

Correlation Graph Fourier Spectrum Model and Its Applications in Rolling Bearing Fault Diagnosis

Yiding Liu¹ Huaming Zhang¹ Peng Xu¹ Lei Feng¹ Xueying Jin¹ Yanlei Liu^{1,*}
Yonggang Xu¹ Yousheng Yang¹

1. CGN (Beijing) New Energy Technology Co., Ltd., Beijing 100070, China

*Corresponding author: liuyl_0505@163.com

Received July 22, 2026 | Revised September 08, 2026 | Accepted September 22, 2026 | Posted Online September 24, 2026

Abstract: The transmission system of wind turbines contains complex nonlinear coupled noise and interference, which poses significant challenges to traditional signal decomposition and deconvolution algorithms. This paper proposes a method based on feature dimensionality enhancement and denoising in the graph spectral domain to extract fault features in wind turbine bearings, referred to as the Correlation Graph Fourier Spectrum Model (CGFS Model). First, a spectrum-based Laplacian matrix is constructed to transform the signal from a one-dimensional representation to the graph spectral domain. A transfer frequency-domain feature analysis method is designed, and a linear graph-spectral order separation strategy is developed to overcome the limitations of traditional methods in connecting the time domain and the graph spectral domain. A hybrid correlation entropy is designed to quantify the characteristics present in different graph spectral orders, enabling precise localization of wind turbine bearing faults. Simulation signals and bearing inner/outer-race fault signals validate the effectiveness and practicality of the proposed model.

Keywords: Graph Fourier spectrum; Hybrid correlation entropy; Wind turbine; Rolling bearing; Fault diagnosis.

1 Introduction

As essential and indispensable components in wind turbine systems, rolling bearings typically operate under harsh service conditions characterized by high pressure and increased temperatures, which renders them highly susceptible to various forms of damage. Failure to timely detect incipient faults or potential failures of these bearings

could trigger severe safety accidents and result in substantial property losses^[1]. Local defects occurring on the outer ring, inner ring or rolling elements of rolling bearings can excite regular periodic pulse signals in the operating process. Vibration responses collected from bearing structures generally contain such fault-related periodic pulse components, hence the accurate extraction and recognition of these fault characteristic pulses from collected vibration signals constitutes the key research focus of rolling bearing fault diagnosis^[2-3].

To extract these weak fault features, various diagnostic methods have been developed from different perspectives. In time-domain analysis, Zhang et al. [4] proposed the pulse signal-to-noise ratio (PSNR) test method by utilizing the periodic pulse characteristics of fault signals. This method models the bearing signal as a periodic pulse with a constant duty cycle. By constructing a generalized likelihood ratio test statistic, fault detection and localization can be achieved without spectrum analysis. Its simplified version, PSNAR, further reduces computational complexity. In frequency-domain analysis, Yu et al. [5] proposed the wavelet-enhanced transient extraction transform (WTET) to address the difficulty that traditional time-frequency analysis methods have in accurately characterizing transient events. This method introduces a group delay estimator into the frequency-domain model. By retaining the principal energy components in the time-frequency plane, it effectively suppresses noise interference and achieves high-resolution characterization of bearing fault impact features. Within the framework of time-frequency domain analysis, Zou et al. [6] developed the time-frequency modulated bispectrum (TFMB) approach to solve the inherent limitation that the bispectrum analysis of modulated signals is highly susceptible to interference under strong noise environments. The short-time Fourier transform is adopted in this approach to maintain complete time-frequency information of vibration signals. Meanwhile, two-dimensional heterogeneous decomposition is integrated to achieve noise suppression, and the modulation Gini index is employed to screen out the optimal frequency slice. Such processing procedure further improves the extraction performance of modulation characteristics. Bao et al. [7] proposed the generalized synchroextracting transform (GSET), which establishes a general high-order instantaneous frequency

solution model and a time-frequency feature extraction operator. This method enables high-precision time-frequency representation of strongly time-varying non-stationary vibration signals and is applicable to the extraction of weak fault impact features in various types of rotating machinery. In convolutional neural network applications, Jin et al. [8] proposed the Kolmogorov–Arnold Convolutional Squeeze-Excitation Network (KACSEN) to address the limited feature extraction capability and high model complexity of traditional convolution kernels. Derived from the Kolmogorov–Arnold representation theorem, the constructed approach configures the convolution kernel in the form of a learnable nonlinear function ensemble. A squeeze-excitation attention mechanism is further embedded to assign adaptive weights to feature channels, which effectively promotes the overall diagnostic precision and model interpretability. For the diagnosis of wind turbine pitch bearing systems, Meng et al. [9] developed a DRN-PISSM network that integrates dilated residual convolution with a physics-informed state-space model. The network takes raw one-dimensional time-domain vibration signals as its direct input and leverages dynamic prior constraints to enhance the robustness of feature extraction under complex operating conditions. With the aim of enhancing diagnostic accuracy and interpretability, Zhang et al. [10] proposed a frequency-slice-based graph framework within the time-frequency domain, where the graph Fourier transform is organically integrated into the framework design. In this specific approach, frequency slices are treated as nodes to construct a Laplacian matrix, and the time-frequency graph Fourier clustering spectrum is further derived from this constructed Laplacian matrix. Subsequently, high-precision spectral kurtosis is utilized to automatically identify and select the optimal reconstruction

frequency band, which remarkably improves the performance of fault diagnosis under strong noise conditions. In the realm of other novel diagnostic frameworks, Kang et al.^[11] developed a high-order Vese–Osher total variation model based on partial differential equations, incorporating a second-order differential term to mitigate the over-smoothing issue inherent in the first-order total variation model. Combined with the Split Bregman iterative algorithm and fast Fourier transform, an efficient solution is achieved. In their approach, the preprocessed angular-domain signal of the sun gear is first denoised, followed by the extraction of third-order cumulant slices to construct a 1.5-dimensional demodulation spectrum, which effectively highlights gear fault features under strong noise interference. For acoustic emission monitoring scenarios where in-band strong noise masks the cyclostationary characteristics of faults, Li et al.^[12] proposed a bidirectional weighted cyclostationary analysis method with in-band noise suppression. By employing a bidirectional weighted cyclic spectral correlation function, the non-fault periodic interference terms are filtered out, while the fault-related cyclic frequency components in the acoustic emission signal are amplified, thereby enabling accurate identification of early-stage weak damage in equipment.

