Characteristic signal based on the combination of empirical mode decomposition method and time series AR model
Extraction method

Li Bin-Bin 1,a, Li Xiang-Dong 1,b*, Pu Xu-Yang 1,c, Kang Zhong-Tao 1,d, Li Zong-Yang 2,e, Zhang Hui-Zhuo 3,f, Chen Chen-Xi 1,g, Hu Yue 1,h

1 Institute of Aerospace Technology, China Aerodynamics Research and Development Center, Mianyang, Sichuan, China
2 Institute of Computational Aerodynamics, China Aerodynamics Research and Development Center
State Key Laboratory of Aerodynamics, Mianyang, Sichuan, China
3High Speed Aerodynamics Research Institute, China Aerodynamics Research and Development Center, Mianyang, Sichuan, China
*aemail:LIBINBIN_cardc@163.com
*bemail:lbb@cardc.cc

Abstract: In the process of signal decomposition by wavelet theory, the wavelet basis function is artificially selected based on experience, and the method based on empirical mode decomposition is decomposed according to the time scale of the signal itself. The article uses two methods to decompose a certain segmented frequency conversion signal to obtain the intrinsic modal component matrix and the wavelet decomposition coefficient matrix, then calculates the Hilbert time spectrum of the two decomposition matrices. The calculations show that the false information generated by the empirical mode decomposition signal is obviously more. Therefore, the empirical mode decomposition method is used to decompose the bearing vibration signal, and the stationary natural mode function obtained is very suitable for establishing an autoregressive AR model to extract the power spectrum of each component for analysis. Finally, the characteristic frequencies of different states of rolling bearings are extracted, which provides support for data-driven fault diagnosis.

1.Introduction
As the structure of mechanical equipment becomes more and more complex, the probability of equipment failure is also greatly increased. Once a lot of industrial equipment breaks down, it will often cause huge loss of life and property. Therefore, it is very necessary to study equipment fault diagnosis. The extraction of characteristic signals is the core issue of mechanical equipment health monitoring and fault diagnosis[1]. One of the current mainstream methods is to use vibration signals as the detection value of fault diagnosis, but in general, vibration signals often have strong nonlinear and non-stationary characteristics. Extracting the characteristic signal in its time series has always been a very difficult problem. In recent years, more and more scholars have carried out research on the fault diagnosis of rotating machinery. Several methods have emerged in the extraction of characteristic signals,
such as the Fourier spectrum analysis method, which is processing some relatively stable signals. It has a great advantage in processing strong non-steady-state signals, but it is incapable of handling strong unsteady-state signals. Later, people proposed improved short-time Fourier method, wavelet theory method, Wiger-Ville distribution and other methods. These methods can better describe the internal information of unsteady signals, but the mathematical methods for signal processing are ultimately unified to Fourier transform, that is, a basis function is selected to transform the signal, and it does not depend on the information of the signal itself. Therefore, the use of these methods to decompose the signal will also bring greater errors.

Empirical Mode Decomposition (EMD) is a new adaptive signal time-frequency processing method creatively proposed by NASA Chinese scientist Huang E and other NASA colleagues in 1998, which is particularly suitable for processing nonlinear non-stationary Signal analysis and processing is considered a major breakthrough in linear and steady-state spectrum analysis based on Fourier transform[1-3]. Starting from the local characteristic time scale of the signal, this method decomposes the complex signal function into the sum of several Intrinsic Mode Functions (IMF). The EMD theory decomposes according to the information of the signal itself, and the obtained IMF component is usually limited in number, and the component signal also contains real physical information, which is very suitable for processing strong nonlinear and non-stationary processes. The AR model is not ideal for the analysis of non-stationary signals, so the AR model cannot be established directly for the vibration signal, but the IMF component obtained by EMD decomposition is stable, and an autoregressive model is established for the IMF component to extract the fault characteristics of the vibration signal. Therefore, this paper combines the EMD and AR models to extract the vibration signal features of rolling bearings. Through the decomposition of a certain segmented frequency conversion simulation signal, it is verified that the EMD decomposition theory has advantages over the wavelet theory. The AR model is used to analyze the power spectrum of each IMF component, and the fault frequency bands of rolling bearings with different health levels under specific motion states are calculated. This can be used as a basis for fault diagnosis and pattern classification.

2. Introduction to the EMD method
EMD is an important part of Hillbert-Huang transform. It can adaptively decompose the signal into several intrinsic modal functions according to the local characteristics of the signal itself, and fundamentally solves the intrinsic basis function, optimal basis selection, constant multi-resolution, and energy brought by the use of basis functions to piece together the signal. Leakage and other issues, so this theory is very suitable for processing nonlinear and non-stationary signals[2-4].

