Du *et al.*
Chinese Journal of Mechanical Engineering (2022) 35:14
https://doi.org/10.1186/s10033-022-00677-5

Chinese Journal of Mechanical Engineering

ORIGINAL ARTICLE

Open Access

Improved Multi-Bandwidth Mode Manifold for Enhanced Bearing Fault Diagnosis

Guifu Du, Tao Jiang, Jun Wang*, Xingxing Jiang and Zhongkui Zhu

Abstract

Variational mode decomposition (VMD) has been proved to be useful for extraction of fault-induced transients of rolling bearings. Multi-bandwidth mode manifold (Triple M, TM) is one variation of the VMD, which units multiple fault-related modes with different bandwidths by a nonlinear manifold learning algorithm named local tangent space alignment (LTSA). The merit of the TM method is that the bearing fault-induced transients extracted contain low level of in-band noise without optimization of the VMD parameters. However, the determination of the neighborhood size of the LTSA is time-consuming, and the extracted fault-induced transients may have the problem of asymmetry in the up-and-down direction. This paper aims to improve the efficiency and waveform symmetry of the TM method. Specifically, the multi-bandwidth modes consisting of the fault-related modes with different bandwidths are first obtained by repeating the recycling VMD (RVMD) method with different bandwidth balance parameters. Then, the LTSA algorithm is performed on the multi-bandwidth modes to extract their inherent manifold structure, in which the natural nearest neighbor (Triple N, TN) algorithm is adopted to efficiently and reasonably select the neighbors of each data point in the multi-bandwidth modes. Finally, a weight-based feature compensation strategy is designed to synthesize the low-dimensional manifold features to alleviate the asymmetry problem, resulting in a symmetric TM feature that can represent the real fault transient components. The major contribution of the improved TM method for bearing fault diagnosis is that the pure fault-induced transients are extracted efficiently and are symmetrical as the real. One simulation analysis and two experimental applications in bearing fault diagnosis validate the enhanced performance of the improved TM method over the traditional methods. This research proposes a bearing fault diagnosis method which has the advantages of high efficiency, good waveform symmetry and enhanced in-band noise removal capability.

Keywords: Variational mode decomposition, Manifold learning, Natural nearest neighbor algorithm, Rolling bearing, Fault diagnosis, Time-frequency signal decomposition

1 Introduction

With the continuous development of industrial modernization and manufacturing informationization, the integration of mechanical equipment and the scale of the system have developed rapidly, posing more severe challenges to fault diagnosis of mechanical equipment. Any glitches or misalignment of single component may affect the normal operation of the entire system and eventually lead to serious accidents. As a key component of the mechanical transmission system, rolling bearings often operate under high-speed and heavy-load conditions and are easily damaged. Therefore, fault diagnosis of rolling bearings is important to avoid catastrophes and reduce economic losses [1–4]. Vibration signals collected from faulty bearings contain periodic transient impulse components that can be used for fault diagnosis. Therefore, extracting the transient components for period calculation is essential for bearing fault diagnosis [5]. However, the vibration signals collected from mechanical equipment often have the properties of nonlinearity and non-stationarity, and contain a large number of narrow-band

*Correspondence: junking@suda.edu.cn
School of Rail Transportation, Soochow University, Suzhou 215131, China

pulse interference and background noise, leading to great challenges for bearing fault detection [6].

A great amount of recent research efforts have been made to explore the fault feature extraction of rolling bearings based on signal processing methods to overcome the challenges of bearing fault diagnosis. As an important technique to deal with non-stationary signals, time-frequency signal decomposition methods can decompose a complex signal into several regular simple modes that can be easily analyzed in the time and frequency domain [7]. Wavelet transform (WT), empirical mode decomposition (EMD) and empirical wavelet transform (EWT) are typically advanced time-frequency signal processing methods that are widely applied to the field of rotating machinery fault diagnosis. The WT translates the original signal by a series of wavelets, each covering a specific frequency sub-band [8]. Nevertheless, the performance of the WT depends on the selected mother wavelet; only the features that are closely related to the mother wavelet could gain relatively high coefficient results, whereas the features that are not similar to the mother wavelet may be ignored [9, 10]. The EMD and some of its variants, such as ensemble EMD (EEMD) and local mean decomposition (LMD), generate a series of complete and almost orthogonal intrinsic mode functions (IMFs) which could imply failure information [11]. However, these methods are all based on recursive decompositions, during which the decomposition errors may be accumulated step by step [12]. The EWT can design a suitable wavelet filter bank according to the processed signal, which overcomes the influence caused by fixed basis functions [13]. However, heavy noise in real applications may weaken the matching capability between the signal and the basis function, and the mode number needs to be preset, leading to the lack of robustness and practicality of the EWT [14].

Variational mode decomposition (VMD) is a completely non-recursive time-frequency signal decomposition method, in which the center frequency and bandwidth of each exacted mode are obtained through searching the optimal solution of constrained variational problem circularly, by which different amplitude-modulated and frequency-modulated (AM-FM) components can be effectively separated [15]. The VMD has been proved to outperform the EMD and its variants in bearing fault diagnosis [16]. However, the success of the VMD depends on some preset parameters, including the number of decomposition modes K and the bandwidth balance parameter α [17, 18]. For the vibration signals measured in different environments, proper parameter values need to be manually selected to extract the bearing fault transient components with less interference. To realize the adaptive setting of the VMD parameters, the mainstream practice is to determine the parameters based on some intelligent algorithms. In Ref. [19], the two parameters K and α in the VMD were optimized synchronously based on the kurtosis index through artificial fish swarm algorithm. Yan et al. [20] applied the genetic algorithm to accelerate the convergence of iterations in the VMD, and implemented a 1.5-dimensional envelope spectrum to detect compound fault information of bearings. Particle swarm optimization (PSO) algorithm was combined with the VMD to realize adaptive selection of the parameters in Ref. [21]. Tang and Wang [22] used Shannon entropy as an index to optimize the parameters of the VMD by the PSO algorithm, by which the weak bearing fault characteristic frequency was extracted successfully. The grasshopper optimization algorithm (GOA) was applied to optimize the VMD parameters based on a measurement index termed weighted kurtosis index, which is constructed by using kurtosis index and correlation coefficient [23]. With the parameters optimized by the above methods, the bearing fault transient components are expected to be extracted with the noise outside the band from the vibration signal by the VMD method [24, 25]. Nevertheless, the extracted narrow-band fault components may ignore some fault-related information and must contain some interfering components and noise. Moreover, in order to obtain the fault-related mode with a high accuracy, the optimization based VMD methods must perform plenty of the VMD trials on the original signal, which is time-consuming [26]. Therefore, the optimization based VMD method needs to balance the accuracy and calculation efficiency in practical applications.

Most recently, in order to escape the inefficient parameter optimization of the VMD and enable the suppression of in-band noise in the application of bearing fault diagnosis, our group proposed a variation of the VMD, called multi-bandwidth mode manifold (Triple M, TM), by combining the VMD and nonlinear manifold learning [27]. This method units a small number of fault-related modes obtained by the recycling VMD (RVMD) with different parameters via a manifold learning algorithm named local tangent space alignment (LTSA). The TM method does not need to optimize the VMD parameters and the fused bearing fault-induced transients show low level of in-band noise. Some encouraging results were achieved in the application to bearing fault diagnosis. However, there are still some issues on the TM learning technique that remain to be addressed to further enhance the performance of the TM method.