The rapid development and widespread application of graph signal processing have brought forward a novel and effective perspective for the analysis and processing of various signals in engineering fields. Specifically, Hammond et al. developed a graph wavelet transform method, which is constructed on the theoretical foundation of spectral graph theory and fully leverages the characteristics of this theory to realize signal analysis. Through the precise definition of scaling operators within the graph spectral domain, the research team constructed a spectral graph wavelet that possesses

excellent localization characteristics in both the graph domain and spectral domain. Meanwhile, they developed a high-efficiency fast algorithm, which is established on the basis of Chebyshev polynomial approximation to improve computational efficiency^[13]. Spielman delivered a systematic and comprehensive review regarding graph spectral analysis and its extensive practical applications in various engineering and scientific fields. In this review, he pointed out that the eigenvalues and eigenvectors obtained from the graph Laplacian matrix are capable of effectively characterizing and reflecting the inherent combinatorial structure of a graph. This critical finding further lays a solid theoretical foundation for the development and practical implementation of graph signal processing^[14]. Shuman et al. studied key definitions of graph empirical mode decomposition, providing a new tool for time-frequency analysis of graph signals^[15]. To achieve effective vertex-frequency analysis on graph structures, Shuman et al.^[16] developed a windowed graph Fourier transform tailored for graph signal processing scenarios. The realization of this transform was accomplished by introducing generalized translation and modulation operators over graph structures; these operators enable the synchronous analysis of graph signals in both the vertex domain and the graph spectral domain, thereby overcoming the limitation of single-domain analysis. From the perspective of algebraic signal processing, Sandryhaila and Moura^[17] developed a comprehensive theoretical framework for discrete signal processing on graphs. In this framework, the graph shift operator serves as a fundamental building block, enabling the definition of key concepts including graph filters, graph convolution, the graph Fourier transform, and frequency response. The effectiveness of graph signal processing methods has been verified in multiple application fields. Ou et

al. first introduced the graph Fourier transform into rolling bearing fault diagnosis. Through the definition of a novel adjacency weight matrix with optimized structural characteristics, the research team discovered that the graph Fourier transform possesses the capability to aggregate impact components and harmonic components into distinct order regions within the graph spectral domain. Building on this important finding, they further proposed a fault diagnosis approach that relies on the extraction of impact components via the graph Fourier transform^[18]. Gao et al. further introduced a weighted horizontal visibility graph to replace the simple path graph. The horizontal visibility algorithm was employed to convert vibration signals into graph structures with clear topological characteristics. This conversion process effectively enhances the clustering performance of the graph Fourier transform, while simultaneously strengthening its anti-interference capability in complex noise environments^[19]. Ou and Yu developed a fault diagnosis approach for rolling bearings by extracting Laplacian energy features from a path graph. They used Laplacian energy as a graph complexity feature and combined it with a Mahalanobis distance classifier to achieve accurate fault identification with only a small number of sampling points and training samples^[20]. In mechanical system reliability analysis, Tang proposed a method that combines graph theory with Boolean functions. By constructing a system reliability graph matrix and using Boolean functions to quantify the failure correlation between components, an effective assessment of the reliability of complex mechanical systems was achieved^[21]. Ding et al. proposed a method based on Laplacian sparsification to address the scalability problem of large-scale spectral graph neural networks. By constructing an ε -sparse approximation of the propagation matrix, they significantly reduced computational

complexity^[22]. Ma et al. combined spectral graph neural networks with large-scale language models and proposed the AutoSGNN framework, enabling automated discovery of the propagation mechanism of spectral graph neural networks^[23].

While graph theory-based methods provide deeper insights into fault diagnosis, they have not yet addressed the constraints imposed by graph construction methods and graph order on feature decoupling. To solve these issues, this paper constructs the Correlation Graph Fourier Spectrum Model (CGFS Model). The proposed spectrum-based Laplacian matrix enables the conversion of signals from a one-dimensional representation to the graph domain, as well as feature dimensionality enhancement and noise reduction based on the graph domain. The designed frequency feature analysis method provides a linear graph order separation scheme, overcoming the limitations of traditional methods in connecting the time and graph domains. To quantify fault features at different graph orders, a hybrid correlation entropy is designed, enabling precise fault localization in wind turbine bearings. Under strong noise, the CGFS method extracts fault features from rolling bearings more effectively than alternatives, achieving improved robustness and diagnostic accuracy.

Different from existing rolling-bearing fault diagnosis methods based on graph Fourier transform (GFT) which directly map raw time-domain vibration signals into graph domain, the proposed CGFS model constructs Laplacian matrix from the frequency-domain amplitude spectrum after FFT transformation, rather than original time series. This frequency-domain graph construction avoids the adverse influence of time-domain large-amplitude random noise outliers on graph topology establishment. Furthermore, previous GFT-based bearing diagnosis works mostly use fixed

graph-order truncation or kurtosis index to select effective graph spectrum components. In contrast, this work designs a hybrid correlation entropy (HCE) indicator combining autocorrelation kurtosis and spectrum energy entropy. It adaptively searches the optimal graph-order segment through sliding window in high-order graph spectral region, realizing the joint measurement of time-domain periodic impulse property and frequency-domain energy distribution characteristics. It overcomes the deficiency that traditional GFT-based methods lack effective quantitative criteria for screening optimal graph-order components, and realizes accurate decoupling and localization of bearing fault features under heavy noise.

The structure and arrangement of the remaining sections in this paper are presented as follows: Section 2 provides a systematic and concise introduction to the fundamental concepts, core principles, and key theoretical frameworks of graph theory. Section 3 details the CGFS method and the calculation of related indicators. Section 4 presents simulation experiments. Section 5 applies the CGFS method to experimental signals from the inner and outer races of bearings. Section 6 concludes the paper.

2. Theoretical Background

Traditional signal processing struggles to extract weak features under complex noise conditions. Graph signal processing, however, can transform the signal into a graph structure. A graph can be denoted as $P = \{D, B\}$, where D represents signal sampling points or data (vertices), and B represents the relationships (edges) between data points. Graph structures are categorized into undirected and directed graphs based on whether the edges have directions, as shown in Fig.1(a) and 1(b), respectively. Based on the principle of connecting each vertex to the

remaining vertices, graphs are classified into nearest-neighbor graphs and complete graphs. In a nearest-neighbor graph, if two vertices d_i and d_j are sufficiently close, they are connected by an edge; otherwise, they are not connected, as shown in Fig.1(c). In a complete graph, which is undirected, each distinct pair of vertices is linked by a single edge, as illustrated in Fig.1(d). In this work, such a graph topology is adopted to achieve a trade-off between computational cost and the preservation of complete fault information.

The fully-connected weighted complete graph is selected in this paper for three main reasons. First, the bearing vibration signal under strong noise contains weak fault impulse components mixed with a large amount of random noise. The nearest-neighbor graph only establishes connections between partial vertices, which easily truncates the correlation between fault-related sampling points and loses weak correlation information of transient fault pulses. Second, for the complete graph, every frequency-domain amplitude vertex establishes weighted edge connections with all other vertices. Using Euclidean distance as edge weight can fully characterize the amplitude difference relationship between different frequency-domain sampling points, which helps the Laplacian matrix to capture subtle amplitude mutations induced by bearing fault impulses in the frequency spectrum. Third, compared with k-nearest neighbor graphs whose graph construction quality is sensitive to neighbor-number hyperparameters, the complete graph avoids the performance fluctuation caused by improper neighbor-number selection. Although the complete graph brings higher computational complexity, we only focus on the high-order part of graph spectrum in subsequent processing, which greatly reduces the actual computation burden. Balancing the capability of capturing weak fault correlation

and parameter robustness, the weighted complete graph is finally adopted in this work.

