International Journal of Innovative
Computing, Information and Control
Volume **15**, Number **4**, August **2019**

ICIC International ©2019 ISSN 1349-4198
pp. 1485–1505

AN ADAPTIVE VMD METHOD BASED ON IMPROVED GOA TO EXTRACT EARLY FAULT FEATURE OF ROLLING BEARINGS

Chengjiang Zhou[1,2], Jun Ma[1,2,*], Jiande Wu[1,2] and Xuyi Yuan[1,2]

[1]Faculty of Information Engineering and Automation
Kunming University of Science and Technology
[2]Engineering Research Center for Mineral Pipeline Transportation of Yunnan Province
No. 727, South Jingming Road, Chenggong District, Kunming 650500, P. R. China
*Corresponding author: mjun@kmust.edu.cn

Received November 2018; revised March 2019

Abstract. *In order to identify the early fault of bearing, an early feature extraction method based on adaptive variational mode decomposition (VMD) is proposed. The method not only improves the local optimal problem of grasshopper optimization algorithm (GOA) but also can adaptively determine the mode number and penalty parameter of VMD. Firstly, the convex-concave decreasing strategy is introduced to adjust the decreasing coefficient of GOA. Then, energy entropy mutual information (EEMI) index is introduced to consider the energy distribution of modes and the dependence between modes and the original signal. Secondly, the optimal parameters of VMD matching with the input signal are obtained by taking the maximum EEMI as the objective function. Finally, the bearing signal is decomposed by the optimized VMD and the sensitive mode with maximum kurtosis is determined, and the fault feature contained in the sensitive mode can be extracted by envelope demodulation. The optimization experiments of 23 sets of benchmark functions show that the convex-concave strategy enhances the balance between exploration and exploitation, and the global and local search ability and stability of the GOA are improved. The experiments on simulation signal and bearing signal show that IGOA-VMD has better decomposition performance than VMD with fixed parameters and GOA-VMD. Therefore, this method provides a new idea and solution for fault feature extraction of bearing and other key components.*

Keywords: Variational mode decomposition (VMD), Improved grasshopper optimization algorithm (IGOA), Energy entropy mutual information (EEMI) index, Bearing, Signal decomposition

1. **Introduction.** Rolling bearing is an important component of the rotating machinery, and it is widely used in various projects. The structure of the rotating machinery is complex, its operating environment is very harsh, and the rolling bearing is easily damaged. Bearing failures can cause major accidents and property losses [1,2]. However, the early warning mechanism and replacement strategy can be implemented in time through early fault detection, and the accident loss will be reduced [3]. Therefore, early fault detection is particularly important.

Envelope demodulation method is often used to detect the fault, and it can retain the resonant frequency band by bandpass filtering [4]. However, the determination of the optimum demodulation band is one of the keys and difficult problems in envelope demodulation. Three effective methods can be adopted. (1) We can detect the pulse series contained in the signal by means of bandpass filter and kurtosis (Kurt), and the popular methods include spectral kurtosis (SK), Infogram, etc. [5]. (2) We can determine the demodulation frequency band by detecting the periodic impulse contained in the signal,

DOI: 10.24507/ijicic.15.04.1485

and the classical methods include minimum entropy deconvolution (MED) and maximum correlation kurtosis deconvolution (MCKD), etc. [6]. (3) Signal decomposition is another effective method that can be selected. The classical methods include wavelet packet decomposition (WPD), empirical mode decomposition (EMD), local mean decomposition (LMD), etc. In addition, variational mode decomposition (VMD) and empirical wavelet transform (EWT) are novel methods proposed in recent years. However, these methods have some inherent drawbacks. Frequency mixing exists in WPD during decomposition [7]. The problems of mode mixing and endpoint effect exist in the decomposition process of EMD and LMD in the practical application [8]. The Fourier segmentation required by EWT strongly depends on the local maximum of the amplitude of the Fourier spectrum, which means that the Fourier segmentation is not reliable [9]. Dragomiretskiy and Zosso proposed VMD in 2014 [10]; the method not only has a solid theoretical basis but also has a robust performance for low sampling rate and noise. However, the mode number and penalty parameter of VMD need to be set in advance, and over-decomposition, under-decomposition and mode mixing may occur when the parameters are not appropriate.

To address these problems, many improved versions of VMD have been reported recently. Lian et al. [11] determined the mode number of VMD through the energy loss coefficient, permutation entropy, frequency domain extreme point and kurtosis, but the decomposition performance of the method was reduced due to the setting of multiple thresholds in the method. Li et al. [12] first determined the mode number of VMD through the peak search and similarity principle, then the modes with similar characteristics were combined to overcome the over-decomposition, and finally, the method was used for bearing fault diagnosis. In order to determine the mode number effectively, Yan et al. [13] treated the modes whose cross-correlation coefficients (Cor) are less than the threshold as invalid modes. The mode number of VMD was determined by setting the threshold artificially in the above methods. The randomness was introduced and the penalty parameter was not considered. To address these problems, Liu et al. [14] decomposed signals by VMD with different parameter combinations, and then the parameter set with the maximum kurtosis value was selected as the optimal parameters. Finally, this method was applied to chatter detection in the milling process. Li et al. [15] took the difference between the estimated bandwidth and the actual bandwidth as the objective function, and then optimized the two parameters of VMD by genetic algorithm (GA). Finally, the bearing degradation process is described by VMD and Gath-Geva clustering. The heuristics algorithm is proved to be effective in parameter selection. Therefore, the intelligent optimization algorithm is considered in this paper.

The optimization algorithm is more and more widely used in signal processing. Guo and Deng [16] introduced a spline interpolation method with shape control parameters into the EMD and used particle swarm optimization (PSO) to find the optimal intrinsic mode function (IMF) and shape control parameters. Li et al. [17] optimized the threshold parameters of MED by the shuffled frog leaping algorithm (SFLA), and then this method was applied to extracting fault characteristics. Shao et al. [18] optimized the key parameters of the deep auto-encoder through the artificial fish swarm algorithm (AFSA), and then used this method for bearing fault diagnosis. In addition, many optimization algorithms have been published in recent years. Saremi et al. [19] proposed the grasshopper optimization algorithm (GOA) in 2017 and compared it with a variety of optimization algorithms. The results show that GOA has outstanding advantages in the optimization of unimodal functions, multimodal functions, and composite functions. However, GOA has some inherent drawbacks, the algorithm is easy to fall into local optimum, and the convergence speed is slow. Ewees et al. [20] initialized and updated the population by using an opposition-based learning (OBL) strategy in the iteration, but the method is much

more time-consuming than GOA. Luo et al. [21] improved GOA through opposition-based learning (OBL), Gaussian mutation and Levy-flight strategy, but the combined approach of the three strategies is much more time-consuming and complex than the original GOA. Therefore, this paper attempts to explore a simple and effective improvement strategy. In the proposed approach, the linear decreasing strategy in the original GOA is replaced by a simple convex-concave decreasing strategy. The decreasing process of parameter c is divided into two stages. The parameter c decreases slowly in the early stage of iteration, which increases the interaction between grasshoppers and the grasshoppers move rapidly in a wide space. In the later stage of iteration, the parameter c decreases rapidly, which reduces the interaction between grasshoppers, and the grasshoppers move slowly in a small range to find the optimal solution. Therefore, the global search ability and local search ability of the algorithm are enhanced.

The selection of the objective function directly determines the optimal solution of the optimization problem. Cor [22], mutual information (MI) [23], Kurt [24], energy coefficient (EC) [25], and energy entropy (EE) [26] are the most commonly used evaluation indices in signal processing. MI can characterize the interdependence and matching degree between two signals [23]. However, if only MI is used as the objective function, some defect information may be lost. EE can characterize the uncertainty of energy distribution. When the bearing is running stable, the energy of the signal depends on the rotating frequency and its harmonics. The signal energy is gradually shifted from the rotating frequency to the fault frequency when an early failure occurs. Therefore, the fault state is closely related to the energy distribution of frequency [26]. However, if only EE is used as the objective function, the modes with high similarity to the original signal may be ignored.

In summary, an adaptive VMD based on IGOA is proposed. Firstly, the nonlinear decreasing strategy is introduced to adjust the decreasing coefficient c of the original GOA, and the balance ability of exploration and exploitation is improved. Then, the energy entropy mutual information (EEMI) index composed of energy entropy and mutual information is introduced to consider the energy distribution of the modes and the dependence between the modes and the original signal. Subsequently, the optimal parameters matched with the input signal can be obtained by taking the maximum EEMI index as the optimization objective. Finally, the fault information can be obtained by analyzing the sensitive mode with the maximum kurtosis. The validity of the proposed method is proved by simulation experiment and testing platform.

The rest of this paper is organized as follows. Theories of VMD and GOA are introduced briefly in Section 2. In Section 3, an adaptive VMD based on IGOA is introduced in detail. The validity of the proposed method is verified by benchmark functions and simulation signals in Section 4. The conclusions of the experiment are summarized in Section 5.

2. **Material and Methods.**

2.1. **Variational mode decomposition.** VMD is a decomposition method based on variational constraint model, which can decompose a multi-component signal $f(t)$ into K meaningful modes $u_k(t)$ $(k = 1, \ldots, K)$ with non-recursively sifting structure [27]. Each mode is closely around a central frequency ω_k, and the bandwidth is estimated through the square L^2 norm of the gradient [28]. The solving process of $u_k(t)$ and ω_k is briefly described as follows.

(1) Determine the mode number K and penalty parameter α in advance and initialize the modes $u_k(t)$ $(k = 1, \ldots, K)$. K and α are usually set to two estimates [3,11,12].

(2) Compute the analytic signal $Au_k(t)$ corresponding to $u_k(t)$ by Hilbert transform, and its one-sided spectrum can be obtained.

$$Au_k(t) = u_k(t) + \frac{j}{\pi}\int_{-\infty}^{+\infty}\frac{u_k(\tau)}{t-\tau}d\tau = \left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t) \tag{1}$$

where $j = \sqrt{-1}$.