The TM method is to learn an inherent manifold structure of the bearing fault transient components by the LTSA algorithm, from a high-dimensional matrix of multi-bandwidth modes that is composed of the

multiple fault-related modes with different bandwidths. The effect of in-band noise suppression firmly relies on a parameter of the LTSA, i.e., the number of nearest neighbors k [27–29]. This is because the LTSA is based on the idea that the local data distribution of the selected neighborhoods for each data is kept from the high-dimensional matrix to the low-dimensional manifold. A conceivable way to improve the noise removal performance is to search the optimal k within a wide range, which is adopted in the original TM method. However, the repeat of the LTSA implementation to determine the proper neighborhood size is a cost of time. In Ref. [30], the local data distributions are changed by adding random noise when constructing the high-dimensional data to produce the best result of the LTSA. Nevertheless, the LTSA algorithm still needs to be repeated in Ref. [30]. The neighborhood sizes for all the data points are the same in the existing LTSA algorithms, which is not sound because the local linearity property and data density between different data areas differ from each other. Therefore, the first issue that this paper addresses is to select proper neighborhood size for each data point to construct personalized local data distribution for the TM feature learning. The natural nearest neighbor (Triple N, TN) algorithm [31] is introduced to adaptively determine the optimal neighbors of each data point, with which there just needs one implementation of the LTSA algorithm. By introducing the TN algorithm, the efficiency and performance of the TM method are improved.

The second issue to be addressed in this study is how to further improve the performance of the manifold feature learned by the TM method to approach the real waveform. Due to that the LTSA algorithm is sensitive for extraction of impulsive features, the waveform of the learned feature would probably be asymmetric in the vertical direction, while the real fault-induced transients are mostly symmetric in the up-and-down direction. It is uncovered that the symmetry of the manifold feature can be improved by synthesizing the first two dimensional data of the LTSA output. Thus, a weight-based feature compensation strategy is designed in this paper to obtain a synthetic TM feature that is representative of the real fault-induced transients.

By addressing the two issues mentioned above, the improved TM method proposed in this paper is expected to outperform the original TM method for bearing fault diagnosis. The remainder of this paper is outlined as follows. Section 2 presents the original TM method succinctly. Section 3 elaborates the improved TM method. The enhanced performance of the proposed method is verified by one simulation case in Section 4 and two experimental cases in Section 5, where the comparisons with the traditional VMD methods are also analyzed. Finally, the conclusions are drawn in Section 6.

2 Multi-Bandwidth Mode Manifold

The multi-bandwidth mode manifold (TM) method aims to reveal the intrinsic waveform structure of the fault-related modes with different bandwidths decomposed from the measured signal. It is realized by conducting manifold learning on the multi-bandwidth modes constructed by the RVMD with different balance parameters. The obtained TM feature indicates the merits of low level of in-band noise and no requirement of optimization of the VMD parameters. The technique of the TM method mainly includes three steps, as illustrated in Figure 1. The following describes these steps succinctly. For more details, please refer to Ref. [27].

Step 1: Decompose the bearing vibration signal $x(t)$ by the RVMD with different bandwidth balance parameters. The RVMD is to repeatedly recycle the residual modes by conducting the VMD with mode number K kept as one, where the first residual mode is the original signal. Given a balance parameter α, the original signal $x(t)$ is firstly decomposed into one extracted mode $u_1(t)$ by the VMD with $K=1$ and one residual mode $u_{r1}(t)$ obtained by:

$$u_{r1}(t) = u_{r0}(t) - u_1(t), \tag{1}$$

where $u_{r0}(t)=x(t)$. Then, the residual mode is further decomposed into a new extracted mode and a new residual mode by the same way. After repeating the above operations p times, p extracted modes $u_i(t)$ $(i=1, 2, \ldots, p)$ and p residual modes $u_{ri}(t)$ $(i=1, 2, \ldots, p)$ are obtained. The value of p is set as 5 in the TM method. The initial center frequency for the extracted mode $u_i(t)$ is the frequency with the largest amplitude in the spectrum of the residual mode $u_{r(i-1)}(t)$. It has been proved that the

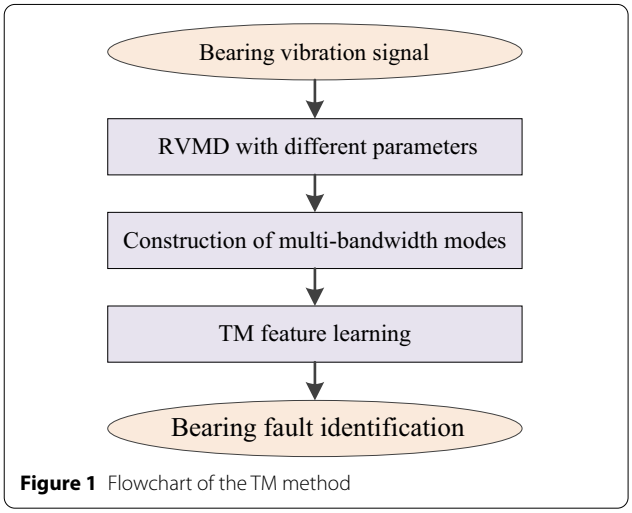

Figure 1 Flowchart of the TM method

decomposition efficiency of the RVMD is much higher than the traditional VMD when producing the same number of extracted modes. By changing the balance parameter α and repeating the RVMD on the original signal, the decomposed modes containing similar frequency contents would have different bandwidths. Ten values of α are selected equidistantly from the range of [100, 5000] in the TM method.

Step 2: Construct the multi-bandwidth modes by selecting the fault-related modes from the decomposed modes. For a specific value of α, the possible fault-induced information is contained in either the extracted modes or the residual modes. The one exhibiting the most salient fault information is called the fault-related mode, which is selected by the Gini index [32]:

$$GI = 1 - 2\sum_{n=1}^{N} \frac{eu^r(n)}{\|eu(n)\|_1}\left(\frac{N-n+0.5}{N}\right), \tag{2}$$

where $eu(n)$ is an N-point discrete format of the envelope of a decomposed mode $u(t)$, $eu^r(n)$ is a reordered version of $eu(n)$, whose elements are arranged from the smallest to the largest, $||\ddot{}||_1$ is the L_1 norm operation. The mode leading to the highest GI value is selected as the fault-related mode corresponding to the specific α. The bandwidths of the fault-related modes obtained with different α differ from one another. The mode with relatively large bandwidth contains more fault-related information than the mode with relatively small bandwidth, while the latter one contains less fault-unrelated components and noise than the former one. As a result, ten fault-related modes are obtained, which constitute a ten-dimensional data matrix called multi-bandwidth modes as below:

$$MBM = \left[u^1\ u^2\ \cdots u^i \cdots\ u^{10}\right]^{\mathrm{T}}, \tag{3}$$

where $u^i \in R^{N\times 1}$ is a vector derived from the ith fault-related mode in the discrete format.

Step 3: Address the LTSA algorithm on the constructed multi-bandwidth modes to learn the TM feature representing the intrinsic waveform structure of the fault-related modes. The LTSA is a widely used manifold learning algorithm for extraction of impulsive features in machinery fault diagnosis [27–29, 33]. The basic principle of the LTSA is to maintain the intrinsic manifold structure with the form of skeleton when reducing the dimensionality of a high-dimensional data, by keeping the local data distribution constructed by its k neighboring points. For the multi-bandwidth modes in Eq. (3), the skeleton structure is the waveform induced by the bearing fault, which is revealed in the output of the LTSA as:

$$WM = [w^1\ w^2\ \cdots w^i \cdots\ w^d]^{\mathrm{T}}, \tag{4}$$

where $w^i \in R^{N\times 1}$ is the ith-dimensional data in the discrete format, and $d << 10$. In the original TM method, d is set as one. Thus, the output is rewritten as:

$$WM = [y_1, y_2, \cdots, y_i, \cdots, y_N], \tag{5}$$

where y_i is the ith data point in w^1. To obtain a fault-induced waveform with a satisfactory effect of in-band noise removal, the parameter k is optimized by the permutation entropy criterion (PE) formulated by:

$$PE(k) = -\frac{\sum_{i=1}^{p!} P_i \ln P_i}{\ln(p!)}, \tag{6}$$

where p is the embedding dimension in the phase space reconstruction of WM, P_i is the probability distribution of the ith permutation of the reconstruction of WM. The output resulting in the minimum PE value is regarded as the TM feature that represents the pure fault transient components.