2.1 EMD algorithm principle
The first thing that needs to be explained is that the necessary condition for the function to satisfy the intrinsic modal function is that the number of extreme points and zero points of the curve are equal or differ by at most 1. Secondly, at any point of the curve, the mean value of the maximum extreme point and the minimum extreme point of the envelope is equal to zero. But in fact, most non-stationary signals do not directly meet the IMF conditions. Huang et al. put forward the following hypothesis: any complex signal is composed of some relatively independent IMF components. Each IMF component can be linear or non-linear. Take the signal \(x(t)\) as an example, the EMD process is as follows:

1. Determine all local maximum and minimum points of the signal.
2. Use a cubic spline to connect all the local maximum points to form the upper envelope, and then use a cubic spline to connect all the local minimum points to form the lower envelope. Find the upper, lower the average value of the envelope \(m_1\), and then calculate:

\[
h_1 = x(t) - m_1
\]

(1)

3. If \(h_1\) is IMF, record \(h_1\) as the first IMF component of \(x(t)\). If \(h_1\) is not IMF, use it as the original data and repeat the above two steps \(n\) times to obtain:

\[
h_{1n} = h_{1(n-1)} - m_{1n}
\]

(2)

If \(h_1\) satisfies the conditions of the IMF, then it is the first-order IMF and denote it as \(c_1\). Where
$m_{1n}$ is the mean envelope of $h_{1(n-1)}$.

(4) Subtract $c_1$ from $x(t)$ to get the residual:

$$r_1 = x(t) - c_1$$

Repeat (1)-(3) with $r_1$ as the original data to obtain the second component $c_2$ of $x(t)$. By analogy, when $r_n$ is a monotonic function and no more IMF components can be extracted, the loop ends and $n$ IMF components are obtained, namely:

$$x(t) = \sum_{i=1}^{n} c_i + r_n$$

Among them, $r_n$ is the residual function, which represents the average trend of the signal.

Using the standard deviation SD of two consecutive processing results as the termination criterion of signal decomposition, the mathematical model is:

$$SD = \sum_{t=0}^{T} \frac{|h_{1(n-1)} - h_{1n}|^2}{h_{1(n-1)}^2}$$

The value of SD is used to control the number of iterations of the process of decomposing the signal, so that the IMF can ensure its linearity and stability without losing its physical meaning.

2.2 Verification of calculation examples based on the comparison of Hilbert-Huang transform and wavelet transform

Take the segmented variable frequency simulation signal $x(t)$ as an example:

$$x(t) = \sin 0.2\pi t_1 + \sin 0.4\pi t_2 + \sin 0.9\pi t_3, [t_1 \in (0,100), t_2 \in (101,200), t_3 \in (201,300)]$$

Figure 1. Time domain diagram of simulated signal $x(t)$

Perform EMD decomposition and wavelet decomposition on the signal $x(t)$ to obtain the IMF component matrix and wavelet coefficient matrix and calculate the Hilbert spectrum of the two. The following figures (a) and (b) are the Hilbert time-frequency map of the simulated signal $x(t)$.

(a) $x(t)$ Hilbert graph based on EMD
(b) $x(t)$ Hilbert graph based on wavelet

Figure 2. Hilbert time-frequency diagram of the simulated signal $x(t)$
Figure 3. Hilber contour map and marginal spectrum of the simulated signal x(t)

Obviously, the Hilber time-frequency diagram based on EMD decomposition has almost no false components, while the Hilber spectrum based on wavelet decomposition produces more false components in the high-frequency part. This is because the wavelet only uses the selected basis function for the low-frequency part of the signal. In order to resolve the decomposition, the high-frequency part is not decomposed, so there are many false redundant components in the graphs of 2-(b) and 3-(b), and the high-frequency part is far more than the low-frequency component. And 2-(a) and 3-(a) basically reflect the time-frequency information of the original signal, 2-(a) you can see that the instantaneous frequencies of the three segmented signals are 0.1Hz, 0.2Hz, 0.45Hz, so The effect of empirical mode decomposition on signal decomposition is significantly better than wavelet algorithm.

3. Basic principles of AR spectrum estimation

AR model is the most basic and most widely used mathematical model in time series analysis methods, and it is a linear model for studying stationary random signals\(^1\,\,3\,\,4\). It is developed on the basis of linear regression model, which condenses the characteristics and working status of the system, and the power spectrum estimation based on AR model parameter modeling can effectively improve the frequency resolution of the power spectrum estimation. The basic idea of the AR model is: first establish an AR model for the time series, and then use the model coefficients to calculate the self-power spectrum of the signal. The mathematical model is:

\[
    x(n) = -\sum_{k=1}^{N} a_k x(n-k) + \omega(n)
\]  

(6)

Where \( a_k \) is the AR model parameter of the autoregressive parameter model. K is the order of the model. \( \omega(n) \) is white noise. When using the autoregressive model, the mathematical model of the power spectral density is:

\[
    P(\omega) = \frac{\sigma^2}{1 + \sum_{k=1}^{N} a_k e^{-j\omega k}}
\]  

(7)

This paper uses AR algorithm to calculate the AR spectrum of each IMF component of the bearing vibration signal.