(3) The analytic signal of each mode is multiplied by $e^{-j\omega_k t}$, and the spectrum of the signal is shifted to the corresponding estimated central frequency ω_k. As a constrained optimization problem, the goal of optimization is to minimize the sum of the spectral bandwidth of all modes.

$$\begin{cases} \min\limits_{\{u_k\},\{\omega_k\}} \left\{ \sum\limits_k \left\| \partial_t \left[\left(\delta(t) + \dfrac{j}{\pi t}\right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} \sum\limits_k u_k = f(t) \end{cases} \tag{2}$$

where δ denotes the Dirac distribution and $*$ is the convolution operator.

(4) Lagrangian multiplier λ and penalty parameter α are introduced to render the reconstruction problem in Equation (2). The constraint problem is transformed into the following non-constrained problem.

$$\begin{aligned} L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t) \right] e^{-j\omega_k t} \right\|^2 \\ + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \tag{3}$$

(5) Solve Equation (4) by using the alternating direction method of multipliers (ADMM) introduced in [29] to obtain the modes u_k and its central frequency ω_k.

$$\widehat{u}_k^{n+1}(\omega) = \frac{\widehat{f}(\omega) - \sum\limits_{i\neq k}\widehat{u}_i(\omega) + \widehat{\lambda}(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2} \tag{4}$$

$$\omega_k^{n+1} = \int_0^\infty \omega|\widehat{u}_k(\omega)|^2 d\omega \Big/ \int_0^\infty |\widehat{u}_k(\omega)|^2 d\omega \tag{5}$$

The detailed steps and descriptions of the VMD algorithm can be found in [10]. There are many methods to determine the VMD parameters at present, but they all have their own shortcomings. Lian et al. [11] determined the mode number through several parameter thresholds, the randomness of the algorithm increases and the method is complicated. Li et al. [12] and Yan et al. [13] determined the mode number by the similarity principle, randomness is introduced and the penalty parameter is not taken into account. The grid search method [14] is an exhaustive search method affected by step length, and its search cost is relatively large. Genetic algorithm [15] is easy to fall into local optimum and has poor stability. Compared with genetic algorithm and other methods, GOA has outstanding advantages in complex function optimization [19]. Therefore, GOA is used to optimize the parameters of VMD.

2.2. **Grasshopper optimization algorithm.** GOA algorithm is an intelligence algorithm, which simulates the swarm behavior of grasshoppers. It has two characteristics. (1) The adults move quickly and have a wide range of motion, and they are mainly used

to explore the search space. (2) The larvae move slowly, and they are used to develop candidate solutions near the target. The mathematical model of GOA is as follows.

$$X_i = S_i + G_i + A_i \tag{6}$$

$$\begin{cases} S_i = \sum_{j=1, j\neq i}^{N} s(d_{ij})\widehat{d_{ij}} \\ G_i = -g\widehat{e_g} \\ A_i = u\widehat{e_w} \end{cases} \tag{7}$$

where X_i defines the position of the i-th grasshopper, S_i is the social interaction, G_i is the gravity force on the i-th grasshopper, and A_i shows the wind advection. In (7), N is the number of grasshoppers, d_{ij} is the distance between the i-th and the j-th grasshopper, calculated as $d_{ij} = |x_j - x_i|$, $\widehat{d_{ij}} = (x_j - x_i)/d_{ij}$ is a unit vector from the i-th grasshopper to the j-th grasshopper. g is the gravitational constant and $\widehat{e_g}$ shows a unity vector towards the centre of earth. u is a constant drift and $\widehat{e_w}$ is a unity vector in the direction of wind. The s function, which defines the social forces, is calculated in (8).

$$s(r) = fe^{\frac{-r}{l}} - e^{-r} \tag{8}$$

where f indicates the intensity of attraction and l is the attractive length scale. Substituting S_i, G_i, and A_i in (6), this equation can be expanded as follows:

$$X_i = \sum_{j=1, j\neq i}^{N} s(|x_j - x_i|)\frac{x_j - x_i}{d_{ij}} - g\widehat{e_g} + u\widehat{e_w} \tag{9}$$

In fact, the model utilized for the swarm is in free space. Therefore, we do not consider gravity (no G_i component) and assume that the wind direction (A_i component) is always towards a target ($\widehat{T_d}$), the improved update method is as follows:

$$X_i^d = c\left(\sum_{j=1, j\neq i}^{N} c\frac{ub_d - lb_d}{2} s\left(\left|x_j^d - x_i^d\right|\right)\frac{x_j - x_i}{d_{ij}}\right) + \widehat{T_d} \tag{10}$$

where ub_d and lb_d are the upper bound and lower bound of the D-dimensional search space, respectively. $\widehat{T_d}$ is the best solution found so far. In order to balance exploration and exploitation of the entire swarm around the target, the parameter c needs to be reduced as the number of iterations increases.

$$c = c_{\max} - l\frac{c_{\max} - c_{\min}}{L} \tag{11}$$

where $c_{\max}$ is the maximum value, $c_{\min}$ is the minimum value, l indicates the current iteration, and L is the maximum number of iterations. In this work, we use 1 and 0.00001 for $c_{\max}$ and $c_{\min}$ respectively [19,28].

GOA is easy to fall into local optimum, and the convergence speed is slow. Ewees et al. [20] updated the population through opposition-based learning (OBL), and Luo et al. [21] improved the GOA by combining OBL, Gaussian mutation and Levy-flight strategies. These improved methods increase the complexity and computation time of the algorithm, so we only improve the decreasing coefficient c to optimize the original GOA, and the improved method is simple and effective.

3. Proposed Method.

3.1. Improved GOA algorithm.

The parameter c is a linear decreasing function in the original GOA, and it is used twice in Equation (10). The outer c balances the exploration and exploitation of the swarm near the target. It is mainly used to reduce the movement of grasshoppers near the target. The inner c reduces the repulsion or attraction forces between grasshoppers proportional to the number of iterations.

For the original GOA, the decreasing function c decreases linearly with increasing iteration number, which means that the interaction force between grasshoppers decreases linearly. Two fatal defects may exist in the linear decreasing function. (1) A wide range of parameter space is omitted due to the small interaction force between grasshoppers in the early iterations. (2) The globally optimal solution is lost because the interaction forces between grasshoppers are relatively large in the later stage of iteration, at which point grasshoppers hover around the target.

In order to adjust the descending rate of decreasing coefficient c, two nonlinear decreasing strategies are discussed in this section. As shown in Figure 1, $c1$ is called a convex-concave strategy, and the improved method based on $c1$ is called IGOA1. $c2$ is known as concave-convex strategy, and the improved method based on $c2$ is called IGOA2.

$$c1 = 0.5 + 0.5 * \cos\left(\frac{\pi * l}{L}\right) \tag{12}$$

$$c2 = c_{\min} + \left(\left(-\left(\frac{2l}{L} - 1\right)\right)^3 + 1\right) * \left(\frac{c_{\max} - c_{\min}}{2}\right) \tag{13}$$

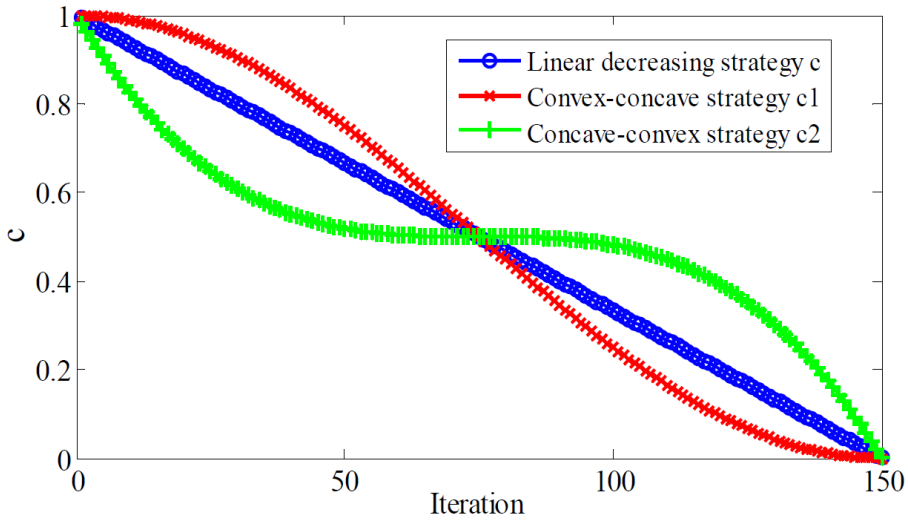


FIGURE 1. Two optimization strategies of c

For the GOA1, the descending rate of $c1$ is very slow in the early iterations, which means that the interaction force between grasshoppers is relatively large. As a result, grasshoppers move quickly in a wide range of parameter space, and larger search space can be explored. The descending speed of $c1$ is very fast in the later stage of iteration, which means that the interaction force between grasshoppers becomes very small. Therefore, its exploration ability is weakened, the search space becomes smaller, and the exploitation ability is enhanced. The optimal solution can be obtained because the grasshoppers move slowly in a small search space near the target. The improved strategy is an ideal optimization strategy in theory.

For the GOA2, the descending rate of $c2$ is very fast in the early iterations, which means that the interaction force between grasshoppers becomes very small. As a result, a wide range of effective search space is lost in the early iterations. The descending speed of $c2$ is very slow in the later stage of iteration, which means that grasshoppers move relatively fast near the target. Grasshoppers may fly over the optimal solution, and the algorithm eventually falls into a local optimum. In theory, the strategy is not advisable and it is used as a control group in this experiment.

