Application of a flat variational modal decomposition algorithm in fault diagnosis of rolling bearings

Haodong Li¹, Ying Xu²,³, Dong An⁴*, Lixiu Zhang⁴, Songhua Li¹ and Huaitao Shi¹

Abstract
Fault diagnosis of rolling bearings can effectively prevent sudden accidents and is an important factor for the safe operation of mechanical systems. However, traditional time–frequency analysis techniques cannot effectively obtain the fault feature information. In this paper, a flat variational modal decomposition denoising method based on wavelet transform and variational modal decomposition is proposed to solve susceptibility of vibration signal to noise interference and easily obtain fault features. In this method, first, a series of mother wavelets with different periods are designed based on tone-burst signals, in the decomposition process of variational modal decomposition. This method is based on the designed mother wavelet along with wavelet correlation coefficient for the elimination of the components that are superfluous and frequent from each intrinsic mode function. Then, the regression coefficients of the denoise components and the original signal are calculated, and we select the corresponding components with higher regression coefficients to reconstruct the signal. The reconstructed signal is taken as the new original signal to be decomposed again by variational modal decomposition, and the relevant components are analyzed by enveloping the spectrum, so as to effectively remove noise interference and ensure accurate acquisition of fault feature frequency. We apply this method to the rolling bearing fault data and a comparative study is made with variational modal decomposition and empirical mode decomposition algorithms. The results show that the signal-to-noise ratio of the signal is improved by 77% and 44% after being processed by the flat variational modal decomposition method, compared to the empirical mode decomposition and the variational modal decomposition methods.

Keywords
Fault diagnosis, variational mode decomposition, smooth wavelet, envelope spectrum

Introduction
Rolling bearing monitoring and diagnosis have always been a focus of engineering and academic studies both at home and abroad. Until now, monitoring and diagnostic technologies based on different mechanisms such as vibration, acoustics, static electricity, temperature, and ferromagnetic spectrum have been continuously developed.¹,² In the fault diagnosis of rolling bearing, vibration analysis is the frequently used method, whose procedure mainly includes the following steps: original signal preprocessing, bearing fault feature extraction, and fault feature recognition. The key of vibration analysis is to extract the feature information from the vibration signal.³

¹School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang, China
²Research Institute for Micronano Detection and Motion Control, Shenyang Jianzhu University, Shenyang, China
³School of Electro-mechanical Engineering, Guangdong University of Technology, Guangdong, China
⁴Research Center for Analysis and Detection Technology, Shenyang Jianzhu University, Shenyang 110168, China

Corresponding author:
Dong An, School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang 110168, China.
Email: andong@sjzu.edu.cn

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (http://www.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).
However, the vibration signal obtained by the data acquisition device contains various noises. Therefore, it is necessary to effectively filter the noise signal, increase the SNR, and extract useful information contained in the noise, so that we can obtain correct analysis results. However, optimal utilization of signal processing techniques and algorithms, extraction of effective bearing fault signatures from vibration signals comprising interfering signals and weak bearing fault signals for detection and diagnosis of faults have always been a challenge for condition monitoring and fault diagnosis of rolling bearings.\(^4\)

Rolling bearing vibration signal frequency band analysis is performed based on various locations of the fault and the degree of failure. Further, the frequency of the fault features generated must also be different. Fault feature frequency has a wide frequency range, but we can use the corresponding analysis to accurately obtain fault features of bearings based on the characteristics of different frequency bands of the signal. In general, spectrum analysis can be performed directly on the collected vibration signal using an appropriate method, and combining the frequency spectrum of the vibration signal, the operating state of the bearing can be confirmed. However, because of the non-stationary and heavy background noise caused by complicated work conditions, the extraction of the early fault feature from practical signals becomes impossible.\(^5,6\) Therefore, a novel method is desired which can identify the weak fault more accurately and distinguish the various fault features adaptively.\(^7\)

In previous researches, many fault signal processing methods have been proposed. Ali et al.\(^8\) use empirical mode decomposition and artificial neural network to realize automatic rolling bearings fault diagnosis. Pang et al.\(^9\) combined Hilbert transform and principal component analysis, and proposed an improved bearing fault diagnosis method. By analyzing the dynamic characteristics of bearings, Farzin and Kim\(^10\) proposed a reliable FDD model reference observer technology based on bearing vibration data modeling and a high-order super-torsional sliding mode observer (HOSTSMO) technology for diagnostic decision-making. As a new time–frequency analysis technique, empirical mode decomposition (EMD) can decompose non-stationary signals into some intrinsic mode functions (IMFs) self-adaptively.\(^11\) For the first proposal put forward in 1998, a great attention raised by EMD could be attributed to its time–frequency and non-stationary processing ability. Wang et al.\(^12\) through the establishment of IMF energy distribution and bearing state mapping for bearing state identification. Cheng et al.\(^13\) presented a method which combined support vector machine (SVM) and EMD for diagnosis of a rolling bearing’s fault envelop spectrum. In 2005, inspired by EMD, Smith\(^14\) proposed another adaptive signal decomposition method, which named local mean decomposition (LMD); researchers also showed significant attention for this method, while numerous LMD-based diagnosis methods were continuously being proposed. Liu and Han\(^15\) proposed a new method for bearing fault diagnosis under the combination of LMD and multi-scale entropy. Liu et al.\(^16\) proposed a new hybrid fault diagnosis algorithm combining the second-generation wavelet de-noising and LMD, and the application of the bearing fault diagnosis of a locomotive was performed. However, these two methods are easily affected by mode aliasing, end effect and sampling frequency because they are both depend on recursive mode decomposition.\(^17\)