4. Feature signal extraction based on EMD-AR

The AR model is mainly used for the analysis of stationary signals, while the bearing vibration signal turns into a more complicated substationary signal, and the analysis effect is poor. Therefore, before establishing the AR model, this article uses the EMD method to preprocess the vibration signal, and decompose the complex signal into the sum of several IMF components. Each IMF component has only one frequency component at any time, and their mean value is zero and relative for single-component signals that are locally symmetric on the time axis. The complex signals are smoothed, and then
combined with the AR power spectrum estimation algorithm to extract the eigenvalues of the vibration signal.

![Time domain diagram of vibration signal](image1)

![First six IMF components of healthy bearing vibration signal](image2)

Figure 4. Time domain diagram of vibration signal and EMD decomposition

This paper takes the vibration signal of rolling bearing under three states of health, inner ring failure, and outer ring failure as an example. The signal is decomposed by EMD to obtain its inherent mode function matrix. The algorithm obtains 14 IMFs according to the SD criterion. Component, through calculation, the energy of the first six IMF components accounted for more than 95% of the total energy, so the AR spectrum analysis of the first six IMF components can basically obtain the characteristic frequency band-power energy value of the original signal. As shown in Figure 4-(a), it is the time domain diagram of the three-channel vibration signal, and 4-(b) is the first six IMF components of the healthy bearing vibration signal after EMD decomposition.

In the same way, the acceleration rotation signal of the outer ring fault and the acceleration rotation signal of the inner ring fault are respectively subjected to EMD decomposition to obtain the first six IMF components of the two, and the AR power spectrum estimation is performed on the IMF, as shown in Figure 5:

![AR power spectrum estimation of the first six IMF components of the vibration signal](image3)

Figure 5. AR power spectrum estimation of the first six IMF components of the vibration signal
Figure 5 shows the AR power spectrum estimation of the first six IMF components of the vibration signal. The power amplitude and energy value of each IMF component in this figure can be used as the characteristic signal. In order to clearly indicate the power amplitude energy change trend of bearing vibration signals of different health levels, the power amplitude energy of the six IMFs are accumulated, as shown in Figure 6:

![Figure 6 The cumulative energy diagram of the six IMF components under three working conditions](image)

It can be seen from Figure 6 that the AR power amplitude energy of the three-channel vibration signal is mainly concentrated below 50000 Hz. In general, the energy of the signal is highest when the inner ring of the bearing fails, followed by the failure of the outer ring, and the vibration signal energy of the bearing in a healthy state is the lowest.

5. Conclusion

(1) As shown in Figure 5, the distribution curves of the first four IMF components of the bearing vibration signals of three different health levels are quite different. The characteristic frequency bands that can be selected in turn are: 20000-50000Hz, 0-40000Hz, 0-20000Hz, 0-6000Hz.

(2) As shown in Figure 6, it is the overall power amplitude energy change trend of the three-channel signal. The frequency bands circled in the figure are about 0-5000Hz, about 5000-13600Hz, and about 27700-42500Hz. The energy value corresponding to each signal in the frequency band is the characteristic signal, which can be used as the input variable of the intelligent classification algorithm such as neural network, and the output corresponding to different fault types.

(3) The empirical mode decomposition method is superior to Fourier transform and wavelet theory in terms of the completeness of mathematical theory, and the false components generated in the signal decomposition process are greatly reduced. Strong nonlinear signals such as vibration signals are not suitable for AR analysis directly. Therefore, EMD decomposition of vibration signals is required first to obtain stable IMF components, and AR analysis of IMF components is carried out to obtain AR analysis of vibration signals indirectly. The combination of time series analysis and EMD effectively extracts characteristic signals, which provides strong support for data-driven failure pattern recognition.

Acknowledgements

This project is supported by the National Natural Science Foundation of China, the fund number is
Mainly study the working conditions and mechanism of gas-liquid coaxial centrifugal nozzle self-excited oscillation. This thesis is one of the phase results produced in the project research.

References

[1] Meng Z, Gu H-Y. Application of AR model under empirical mode decomposition to extract fault features of rotating machinery [J]. Journal of Yanshan University, 35(4): 342-346.

[2] Zhang L-L, Xiao J. A case course of mechanical fault diagnosis technology based on MATLAB [M]. Beijing: Higher Education Press.

[3] Cheng J-S, Yu D-J, Yang Y. Fault diagnosis method of automobile transmission gear based on EMD and AR model [J]. Automotive Engineering, 27 (1): 101-110.

[4] Cui J-G, Zheng X-Q, Li Z-H, et al. Application of empirical mode decomposition and AR in aircraft health diagnosis [J]. Computer Engineering and Applications, 47(14): 204-206.

[5] Chen C-Z, Sun M-X, Zhou B, etc. HHT improvement and its application in wind turbine fault diagnosis [J]. East China Electric Power, 42(6): 1123-1128.

[6] Qin N, Jin W-D, Huang Jin, et al. High-speed train bogie fault feature extraction based on EEMD sample entropy [J]. Journal of Southwest Jiaotong University, 49(1): 27-32.