3.2. **Construction of optimization model.** The dependence between the original signal and the modes obtained by VMD can be characterized by mutual information (MI). When the fault state changes, the energy entropy (EE) can not only reflect the energy distribution of the frequency domain but also characterize the transfer of energy in the modes [26]. Two factors are considered comprehensively, and the energy entropy mutual information (EEMI) index composed of energy entropy and mutual information is proposed. Considering the minimum mode number is 2, the sum of EEMI of the first two modes is taken as the optimization objective. The optimization problem is described as follows:

$$\begin{cases} \mathit{fitness} = \min\limits_{\gamma=(K,\alpha)} \left\{ -\sum\limits_{i=1}^{2} EEMI \right\} \\ K \in [2, 10] \\ \alpha \in [1000, 10000] \end{cases} \tag{14}$$

where *fitness* represents the objective function, $\gamma = (K, \alpha)$ is the parameter group of VMD to be optimized.

$$EEMI = EE * MI \tag{15}$$

EEMI is the product of EE and MI for each IMF, where EE can be regarded as the weight of MI. The EE of each mode is calculated as follows:

$$E_i = \int_{-\infty}^{+\infty} |imf_i(t)|^2 dt, \quad i = 1, 2, \ldots, K \tag{16}$$

where $imf_i(t)$ $(i = 1, \ldots, K)$ are the modes of different frequency bands, and $E = \{E_1, E_2, \ldots, E_K\}$ is the energy distribution of the vibration signals in the frequency domain. In order to facilitate the analysis, the energy is normalized, σ_i is energy coefficient (EC).

$$\sigma_i = E_i / E \tag{17}$$

where $E = E_1 + \cdots + E_K$, the EE of IMF is defined based on Shannon entropy.

$$EE = -\sum_{i=1}^{N} (\sigma_i) \ln(\sigma_i) \tag{18}$$

MI is used to measure the dependence between the original signal and modes, and it is more effective than the cross-correlation coefficient (Cor) [23]. The definition of MI is given in the discrete domain, and it is equivalent to Formula (20).

$$MI(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{19}$$

$$MI(X;Y) = H(Y) - H(Y|X) \tag{20}$$

where $p(x)$ and $p(y)$ denote the edge probability distribution function of X and Y respectively, $p(x, y)$ denotes the joint probability distribution function of X and Y. $H(Y)$ is the edge entropy of Y, and $H(Y|X)$ is the conditional entropy.

3.3. **Parameter adaptive VMD.** As shown in Figure 2, the detailed steps of the proposed method are summarized as follows.

(1) Set the ranges of the VMD parameters, and initialize the population X and IGOA parameters including search agents N and maximum iteration number L. Here the mode number $K \in [2, 10]$, penalty parameter $\alpha \in [1000, 10000]$, $N = 30$, $L = 10$ [28], population X can be initialized by these parameters.

(2) Decompose the experiment signal using VMD and calculate EEMI of all IMFs, the fitness $-\sum_{i=1}^{2} EEMI$ can be obtained. Save the best fitness and corresponding position.

(3) $l = l + 1$, update parameter $c1$ through the improved strategy (12).

(4) Normalize the distance between search agents to [1,4], and update the location of the search agent through an improved strategy (12), where c is replaced by $c1$.

(5) Determine if all search agents have been updated, and if not, perform step (4). Otherwise, go to step (6).

(6) Decompose the experiment signal using VMD and calculate fitness $-\sum_{i=1}^{2} EEMI$, save the best fitness and corresponding position.

(7) Update the population X through the best search agent.

(8) Determine if the maximum number of iterations is reached, and if not, perform step (3) to step (7). Otherwise, the best fitness and parameter combination can be obtained.

(9) Decompose the experiment signal using optimized VMD and take the IMF with maximum kurtosis as the sensitive mode.

(10) Analyze the sensitive mode through Hilbert envelope demodulation.

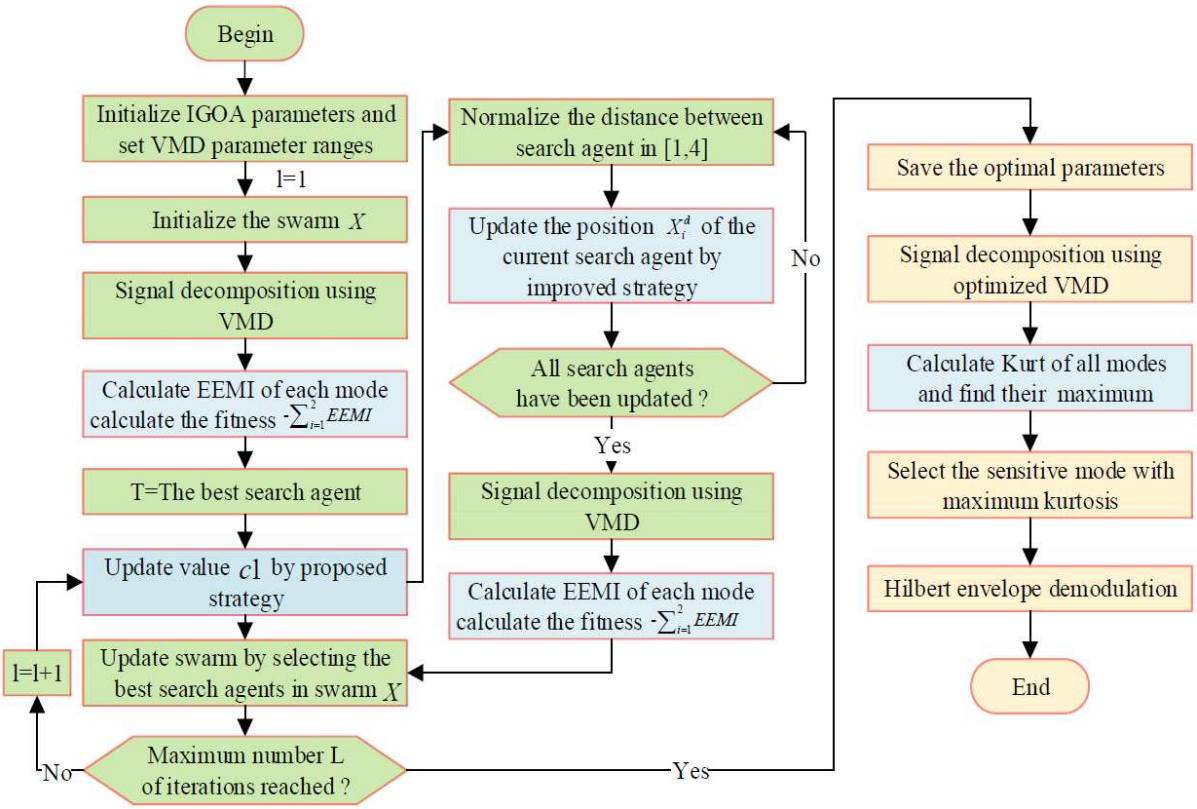


FIGURE 2. The optimization process of the proposed IGOA-VMD model

4. Experimental Study.

4.1. **Function optimization experiment.** In order to evaluate the optimization performance of the proposed method, we optimized 23 benchmark functions by IGOA1 and IGOA2. The detailed description of the functions can be found in [20]. All experiments are implemented using MATLAB R2014a under the environment of Intel(R) Core(TM) i5-5200U CPU @ 2.20GHz 2.19GHz, 8GRAM. In order to avoid random error, we have run 30 times for every method on each function. Average fitness (Avg), standard deviation (STD), success rate (SR) and average time (Time) are used to evaluate optimization results, as shown below.

$$Avg = \mu_F = \frac{1}{N_r}\sum_{i=1}^{N_r} F_i^* \tag{21}$$

$$STD = \sqrt{\frac{1}{N_r - 1}\sum_{i=1}^{N_r} (F_i - \mu_F)^2} \tag{22}$$

$$SR = \left(\frac{\text{Number of times reached VTR}}{N_r}\right) \tag{23}$$

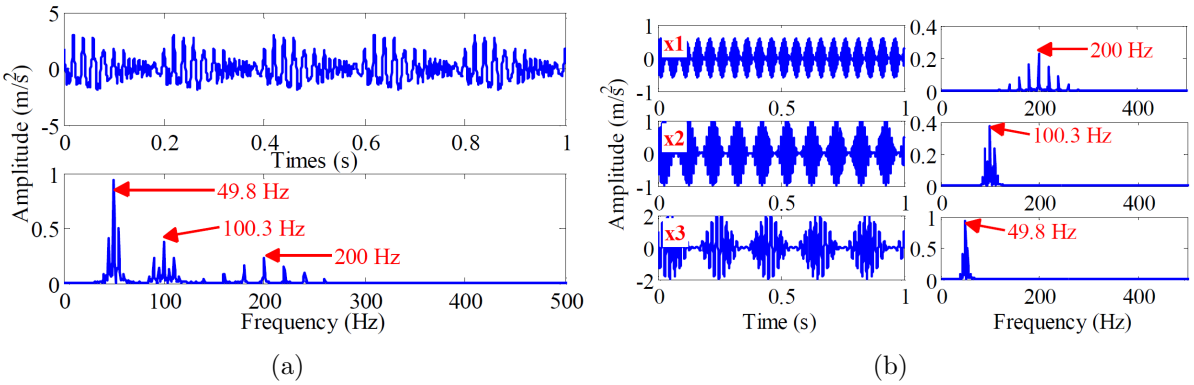
where Avg represents the average of best fitness in 30 runs, STD represents the dispersion degree of the best fitness, and SR is the ratio of the number of times an algorithm reached the required value-to-reach (VTR = 10E-5).