Variational mode decomposition is a recently proposed by Dragomiretskiy and Zosso, which can decompose signals adaptively. It eliminates mode mixing through performing decomposition by solving a constrained variational problem. Viswanath et al.\(^18\) proposed a new method to detect spike of disturbed power signal based on VMD, and experiment shows that the proposed methodology has a great perform in case of single tone signals. Upadhyay and Pachori\(^19\) detected instantaneous voiced/non-voiced of speech signals using VMD and found that the proposed method at different signal-to-noise ratios indicated the effectiveness. Chang et al.\(^20\) proposed a decomposition based on VMD, singular value decomposition (SVD) and the convolutional neural network (CNN) local weak feature information and planetary gear fault diagnosis methods, to accurately identify the different fault types. Zhipeng et al.\(^21\) based on VMD–SVD proposed an adaptive fault diagnosis method of rotating machinery, freed from the prior knowledge and the disadvantages of manual intervention. In Aneesh et al. and Salim,\(^22,23\) a comparison between VMD and other signal decomposition methods such as EMD, EWT, EEMD, etc. were performed, and the experimental results show that both in signal decomposition and feature extraction, the VMD is superior to these traditional decomposition methods. In this paper, based on wavelet transform and variational modal decomposition algorithm, we have proposed a flat variational modal decomposition denoising method. This method constructs a flat wavelet function during the decomposition of the VMD and applies the constructed flat wavelet denoising for the elimination of the components that are superfluous and frequent from IMF. Finally, the regression coefficients between the denoised components and the original signal are calculated, and the components with high regression coefficients are selected for reconstruction. The reconstructed signal is again subjected to VMD as the new original signal, and envelope analysis of the relevant components is performed. The main innovation of this paper is not limited to the traditional VMD signal analysis method, but the introduction of smooth wavelet operator in the decomposition process, and the design of wavelet
operator to achieve denoising. In the simulation experiment, we found that the FVMD method can effectively solve the problem of complete removal of noise due to mode mixing. Compared with the traditional VMD and the EMD algorithms, the FVMD method has obviously improved on the SNR and kurtosis.

This rest part of paper is organized as follows: The next section introduces the principle of VMD and methods for related parameter selection; the subsequent section constructs a new smooth wavelet function, and the proposed smooth wavelet solution for eliminating IMF redundant noise is discussed; then, the algorithm flow of FVMD is discussed in detail and its superior performance with simulation signals is verified. Furthermore, the ability of the proposed method on rolling bearing diagnosis is validated through analysis; the penultimate section deals with the acquisition of fault characteristic frequency of rolling bearing inner ring and outer ring; finally, we summarized the whole manuscript and given our conclusions in the last section.

Variational mode decomposition

The overall framework of the VMD is a constrained variational problem. The VMD algorithm assumes that each modal component closely surrounds a certain center frequency, transforming the modal bandwidth determination process into a constrained variational problem, thereby separating the modal components through its solution. In the VMD algorithm, each component consists of a series of amplitude-modulated and frequency-modulated signals (AM-FM signals or modes), and its expression is given by

$$u_k(t) = A_k(t)\cos(\varphi_k(t))$$ (1)

In the expression: $u_k(t)$ is the harmonic signal, $A_k(t)$ is the instantaneous amplitude of the signal, $\varphi_k(t)$ is the phase of the signal. The corresponding constrained variational model can be described as seeking K-th IMF components with a specific sparseness, so that the sum of the estimated bandwidths of the components is minimized. The constraint is that the sum of the components is equal to the original signal $f(t)$. Following are the steps involved in construction of the model:

1. Each mode undergoes a Hilbert transformation and calculates the analytical signal to obtain the unilateral spectrum corresponding to the component, given by

$$\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)$$ (2)

where, $\delta(t)$ is the Dirac distribution and * means convolution.

2. Estimation of the center frequency $w_k$ of each IMF component and multiplying it with the exponential signal $e^{-j\omega_k t}$.

$$\left[\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)\right] e^{-j\omega_k t}$$ (3)

3. Calculating the square norm $L^2$ of the gradient of the demodulated signal, and the bandwidth of each IMF is estimated. Finally, a constrained variational model is constructed as follows

$$\min_{\{u_k\};\{w_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[ \left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)\right] e^{-j\omega_k t}\right\|^2 \right\}$$

$$\text{s.t.} \sum_{k=1}^{K} u_k(t) = f(t)$$ (4)

where $\delta(t)$ is the Dirac distribution, $j$ is an imaginary unit, * is the convolution, $\partial_t$ denotes partial derivative of function, $u_k(t) = \{u_1, u_2, u_3, \ldots, u_K\}$ and represents the K-th IMF components after decomposition, $w_k(t) = \{w_1, w_2, w_3, \ldots, w_K\}$ represents central frequency of each component.
In order to solve the above variation problem, a penalty factor \(a\) (the default value is \(a = 2000\)) and a Lagrangian multiplier \(\lambda\) are introduced to change the constrained variational problem into an unconstrained variational problem, and the following form of augmented Lagrangian expression is obtained

\[
L(\{u_k\}, \{w_k\}, \lambda) = a \sum_{k=1}^{K} \left\| \frac{\partial}{\partial t} \left[ \left( \frac{\delta(t)}{\pi t} + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jn\omega t} \right\|^2 + \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|^2_2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle
\]

(5)

where \(\lambda\) is the Fourier transform, \(n\) is the number of iterations, \(s\) is the fidelity factor.

