite fault characteristics and accurately identify the fault type. The validity of the proposed method is verified.

The proposed method has strong noise robustness and can effectively eliminate modal aliasing. This paper can effectively identify and extract the composite fault features of gear pitting wear. However, the correctness and accuracy of composite fault diagnosis of other mechanical equipment need to be further explored.

Acknowledgements

The authors gratefully acknowledge the support of project funded by the Science and Technology Research Projects of Education Department of Liaoning Province of China (LG201921), Liaoning BaiQianWan Talents Program (2020091031), Natural Science Foundation of Liaoning Province of China (2020-MS-216), State Key Laboratory of Mechanical Transmissions (SKLMT-KFKT-201605), and China Postdoctoral Science Foundation (2017M610496).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Jingyue Wang https://orcid.org/0000-0002-9851-408X

References

1. Singh A and Parey A. Gearbox fault diagnosis under fluctuating load conditions with independent angular re-sampling technique, continuous wavelet transform and multilayer perceptron neural network. IET Sci Meas Technol 2017; 11(2): 220–225.

2. Yang F, Shen X and Wang Z. Multi-Fault diagnosis of gearbox based on improved mulitpoint optimal minimum entropy deconvolution. Entropy 2018; 20(8): 611.

3. Xu J, Zhao R, Zhu Y, et al. Application of local mean decomposition in composite fault diagnosis of rotating machinery. Noise Vib Control 2012; 5: 144–149.

4. Zhang D, Yu D and Zhang W. Energy operator demodulating of optimal resonance components for the composite faults diagnosis of gearboxes. Meas Sci Technol 2015; 26(11): 115003.

5. Wang Z, Wang J, Cai W, et al. Application of an improved ensemble local mean decomposition method for gearbox composite fault diagnosis. Complexity 2019; 2019(7): 1–17.

6. Isham MF, Leong MS, Lim MH, et al. Intelligent wind turbine gearbox diagnosis using VMDEA and ELM. Wind Energy 2019; 22(6): 813–833.

7. Li J, Cheng J, Huang Z, et al. A gearbox compound fault diagnosis method based on SWD-AVDIF. Noise Vib Control 2019; 39(01): 171–176.

8. Yang X, Sheng P, Jing S, et al. Gear fault diagnosis based on EEMD and Choi Williams distribution. Mech Trans 2019; 43(04): 120–124.

9. Yang Y, Zeng M and Cheng J. Research on the scale decomposition method of local features and its component criteria. China Mech Eng 2013; 02: 195–201.

10. Liang X. Research on windowed Fourier transform harmonic detection algorithm and its interpolation improvement. J Electr Power Sci Tech 2007; 03: 40–44.

11. Wang D and Zhang W. Fault diagnosis of rolling bearings based on wavelet packet transform. China Mech Eng 2012; 23(3): 295–298.

12. Huang NE, Shen Z, Long SR, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc A 1998; 454(1971): 903–995.

13. Cheng JS, Yu DJ, Tang JS, et al. Local rub-impact fault diagnosis of the rotor systems based on EMD. Mech Mach Theory 2009; 44: 784–791.

14. Yang B and Suh CS. Interpretation of crack-induced rotor non-linear response using instantaneous frequency. Mech Syst Signal Process 2004; 18(3): 491–513.

15. Wu Z and Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adapt Data Anal 2009; 01(01): 1–41.

16. Iatsenko D, Mcclintock PV and Ste-Fanovska A. Nonlinear mode decomposition: a noise-roubst adaptive decomposition method. Phys Rev E 2015; 9(3): 032916.

17. Feng Z and Qin Y. Fault diagnosis of planetary gearbox based on Hilbert vibration decomposition and high-order energy operator. Vib Shock 2016; 35(5): 47–54.

18. Gengel E and Pikovsky A. Phase demodulation with iterative Hilbert transform embeddings. Signal Process 2019; 165: 115–127.

19. Shi P, Ding X, Li G, et al. An improved EMD method and its application in fault diagnosis of rotating machinery. J Vib Shock 2013; 32(4): 185–190.

20. Liu H, Chen T and Qu L. Energy operator demodulation method and its application in mechanical signal demodulation. J Mech Eng 1998; 34(5): 85–90.

21. Wang T, Feng Z and Hao R. Research on fault diagnosis of rolling bearings based on teager energy operator. Vib Shock 2012; 31(2): 1–5.

22. Iatsenko D, Mcclintock P and Stefanovska A. Extraction of instantaneous frequencies from ridges in time–frequency representations of signals. Signal Process 2016; 125: 290–303.

23. Zhang Y, Tang B and Deng L. Multi-component signal instantaneous parameter estimation and application based on wavelet ridge. J Mech Eng 2014; 20(10): 1–8.