In this study, the maximum number of iterations (Maxiter) is set to 150, and the population size is 50. The optimization results obtained by the three strategies are shown in Table 1. All Avg indexes obtained by IGOA1 are smaller than that obtained by GOA except F16, which indicates that the global search ability of IGOA1 is improved compared with GOA. The Avg indexes obtained by IGOA2 are less than or equal to that obtained by GOA except F1, F4, F6, F9, F11, F13, F22, and F23, which indicates that the concave-convex strategy is not feasible for the above functions. The STD indexes obtained by IGOA1 are smaller than that obtained by GOA except F20, which indicates that the stability of IGOA1 is improved compared with GOA. The STD indexes obtained by IGOA2 are less than or equal to that obtained by GOA for half of the functions, which shows that the stability of IGOA2 is fairly close to GOA. The SR indexes obtained by IGOA1 are all larger than that obtained by GOA and IGOA2, which indicates that the optimal performance of IGOA1 is the best compared with the other two methods. The optimal performance of IGOA2 is the worst, and it will not be discussed in subsequent experiments. In addition, the time consumption of the three strategies is almost equal, and the introduction of the improved strategies does not increase the time complexity. In summary, the experimental results coincide with theory analysis. It shows that the convex-concave strategy (GOA1) can increase the search space and the moving speed of grasshoppers in the early iterations, and the global exploration capability of the algorithm is improved. Meanwhile, the strategy (GOA1) can reduce the moving speed and search space of grasshoppers near the target, and the local exploitation capability is improved. Therefore, the optimal performance of IGOA1 is the best, and it is used in the later experiments. In the following sections, IGOA1 is called IGOA.

4.2. **Simulation signal analysis.** In order to verify the validity of the proposed method, we analyze a multi-component amplitude modulation and frequency modulation (AM-FM) signal $x(t)$ by the proposed method. The sampling frequency is 3000Hz and the sampling time is 1s. The time domain and frequency domain waveform of $x(t)$ and its

TABLE 1. Comparison between IGOA and original GOA

F	GOA				IGOA1				IGOA2			
	Avg	STD	Time	SR	Avg	STD	Time	SR	Avg	STD	Time	SR
F1	1.47E-04	4.33E-04	15.48	76.6	1.58E-09	1.96E-09	16.76	100	1.37E-02	7.5E-03	16.31	0
F2	1.57E+00	1.72E+00	15.86	0	7.86E-01	1.14E+00	16.70	0	7.03E-01	1.07E-01	16.39	0
F3	2.50E+00	2.02E+00	15.20	30	1.60E-03	4.40E-03	16.13	73.3	6.52E-02	1.49E-01	16.38	0
F4	3.70E-03	6.10E-02	15.42	0	2.51E-05	2.16E-05	16.94	100	6.25E-02	4.00E-02	16.65	0
F5	7.51E+03	2.31E+04	15.53	0	7.44E+02	1.26E+03	16.66	3.33	4.71E+03	1.64E+04	16.67	0
F6	4.15E-05	1.53E-04	15.25	93.3	2.33E-09	3.91E-09	16.53	100	9.80E-03	6.90E-03	16.82	0
F7	7.43E-02	1.31E-01	15.15	0	3.68E-02	5.51E-02	16.50	0	7.38E-02	1.69E-01	16.35	0
F8	−1.55E+03	2.40E+02	15.38	100	−1.59E+03	1.58E+02	15.76	100	−1.56E+03	2.00E+02	16.83	100
F9	7.28E+00	4.77E+00	15.16	0	6.30E+00	3.81E+00	15.80	0	9.10E+00	4.98E+00	16.56	0
F10	1.21E+00	9.42E-01	15.17	0	1.12E+00	8.50E-01	15.86	10	1.08E+00	1.00E+00	16.01	0
F11	2.41E-01	1.26E-01	15.68	0	1.93E-01	1.06E-01	16.43	0	2.73E-01	1.43E-01	16.21	0
F12	1.41E-01	2.63E-01	16.04	20	9.42E-02	2.32E-01	17.64	23.3	1.63E-01	2.38E-01	16.98	3.33
F13	5.10E-03	6.70E-03	16.02	36.6	2.60E-03	4.90E-03	16.90	53.3	8.50E-03	8.00E-03	16.89	0
F14	1.55E+00	1.24E+00	6.188	0	1.09E+00	5.42E-01	7.133	0	1.55E+00	1.39E+00	6.53	0
F15	1.30E-02	1.36E-02	10.65	0	9.10E-03	1.26E-02	10.84	0	1.11E-02	1.32E-02	10.65	0
F16	−1.031	5.01E-11	5.40	100	−1.031	1.06E-15	6.24	100	−1.031	4.70E-09	5.23	100
F17	4.04E-01	2.44E-02	5.74	0	3.97E-01	9.06E-14	5.15	0	3.97E-01	9.70E-09	5.34	0
F18	3.000	5.43E-10	5.46	0	3.000	2.27E-14	5.37	0	3.0000	4.24E-08	5.32	0
F19	−3.831	7.30E-03	10.85	100	−3.862	1.45E-04	10.51	100	−3.8292	1.41E-01	10.32	100
F20	−3.258	6.05E-02	16.16	100	−3.269	6.76E-02	16.21	100	−3.2577	7.67E-02	16.02	100
F21	−5.478	3.2608	10.86	100	−6.732	3.1632	10.62	100	−5.7341	3.5122	10.64	100
F22	−7.299	3.6401	10.91	100	−7.567	3.5774	11.50	100	−6.9929	3.5982	10.57	100
F23	−7.494	3.8247	10.11	100	−7.702	3.5914	10.46	100	−6.8863	3.7679	10.12	100



(a) (b)

FIGURE 3. (a) Simulated signal $x(t)$ and its frequency-domain waveform, and (b) its three components

component signals are shown in Figure 3. The dominant frequencies of the three component signals are 200Hz, 100Hz, and 50Hz, respectively.

$$\begin{cases} x(t) = x_1(t) + x_2(t) + x_3(t) \\ x_1(t) = (0.3 + 0.3\sin(40\pi t)) \times \cos(400\pi t + 1.5\sin(40\pi t)) \\ x_2(t) = (0.5 + 0.5\sin(20\pi t)) \times \cos(200\pi t + 0.5\sin(10\pi t)) \\ x_3(t) = (1 + \sin(10\pi t)) \times \cos(100\pi t + 0.3\sin(10\pi t)) \end{cases} \tag{24}$$

In order to explore the effect of the penalty parameter α on the decomposition results, α is set to 200 and 10000 respectively. Since the number of simulation components is 3, the mode number K is set to 3. The decomposition results are shown in Figure 4,

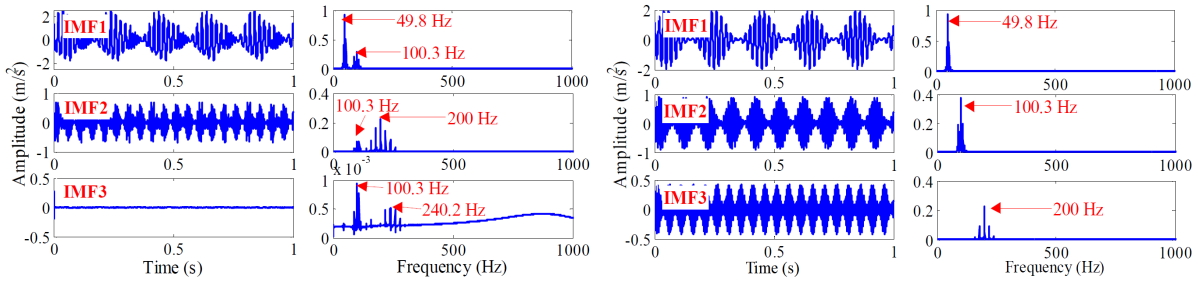


FIGURE 4. The effect of parameter (a) $\alpha = 200$ and (b) $\alpha = 10000$ on the result

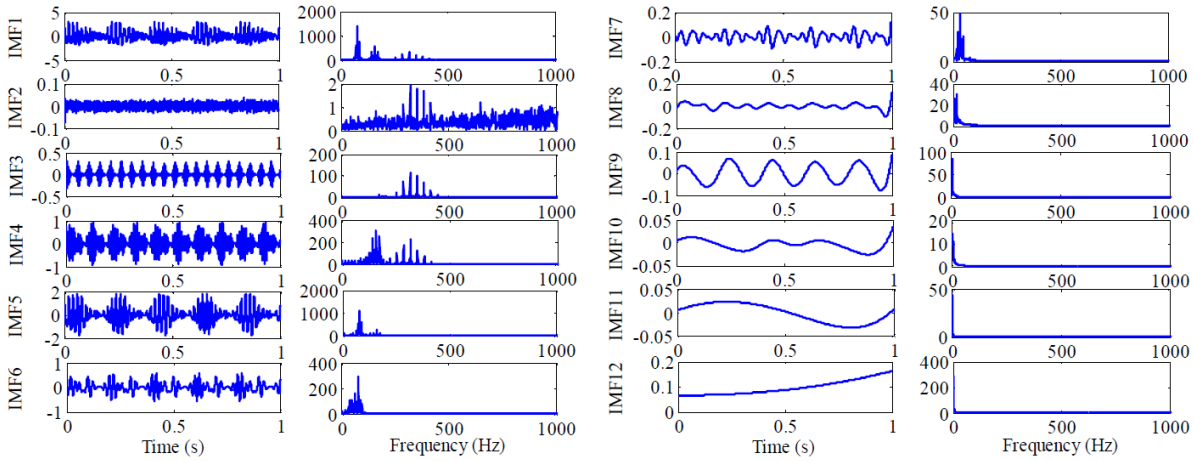


FIGURE 5. The decomposition results of simulated signal $x(t)$ obtained by EEMD

when $\alpha = 200$, 100.3Hz can be observed simultaneously in IMF1, IMF2, IMF3, and 200Hz appears simultaneously in IMF2 and IMF3. Both mode mixing and frequency mixing occur during the decomposition process. In addition, the bandwidth of each mode becomes wider compared to the original simulation components (Figure 3(b)), and the time domain waveform is seriously distorted. When $\alpha = 10000$, the bandwidth of each mode becomes narrower compared to the original simulation components (Figure 3(b)), which results in the loss of some effective information contained in the original signal.