3 Improved Multi-Bandwidth Mode Manifold

The TM method is a promising tool to extract the bearing fault features under noisy working conditions. It overcomes the difficulty of parameter setting in the VMD method by uniting the modes containing similar fault contents that are obtained with different parameters. The computational efficiency and feature extraction performance of the TM method have been enhanced as compared to the traditional VMD methods. This paper intends to further improve the TM method by developing the LTSA algorithm. The improved TM method addresses the issues in two aspects: construction of personalized local data distribution and formation of symmetric TM feature, which are described in detail in the following.

3.1 Construction of Personalized Local Data Distribution

In the original TM method, the local data distributions of the high-dimensional data of multi-bandwidth modes being established in Eq. (3) are constructed by the same number of neighboring data points via the traditional k-nearest neighbor algorithm in the LTSA. The multi-bandwidth modes is consist of the fault-induced transient components and the fault-unrelated components. The fault-induced transient components distribute in the sparse areas because they have impulsive characteristic, while the fault-unrelated components distribute in the dense areas because they are regarded as noise. If k is larger than the data number of one fault impulse in the sparse areas, some noise data points in the dense areas are selected as the neighbors of the impulse data points, which corrupts the inherent regularity of the fault

impulse, leading to that the fault-induced transients in the learned TM feature are not dominated or even submerged by the noise. If k is smaller than the data number of one fault impulse in the sparse areas, the noise data points in the dense areas could not obtain enough neighbors to show their difference from the impulse data points, resulting in that some noise components are regarded as the manifold structure that is kept in the TM feature. Due to the different local linearity property and data density between the impulse areas and noise areas, the local data distributions of different data points should be represented by different number of neighboring data points. Moreover, the determination of the neighborhood size in the original TM method is time-consuming because the LTSA algorithm must be repetitively conducted with a series of neighborhood sizes.

This paper proposes to construct personalized local data distribution for each data point by introducing the natural nearest neighbor (TN) algorithm to the LTSA. The TN algorithm is a scale-free nearest neighbor method that does not preset specific scale which determines the performance of manifold learning, such as the neighborhood size k in the traditional k-nearest neighbor algorithm. If the traditional k-nearest neighbor algorithm is regarded as an active neighbor search process, the TN algorithm is a completely passive neighbor confirmation process. The idea of the TN algorithm is to assign the neighbor number of each data according to the density of the data area for construction of personalized local data distribution. The specific steps of the TN algorithm is as follows:

1) Calculate the distances between each data point z_i and other data points.

2) Given the initial value of k, record the nearest k data points for each point z_i.

3) For each data point z_i, the data points that take z_i as one of the nearest k points are considered as the natural nearest neighbors of z_i.

According to the TN algorithm, the impulse data points will have less neighbor numbers than the noise data points. On the other hand, due to that the amplitudes of the fault-induced transients are usually larger than the noise, the impulse data points are difficult to be treated as the neighbors of the noise data points, and vice versa. Thus, the difference of local data distribution between the fault-related transient components and the fault-unrelated components is strengthened. Therefore, the TN algorithm helps to select proper neighborhood size for each data points and construct personalized local data distribution, which is beneficial for the TM feature learning by the LTSA technique. The LTSA based on the TN is called TN-LTSA in this paper. Some data points may also be misled as the neighbors of improper areas by the TN algorithm. In this condition, the inconsistency between the constructed local data distribution and those in the same area is increased, which will be alleviated in the dimensionality reduction process of the LTSA method. By introducing the TN algorithm to the LTSA, the performance of the TM feature is expected to be improved.

Due to the fact that the construction of the local data distribution is only related to the local data density in the TN algorithm, the parameter k has little effect on the feature learning performance. Therefore, the TN-LTSA only needs to be performed once for the TM feature learning, which improves the efficiency of the TM method significantly. The range of k value is generally between 10 and 50. Without loss of generality, k is set as 30 in this paper.

3.2 Formation of Symmetric TM Feature

In the original TM method, the manifold output is a one-dimensional vector and is regarded as the TM feature representing the fault transient components. However, the TM feature would probably be asymmetric in the up-and-down direction, which is not the real waveform pattern of the fault transient components. There are many reasons for this problem, including but not limited to unreasonable local information extraction, improper local space construction, and theoretical limitations of the LTSA on the processing of one-dimensional signals.

In order to alleviate the asymmetry phenomenon of the TM feature, a weight-based feature compensation strategy is proposed in this paper to form a synthetic TM feature. d is set as two in the improved TM feature. Then, the TN-LTSA output is written as

$$WM = [w^1 \ w^2 \ ]^{\mathrm{T}}. \tag{7}$$

The two vectors w^1 and w^2 are actually two eigenvectors of an alignment matrix constructed in the TN-LTSA, whose corresponding eigenvalues are λ_1 and λ_2, and $\lambda_1 < \lambda_2$. The smaller the eigenvalue is, the lower the affine error of the manifold feature in the corresponding dimension will be. Therefore, w^1 is more similar to the fault-related transients than w^2. It is discovered that, when w^1 is asymmetric in the vertical direction, w^2 has complementary waveform pattern as compared to w^1. This motivates us to combine the second vector to compensate the asymmetry of the first vector. Considering the different amplitude properties of the two vectors, the two eigenvalues are used as the weight coefficients of the opposite eigenvectors. The synthetic TM feature is formed by the weight-based feature compensation strategy as:

$$w = \frac{\lambda_1 w^2 \pm \lambda_2 w^1}{\left\|\lambda_1 w^2 \pm \lambda_2 w^1\right\|_2}, \tag{8}$$

where the plus or minus sign is determined according to the respective waveform, $||\ddot{}||_2$ is the L_2 norm operation. The symmetry of the synthetic TM feature is much enhanced, hence is more suitable to represent the fault transient components than the feature obtained in the original TM method.

3.3 Summary of the Improved TM Method

The flowchart of the improved TM method are illustrated in Figure 2, and the specific procedures are briefly described as follows.

Step 1: Perform the RVMD on the original signal repetitively with 10 values of the bandwidth balance parameter α, which is selected from the range of [100, 5000] equidistantly.

Step 2: Construct the multi-bandwidth modes by selecting the fault-related modes from the decomposed modes based on the Gini index.

Step 3: Employ the proposed TN-LTSA on the matrix of multi-bandwidth modes with $k=30$ and $d=2$. The manifold output includes two-dimensional data with complementary waveform.

Step 4: Form the symmetric TM feature by the weight-based feature compensation strategy via Eq. (8). The possible bearing fault characteristic period is expected to be easily identified in the obtained symmetric TM feature.