To study the information embedded in graph structures, algebraic and analytical methods are combined, forming graph theory. Graph storage primarily involves the adjacency matrix M and the Laplacian matrix G . The adjacency matrix is a core matrix representation of a graph, directly describing the adjacency relationships between vertices. Different graph structures have different rules for assigning matrix elements. For an undirected and unweighted graph, if an edge connects vertices v_i and v_j , then $m_{ij} = m_{ji} = 1$; otherwise, $m_{ij} = m_{ji} = 0$. Diagonal

elements are $m_{ii} = 0$. In practical applications, edges are usually weighted, and appropriate weights are selected according to signal characteristics. To capture edge-wise interactions while respecting vertex-wise differences, the Euclidean distance is adopted as the weighting scheme for wind turbine bearing fault diagnosis in this work. The relevant computational formulation is illustrated in Equation (1), in which $x(i)$ and $x(j)$ represent the signal amplitudes corresponding to vertices v_i and v_j respectively.

$$m_{ij} = \|x(i) - x(j)\|_2 \quad (1)$$

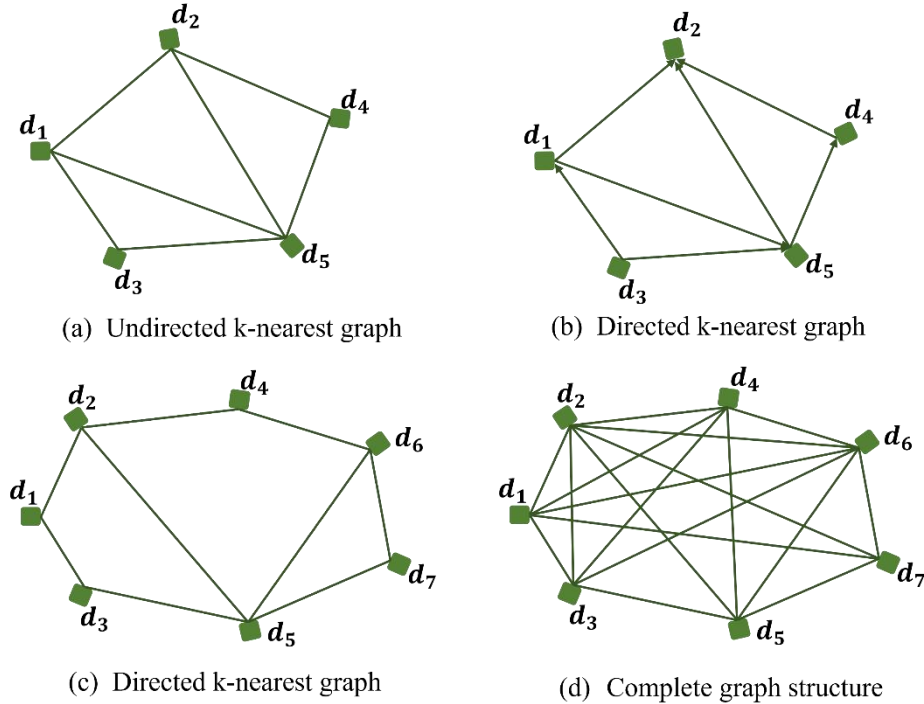


Fig. 1. Common graph structures

In graph theory, the degree matrix Z is a diagonal matrix that encodes the degree distribution of vertices. This matrix acts as a fundamental constituent for the construction of the Laplacian matrix, and its dimensional scale is consistent with the total number of graph vertices. Regarding the weighted complete graph adopted in this work, the

diagonal entry Z_{ii} corresponds to the weighted degree of vertex v_i , which is defined as the summation of weights for all edges connected to vertex v_i :

$$z_{ii} = \sum_{j=1}^n m_{ij} \quad (2)$$

$$Z = \begin{bmatrix} z_{11} & 0 & 0 & 0 \\ 0 & z_{22} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & z_{xx} \end{bmatrix} \quad (3)$$

The Laplacian matrix G can be computed by combining the adjacency matrix and the degree matrix, as illustrated in Equation (4). Both the adjacency matrix and the degree matrix belong to real symmetric matrices, so the Laplacian matrix naturally inherits the inherent mathematical properties of the above two matrices. The Laplacian matrix constructed from an undirected graph satisfies the property of symmetric positive semi-definiteness. All eigenvalues solved from this matrix are non-negative real values, and the corresponding eigenvectors maintain mutual orthogonality. Such mathematical characteristics enable the Laplacian matrix to be directly applicable to classical linear algebra approaches, including spectral decomposition and quadratic form analysis.

$$G = Z - M \quad (4)$$

$$Gl_s = \lambda_s l_s \quad (5)$$

In Equation (5), the variable s runs from 1 to n , indicating both the dimension and the order, where n denotes the total vertex count of the graph. The column vectors $l_1, l_2, \dots, l_n$ form the eigenvectors of the Laplacian matrix, collectively denoted as $L = [l_1, l_2, \dots, l_n]$.

3. Correlation Graph Fourier Spectrum Model

Traditional time-domain decomposition, spectral filtering, and time-frequency analysis methods are easily interfered with by complex signal components, resulting in irrelevant components that hinder fault feature extraction. The graph spectral domain uses the Laplacian matrix to perform high-dimensional analysis of time-domain signals, which helps address the challenges faced by

traditional analysis methods. To address the limitations of conventional analysis methods, the Correlation Graph Fourier Spectrum Model (CGFS Model) is introduced in this work. The main procedure is outlined below:

- 1) Collect vibration acceleration signals from the wind turbine bearing system and store the data.
- 2) Calculate the Fourier spectrum of the input signals using FFT and retain the amplitude and phase information in the frequency domain.
- 3) Construct a Laplacian matrix from the amplitude information and perform eigen decomposition to obtain a sequence of eigenvalues.
- 4) Construct a graph signal structure using the eigenvalue sequence, and use the graph Fourier transform to convert the graph signal into the spectral domain.
- 5) Extract the highest-order segment and create key feature spectral bands
- 6) Configure a sliding-window-based spectral filtering process, and use mixed correlation entropy to identify fault information within the window.
- 7) Extract the spectral order segment corresponding to the highest negative entropy, calculate the envelope spectrum, and diagnose the fault.

The flowchart is shown in Fig. 2.