We compared the decomposition performance of EEMD and VMD. The noise amplitude and ensemble number of EEMD are 0.3 and 500 [30]. The decomposition results obtained by EEMD are shown in Figure 5, the IMF3, IMF4, and IMF5 are similar to the simulation components $x_1(t)$, $x_2(t)$, and $x_3(t)$, respectively. However, the mode mixing and waveform distortion occur in the time domain. In addition, the central frequencies of IMF3, IMF4, and IMF5 are not equal to that of the corresponding simulation components, and frequency mixing occurs in the frequency domain.

We optimize the VMD method by GOA and IGOA respectively, and the obtained fitness curve is shown in Figure 6. The best fitness obtained by IGOA is -1.231 when the number of iterations is 5, and the optimal parameters are $K = 3$, $\alpha = 5210$. The best fitness obtained by GOA is -1.212 when the number of iterations is 7, and the optimal parameters are $K = 3$, $\alpha = 9661$. Therefore, the convergence speed of IGOA is faster and the optimization accuracy is higher compared with the original GOA.

The simulation signal $x(t)$ is decomposed by IGOA-VMD and GOA-VMD respectively, and the results are shown in Figure 7. The results obtained by the two methods are

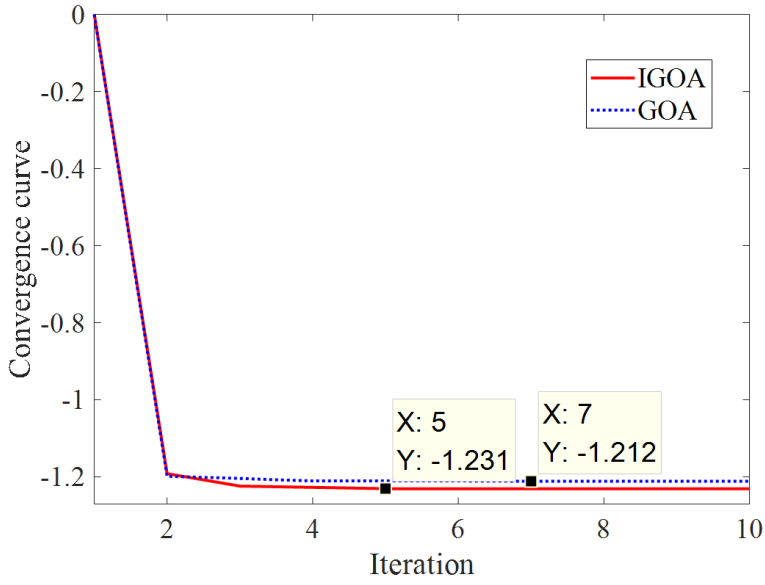


FIGURE 6. Convergence curve of IGOA and GOA for VMD parameter optimization

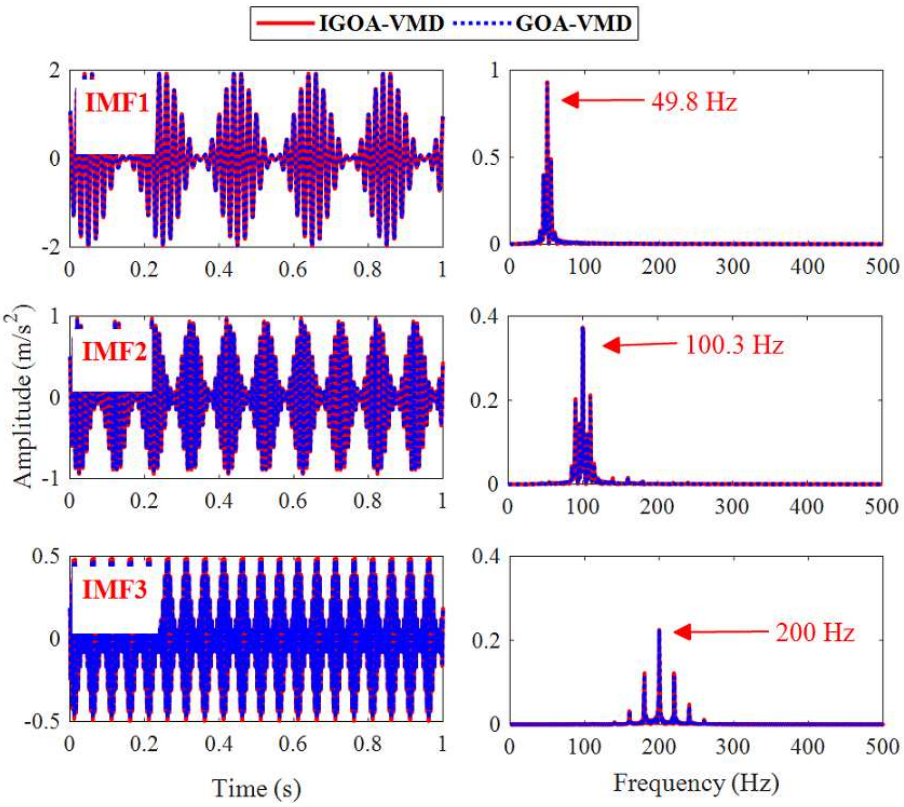


FIGURE 7. The decomposition results of simulated signal $x(t)$ obtained by IGOA-VMD and GOA-VMD

plotted on the same plot. Since the two sets of decomposition results are very similar, we evaluate the decomposition results by analyzing the multiple indexes of the modes, including Cor, MI, EC, EE, and EEMI. As shown in Figure 8, the difference between the indexes Cor, EE, EC obtained by IGOA and that obtained by GOA is very small, while the indexes MI and EEMI obtained by IGOA are larger than that obtained by GOA. It is indicated that MI and EEMI are sensitive to the two parameters of VMD, and the EEMI

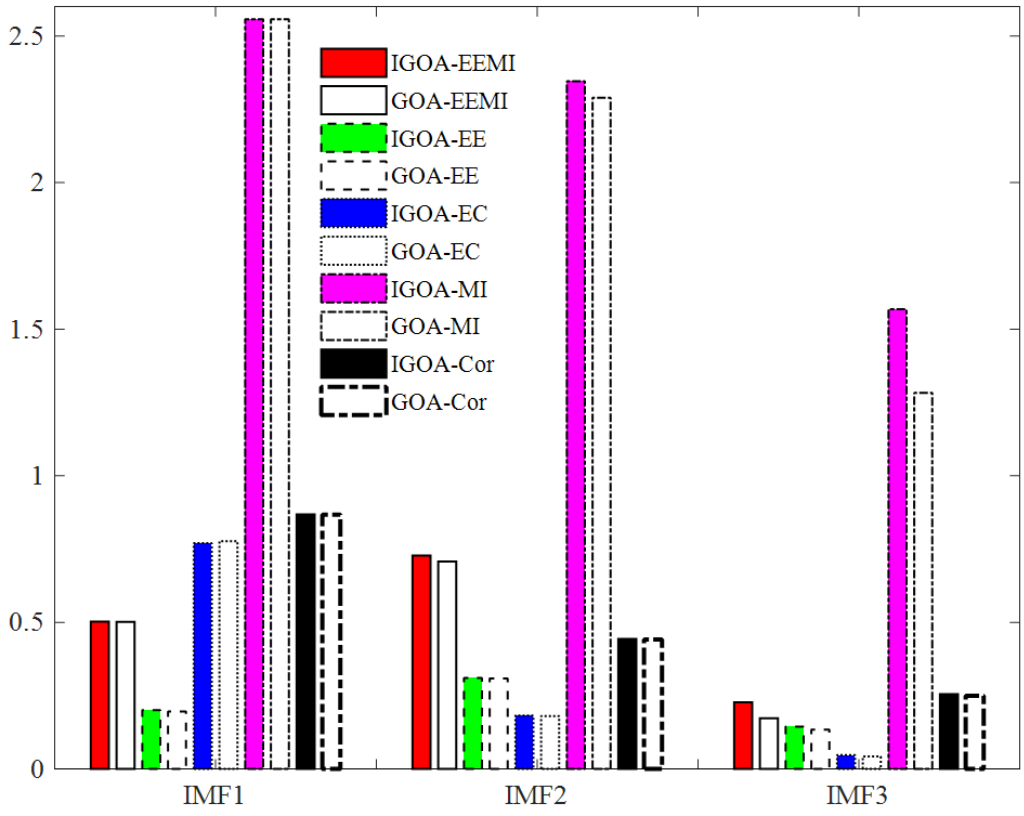


FIGURE 8. Decomposition results evaluation of simulated signal $x(t)$

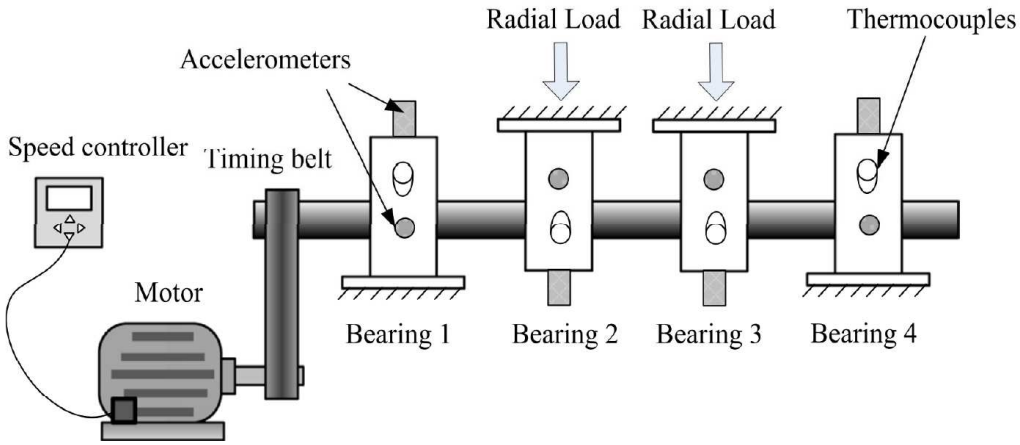


FIGURE 9. Illustration of the bearing experiment platform

index is effective. The results show that the GOA-VMD method is effective, and IGOA improves the decomposition performance of VMD compared with the original GOA.