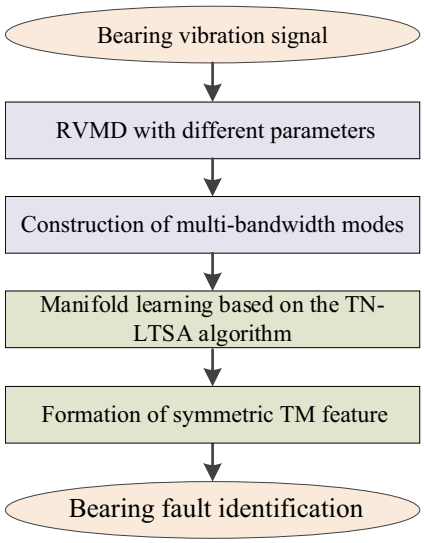

Figure 2 Flowchart of the improved TM method

4 Simulation Analysis

In order to verify the enhanced performance of the improved TM method for extraction of bearing fault-related transients, a test signal consisting of white noise and periodic impulses is simulated by considering a free vibration model with damping as follows:

$$\begin{aligned} x(n) =& A\sum\nolimits_r e^{-\beta(n-rFs/f_m-\tau_r)/Fs}\sin(2\pi f_1 \\ &-(n-rFs/f_m-\tau_r)/Fs)+\xi(n), \end{aligned} \tag{9}$$

where $\beta=200$ denotes the attention factor, r is the number of the impulses, $Fs=2000$ Hz is the sampling frequency, $f_m=10$ Hz is the fault characteristic frequency, $f_1=200$ Hz denotes the resonant frequency, τ_r is used to simulate the randomness caused by slippage, which is subject to a discrete uniform distribution and is ranged among $(-\pi, \pi]$, $\xi(n)$ represents the white noise that results in the signal-to-noise ratio (SNR) as -9 dB. A sample with 1 s time length is used for analysis.

The simulated pure signal, noisy signal and their spectra are shown in the Figure 3. The pure signal consists of exponentially decaying pulses that last for a short period of time. However, the periodic impulses are drowned by strong white noise in the noisy signal, making it difficult to be identified in the waveform and spectrum of the noisy signal.

The proposed improved TM method is firstly performed on the simulated noisy signal. First, the RVMD is carried out on the original signal with 10 values of the bandwidth balance parameter α. In each RVMD, all extracted modes and residual modes are considered as the candidates of fault-related modes after 5 times recycling. The spectra of the 5 extracted modes with α being 5000 are presented in Figure 4, where the background spectrum is of the original noisy signal. It can be seen that the first mode extracted contains most of the pure

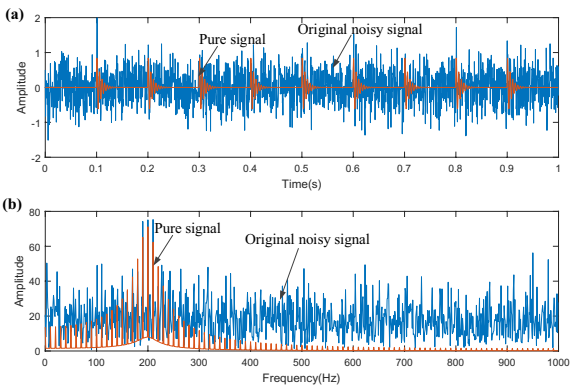

Figure 3 Simulated pure and noisy signals: **a** waveforms and **b** spectra

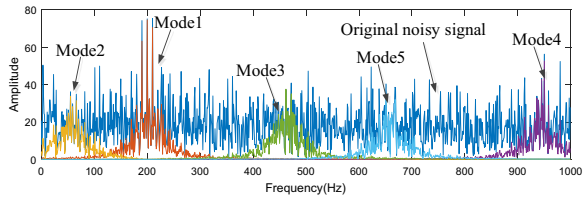

Figure 4 Spectra of the extracted modes obtained by the RVMD for the signal in Figure 3

signal information, partly owing to the strategy that the RVMD sets the initial center frequency as the frequency that has the highest amplitude in the analyzed signal spectrum. Then the first mode extracted is picked out as a fault-related mode by the Gini index, and a ten-dimensional data matrix is formed together with other fault-related modes having different bandwidths. Figure 5 shows the waveforms and spectra of the first, fifth and tenth target modes sorted in ascending order of α values. All target modes were extracted in the first recycling round of RVMD, but have different bandwidths due to variant α values. The target mode with small bandwidth contains relatively less noise and fault-unrelated components, while the target mode with large bandwidth contains more fault-related information. To reveal the inherent regularity of the high-dimensional data matrix, the TN-LTSA is performed on the multi-bandwidth modes. Given the initial value of the nearest neighbor number k as 30, the number of neighbors for each data point is reallocated by the TN algorithm, as illustrated in Figure 6a. The number of neighbors for each data point is determined by the density of its area, which maintains the uniqueness of the impulse data points and the randomness of the noise data points. The waveforms of the two eigenvectors of the output manifold feature are shown in Figure 6b and c, respectively. The corresponding eigenvalues are 0.082 and 0.101, respectively. Due to proper construction of personalized local data distribution by the TN algorithm, the transient components are retained and the noise is mostly eliminated in the two dimensional data. However, both the features have the problem of asymmetric waveform in the up-and-down direction. The synthetic TM feature is finally obtained by the proposed weight-based feature compensation strategy, as presented in Figure 6d, where the asymmetry problem is effectively alleviated, and the periodicity of the fault impulses is clearly manifested.

As comparisons, the traditional TM method and the PSO-based VMD (PSO-VMD) method [26] are used

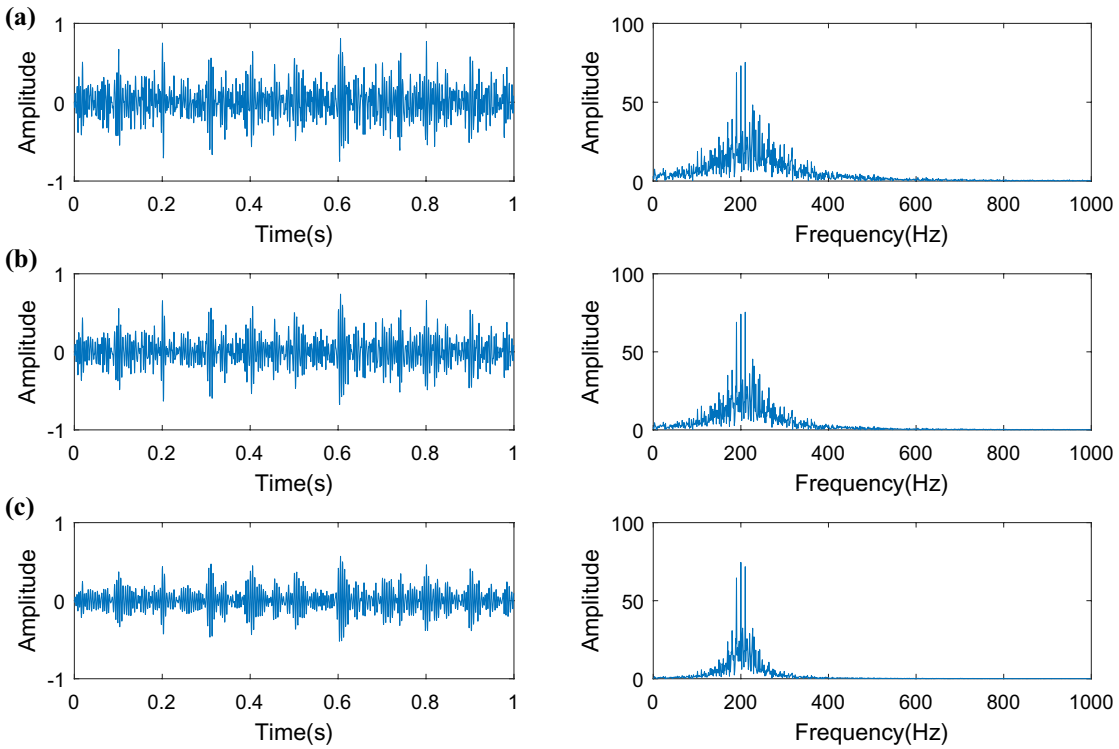

Figure 5 Waveforms and spectra of the **a** first, **b** fifth and **c** tenth target modes obtained by the RVMD with different bandwidth balance parameters for the signal in Figure 3