3.1 Frequency Domain Spectral Model Theory Based on Graphical Fourier Transform

The emergence of graph theory has introduced a notion analogous to "frequency" into data analysis, thereby facilitating the ongoing advancement and refinement of

graph-based methodologies and giving rise to the emerging discipline of graph signal processing (GSP). Serving as the fundamental theoretical framework of GSP, the graph Fourier transform (GFT) implements orthogonal projection of graph signals onto the eigenbasis of the Laplacian matrix. This operation effectively converts the expression form of graph signals from the vertex domain into the graph spectral domain. The internal working principle of GFT maintains a high degree of consistency with the spectral decomposition technique adopted in traditional Fourier transform. Since the original time-domain signal $x(n)$ is inevitably contaminated by diverse noise

interference in practical acquisition scenarios, the fast Fourier transform (FFT) is firstly implemented on $x(n)$. Such processing converts the discrete time-domain signal into the frequency domain, and the corresponding mathematical definition is presented as follows:

$$X(f) = \sum_{n=0}^{N-1} x(n) \cdot e^{-\frac{2\pi}{N}kn}, k = 0, 1, 2, \dots, N-1 \quad (6)$$

where N denotes the total length of the discrete sequence. As a result, a frequency-domain signal $X(f)$ is obtained, which maintains complete amplitude and phase information simultaneously.

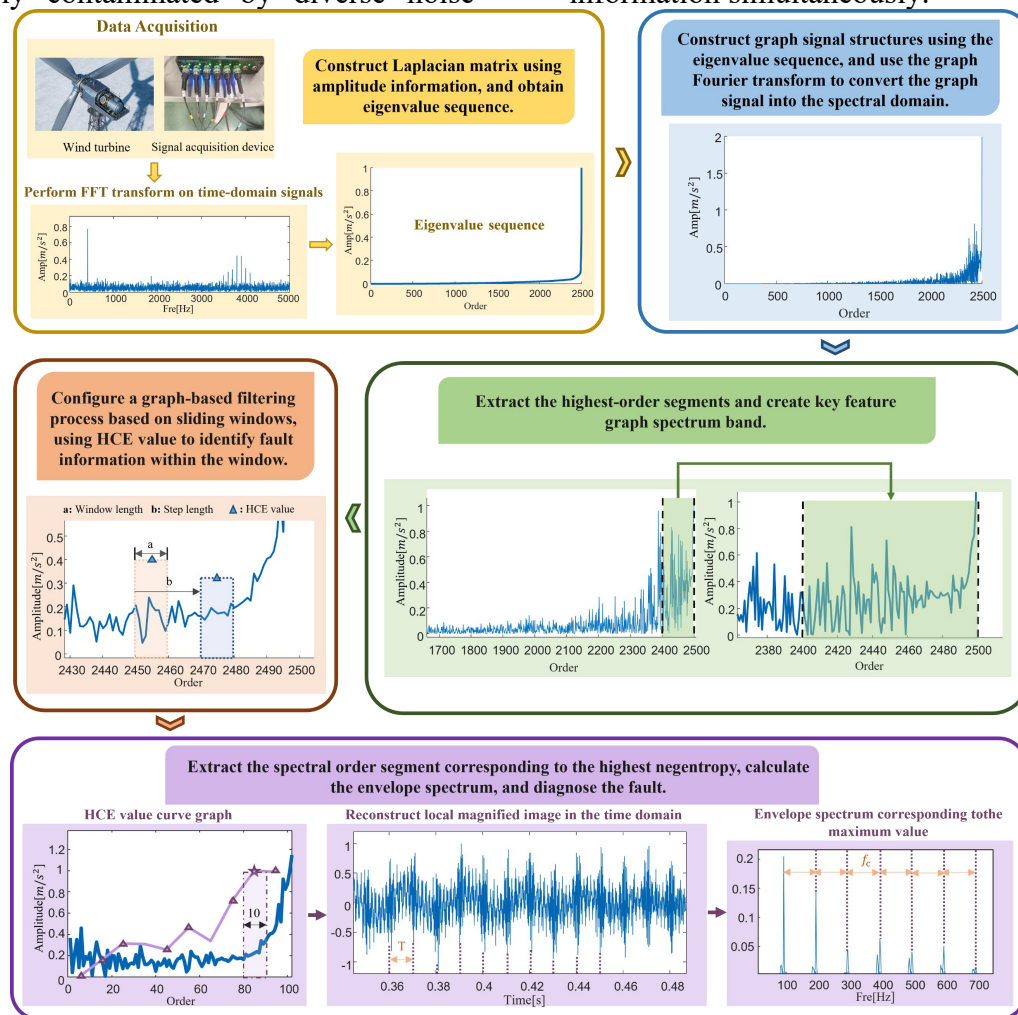


Fig. 2. Flow chart of the Correlation Graph Fourier Spectrum Model

Once the original discrete signal has been transformed into the frequency domain, a Laplacian matrix is established based on the amplitude information. Concurrently, the phase information of $X(f)$ is extracted to facilitate the construction of a graph signal. Because the frequency domain is capable of decomposing complex non-stationary time-domain signals into harmonic components at various frequencies, it provides an intuitive representation of the signal's frequency composition and energy distribution. Accordingly, this study applies the graph Fourier transform (GFT) to the constructed graph signal, yielding a frequency-domain spectrum that supports the extraction of periodic fault characteristics.

$$Re(f) = ABS(X(f)) \quad (7)$$

The above formula $Re(f)$ represents a frequency domain signal with only amplitude information.

$$\hat{g}(k) = \sum_{n=0}^{N-1} Re(n) \cdot l_k(n), k = 0, 1, 2, \dots, N-1 \quad (8)$$

$$\hat{g} = L^T \cdot Re \quad (9)$$

In the above formulas, $\hat{g}(k)$ represents the frequency-domain spectrum sequence, and l_k is the eigenvector corresponding to the frequency-domain spectrum.

By selecting an appropriate order in the frequency-domain spectrum $\hat{g}'(k)$, an inverse graph Fourier transform can be performed:

$$Re'(p) = \sum_{k=0}^{N-1} l_p(p) \cdot \hat{g}'(k), k = 0, 1, 2, \dots, N-1 \quad (10)$$

$$Re' = L \cdot \hat{g} \quad (11)$$

After the frequency-domain spectrum has been successfully acquired through appropriate signal processing techniques, a subset of representative orders is carefully

selected to conduct inverse transform analysis. This analytical process generates a new set of amplitude information, denoted as $Re'(p)$, which accurately reflects the adjusted amplitude characteristics corresponding to the selected representative orders. The updated amplitude information is further fused with the phase components derived from the original signal to construct a modified frequency-domain signal denoted as $X'(f)$. In this constructed signal, the inherent phase characteristics stay consistent with those possessed by the original frequency-domain signal. the inverse fast Fourier transform (IFFT) is performed to reconstruct the corresponding time-domain signal $x'(n)$. The mathematical expression of this signal reconstruction procedure is given below.

$$x'(n) = \frac{1}{N} \sum_{k=0}^{N-1} X'(f) \cdot e^{\frac{2\pi}{N} k \cdot n \cdot j}, n = 0, 1, 2, \dots, N-1 \quad (12)$$

Finally, bearing fault diagnosis is achieved by computing the Hilbert envelope spectrum of the reconstructed signal.