4.3. **Early fault signals analysis.** The run-to-failure bearing data are derived from Intelligent Maintenance System (IMS) center [31]. The experimental platform is shown in Figure 9, and four Rexnord ZA-2115 bearings are mounted on the same spindle. The rotational speed of the spindle is 2000 rpm, and the rotating frequency of the bearing is 33.33Hz (2000/60 = 33.33Hz). The PCB353B33 sensors are installed on the bearing housing, and the vibration signal is collected once every 10 minutes via DAQ-6062E data acquisition card. The sampling frequency is 20KHz and the data length is 20480 points.

The parameters of the bearing are shown in Table 2. By using the following formula [15] and bearing parameters, we can get the theoretical value of outer race defect frequency 236.4Hz.

$$F_o = \frac{Z}{2}\left(1 - \frac{d}{D}\cos\theta\right) f_r \tag{25}$$

The experimental signal is from Bearing 1 of 2nd_test, and the outer race defect as shown in Figure 10. The number of data files is 984, #1∼#978 (where # means the number of files) can be used to characterize the whole life cycle of the outer ring, while #978-#984 is vague and distinctive. The root mean square (RMS) of the data is shown in Figure 11, and the performance degradation process of the bearing can be divided into four stages. It is rational for #1 to #520 for normal stage, #521 to #699 for slight fault stage, #700 to #850 for serious stage, and #851 to end for fault stage [15].

As shown in Figure 12, we cannot find the defect characteristic frequency of the #560 signal from the frequency domain waveform. Therefore, we analyze the #560 signal by

TABLE 2. The bearing structure factor of Rexnord ZA-2115

Ball diameter (d)	Number of balls (Z)	Pitch diameter (D)	Contact angle (θ)
8.407mm	16	71.501mm	15.17

FIGURE 10. Outer race defect of set number 2

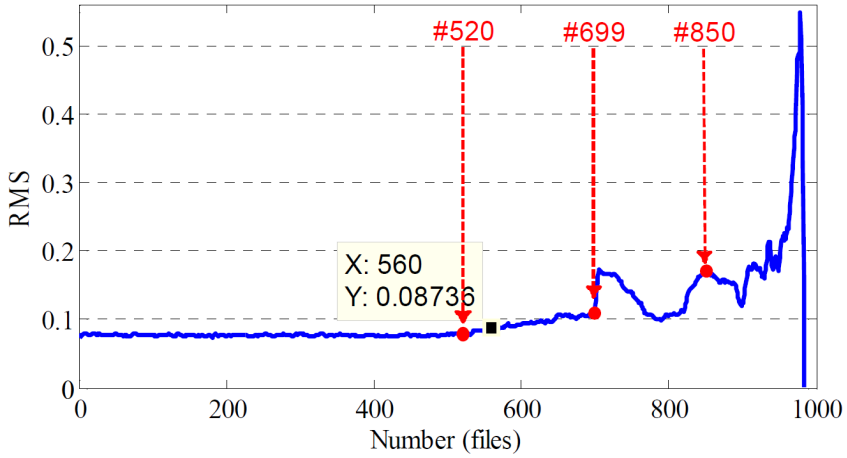


FIGURE 11. The RMS of the data

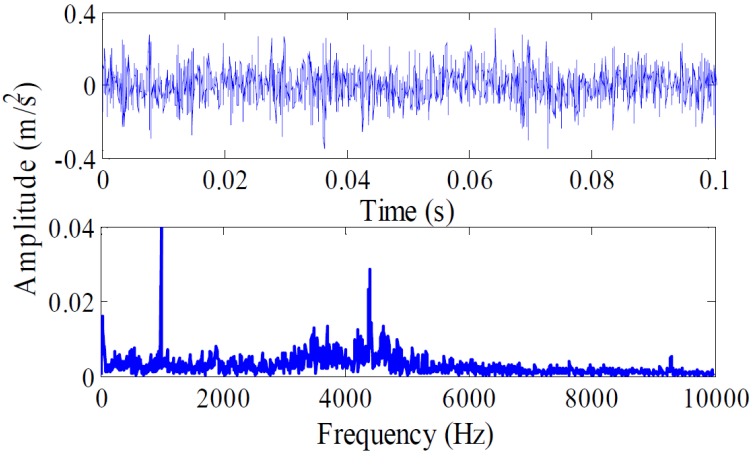


FIGURE 12. Time-domain and frequency-domain waveform of #560

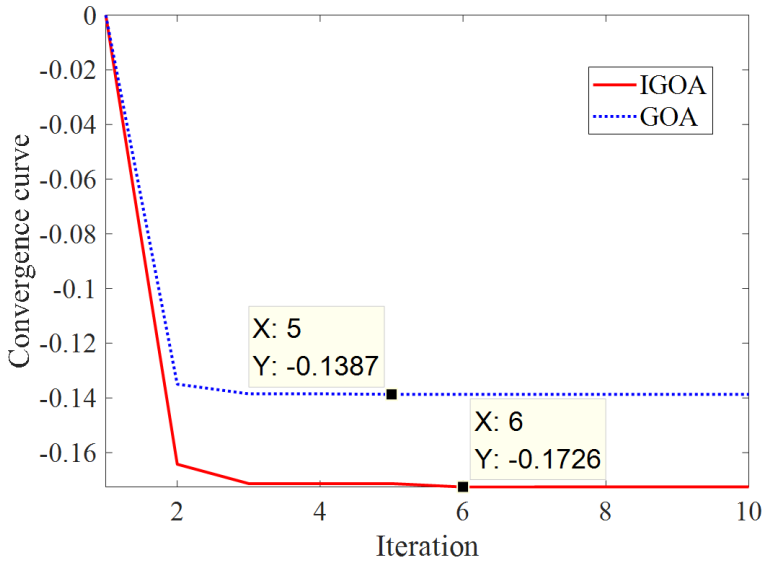


FIGURE 13. Convergence curve of IGOA and GOA for #560

the proposed method. The fitness curve is shown in Figure 13. The best fitness obtained by IGOA is -0.1726 when the number of iterations is 6, and the optimal parameters are $K = 2$, $\alpha = 1029$. The best fitness obtained by GOA is -0.1387 when the number of iterations is 5, and the optimal parameters are $K = 2$, $\alpha = 2287$. The optimal fitness of IGOA is smaller, so the optimized performance of IGOA is improved compared to the original GOA.

We decompose the #560 signal by the optimized VMD, and the results are shown in Figure 14. Since the decomposition results are very similar, the modes and their spectra are plotted on the same plot. The amplitude of the modes and their spectra obtained by IGOA is larger than that obtained by GOA, because the penalty parameter α obtained by IGOA is smaller than that obtained by GOA. The smaller the penalty parameter α is, the wider the frequency band of IMF is.

In order to compare the effect of IGOA and GOA on the decomposition performance of VMD, we evaluate the decomposition results through Cor, MI, EC, EE, and EEMI, and the results are shown in Figure 15. The indexes Cor, MI, and EEMI of IMF1 and IMF2 obtained by IGOA are increased compared with that obtained by original GOA. It shows

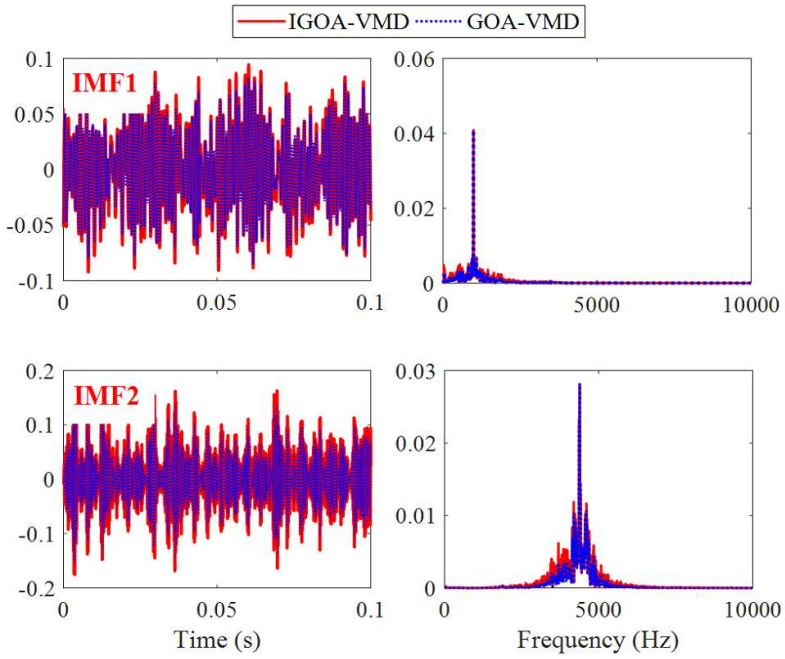


FIGURE 14. The decomposition results of signal #560 obtained by IGOA-VMD and GOA-VMD