Iteration of \(u_k^{n+1}\), \(w_k^{n+1}\) and \(\lambda^{n+1}\) is performed using an alternating direction method of multiplier (ADMM) algorithm. Finally, the solution to the constraint variation problem is given in equations (2) to (4). The search for “saddle point” resulted in the corresponding variable update expression is given by

\[
\hat{u}_k^{n+1}(w) = \frac{f(w) - \sum_{i=1}^{K-1} \hat{u}_i^{n+1}(w) - \sum_{i=k+1}^{K} \hat{u}_i^{n}(w) + \hat{\lambda}(w)}{1 + 2a(w - w_k)^2}
\]

(6)

\[
w_k^{n+1} = \frac{\int_0^\infty w |\hat{u}_k^{n+1}(w)|^2 \, dw}{\int_0^\infty |\hat{u}_k^{n+1}(w)|^2 \, dw}
\]

(7)

\[
\hat{\lambda}^{n+1}(w) = \hat{\lambda}(w) + \tau \left( \frac{f - \sum_{k=1}^{K} \hat{u}_k^{n+1}(w)}{\sum_{k=1}^{K} \left( \frac{\left\| \hat{u}_k^{n+1}(w) - \hat{u}_k^n(w) \right\|^2}{\|\hat{u}_k^n(w)\|^2_2} \right)} \right)
\]

(8)

where \(\wedge\) is the Fourier transform, \(n\) is the number of iterations, \(\tau\) is the fidelity factor.

Iteration of the variational model is performed while the frequency center and bandwidth of each IMF component are updated until the iterative stopping condition is satisfied

\[
\sum_{k=1}^{K} \left( \frac{\left\| \hat{u}_k^{n+1}(w) - \hat{u}_k^n(w) \right\|^2}{\|\hat{u}_k^n(w)\|^2_2} \right) < \varepsilon
\]

(9)

The corresponding algorithm process can be described as follows:

1. Initialization of \(\{\hat{u}_k^1\}, \{\hat{w}_k^n\}, \{\hat{\lambda}\}\) and \(n\).
2. According to the formulae (6) and (7) \(u_k, w_k\) are updated in the frequency domain.
3. The noise margin parameter \(\lambda\) is constantly updated. When \(\lambda = 0\), noise reduction can be performed in a strong noise environment.
4. Repeat steps (2) through (3) until it can satisfy the iteration stop condition, then stopping iteration and obtaining the K IMF components.

When using the VMD algorithm for signal decomposition, it is necessary to manually set the number of decompositions in advance. The study found that in the decomposition, when K is too small, multiple ‘modes’ in the original signal may co-exist in the same IMF component, or some ‘modes’ cannot be recognized effectively, resulting in under-decomposition or leakage decomposition. When K is too large, a certain ‘mode’ in the signal may be ‘pulled’ into multiple IMF components, resulting in excessive decomposition. To address this problem, in this paper, we propose a mean feature method based on instantaneous frequency of components to determine the decomposition modulus K. This method first performs Hilbert transform on the VMD-processed matrix and calculates the instantaneous frequency of the transformed components. In this case, the instantaneous frequency represents the first derivative of the instantaneous phase time function. Finally, the average value of the
instantaneous frequency of each component is calculated, and K is determined according to this value. To make
the above method easier to understand, we introduce a random signal \( A(t) \)
\[
v_1 = \cos(2 \pi f_1 t) \\
v_2 = \frac{1}{4} \cos(2 \pi f_2 t) \\
v_3 = \frac{1}{16} \cos(2 \pi f_3 t)
\]
\[\text{(10)}\]
\[A(t) = v_1 + v_2 + v_3 + 0.1 \times \text{randn(size}(v_1))\]
\[\text{(11)}\]
The signal \( A(t) \) is introduced in the MATLAB program to validate the method. The results are shown in
Figure 1.
As seen from Figure 1, when \( K = 5 \), the curve bends largely downwards. This is because, to obtain the optimal
solution for modality, the VMD algorithm searches the frequency domain in cycles. If the number of decom-
positions is too large, the component will break and flocculate, especially at high frequencies. This way, even at
high frequencies, the average instantaneous frequency is lower, which is also the root cause of the downturn. If the
decomposition continues, the signal will be destroyed, as shown in Figure 1(5), and the average instantaneous
frequency drops rapidly. Therefore, the optimal decomposition number of the above signal is \( K = 4 \).

**Flat wavelet denoising algorithm**

Wavelet transform\(^{25}\) (WT) is a method that can analyze and transform locally in time (space) frequency, and can
gradually multi-scale the signal (function) through the telescoping translation operation, and thereby achieving
time division at high frequencies and frequency subdivision at low frequencies, which can automatically adapt to
the requirements of time–frequency signal analysis. The expression of the wavelet transform is shown as follows.
It can be seen as the sum of the whole signal \( x(t) \) multiplied by a wavelet function (mother wavelet \( \psi(t) \))
\[
W = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} x(t) \psi \left( \frac{t - l}{s} \right) dt
\]
\[\text{(12)}\]

![Figure 1. The average value of the instantaneous frequency of the component.](image-url)
where $W$ represents wavelet coefficients, $s$ and $l$ are scale and shift, respectively. It must be emphasized that the key of the success of wavelet transform or at least its best result is to select a mother wavelet that maximally matches the shape of the signal under analysis. Narrowband tone-burst signals have good locality and are distributed on various scales. However, the energy of noise is concentrated on small scales, and its wavelet coefficient rapidly decays as the scale increases. The expression of the tone-burst signal $X(t)$ is simply a sine wave modulated by windows such as Hamming window:

\[
X(t) = \sin(\omega t + \theta) \left(0.08 + 0.46 \left(1 - \cos\left(\frac{\omega t}{N}\right)\right)\right)
\]

where $\omega$, $t$, $\theta$, and $N$ represent the circular central frequency, time, phase, and number of cycles, respectively. In this paper, a family of analytic tone-burst signals with different number of cycles has been designed as the mother wavelet. Figure 2 shows the waveform of tone wavelet. Although it approximately matches the shape of the analysis signal, its performance at the edge of the signal is not satisfactory. Therefore, the threshold tone wavelet function is designed in this paper.