Spectral clustering methods possess prominent comprehensive advantages in signal processing fields, featuring simple operational procedures for practical implementation and stable and efficient clustering performance when processing various types of signal data. For this reason, such clustering approaches gain widespread popularity and are extensively employed in fundamental theoretical research as well as realistic engineering practice scenarios. The Laplacian matrix owns inherent structural characteristics that correspond well to clustering attributes. On this basis, the Laplacian matrix is taken as the basis function of the graph Fourier transform (GFT) in the theoretical framework of this research. The GFT algorithm enables the decomposition of intricate graph signals into linear superposition forms composed of

eigenvectors with diverse orders. A strict one-to-one matching relationship is further built between vertex attribute information carried by graph signals and order feature information contained in the frequency-

domain spectrum. Such inherent correlation endows the frequency-domain spectral model with abundant underlying feature information, which is indispensable for subsequent signal feature extraction and fault diagnosis analysis.

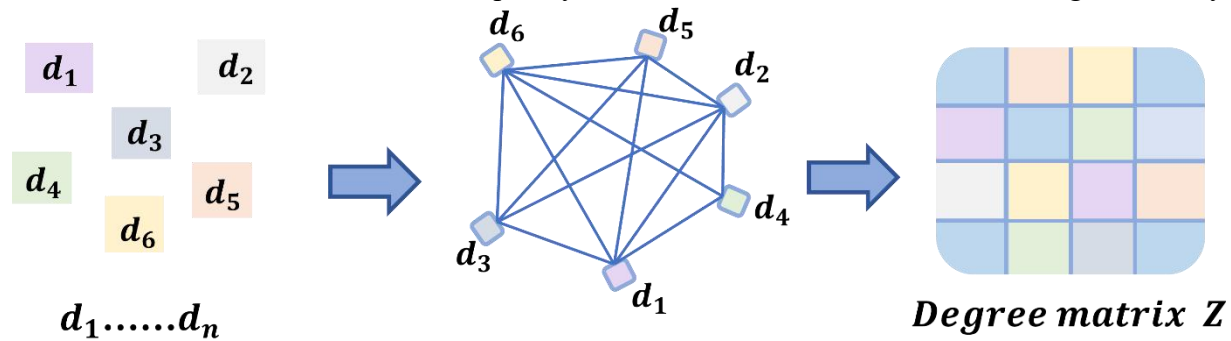


Fig. 3. Constructing a spectrum based on frequency

3.2 Spectral Filtering Based on Eigenvalue Sequence Order

In the conversion from the frequency domain to the spectral domain, not all orders are suitable for fault diagnosis analysis. Effective fault-related information is not uniformly distributed across all orders but is rather clustered within specific order intervals, with core fault features typically concentrated in relatively high-order ranges. Eigenvalues and frequency-domain spectra are closely correlated, and the importance of eigenvalue order is highly consistent with the importance of the corresponding frequency-domain spectral order. The value of the eigenvalue gradually increases with increasing order, reflecting the increasing importance of the corresponding eigenvector. In other words, the later the order of the eigenvector, the more complete the fault-related feature information it contains.

From the perspective of graph spectral analysis, low-order graph spectral components correspond to the global smooth and low-frequency trend components of graph-constructed frequency-domain signals, which are mainly dominated by harmonic background components and Gaussian noise.

By contrast, high-order graph spectral bases possess strong local variation characterization capability, which are sensitive to abrupt amplitude fluctuations in frequency-domain spectrum. Bearing fault impulses will produce periodic side-band modulations and local sharp amplitude mutations in the frequency spectrum after FFT. These local spectral mutations can be sparsely represented and concentrated in high-order graph spectral components after graph Fourier transform. Therefore, the highest-order graph spectral segments can concentrate the transient fault impulse information, while suppressing the smooth harmonic and random noise components. Considering the signal length and computational overhead, the last 100 high-order graph spectral components are intercepted as the candidate feature band for subsequent analysis.

To improve algorithm efficiency and diagnostic accuracy, this paper selects the last 100 orders, as shown in Fig. 4. The figure also shows that noise in the reconstructed frequency-domain waveform within the [2401–2500] order range is suppressed, while the fault feature frequency band is still

selected. This result demonstrates that the proposed CGFS method can effectively

reduce signal noise.

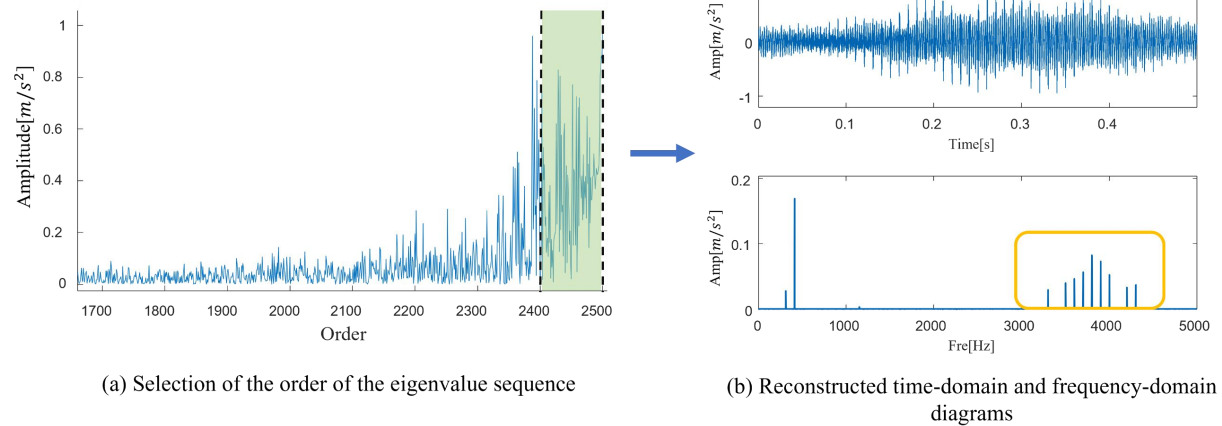


Fig. 4. Order selection effect diagram

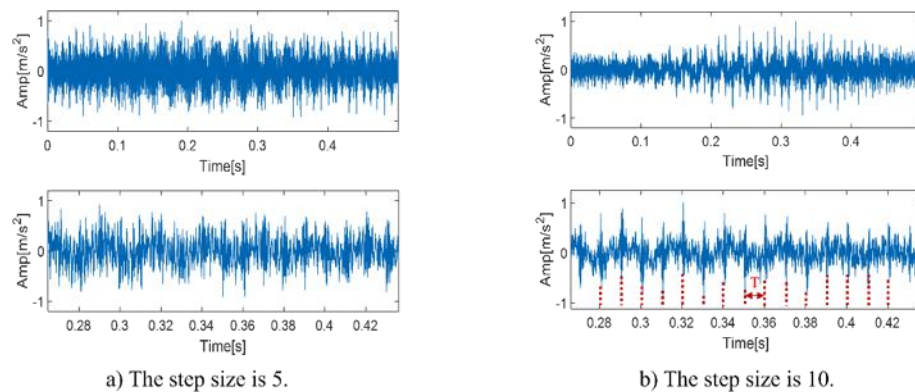


Fig. 5. Reconstructed time-domain waveform and locally amplified diagram (window size fixed at 10)