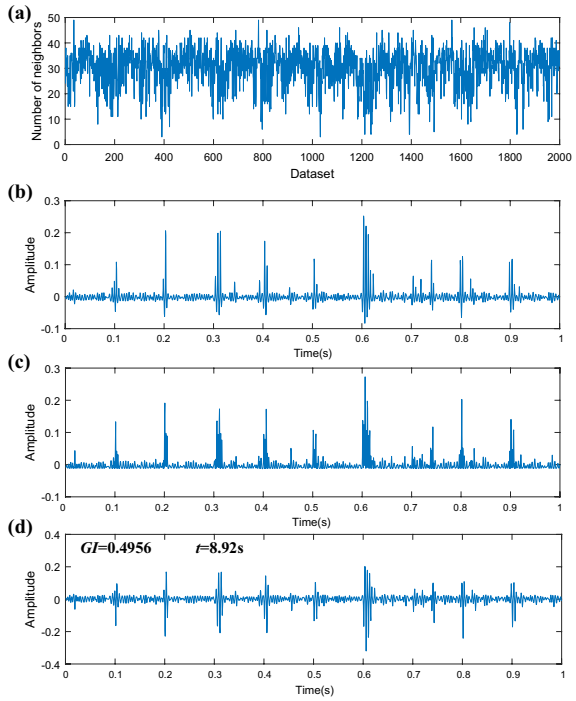

Figure 6 Results of the simulated noisy signal in Figure 3 analyzed by the proposed method: **a** reallocated neighbor numbers by the TN algorithm, **b** waveform of the first-dimensional data, **c** waveform of the second-dimensional data and **d** waveform of the synthetic TM feature

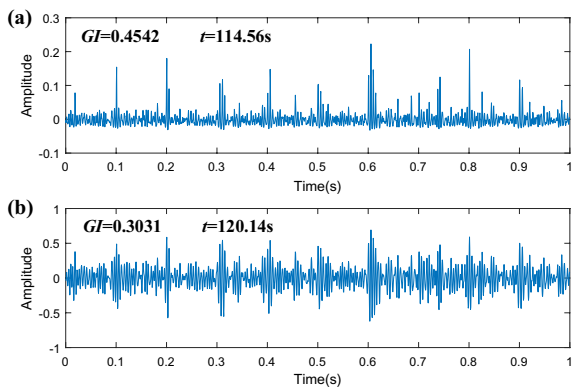

Figure 7 Waveforms of **a** the feature obtained by the original TM method and **b** the optimal target mode obtained by the PSO-VMD method for the signal in Figure 3

to analyze the simulated noisy signal in Figure 3. When conducting the original TM method, the parameter k is optimized from the range of [10, 50]. The TM feature extracted is shown in Figure 7a, in which the in-band noise is less than the fault-related modes obtained by the RVMD shown in Figure 5, while is a bit larger than the first dimensional data of the manifold feature extracted by the TN-LTSA shown in Figure 6b. Moreover, the asymmetry problem is obvious in Figure 7a. In the PSO-VMD method, the bandwidth balance parameter α and the number of modes K are optimized via the PSO algorithm, whose ranges are from 100 to 5000 and from 3 to 6, respectively. The waveform of the optimal mode is shown in Figure 7b. It can be seen that the noise of the optimal mode is much larger than the original and improved TM methods, hindering the identification of the periodic transients.

To quantify the performance and efficiency of the three methods, the Gini index and computation time of each method are calculated and given in the corresponding results in Figures 6d, 7a and b, respectively. The proposed improved TM method achieves the result with the highest GI value and spends much less time than the other methods, indicating that the improved TM method outperforms the original TM method and the PSO-VMD method in both performance and efficiency for extraction of fault transient components.

5 Experimental Verification

In order to verify the enhanced performance of the improved TM method for rolling bearing fault diagnosis, two groups of bearing dataset, including the NASA bearing dataset and the bearing test rig dataset obtained in our research group are analyzed in this section.

5.1 NASA Bearing Dataset

The NASA bearing dataset are lifecycle vibration signals of a bearing run-to-failure test provided by the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati [34]. A severe outer-race fault was found in a bearing at the end of one test, whose characteristic period is calculated to be $T_O = 0.004$ s. For more details on this experiment, please refer to Ref. [35].

The waveform and spectrum of a vibration signal measured at the early stage of the outer-race defect is shown in Figure 8. It can be seen that the in-band noise and harmonic components mask the fault transient components, making it difficult to identify the existence of the bearing fault. The proposed improved TM method is firstly introduced to analyze the vibration signal in Figure 8. By using the TN algorithm, the data points in the high-dimensional data of the multi-bandwidth modes have different neighbor numbers for construction of personalized local data distribution, as is illustrated by Figure 9a. The impulse data points have relatively larger neighbor numbers than the noise data points, which strengthens the divergence of local data distribution between the fault transients and noise. By synthesizing the two vectors of the manifold output

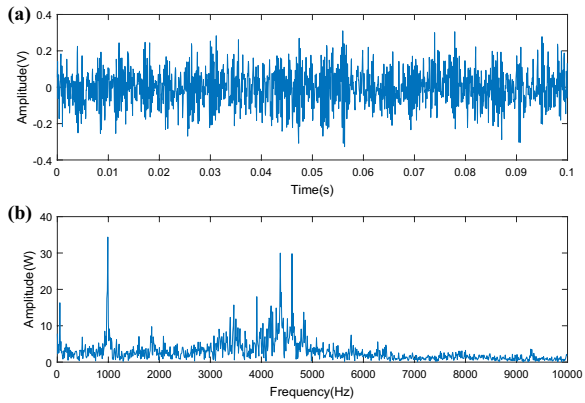

Figure 8 Vibration signal with an incipient outer-race defect: **a** waveform and **b** spectrum

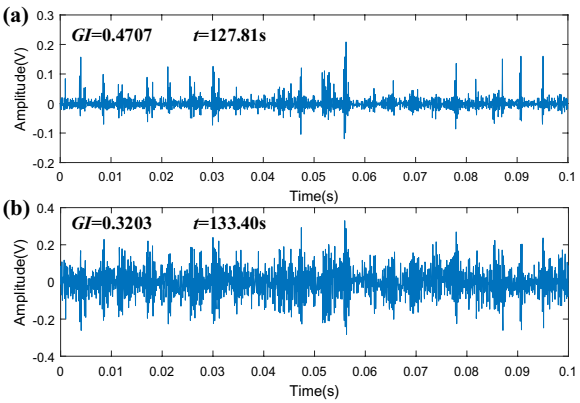

Figure 10 Waveforms of **a** the feature obtained by the original TM method and **b** the optimal target mode obtained by the PSO-VMD method for the signal in Figure 8