Noisy signal is produced after wavelet transform, and the noise cannot be completely removed. But its wavelet coefficient has a strong correlation at each scale, especially in the vicinity of sudden change of the signal, and this correlation is more obvious. However, the noise-corresponding wavelet coefficients have no obvious correlation among the scales. Therefore, the correlation between the corresponding points of the wavelet coefficients at different scales can be used to determine whether the point is the signal coefficient or the noise coefficient, so that it can be chosen. At the same time, after the noisy signal is denoised by the wavelet coefficient correlation, some noise corresponding wavelet coefficients may still remain in the vicinity of the signal mutation, so threshold denoising and smoothing processing are also required. Therefore, in this paper, we propose a cross-correlation threshold smoothing denoising algorithm based on wavelet correlation coefficients.

Consider a noisy signal, $f(k) = s(k) + n(k)$, where, $s(k)$ is the original signal, $n(k)$ is Gaussian white noise. We applied wavelet transform on the signal $f(k)$ so as to obtain the wavelet coefficient $w_{ik}$, which represents the wavelet coefficient at the position $k$ on the scale $i$. Wavelet coefficient $w_{ik} = u_{ik} + v_{ik}$, where $u_{ik}$ and $v_{ik}$ are the wavelet coefficient corresponding to original signal and the noise signal, respectively. Assuming that the number of scales participating in the wavelet coefficient correlation operation is $L$, the correlation coefficient at the position $k_1$ on the scale $i$ can be defined as

\[
C_{Wi,k_1} = w_{i,k_1} \cdot w_{i+1,k_1}, \quad i < L + 1, k_1 = 1, 2, 3, \ldots, k
\]

Here, $C_{Wi}$ represents the wavelet energy at scale $i$ position $k$. In order to make the correlation coefficient of each scale comparable to the wavelet coefficient, the energy of $C_{Wi}$ is normalized. The expression of normalized correlation coefficient can be defined as follows

\[
\tilde{w}_{i,k_1} = C_{Wi,k_1} \left[\frac{P_{Wi}}{P_{CWi}}\right]^{1/2}
\]

where $P_{Wi} = \sum_k w_{ik}^2$, $P_{CWi} = \sum_k C_{Wi,k}^2$, $P_{Wi}$, $P_{CWi}$ represent respectively the energy of wavelet coefficients and correlation coefficients corresponding to the scale $i$.

Figure 2. Five-cycle tone-burst.
Comparison of the absolute value of normalized correlation coefficient and wavelet coefficient is done, by which: if \(|\tilde{w}_{i,k_1}| > |\tilde{w}_{i,k_1}|\), the wavelet coefficient at point \(k_1\) on the scale, \(i\) is considered to belong to the wavelet coefficient of the original signal, and \(w_{i,k_1}\) is assigned to \(u_{i,k_1}\); if \(|\tilde{w}_{i,k_1}| < |\tilde{w}_{i,k_1}|\), the wavelet coefficient at point \(k_1\) on the scale, \(i\) is considered to belong to the wavelet coefficient of the noise signal, and \(w_{i,k_1}\) is assigned to \(v_{i,k_1}\). The above steps are repeated until the variance of \(v_{i,k}\) satisfies a predetermined noise level-related threshold

\[
T_{H_j} = \|\psi\|^2 \sigma^2 / j
\]

Here, this paper refers to the threshold of noise level defined in Zhao et al.,\(^{29}\) where \(\|\psi\|\) represents the norm of the generating wavelet, and \(\sigma^2\) represents the noise variance on the scale \(i\), and the \(w_{i,k}\) is the wavelet coefficient of the useful signal. But the signal will inevitably deviate from the original signal after denoising by wavelet coefficient correlation. In this study, we take the scale wavelet coefficient \(w_{i,k}\) as a signal variable and assign it to \(x(k)\). At the same time, the original noisy signal \(f(k)\) is treated as another signal variable and is assigned to \(y(k)\). Since \(x(k)\) is obtained by \(y(k)\) after wavelet transform, both have similar signal characteristics. Next, we perform a cross-correlation analysis of \(x(k)\) and \(y(k)\), and calculate the maximum corresponding to the number of correlations, which is the point corresponding to the most similar position of \(x(k)\) and \(y(k)\). The error amount of the two signals is obtained by comparing this point with the length of the original signal. The wavelet coefficient must be corrected according to the error value to obtain the new wavelet coefficient \(w_{i,k}^*\). For the new wavelet coefficients after error correction, some noise corresponding wavelet coefficients will still remain in the vicinity of the signal mutation. To address this problem, in this paper, we use the threshold signal denoising method to process the wavelet coefficients after the error has been corrected,\(^{30}\) so that the noise signal near the signal mutation can be filtered and the wavelet can be reconstructed. After completion of the above steps, the wavelet coefficients will have basically filtered out the noise and correspond to the edge of the signal, and thereby achieving a flat coefficient.