Window size and sliding step size are two key adjustable parameters in the algorithm, and their settings directly affect the effectiveness of the analysis results and computational efficiency. If the window is set too large, although it can cover a wider frequency band, it will include more interference components unrelated to the fault in the calculation, resulting in a decrease in the signal-to-noise ratio of the reconstructed signal and affecting diagnostic clarity. Conversely, if the window is too small, it may not completely capture the distribution pattern of fault features in the

frequency-domain spectrum, resulting in insufficient feature extraction and reduced diagnostic reliability. Similarly, the choice of sliding step size requires careful consideration: a step size that is too large improves computation speed but may skip important feature intervals, causing omission of key frequency-domain spectral orders; a step size that is too short ensures detailed analysis but significantly increases the computational load due to excessive window overlap, reducing algorithm efficiency. A balance must be struck between feature

completeness, anti-interference ability, and computational cost. As shown in Fig.5, keeping the window size at 10 and comparing step sizes of 5 and 10, the results show that

when the step size is 10, periodic pulse information is clearly observed in the frequency-domain spectrum.

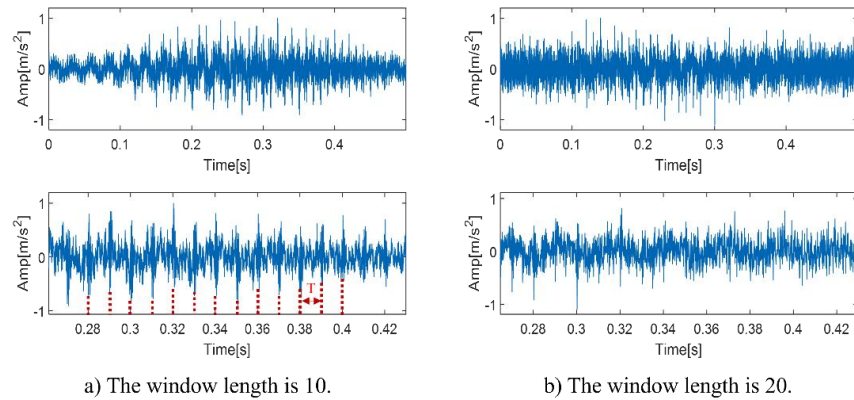


Fig. 6. Reconstructed time-domain waveform and locally amplified diagram (step size fixed at 10)

As shown in Fig. 6, keeping the step size at 10 and comparing window lengths of 10 and 20, the results show that when the window length is 10, periodic pulse information is clearly observed in the frequency-domain plot. In summary, through controlled variable experiments and effect verification, this paper selects a window length and sliding step size of 10 as the optimal parameter configuration. This setting effectively controls redundant calculations and noise interference while ensuring that fault characteristics are fully captured.

3.3 Optimal order selection scheme based on Hybrid Correlation Entropy

To select the optimal reconstruction order, this section sets a sliding window on the key frequency-domain spectral bands and uses a hybrid correlation entropy indicator to quantify and label the feature information

within the window. As shown in Fig. 7, the window slides over the selected 100 orders, with both window length and sliding step size set to 10.

Fault impulses in rotating machinery generally exhibit periodic behavior, and the vibration signals of faulty gears or bearings can be modeled as periodic impulse sequences. Existing diagnostic algorithms aim to extract informative fault signatures from noisy measurements and to quantify fault characteristics. Autocorrelation is effective in extracting and reinforcing periodic fault components, whereas entropy serves as a measure of system stability. Accordingly, we propose a Hybrid Correlation Entropy (HCE) that combines autocorrelation with energy spectrum entropy to quantify the feature information in the order domain.

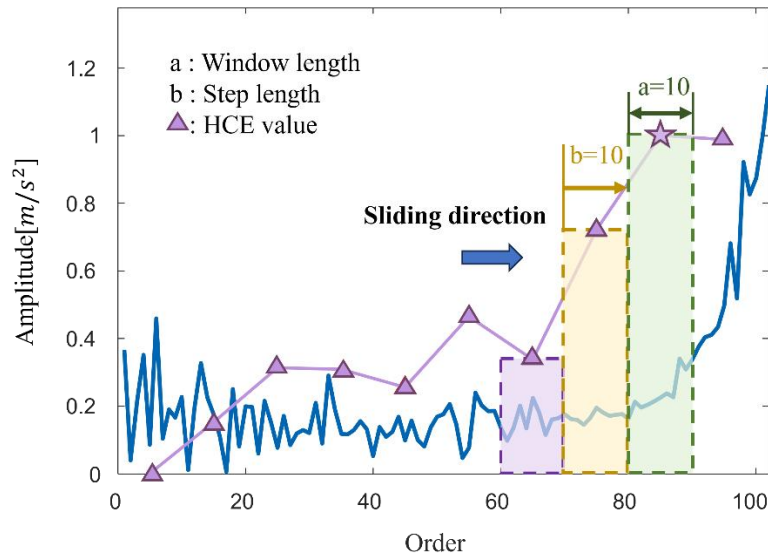


Fig. 7. Schematic diagram of key feature map with windowing processing

Envelope autocorrelation kurtosis accounts for both the amplitude and periodic characteristics of vibration signals while suppressing random interference noise. It serves as a metric for the adaptive selection of the optimal demodulation band, thereby enabling the effective identification of incipient weak faults in rolling bearings under high background noise. The calculation procedure for the autocorrelation kurtosis is detailed as follows, commencing with the computation of the Hilbert envelope of the original signal.

$$e[n] = |\mathcal{H}\{x[n]\}|, n = 0, 1, \dots, N-1 \quad (13)$$

Then calculate the squared envelope signal:

$$y[n] = e^2[n] \quad (14)$$

Compute the unbiased autocorrelation function of $y(n)$:

$$R[m] = \frac{1}{N-|m|} \sum_{n=0}^{N-|m|-1} y[n] y[n+|m|] \quad (15)$$

Here, $m = -(N-1), \dots, N-1$ represents all values of the return delay from $-(N-1)$ to $N-1$.