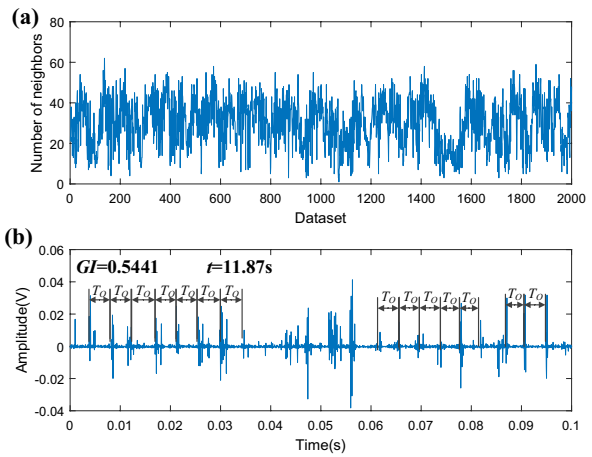

Figure 9 Results of the vibration signal in Figure 8 analyzed by the proposed method: **a** the reallocated neighbor numbers by the TN algorithm and **b** the waveform of the synthetic TM feature

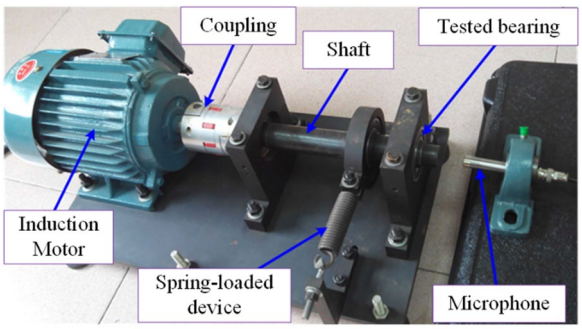

Figure 11 Bearing test rig established in our research group

based on the weight-based feature compensation strategy, the synthetic TM feature is achieved and displayed in Figure 9b. It can be seen that the synthetic TM feature is symmetric, where the fault transient components are preserved obviously by the improved TM method while the noise is greatly suppressed. As comparisons, the vibration signal in Figure 8 is also analyzed by the original TM method and the PSO-VMD method. The resultant waveform of the TM feature is shown in Figure 10a, where the survived in-band noise is notable and the asymmetry problem in the up-and-down direction exists. The waveform of the optimal mode obtained via the PSO-VMD is displayed in Figure 10b, where the in-band noise is much heavier than those in the two other results. The Gini value of each result and the computation time spent in each method are calculated and recorded in the corresponding figure. The improved TM method exhibits the highest *GI* value and the smallest computation time, demonstrating its outstanding advantages in retaining fault transient components and improving computing efficiency.

5.2 Bearing Test Rig Dataset

A simplified bearing test rig was established in our research group for collection of bearing data with seeded faults, as shown in Figure 11. The rotor-bearing system was driven by an induction motor. A spring-loaded device was designed as a radial loader of the rotor-bearing system. The tested bearing (type: N306E) was installed at the right end of the shaft. A seeded defect that is a single slit with a width of 0.5 mm was machined on the outer raceway and inner raceway of the tested bearings separately. The acoustic signals of the tested bearings were acquired by a microphone (Model: INV9206) near the tested bearing under a noisy surrounding environment. The shaft

rotating speed was 1464.6 r/min and the sampling frequency of the data acquisition system was set to be 20 kHz. Therefore, the characteristic periods of the bearing outer-race and inner-race defects are calculated to be $T_O = 0.010$ s and $T_I = 0.007$ s, respectively.

5.2.1 Seeded Outer-Race Defect

The bearing acoustic signal with the outer-race defect is first analyzed. The waveform and spectrum of the original signal are displayed in Figure 12, where the transient impulses are seriously corrupted by the background noise. The improved TM method, the original TM method and the PSO-VMD method are used to analyze the signal in Figure 12 successively. Figure 13a presents the number of neighbors for each data point in the improved TM method, where the neighbor numbers of the noise data points are significantly larger than those of the impulse data points. The two vectors of the TN-LTSA output are synthesized according to the weight-based feature compensation strategy, and waveform of the synthetic TM feature is displayed in Figure 13b. It is obvious that the in-band noise is largely suppressed by the TN-LTSA algorithm, and the synthetic TM feature is symmetric due to the weight-based feature compensation strategy. The average period of the transients in Figure 13b is 0.097 s, which is close to $To = 0.010$ s, proving that the bearing has a defect in the outer raceway. The *GI* value of the synthetic TM feature is 0.5170. And the computing time of the improved TM method on this signal is 9.93 s. The result of the original TM method is displayed in Figure 14a, where a small amount of in-band noise is retained and the waveform is asymmetric. The *GI* value of the TM feature is 0.4319, which is smaller than that of the synthetic TM feature. The computing time of the original TM method on this signal is 98.05 s, which is longer than that of the improved TM method. The result of the PSO-VMD method is displayed in Figure 14b, in which the in-band noise is still heavy and the recognition of the transient components is affected. The *GI* value of the optimal target mode in Figure 14b is 0.3732, which is the smallest. The computing time of the PSO-VMD method on this signal is 103.72 s, which is the longest. Therefore, the proposed method has enhanced fault identification capability and computing efficiency.

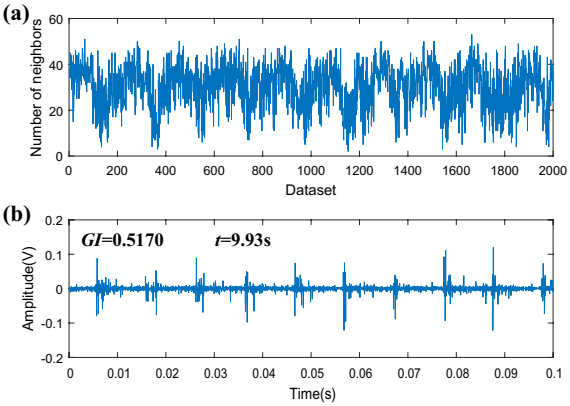

Figure 13 Results of the acoustic signal in Figure 12 analyzed by the proposed method: **a** the reallocated neighbor numbers by the TN algorithm and **b** the waveform of the synthetic TM feature

5.2.2 Seeded Inner-Race Defect

To further confirm the enhanced performance of the proposed method in bearing fault diagnosis, the bearing acoustic signal with the inner-face defect is also

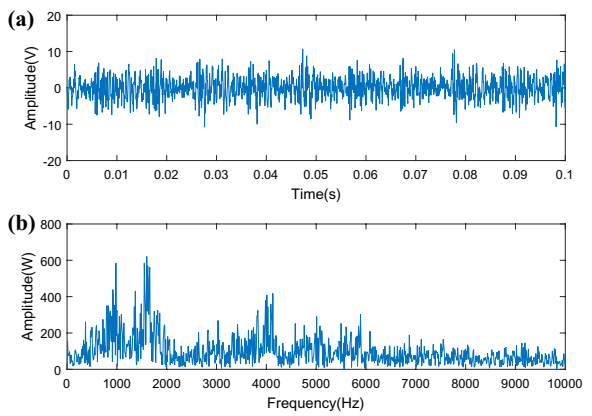

Figure 12 Acoustic signal with a seeded outer-race defect: **a** waveform and **b** spectrum

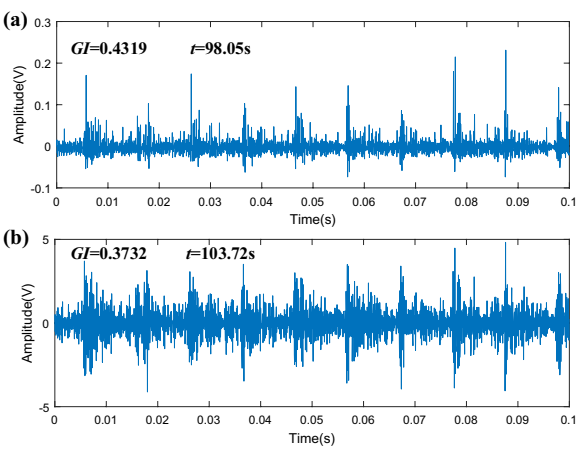

Figure 14 Waveforms of **a** the feature obtained by the original TM method and **b** the optimal target mode obtained by the PSO-VMD method for the signal in Figure 12