**Flat variational modal decomposition**

From ‘Variational mode decomposition’ section, it is seen that the IMF component obtained from the VMD method contains the main fault feature information. Spectrum analysis of the IMF component can effectively extract fault features. But the IMF component also contains a large number of noise signals. Under strong noise background, useful information may be submerged in frequency spectrum analysis of IMF components, thereby reducing the accuracy of fault diagnosis. To overcome these restrictions, in this paper, we propose a flat variational modal decomposition method based on wavelet coefficient correlation denoising. In FVMD, we applied the wavelet transform discussed in the previous section to IMFs during the shifting process just like the following equation

\[
W_{i,k} = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} \text{IMF}(t) \psi_{i,k}(t) dt
\]

where \(W_{i,k}\) is the wavelet coefficient at position \(k\) on scale \(i\), \(\psi_{i,k}(t)\) is the mother wavelet (The smooth wavelet is designed in the third section). For the selection of mother wavelet, instead of using standard mother wavelets, in FVMD, a family of analytic tone-burst signals with different number of cycles was designed as a mother wavelet which best matches the fault signature signal. Then, using the flat wavelet denoising method described in ‘Flat wavelet denoising algorithm’ section, the unwanted components in each wavelet coefficient are eliminated. After denoising, estimating the new wavelet coefficients and reconstruct a new IMF subsequently, called flat intrinsic mode function (FIMF). The key to FVMD is signal reconstruction. After FIMF is obtained, the regression coefficient between the reconstructed FIMF and the original signal needs to be calculated. As the reconstructed FIMF does not completely contain useful signals, some FIMFs may even be spurious components. Therefore, in FVMD, we select only the components with higher regression coefficient of FIMF for signal reconstruction. The overall implementation process of FVMD is as follows:

1. Obtain IMF according to VMD which has been discussed in ‘Variational mode decomposition’ section. The choice of the decomposition modulus \(K\) follows the method described in ‘Variational mode decomposition’ section.
2. The flat wavelet coefficients are calculated according to the mother wavelet based on tone-burst signal and the smoothing denoising method described in ‘Flat wavelet denoising algorithm’ section

\[ W_{tk} = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} \text{IMF}(t) \psi \left( \frac{t - l}{s} \right) dt \]  \hspace{1cm} (18)

where \( s \) and \( l \) are the scales and displacements.

3. The threshold denoising method described in ‘Flat wavelet denoising algorithm’ section is used to perform threshold filtering on the wavelet coefficients in step (2), to filter out the noise signal near the signal mutation and obtain smooth wavelet coefficients \( W_{ij} \).

4. Reconstruct FIMF from the smooth coefficients through the inverse wavelet transform.

5. The regression coefficient between the FIMF component and the original signal in step (4) is calculated and the FIMF with a higher correlation coefficient is selected to reconstruct signal. Then, the reconstructed signal is used as the original signal to perform VMD, and the envelope spectrum analysis of the decomposed FIMF is performed so as to clearly extract fault information.

The FVMD-based fault diagnosis process can be represented as a flowchart as shown in Figure 3.

### Experiment analysis

**Data collection**

In order to verify the validity of the method proposed in this paper for the analysis of measured vibration signals, we analyzed a rolling bearing fault signal of inner ring and outer ring. The experimental data for bearing were provided by the Case Western Reserve University. The experimental equipment is shown in Figure 4. This experiment takes the bearing of the motor drive end as the diagnostic object. Single point damage is introduced into the inner ring and outer ring of the test bearing by means of electro discharge machining to simulate the three faults of the bearing. with fault diameters of 0.1778 mm, depth of 0.2794 mm, the data acquisition system consists of an acceleration sensor mounted on the upper side of the motor drive end with a sampling frequency of 12 kHz per channel. The motor speed was 1797 r/min and the length of the sample was 10240. In Ali et al., there is a detailed description about the experimental data collection. In this study, we only consider the bearing states covering outer race fault (ORF) and inner race fault (IRF). The rolling bearing inner ring damage fault characteristic frequency is \( f_1 = 162 \) Hz, and the outer ring damage fault characteristic frequency is \( f_2 = 107 \) Hz. Their time domain waveforms and frequency domain waveforms are shown in Figure 5.

![Figure 3. Flat variational modal decomposition algorithm fault diagnosis process.](image-url)
In this paper, the performance of FVMD algorithm is analyzed from the aspects of signal decomposition effect and signal de-noising, and the effectiveness and superiority of the FVMD method are illustrated by comparison with the EMD and the VMD methods.

**Experiment with the proposed method**

When FVMD is employed to decompose signals, first we must set three parameters, which are the balancing parameter of the data-fidelity constraint $D$, decomposing mode number $K$ and time-step of the dual ascent $W$. In this study, $D$ was set with the default value 2000, while $W$ was set as 0.1 to avoid distortion. The decomposition modulus $K$ was chosen according to the method described in ‘Variational mode decomposition’ section. The inner ring fault signal is selected to perform VMD decomposition with 10,240 signal points. The mean value of the instantaneous frequency of the components under different $K$ values is shown in Figure 6.

Starting from $K = 5$, the mean value curve of the instantaneous frequency of the component decreases rapidly. This paper considers that over-decomposition has occurred. Therefore, the modal number is selected as 4. The fault signal of the outer ring and the inner ring of the rolling bearing is decomposed by VMD in the time domain as shown in Figure 7.
It can be seen from the time domain diagram that the signal decomposed by the VMD method has impulse components, and the periodic characteristics are not conspicuous. In addition, there is still a lot of noise interference, and the state information of the bearing is basically submerged. Therefore, the useful bearing state information cannot be obtained through the time domain waveform. The signal is further enveloped for spectral analysis, and the results are shown in Figure 8. Although there is a frequency component $f_c$ corresponding to the fault of the rolling body in the figure, there is a lot of noise interference in the signal, and the fault signal is not obvious. Especially in the fourth modal component, the fault signal is basically submerged in the noise signal, and it is difficult to make an accurate judgment on the bearing state.

Figure 6. The average value of the instantaneous frequency of the input signal.