Set the initial delay m_1 , the final delay m_2 , and the segment length L ,

$$m_1 = \lfloor 0.010 \cdot N \rfloor + 1 \quad (16)$$

$$m_2 = \lfloor 1.55 \cdot N \rfloor - N \quad (17)$$

$$L = m_2 - m_1 + 1 = \lfloor 1.55N \rfloor - N - \lfloor 0.010N \rfloor \quad (18)$$

Calculate its truncated subsequence:

$$R_{cut}[k] = R[m_1 + k], k = 0, 1, \dots, L-1 \quad (19)$$

Subsequently, calculate the mean of the autocorrelation sequence:

$$\mu_R = \frac{1}{L} \sum_{k=0}^{L-1} R_{cut}[k] \quad (20)$$

Finally, calculate the kurtosis of the truncated autocorrelation.

$$\sigma_R^2 = \frac{1}{L} \sum_{k=0}^{L-1} (R_{cut}[k] - \mu_R)^2 \quad (21)$$

$$\mu_4 = \frac{1}{L} \sum_{k=0}^{L-1} (R_{cut}[k] - \mu_R)^4 \quad (22)$$

$$K_{AC} = \frac{\mu_4}{(\sigma_R^2)^2} \quad (23)$$

Next, the spectral entropy of the energy spectrum is computed to capture fault-induced energy concentration and to facilitate reliable detection of weak bearing faults in the presence of strong noise, since it quantifies the spread of the spectral energy distribution. In bearing fault cases, the occurrence of the characteristic frequency and its harmonics leads to an increase in the entropy value.

As a complement to the frequency-domain features, we do not directly use the envelope spectrum but rather measure the frequency-domain energy distribution of the squared original signal. The number of FFT points is determined as the first step.

$$N_{fft} = 2 \times \lfloor \frac{N}{2} \rfloor \quad (24)$$

After applying zero-padding to x for the FFT, the energy spectrum is computed over the positive frequencies, omitting the DC and Nyquist terms.

$$S[k] = \sum_{n=0}^{N_{fft}-1} x^2[n] e^{-j \frac{2\pi}{N_{fft}} kn}, k = 0, 1, \dots, N_{fft} - 1 \quad (25)$$

$$P[i] = |S[i + 1]|^2, i = 0, 1, \dots, N_p - 1 \quad (26)$$

where $N_p = N_{fft}/2 - 1$, and μ_p is defined as the mean of the energy spectrum:

$$\mu_p = \frac{1}{N_p} \sum_{i=0}^{N_p-1} P[i] \quad (27)$$

This paper proposes a logarithmically weighted information metric, denoted as G , based on the ratio relative to the mean, which is computed only for positive spectral lines ($P[i] > 0$).

$$G = \frac{1}{N_p} \sum_{i \in J} \frac{P[i]}{\mu_p} \cdot \ln \left(\frac{P[i]}{\mu_p} \right) \quad (28)$$

Unlike the standard Shannon entropy, this metric is a customized approximation of the average negentropy and can be negative.

The proposed Hybrid Correlation Entropy is then defined as the product of the kurtosis and this information measure.

$$HCE = K_{AC} \times G \quad (29)$$

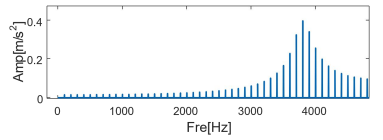
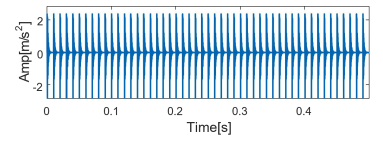
The HCE jointly enhances the periodic impulse intensity in the time domain and the distribution in the frequency domain. This coupling magnifies fault signatures by exploiting the complementarity of both domains, thus improving the metric's sensitivity to incipient weak faults. As a result, this composite metric yields a more reliable and stable indication of fault existence than either domain-based measure alone. A larger entropy value reflects not only stronger periodic impulses in the time domain but also a more complex frequency-domain spectrum that departs markedly from a Gaussian shape. The synergistic reinforcement between these two aspects suggests a higher likelihood and more advanced stage of bearing faults, whereas a lower entropy points to a signal that more closely resembles the random and stationary behavior of a healthy state.

4. Simulation verification

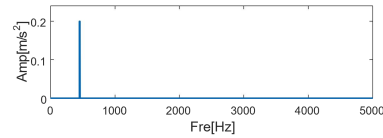
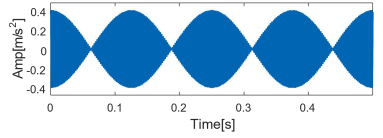
The feasibility and superiority of the proposed CGFS method are validated via simulation experiments. In actual operating conditions, large-scale mechanical equipment is commonly affected by intense background noise and multiple uncertain interference factors. This chapter introduces local impulse interference and noise interference into the simulation signal to simulate random and accidental impacts experienced by the equipment during actual operation. The sampling frequency adopted

for the simulated signal in this experiment is 10,000 Hz. The simulated signal consists of five individual signal components, as presented in the following content:

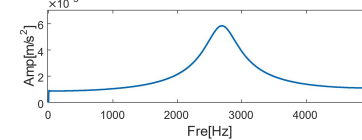
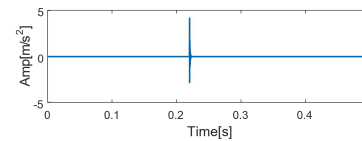
$$\left\{ \begin{array}{l} x_1(t) = 3.8 \sum_{k=1}^K e^{-2\pi g f_n(t-kT)} \sin[2\pi f_n \sqrt{1-g^2}(t-kT)] \\ x_2(t) = 0.4 \cos(8\pi t) \cdot \cos(900\pi t) \\ x_3(t) = 0.8 \sin(77\pi t) \sin(3200\pi t + \sin(175\pi t)) + 0.8 \cos(4\pi t) \cos(111\pi t) \\ x_4(t) = 5e^{-0.2\pi \times 2700 t} \sin(2\pi \times 2700 \sqrt{1-0.1^2} t) \\ x(t) = x_1(t) + x_2(t) + x_3(t) + x_4(t) + g(t) \end{array} \right. \quad (30)$$



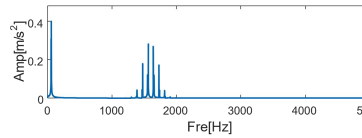
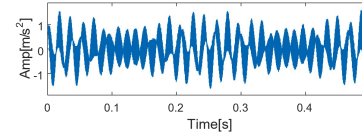
(a) $x_1(t)$



(c) $x_3(t)$



(b) $x_2(t)$



(d) $x_4(t)$

Fig. 8. Each component signal

The four aforementioned signal components are superimposed to form a composite signal. Gaussian white noise with a signal-to-noise ratio of -10 dB is further added to this composite signal, so as to simulate the complex interference environment existing in practical engineering operation scenarios. The time-domain waveform, frequency-domain spectrum and envelope spectrum of the noisy signal $x(t)$ are depicted in Fig. 9. It is difficult to clearly distinguish the periodic impulse components from the time-domain waveform. In the frequency-domain

spectrum, inherent fault feature components are severely concealed by the noise floor, resulting in ineffective discrimination of complete sideband structures associated with fault features. As reflected by the envelope spectrum displayed in Fig.9 (b), only a single dominant amplitude peak appears within the overall spectra. The attribute of this frequency component cannot be confirmed as the genuine fault characteristic frequency, which leads to the inability to complete reliable fault identification and classification tasks.