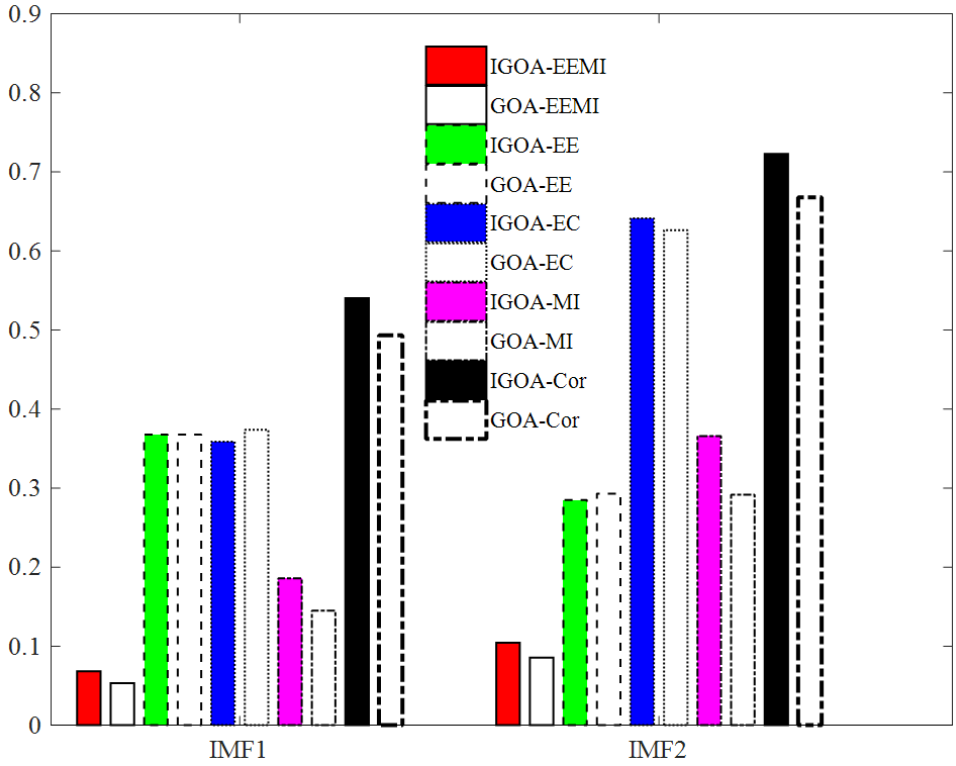


FIGURE 15. Decomposition results evaluation of signal #560

that the correlation between the obtained IMF and the original signal increases through the IGOA-VMD method, and the obtained IMF contains more information about the original signal. Meanwhile, the EC of IMF1 obtained by IGOA is relatively small, while the EC of IMF2 obtained by IGOA is larger than that obtained by GOA. The results

show that the results obtained by IGOA can better reflect the energy transfer of the defect signals.

For comparative analysis, we decompose the #560 signal by the VMD with fixed parameters ($K = 4$, $K = 2000$) [3,12]. 4 IMFs are obtained, and their time domain and frequency domain waveforms are shown in Figure 16. The mode mixing between the four IMFs is obvious in the time domain waveform, and the peaks of the central frequencies of IMF2 and IMF4 are not significant compared to the decomposition results of the optimized VMD (Figure 14). Moreover, the frequency bands obtained by the VMD with fixed parameters overlap with each other, and the information contained in the specific frequency bands cannot be separated effectively. The results show that the proposed optimization model is effective and the introduction of EEMI index is feasible.

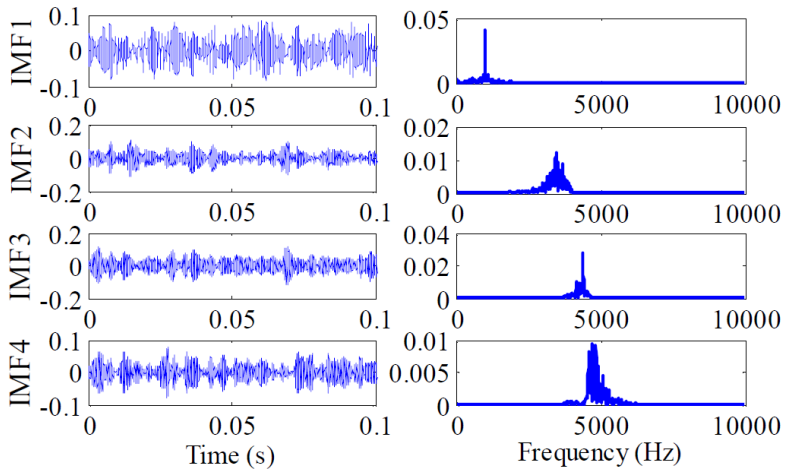


FIGURE 16. The decomposition results of signal #560 obtained by VMD ($K = 4$, $K = 2000$)

TABLE 3. Kurtosis of modes obtained by VMD

Method	IMF1	IMF2	IMF3	IMF4
VMD	2.2491	**3.4436**	2.6955	2.7607
GOA-VMD	2.2361	**3.1103**	–	–
IGOA-VMD	2.3886	**3.4206**	–	–

The kurtosis of the modes obtained by the above three methods is shown in Table 3. The kurtosis values of IMF1 and IMF2 obtained by IGOA-VMD are the largest compared with GOA-VMD, and the effectiveness of IGOA is proved again. For the three methods, the kurtosis of IMF2 is the largest. So IMF2 is the sensitive mode in the three methods.

The envelope demodulation of the sensitive modes obtained by the three methods is shown in Figure 17. In Figure 17(a), the maximum spectral peaks of the envelope spectrum appear at 58.59Hz (2Fr), 29.3Hz (Fr), and 136.7Hz. The first two frequencies are close to the rotating frequency (Fr) and its second harmonic (2Fr) of the bearing, and the 136.7Hz is close to the rotating frequency fourth harmonic (4Fr) or rolling element fault frequency (139.9Hz). Therefore, the recognition of early faults is extremely difficult, and the recognition results may be erroneous. The other three frequencies with large envelope amplitude are 195.3Hz, 244.1Hz and 302.7Hz. The frequency 195.3Hz is close to the sixth harmonic of rotating frequency (6Fr). The frequencies 244.1Hz and 302.7Hz are close to the outer ring fault frequency (Fo) and the modulation frequency (Fo+2Fr) between Fo

and rotating frequent, respectively. However, these effective frequencies are not obvious due to the influence of many interference frequencies. Therefore, it is difficult to identify early faults through the VMD ($K = 4$, $K = 2000$) with fixed parameters.

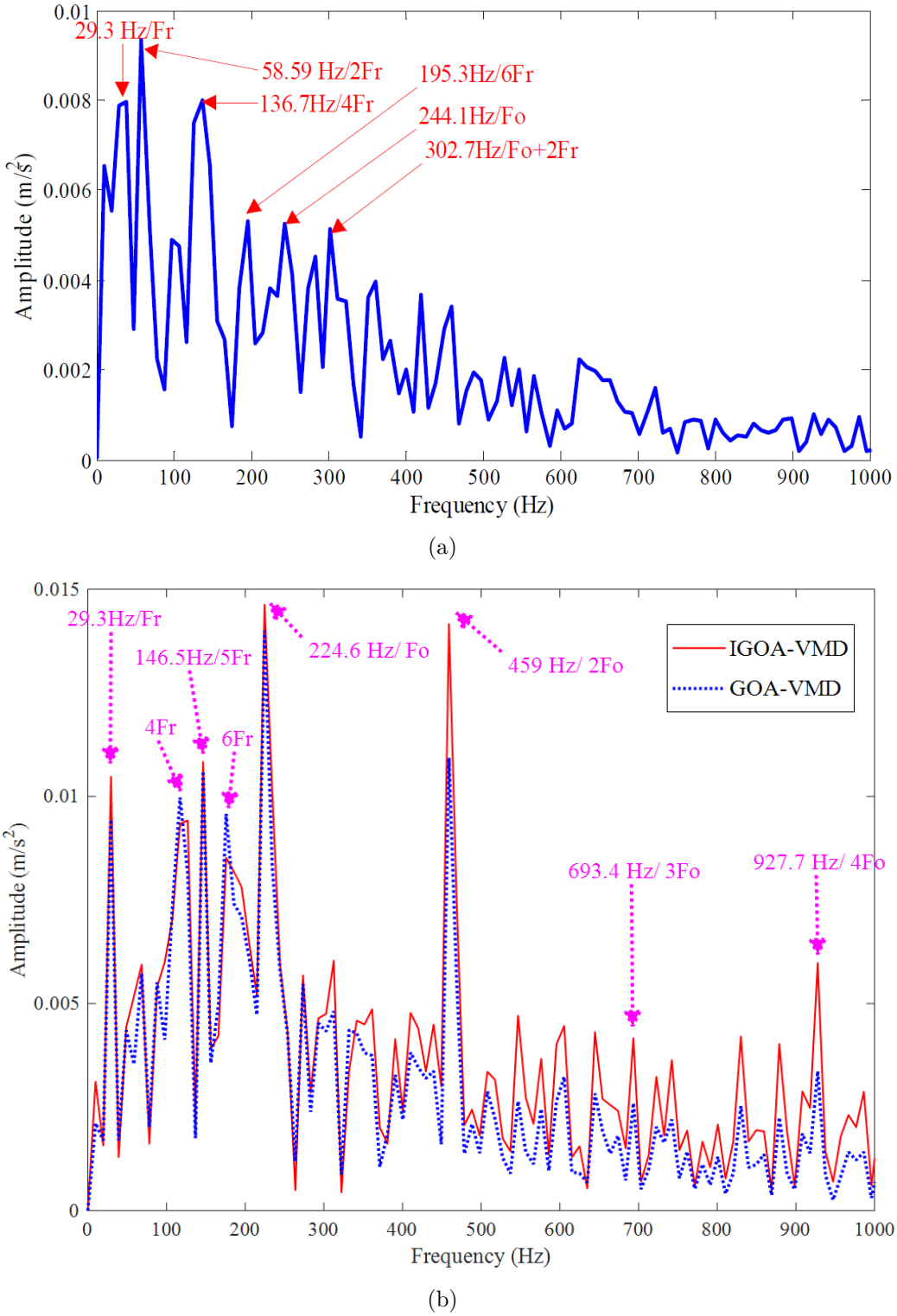


FIGURE 17. The envelope spectrum of the sensitive IMF obtained by (a) VMD with assigned parameters, and (b) VMD based on IGOA and GOA