Figure 7. Time domain spectrum of each modal component after VMD processing. (a) Inner ring. (b) Outer ring. VMD: variational modal decomposition.
The signal is analyzed using the proposed method. First, the modal parameter is set using the component instantaneous frequency index, and the number of decomposition modes is determined to be 4. Then the parameters D and W are set using the VMD method to process the measured signal to obtain four IMF components. Then, the four IMF components are smoothened and denoised by the proposed method to obtain the smoothened FIMF. In order to ensure that the reconstructed signal can retain more information, finally, the regression coefficients of each FIMF component and the original signal after smooth denoising are calculated and the results are shown in Tables 1 and Table 2, and for the regression coefficient calculation method, the reader is referred to the algorithm in Montoril et al.\textsuperscript{32}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Envelope spectrum of each modal component after VMD processing. (a) Inner ring. (b) Outer ring. VMD: variational modal decomposition.}
\end{figure}

\begin{table}[h]
\centering
\caption{Regression coefficients of each component of the inner ring.}
\begin{tabular}{lcccc}
\hline
\textbf{FIMF} & 1     & 2     & 3     & 4     \\
\hline
Regression coefficients & 0.8405 & 0.6571 & 0.3031 & 0.2799 \\
\hline
\end{tabular}
\end{table}

\textit{FIMF: flat intrinsic mode function.}
As shown in Table 1 and 2, the regression coefficients of IMF1 and IMF2 are large, indicating that they retain the most significant characteristics in the original signal. So, these two IMF components are extracted for original signal reconstruction, which are decomposed and executed by the VMD. The envelope spectrum analysis was performed and the results are shown in Figure 9.

As shown in Tables 1 and 2, the regression coefficients of IMF1 and IMF2 are large, indicating that they retain the most significant characteristics in the original signal. So, these two IMF components are extracted for original signal reconstruction, which are decomposed and executed by the VMD. The envelope spectrum analysis was performed and the results are shown in Figure 9.

As shown in Figure 9, Figure 9(a) corresponds to the inner ring fault of the rolling bearing. Figure 8(a) shows a peak at 163.83 Hz, which is closer to the inner ring fault 162 Hz, and multiplier generation occurs, which is attributed to the correctness of the inner ring fault. It can be seen from Figure 9(b) in the bearing outer ring fault that a peak appears at 106.56 Hz, which is closer to the bearing outer ring fault frequency 107 Hz, and multiplier generation occurs here too, indicating that the bearing outer ring is faulty. It can be seen from the figure that after
FVMD algorithm processing, the noise of the original signal is basically removed, and the fault information can be obviously extracted. Furthermore, there are local defects on the rolling elements of the bearing, and the analysis results are completely consistent with the actual situation. The actual diagnosis shows that the method in this paper can effectively remove redundant noise interference and ensure accurate fault diagnosis.

In order to verify the advantages of the method described in this paper, the EMD method and the VMD method are compared. In order to facilitate the comparative analysis, we take only the first four components after EMD and VMD processing.

It can be seen from the spectrum of EMD analysis (as shown in Figure 10) that there is a certain aliasing of the frequencies of the EMD components, which may cause the feature information to be decomposed into different modal components, so that it is submerged in the noise and weak fault information cannot be obtained. Therefore, the EMD method cannot completely remove the influence of noise and has some disadvantages. In the envelope spectrum of EMD which is shown in Figure 11, the frequency doubling of the first and the second

![Figure 10. Spectrogram of each modal component after EMD processing. EMD: empirical mode decomposition.](image1)

![Figure 11. Envelope spectrum of each modal component after EMD processing. EMD: empirical mode decomposition.](image2)
components in the EMD decomposition signal is very obvious, but it also contains other frequency components and is accompanied by noise interference. The characteristic line is subtle. Using FVMD, the 162 Hz and the 107 Hz lines can be quickly found, which are consistent with the fault frequency of the inner and the outer rings of the bearing.

In order to quantitatively analyze the denoising effect of the signal after noise reduction in the above two methods, the kurtosis value and the signal-to-noise ratio are selected as the evaluation indices after noise reduction. The kurtosis value indicates the fault characteristic information contained in the reconstructed signal. The larger the absolute value of the kurtosis index is, the higher the number of components containing the fault impact in the IMF, and the easier it is to extract more fault information. The signal-to-noise ratio reflects the denoising ability of the signal. The larger the signal-to-noise ratio, the better the denoising effect. The kurtosis values of the modal components after the fault signal has been denoised by the three methods are shown in Tables 3 to 5.

It can be seen from Table 3 to 5 that after the original fault signal is processed by the FVMD method, the kurtosis values of the modal components are significantly higher than the EMD method and the VMD method, indicating that the FVMD method can effectively remove noise interference and retain more multiple original fault information. It can be seen from Table 6 that the signal-to-noise ratio of the signal after the FVMD processing has been significantly improved. Compared with the EMD and the VMD methods, the signal-to-noise ratio has increased by 77% and 44%, respectively, proving the effectiveness of the FVMD method.

In order to make the results more convincing, we selected the fault signal under different working conditions for FVMD, EMD and VMD experiments, and calculated the kurtosis value of the reconstructed signal and the signal-to-noise ratio of the reconstructed signal after denoising. The results are shown in Tables 7 to 10.

It can be seen from Tables 7 and 8 that at different speeds, the signal-to-noise ratio of the original signal has been significantly improved after being processed by the FVMD method, and is 77% higher than the EMD algorithm and 44% higher than the VMD algorithm. This indicates that the FVMD method can effectively remove noise interference. At the same time, it can be seen from Tables 9 and 10 that at different speeds, the kurtosis value processed by FVMD algorithm is much higher than that of EMD and VMD algorithm, indicating the FVMD method can retain the fault information in the original signal while removing noise interference.

### Table 3. The kurtosis value of each modal component after EMD processing.

|       | Inner ring | Outer ring |
|-------|------------|------------|
| IMF1  | 3.6127     | 3.1477     |
| IMF2  | 2.2063     | 2.2003     |
| IMF3  | 1.3908     | 2.1284     |
| IMF4  | 2.1715     | 1.8793     |

EMD: empirical mode decomposition.