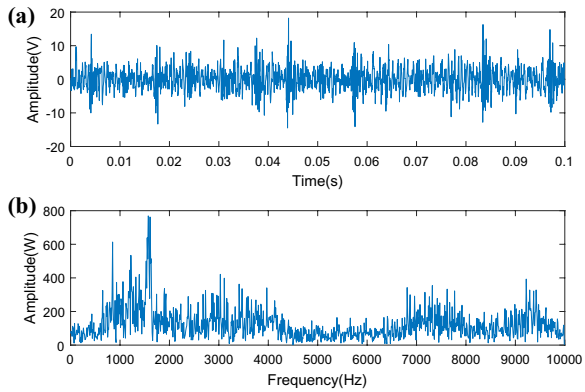

Figure 15 Acoustic signal with a seeded inner-race defect: **a** waveform and **b** spectrum

analyzed by the above three methods. Figure 15 shows the waveform and spectrum of the acoustic signal with the inner-race fault. The transient components in the waveform are contaminated by heavy noise. In the spectrum, no modulation resonance band spaced at the fault characteristic frequency aroused by fault impacts can be identified. When introducing the improved TM method to reveal the nonlinear inherent structure from the high-dimensional multi-bandwidth modes, the number of neighbors per data point is assigned based on the data density by the TN algorithm, as displayed in Figure 16a. Through the proposed compensation strategy, the synthetic TM feature is obtained and is shown in Figure 16b, where the in-band noise is almost completely removed and the transient impulses can be easily identified. The average period of the transients in Figure 16b is 0.068 s, which is close to $T_i = 0.007$ s, proving that the bearing has a defect in the inner raceway. The results of the original TM and the PSO-VMD methods are shown in Figure 17 for comparisons. The TM feature shown in Figure 17a is asymmetric and the optimal target mode in Figure 17b contains notable in-band noise. As can be seen in the figures of different results, the improved TM method achieves the highest *GI* value and costs the shortest computation time, demonstrating again that the proposed improved TM method is superior to the traditional methods for bearing fault diagnosis.

6 Conclusions

(1) Aiming to overcome the shortcomings of the original TM method, this paper proposes an improved TM method for enhanced bearing fault diagnosis. The proposed method improves the computing efficiency significantly and achieves impulse-enhanced and symmetrical fault feature of the rolling bearings.

(2) The TN algorithm is introduced to the LTSA algorithm to construct personalized local data distribution for each data point in the high-dimensional data of multi-bandwidth modes. The TN algorithm selects neighbors according to the data density, which is scale free and reasonable for local space construction. Therefore, the TN-LTSA only needs to be performed once to improve the efficiency, and the performance for extraction of bearing fault transients is enhanced.

(3) By considering the complementary waveform property of the two vectors of the TN-LTSA output, a weight-based feature compensation strategy is proposed to form a synthetic TM feature that is symmetric in the up-and-down direction. The synthetic

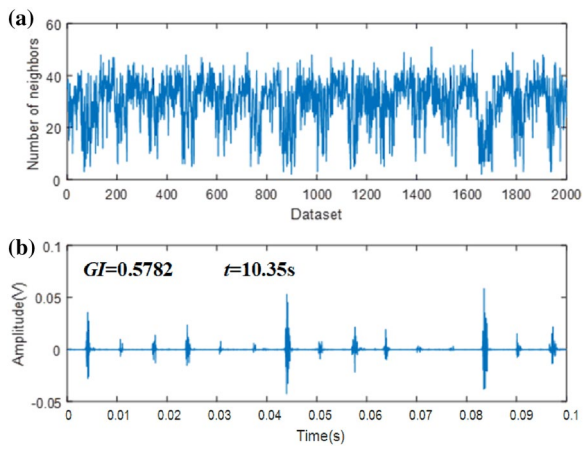

Figure 16 Results of the acoustic signal in Figure 15 analyzed by the proposed method: **a** the reallocated neighbor numbers by the TN algorithm and **b** the waveform of the synthetic TM feature

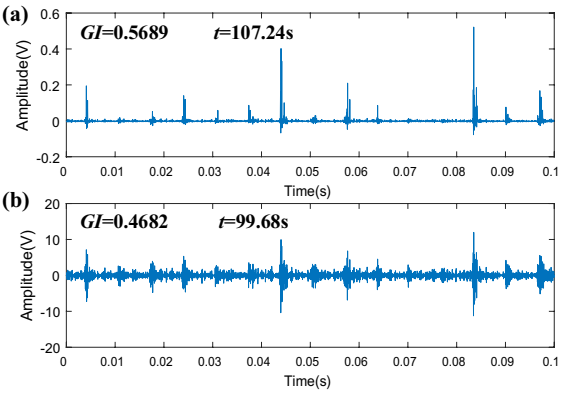

Figure 17 Waveforms of **a** the feature obtained by the original TM method and **b** the optimal target mode obtained by the PSO-VMD method for the signal in Figure 15

TM feature obtained by the improved TM method is more representative of the fault transient components than the TM feature obtained by the original TM method.

(4) The enhanced performance of the improved TM method is validated by one simulation study and two applications to bearing fault diagnosis, by comparing with the original TM method and a traditional VMD method.

Acknowledgements
Not applicable.

Authors' contributions
JW was in charge of the whole trial; GD and TJ wrote the manuscript; JW substantively revised the manuscript; XJ assisted with sampling and laboratory analyses; ZZ assisted with design of the experiment. All authors read and approved the final manuscript.

Authors' Information
Guifu Du, born in 1990, is currently an associate professor at *School of Rail Transportation, Soochow University, China*. He received his B.E. and Ph.D. degrees from *China University of Mining and Technology, China*, in 2012 and 2017, respectively. His research interests include condition monitoring and fault diagnosis of rail vehicle transmission system, control and protection of DC traction power system and railway electrification.

Tao Jiang, born in 1994, is currently a master candidate on vehicle engineering at *School of Rail Transportation, Soochow University, China*. He received the B.S. degree on vehicle engineering from *Soochow University, China*, in 2017. His research interests include signal processing and machinery fault diagnosis.
Jun Wang, born in 1987, is currently an associate professor at *School of Rail Transportation, Soochow University, China*. He received his B.S. degree from *Wuhan University of Technology, China*, and Ph.D. degree from *University of Science and Technology of China,* in 2010 and 2015, respectively. From July 2015 to June 2017, he was with *University of Nebraska-Lincoln, USA*, where he was a Senior Research Associate at *Department of Electrical and Computer Engineering*. His research interests include signal processing, condition monitoring, fault diagnosis and intelligent maintenance of mechanical and electrical systems.

Xingxing Jiang, born in 1989, is currently an Associate Professor at *School of Rail Transportation, Soochow University, China*. He received his B.S. and Ph.D. degrees from *Nanjing University of Aeronautics and Astronautics, China*, in 2012 and 2016, respectively. His research interests include machinery condition monitoring and fault diagnosis, and time-frequency analysis.

Zhuzhong Kui, born in 1974, is currently a professor at *School of Rail Transportation, Soochow University, China*. He received his B.S. degree and M.S. degree from *Hefei Polytechnic University, China*, in 1997 and 2002, respectively. Then, he received his Ph.D. degree from *University of Science and Technology of China*, in 2005. His research interests include vehicle system dynamics and control, vibration measurement and signal processing.