In Figure 17(b), the envelope spectra of the sensitive modes obtained by IGOA-VMD and GOA-VMD are plotted on the same plot. The maximum spectral peaks of the envelope spectrum appear at frequencies 224.6Hz and 459Hz, which are close to the outer ring fault frequency (Fo) and its second harmonic (2Fo) respectively. Therefore, the early defect of the outer ring can be identified accurately by two methods. The other four frequencies with large envelope amplitude are 29.3Hz, 117.2Hz, 146.5Hz and 175.8Hz. These frequencies are approximately equal to the rotating frequency and its high frequency harmonic. It can be seen from Figure 17(b) that the amplitude of the envelope spectrum obtained by the IGOA-VMD is larger than that obtained by the GOA-VMD. It is worth noting that two large spectral peaks appear at 693.4Hz and 927.7Hz in the envelope spectrum obtained by the IGOA-VMD. These two frequencies are close to the third harmonic (693.4Hz/3Fo) and fourth harmonic frequencies (927.7Hz/4Fo) of the outer ring frequency. However, these two effective frequencies are almost submerged by the interference frequencies in the envelope spectrum obtained by the GOA-VMD. The results show that the fault information contained in the sensitive mode obtained by IGOA-VMD is more complete than that obtained by GOA-VMD and VMD. Therefore, the proposed method is very effective.

5. **Conclusions.** An early fault diagnosis method based on adaptive VMD is proposed in this paper. Firstly, the linear decreasing coefficient of GOA is replaced by the convex-concave decreasing strategy, and the effectiveness of the improved strategy (IGOA) is verified by 23 sets of benchmark functions. Secondly, the energy-entropy mutual information (EEMI) index is taken as the objective function, and the optimal parameters of VMD matching with the input signal can be obtained. The results show that EEMI is effective. Finally, the sensitive modes obtained by VMD, GOA-VMD and IGOA-VMD are analyzed by Hilbert envelope demodulation. It shows that the decomposition performance of IGOA-VMD is the best. The main contributions of this approach are summarized below. (1) The local optimum problem and the balance problem between exploration and exploitation in GOA have been improved by introducing the convex-concave decreasing strategy and the EEMI index. (2) The mode number and penalty parameter of VMD are determined adaptively by IGOA according to the signal characteristics, and the randomness of parameter selection is avoided. Next, the proposed method will be applied to early fault diagnosis and composite fault diagnosis of bearings in the practical industrial environment. In addition, the proposed IGOA is expected to be applied to the optimization of other intelligent diagnosis methods.

Acknowledgement. This work is supported by the National Natural Science Foundation of China (nos. 61663017 and 51765022). The authors sincerely thank the team for their guidance, and thank the University of Cincinnati for their bearing datasets. The authors sincerely express thanks to the reviewers for taking the time to review the paper in a busy schedule.

REFERENCES

[1] D. Abboud, M. Elbadaoui, W. A. Smith and R. B. Randall, Advanced bearing diagnostics: A comparative study of two powerful approaches, *Mechanical Systems and Signal Processing*, vol.114, pp.604-627, 2019.

[2] J. Li, J. Zhang, M. Li and Y. Zhang, A novel adaptive stochastic resonance method based on coupled bistable systems and its application in rolling bearing fault diagnosis, *Mechanical Systems and Signal Processing*, vol.114, pp.128-145, 2019.

[3] J. Ma, J. Wu and X. Wang, Incipient fault feature extraction of rolling bearings based on the MVMD and Teager energy operator, *ISA Transactions*, vol.80, pp.297-311, 2018.

[4] Z. Feng, H. Ma and M. J. Zuo, Amplitude and frequency demodulation analysis for fault diagnosis of planet bearings, *Journal of Sound and Vibration*, vol.382, pp.395-412, 2016.
[5] A. Moshrefzadeh and A. Fasana, The Autogram: An effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis, *Mechanical Systems and Signal Processing*, vol.105, pp.294-318, 2018.
[6] Y. Miao, M. Zhao, J. Lin and Y. Lei, Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings, *Mechanical Systems and Signal Processing*, vol.92, pp.173-195, 2017.
[7] E. G. Plaza and P. J. N. López, Application of the wavelet packet transform to vibration signals for surface roughness monitoring in CNC turning operations, *Mechanical Systems and Signal Processing*, vol.98, pp.902-919, 2018.
[8] L. Wang, Z. Liu, Q. Miao and X. Zhang, Complete ensemble local mean decomposition with adaptive noise and its application to fault diagnosis for rolling bearings, *Mechanical Systems and Signal Processing*, vol.106, pp.24-39, 2018.
[9] Y. Hu, X. Tu, F. Li and G. Meng, Joint high-order synchrosqueezing transform and multi-taper empirical wavelet transform for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions, *Sensors*, vol.18, no.1, 2018.
[10] K. Dragomiretskiy and D. Zosso, Variational mode decomposition, *IEEE Trans. Signal Processing*, vol.62, no.3, pp.531-544, 2014.
[11] J. Lian, Z. Liu, H. Wang and X. Dong, Adaptive variational mode decomposition method for signal processing based on mode characteristic, *Mechanical Systems and Signal Processing*, vol.107, pp.53-77, 2018.
[12] Z. Li, J. Chen, Y. Zi and J. Pan, Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive, *Mechanical Systems and Signal Processing*, vol.85, pp.512-529, 2017.
[13] X. Yan, M. Jia and Z. Zhao, A novel intelligent detection method for rolling bearing based on IVMD and instantaneous energy distribution-permutation entropy, *Measurement*, vol.130, pp.435-447, 2018.
[14] C. Liu, L. Zhu and C. Ni, Chatter detection in milling process based on VMD and energy entropy, *Mechanical Systems and Signal Processing*, vol.105, pp.169-182, 2018.
[15] Y. Li, H. Li, B. Wang and H. Gu, Rolling element bearing performance degradation assessment using variational mode decomposition and Gath-Geva clustering time series segmentation, *International Journal of Rotating Machinery*, vol.2017, pp.1-12, 2017.
[16] T. Guo and Z. Deng, An improved EMD method based on the multi-objective optimization and its application to fault feature extraction of rolling bearing, *Applied Acoustics*, vol.127, pp.46-62, 2017.
[17] J. Li, J. Jiang, X. Fan et al., A new method for weak fault feature extraction based on improved MED, *Shock and Vibration*, vol.2018, no.2, pp.1-11, 2018.
[18] H. Shao, H. Jiang, H. Zhao and F. Wang, A novel deep autoencoder feature learning method for rotating machinery fault diagnosis, *Mechanical Systems and Signal Processing*, vol.95, pp.187-204, 2017.
[19] S. Saremi, S. Mirjalili and A. Lewis, Grasshopper optimisation algorithm: Theory and application, *Advances in Engineering Software*, vol.105, pp.30-47, 2017.
[20] A. A. Ewees, M. A. Elaziz and E. H. Houssein, Improved grasshopper optimization algorithm using opposition-based learning, *Expert Systems with Applications*, vol.112, pp.156-172, 2018.
[21] J. Luo, H. Chen, Q. Zhang, Y. Xu, H. Huang and X. Zhao, An improved grasshopper optimization algorithm with application to financial stress prediction, *Applied Mathematical Modelling*, vol.64, pp.654-668, 2018.
[22] J. Chen, D. Zhou, C. Lyu and C. Lu, An integrated method based on CEEMD-SampEn and the correlation analysis algorithm for the fault diagnosis of a gearbox under different working conditions, *Mechanical Systems and Signal Processing*, vol.113, pp.102-111, 2018.
[23] W. Gao, L. Hu and P. Zhang, Class-specific mutual information variation for feature selection, *Pattern Recognition*, vol.79, pp.328-339, 2018.
[24] T. Lin, G. Chen, W. Ouyang, Q. Zhang, H. Wang and L. Chen, Hyper-spherical distance discrimination: A novel data description method for aero-engine rolling bearing fault detection, *Mechanical Systems and Signal Processing*, vol.109, pp.330-351, 2018.
[25] R. Akhand and S. H. Upadhyay, An integrated approach to bearing prognostics based on EEMD-multi feature extraction, Gaussian mixture models and Jensen-Rényi divergence, *Applied Soft Computing*, vol.71, pp.36-50, 2018.

[26] Z. Zhang, H. Li, G. Meng, X. Tu and C. Cheng, Chatter detection in milling process based on the energy entropy of VMD and WPD, *International Journal of Machine Tools and Manufacture*, vol.108, pp.106-112, 2016.

[27] M. Zhang, Z. Jiang and K. Feng, Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump, *Mechanical Systems and Signal Processing*, vol.93, pp.460-493, 2017.

[28] X. Zhang, Q. Miao, H. Zhang and L. Wang, A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery, *Mechanical Systems and Signal Processing*, vol.108, pp.58-72, 2018.

[29] R. T. Rockafellar, A dual approach to solving nonlinear programming problems by unconstrained optimization, *Mathematical Programming*, vol.5, no.1, pp.354-373, 1973.

[30] X. Zhang, Z. Liu, Q. Miao and L. Wang, An optimized time varying filtering based empirical mode decomposition method with grey wolf optimizer for machinery fault diagnosis, *Journal of Sound Vibration*, vol.418, pp.55-78, 2018.

[31] Y. Li, X. Liang, M. Xu and W. Huang, Early fault feature extraction of rolling bearing based on ICD and tunable Q-factor wavelet transform, *Mechanical Systems and Signal Processing*, vol.86, pp.204-223, 2017.