### Table 4. The kurtosis value of each modal component after VMD processing.

|       | Inner ring | Outer ring |
|-------|------------|------------|
| IMF1  | 4.9700     | 4.2661     |
| IMF2  | 4.7733     | 3.5155     |
| IMF3  | 3.8527     | 3.3059     |
| IMF4  | 2.7660     | 3.7531     |

VMD: variational modal decomposition.

### Table 5. The kurtosis value of each modal component after FVMD processing.

|       | Inner ring | Outer ring |
|-------|------------|------------|
| IMF1  | 9.9825     | 8.0434     |
| IMF2  | 7.8406     | 6.6453     |
| IMF3  | 4.2153     | 5.9034     |
| IMF4  | 2.9214     | 3.7333     |

FVMD: flat variational modal decomposition.
Table 6. Comparison of signal-to-noise ratio after denoising.

|               | Inner ring | Outer ring |
|---------------|------------|------------|
| original signal | 3.9705     | 4.5515     |
| EMD           | 5.5614     | 5.9836     |
| VMD           | 6.8327     | 7.3544     |
| FVMD          | 9.8436     | 10.5909    |

VMD: variational modal decomposition; FVMD: flat variational modal decomposition; EMD: empirical mode decomposition.

Table 7. Inner ring signal-to-noise ratio at different speeds.

| Speed          | Original signal | EMD     | VMD     | FVMD     |
|----------------|-----------------|---------|---------|----------|
| 1797 r/min     | 3.9705          | 5.5614  | 6.8327  | 9.8436   |
| 1772 r/min     | 3.6238          | 5.1425  | 7.4608  | 9.1022   |
| 1750 r/min     | 4.0145          | 5.9546  | 8.6390  | 10.5396  |
| 1730 r/min     | 3.9195          | 6.1015  | 8.8521  | 10.7996  |

VMD: variational modal decomposition; FVMD: flat variational modal decomposition; EMD: empirical mode decomposition.

Table 8. Outer ring signal-to-noise ratio at different speeds.

| Speed          | Original signal | EMD     | VMD     | FVMD     |
|----------------|-----------------|---------|---------|----------|
| 1797 r/min     | 4.5515          | 5.9836  | 7.3544  | 10.5909  |
| 1772 r/min     | 4.1563          | 6.0248  | 8.7408  | 10.6638  |
| 1750 r/min     | 3.9706          | 5.4361  | 7.8868  | 9.6218   |
| 1730 r/min     | 4.0513          | 5.7659  | 8.3652  | 10.2056  |

VMD: variational modal decomposition; FVMD: flat variational modal decomposition; EMD: empirical mode decomposition.

Table 9. The kurtosis value of the inner ring reconstruction signal at different speeds.

| Speed          | Original signal | EMD     | VMD     | FVMD     |
|----------------|-----------------|---------|---------|----------|
| 1797 r/min     | 2.5385          | 3.8477  | 5.0443  | 8.3871   |
| 1772 r/min     | 2.7724          | 3.1348  | 4.3562  | 9.0514   |
| 1750 r/min     | 3.0014          | 4.1020  | 4.9689  | 9.1536   |
| 1730 r/min     | 2.4381          | 3.7962  | 4.7579  | 8.9825   |

VMD: variational modal decomposition; FVMD: flat variational modal decomposition; EMD: empirical mode decomposition.

Table 10. The kurtosis value of the outer ring reconstruction signal at different speeds.

| Speed          | Original signal | EMD     | VMD     | FVMD     |
|----------------|-----------------|---------|---------|----------|
| 1797 r/min     | 1.9679          | 3.1583  | 4.3562  | 8.8071   |
| 1772 r/min     | 2.0169          | 3.6532  | 4.0109  | 8.1536   |
| 1750 r/min     | 2.3566          | 3.8565  | 4.5914  | 9.0107   |
| 1730 r/min     | 2.5971          | 4.0016  | 5.1384  | 9.3112   |

VMD: variational modal decomposition; FVMD: flat variational modal decomposition; EMD: empirical mode decomposition.
Thus, the superiority of the FVMD algorithm in noise reduction is verified by comparison with EMD and VMD algorithms.

**Conclusion**

Considering the susceptibility of vibration signal-to-noise interference, which may cause the normal signal to be submerged in the interference, in this paper, we proposed a new smoothing denoising algorithm based on the traditional VMD algorithm. In this paper, the smoothing wavelet operator is designed and innovatively introduced into the decomposition process in the process of VMD decomposition. In the traditional VMD decomposition process, the fault signal is denoised by the construction of a new set of mother wavelets, and the threshold signal denoising method is used to process the noise at the signal mutation to obtain the FIMF. Then, by calculating the regression coefficient, the FIMF component was selected to reconstruct the new signal, and then the signal was de-noised by performing the VMD decomposition again. The algorithm is applied to the bearing fault diagnosis experiment and compared with the traditional EMD and VMD algorithms. The simulation results show that the proposed algorithm has a significant improvement in signal-to-noise ratio after denoising, compared with the traditional algorithm. The signal-to-noise ratio of the original signal has been significantly improved after being processed by the FVMD method, which is 77% higher than the EMD algorithm and 44% higher than the VMD algorithm. Simulations and experimental results have proved the effectiveness of the proposed method in this paper. In terms of application prospect, the smooth wavelet designed in this paper can be applied to a variety of complex objects. For different signal types, the algorithm in this paper can be tried to be extended to a broader field.

**Data availability**

Data from this manuscript may be made available upon request to the corresponding author.

**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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article:

This study was supported by the Basic Research Projects for Colleges and Universities of Liaoning Province (LJZ2017035); Science and Technology fund of Liaoning Province (20180550002); Liaoning Province Key Research and Development Project (2017225016); National Science Foundation of China (51705341).