Funding
Supported by National Natural Science Foundation of China (Grant Nos. 51805342,51875376, 52007128), Jiangsu Provincial Natural Science Foundation of China (Grant No. BK20180842), China Postdoctoral Science Foundation (Grant Nos. 2021M692354, 2018M640514), Suzhou Prospective Research Program of China (Grant No. SYG201932), and Jiangsu Provincial Natural Science Fund for Colleges and Universities of China (Grant No. 18KJB470022).

Competing Interests
The authors declare no competing financial interests.

Received: 24 March 2020 Revised: 16 April 2021 Accepted: 10 January 2022
Published online: 22 February 2022

References

[1] A Hu, L Xiang, S Xu, et al. Frequency loss and recovery in rolling bearing fault detection. *Chinese Journal of Mechanical Engineering*, 2019, 32: 35, https://doi.org/https://doi.org/10.1186/s10033-019-0349-3.
[2] W Guo, X Jiang, N Li, et al. A coarse TF ridge-guided multi-band feature extraction method for bearing fault diagnosis under varying speed conditions. *IEEE Access*, 2019, 7: 18293-18310.
[3] Y Mao, S Qin, Y Qin. Demodulation based on harmonic wavelet and its application into rotary machinery fault diagnosis. *Chinese Journal of Mechanical Engineering*, 2009, 22(3): 107-113.
[4] X Zhang, N Hu, Z Cheng, et al. Enhanced detection of rolling element bearing fault based on stochastic resonance. *Chinese Journal of Mechanical Engineering*, 2012, 25(6): 1287-1297.
[5] G Chen, L Qie, A Zhang, et al. Improved CICA algorithm used for single channel compound fault diagnosis of rolling bearings. *Chinese Journal of Mechanical Engineering*, 2016, 29(1): 204-211.
[6] Y Hu, S Zhang, A Jiang, et al. A new method of wind turbine bearing fault diagnosis based on multi-masking empirical mode decomposition and fuzzy c-means clustering. *Chinese Journal of Mechanical Engineering*, 2019, 32:46, https://doi.org/https://doi.org/10.1186/s10033-019-0356-4.
[7] J Lian, Z Liu, H Wang, et al. Adaptive variational mode decomposition method for signal processing based on mode characteristic. *Mechanical Systems and Signal Processing*, 2018, 107: 53-77.
[8] Z Chen, J Xu, D Yang. New method of extracting weak failure information in gearbox by complex wavelet denoising. *Chinese Journal of Mechanical Engineering*, 2008, 21(4): 87-91.
[9] P K Kankar, S C Sharma, S P Harsha. Fault diagnosis of rolling element bearing using cyclic autocorrelation and wavelet transform. *Neurocomputing*, 2013, 110: 9-17.
[10] H Hong, M Liang. Fault severity assessment for rolling element bearings using the Lempel–Ziv complexity and continuous wavelet transform. *Journal of Sound and Vibration*, 2009, 320: 452-468.
[11] G Shen, L Tao, Z Chen. Gearbox fault diagnosis based on empirical mode decomposition. *Chinese Journal of Mechanical Engineering*, 2004, 17(3): 454-456.
[12] X Jiang, J Wang, C Shen, et al. An adaptive and efficient variational mode decomposition and its application for bearing fault diagnosis. *Structural Health Monitoring*, 2020. https://doi.org/10.1177/1475921720970856.
[13] J Gilles. Empirical wavelet transform. *IEEE Transactions on Signal Processing*, 2013, 61(16): 3999-4010.
[14] J Zheng, H Pan, S Yang, et al. Adaptive parameterless empirical wavelet transform based time-frequency analysis method and its application to rotor rubbing fault diagnosis. *Signal Processing*, 2017, 130: 305-314.
[15] K Dragomiretskiy, D Zosso. Variational mode decomposition. *IEEE Transactions on Signal Processing*, 2014, 62(3): 531-544.
[16] Z Li, J Chen, Y Zi, et al. Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive. *Mechanical Systems and Signal Processing*, 2017, 85: 512-529.
[17] A Humeau-Heurtier, P Abraham, S Henni. Bi-dimensional variational mode decomposition of laser speckle contrast imaging data: A clinical approach to critical limb ischemia. *Computers in Biology and Medicine*, 2017, 86: 107-112.
[18] W Mou, L Shi, Y Cai, et al. IC engine fault diagnosis method based on KVMD-PWVD and LNMF. *Journal of Vibration and Shock*, 2017, 36(2): 45-51.
[19] J Zhu, C Wang, Z Hu, et al. Adaptive variational mode decomposition based on artificial fish swarm algorithm for fault diagnosis of rolling bearings. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 2015, 231(4): 635-654.
[20] X Yan, M Jia, L Xiang. Compound fault diagnosis of rotating machinery based on OVMD and a 1.5-dimension envelope spectrum. *Measurement Science and Technology*, 2016, 27(7): 075002.

[21] C Yi, Y Lv, Z Dang. A fault diagnosis scheme for rolling bearing based on particle swarm optimization in variational mode decomposition. *Shock and Vibration*, 2016, 2016: 9372691.
[22] G Tang, X Wang. Parameter optimized variational mode decomposition method with application to incipient fault diagnosis of rolling bearing. *Journal of Xi'an Jiaotong University*, 2015, 49(5): 73-81.
[23] X Zhang, Z Liu, Q Miao, et al. A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery. *Mechanical Systems and Signal Processing*, 2018, 108: 58-72.
[24] S Zhang, Y Wang, S He, et al. Bearing fault diagnosis based on variational mode decomposition and total variation denoising. *Measurement Science and Technology*, 2016, 27(7): 1-10.
[25] Y Miao, M Zhao, J Lin. Identification of mechanical compound-fault based on the improved parameter-adaptive variational mode decomposition. *ISA transactions*, 2018, 84: 82-95.
[26] X Wang, Z Yang, X Yan. Novel particle swarm optimization-based variational mode decomposition method for the fault diagnosis of complex rotating machinery. *IEEE/ASME Transactions on Mechatronics*, 2018, 23(1): 68-79.
[27] T Jiang, J Wang, C Shen, et al. Multi-bandwidth mode manifold for fault diagnosis of rolling bearings. *IEEE Access*, 2019, 7: 179620-179633.
[28] Y Wang, G Xu, L Liang, et al. Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis. *Mechanical Systems and Signal Processing*, 2015, 54-55: 259-276.
[29] J Wang, Q He. Wavelet packet envelope manifold for fault diagnosis of rolling element bearings. *IEEE Transactions on Instrumentation and Measurement*, 2016, 65(11): 2515-2526.
[30] J Wang, G Du, Z Zhu, et al. Fault diagnosis of rotating machines based on the EMD manifold. *Mechanical Systems and Signal Processing*, 2020, 135: 106443.
[31] Q Zhu, J Feng, J Huang. Natural neighbor: A self-adaptive neighborhood method without parameter K. *Pattern Recognition Letters*, 2016, 80: 30-36.
[32] D Wang. Some further thoughts about spectral kurtosis, spectral L2/L1 norm, spectral smoothness index and spectral Gini index for characterizing repetitive transients. *Mechanical Systems and Signal Processing*, 2018, 108: 360-368.
[33] J Wang, Q He, F Kong. Multiscale envelope manifold for enhanced fault diagnosis of rotating machines. *Mechanical Systems and Signal Processing*, 2015, 52-53: 376-392.
[34] J Lee, H Qiu, G Yu, et al. Rexnord Technical Services. IMS, University of Cincinnati. "Bearing Data Set", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA. 2007. https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository.
[35] H Qiu, J Lee, J Lin, et al. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of Sound and Vibration*, 2006, 289(4-5): 1066-1090.