**ORCID iD**

Dong An https://orcid.org/0000-0002-7329-8704

**References**

1. Li C, Liang M and Wang T. Criterion fusion for spectral segmentation and its application to optimal demodulation of bearing vibration signals. *Mech Syst Signal Process* 2015; 64–65: 132–148.
2. Liu Y., Xin X., Zhao Y., Ming S., Ma Y., Han J., Study on coupling fault dynamics of sliding bearing-rotor system, *J. Comput. Nonlinear. Dynam* 2019 14: 1–11.
3. Jinde Z, Zhanwei J, Ziwei P, et al. VMD based adaptive multiscale fuzzy entropy and its application to rolling bearing fault diagnosis. *IEEE* 2016; 18: 1–4.
4. Lei Y, Lin J, He Z, et al. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech Syst Signal Process* 2013; 35: 108–126.
5. Yan R, Gao RX and Chen X. Wavelets for fault diagnosis of rotary machines: a review with applications. *Signal Process* 2014; 96: 1–15.
6. Sun H, He Z, Zi Y, et al. Multiwavelet transform and its applications in mechanical fault diagnosis – a review. *Mech Syst Signal Process* 2014; 43: 1–24.
7. Li Z, Chen J, Zi Y, et al. Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive. *Mech Syst Signal Process* 2017; 85: 512–529.
8. Ali JB, Fnaiech N, Saïdi L, et al. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl Acoust* 2015; 89: 16–27.
9. Pang B, Tang G and Tian T. Rolling bearing fault diagnosis based on an improved HTT transform. *Sensors* 2018; 4: 1203–1222.
10. Farzin P and Kim JM. Bearing fault diagnosis by a robust higher-order super-twisting sliding mode observer. *Sensors* 2018; 18: 1128–1150.
11. Huang NE, Shen Z, Long SR, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proc Math Phys Eng Sci* 1998; 454: 903–995.
12. Wang LY, Wang Q, Zhang MJ, et al. A grey fault diagnosis method for rolling bearings based on EMD. *J Vib Shock* 2014; 33: 197–202.
13. Cheng JS, Yu DJ and Yang Y. A fault diagnosis approach for roller bearing based on SVM and EMD envelope spectrum. *Syst Eng Theory Pract* 2005; 9: 131–136.
14. Smith JS. The local mean decomposition and its application to EEG perception data. *J R Soc Interface* 2005; 2: 443–454.
15. Liu H and Han M. A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings. *Mech Mach Theory* 2014; 75: 67–78.
16. Liu Z, He Z, Guo W, et al. A hybrid fault diagnosis method based on second generation wavelet de-noising and local mean decomposition for rotating machinery. *ISA Trans* 2016; 61: 211–220.
17. Aijun HU. New process method for end effects of HILBERT-HUANG transform. *J Mech Eng* 2008; 44: 154–158.
18. Viswanath A, Jose KJ, Krishnan N, et al. Spike detection of disturbed power signal using VMD. *Proc Comput Sci* 2015; 46: 1087–1094.
19. Upadhyay A and Pachori RB. Instantaneous voiced/non-voiced detection in speech signals based on variational mode decomposition. *J Franklin Inst* 2015; 352: 2679–2707.
20. Chang L, Gang C and Xihui C. Planetary gears feature extraction and fault diagnosis method based on VMD and CNN. *Sensors* 2018; 5: 1523–1543.
21. Zhipeng W, Li-min J and Yong Q. Adaptive diagnosis for rotating machineries using information geometrical kernel-ELM based on VMD-SVD. *Entropy* 2018; 20: 73–91.
22. Anesh C, Kumar S, Hisham PM, et al. Performance comparison of variational mode decomposition over empirical wavelet transform for the classification of power quality disturbances using support vector machine. *Proc Comput Sci* 2015; 46: 372–380.
23. Salim L. Comparative study of ECG signal denoising by wavelet thresholding in empirical and variational mode decomposition domains. *Healthcare Technol Lett* 2014; 1: 104.
24. Dragomiretskiy K and Zosso D. Variational mode decomposition. *IEEE Trans Signal Proc* 2014; 62: 531–544.
25. Argoul F, Arneodo A, Elezgaray J, et al. Wavelet transform of fractal aggregates. *Phys Lett A* 2017; 135: 327–336.
26. Abbate A, Koay J, Frankel J, et al. Signal detection and noise suppression using a wavelet transform signal processor: application to ultrasonic flaw detection. *IEEE Trans Ultrason Ferroelect Freq Contr* 1997; 44: 14–26.
27. Rostami J, Chen J and Tse P. A signal processing approach with a smooth empirical mode decomposition to reveal hidden trace of corrosion in highly contaminated guided wave signals for concrete-covered pipes. *Sensors* 2017; 17: 302.
28. Yuan WJ, Liu YD and Zhai CP. Modified denoising method based on wavelet correlation at scales and threshold. *Commun Technol* 2018; 2: 284–288.
29. Zhao R-Z, Qu HZ and Song GX. A threshold filtering algorithm based on the region relativity of the wavelet coefficients. *J Xidian Univ* 2001; 28: 324–327.
30. Vargas RN. Seismic trace noise reduction by wavelets and double threshold estimation. *IET Signal Proc* 2018; 11: 1069–1075.
31. Wenlong Fu, Jianzhong Z and Yongchuan Z. Fault diagnosis for rolling element bearings with VMD time-frequency analysis and SVM. *IEEE* 2015; 22: 69–72.
32. Montoril MH, Morettin PA and Chiani C. Wavelet estimation of functional coefficient regression models. *Int J Wavelets Multiresol Inform Process* 2018; 16: 8–39.