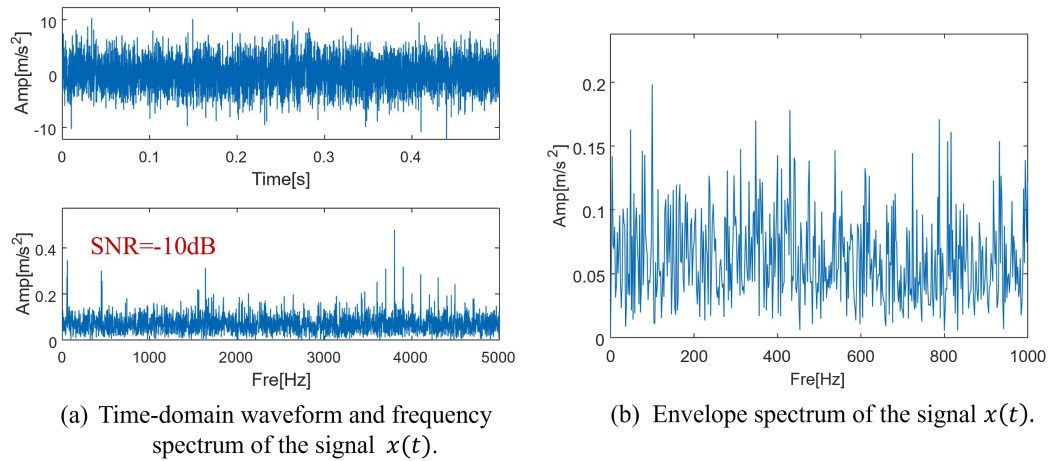


Fig. 9. Time-domain waveform, spectrum, and envelope spectrum of the signal $x(t)$

The proposed CGFS method is adopted to conduct signal processing and quantitative analysis for the simulated signal of bearing outer race fault. Key frequency-domain spectral bands distributed within the last 100 orders are screened out for subsequent analysis, as intuitively demonstrated in Fig.10. As observed in subgraph (a), a sliding window with a window length of 10 and a step size of 10 is utilized to compute the corresponding HCE indicator values. The optimal order band matched with the maximum indicator value is determined as [2491–2500]. Subsequently, Fig. 11(a) compares the reconstructed waveform with the original waveform over the time interval ranging from 0.36 to 0.46 s, where distinct periodic impulse components can be clearly observed. After signal reconstruction, the corresponding Hilbert envelope spectrum is presented in Fig. 11 (b). The 1st to 7th harmonic components of the bearing fault characteristic frequency can be clearly identified in the overall spectrum. Background interference noise is suppressed to a satisfactory degree, bringing considerable enhancement to the overall signal-to-noise ratio. The above analysis verifies that the proposed CGFS method possesses excellent capability for effective

extraction of fault feature information.

The Fast Kurtogram serves as a crucial and widely used tool in the field of rotating machinery fault diagnosis, playing an irreplaceable role in fault feature extraction and early fault identification. This algorithm takes spectral kurtosis as its core evaluation indicator and adopts a binary tree-based alternating frequency band decomposition strategy to construct a multi-resolution filter bank with hierarchical frequency division capabilities. It refines the frequency resolution layer by layer through iterative decomposition, and calculates the kurtosis value of each divided sub-band to characterize the impulsiveness of signals in different frequency ranges. By integrating the Fast Fourier Transform (FFT) into the entire computational process, the algorithm effectively overcomes the inefficiency of traditional Kurtogram in large-scale data processing scenarios, which substantially boosts computational efficiency while maintaining analytical accuracy. Such advantages make it well-suited for real-time monitoring of rotating machinery and large-volume signal data processing scenarios. For this purpose, the Fast Kurtogram algorithm is chosen as the comparison baseline to evaluate the performance of the proposed approach,

and the relevant analytical results are presented in Fig. 12. The fault-related information identified by this algorithm is concentrated in the 5th decomposition layer. Following the extraction of signals from this specific frequency band, the signal reconstruction process is carried out in sequence, and the corresponding envelope spectrum is further calculated through relevant mathematical operations. Even after

completing all these processing steps, the characteristic frequency of the outer ring fault still fails to be clearly identified in the obtained envelope spectrum. This phenomenon indicates that the Fast Kurtogram method fails to clearly extract effective fault features under strong noise conditions, limiting its application in complex interference environments.

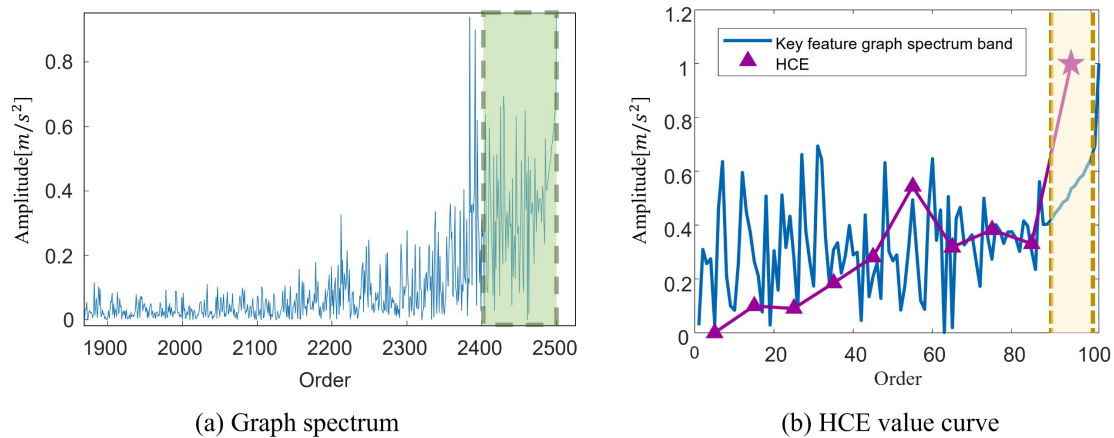


Fig. 10. Spectra and HCE value curves obtained using the CGFS method

The Autogram algorithm belongs to a class of adaptive spectral kurtosis analysis techniques that are dedicated to the fault diagnosis of rotating machinery. This approach incorporates inherent adaptive frequency division strategies and iterative optimization mechanisms, which can further strengthen the extraction performance of latent fault feature information. This algorithm adaptively partitions the entire frequency domain into multiple sub-bands according to the inherent characteristics of the input signal, while continuously performing iterative optimization on kurtosis-related indices to refine the evaluation accuracy of signal impulsiveness. By conducting comprehensive analysis on inherent signal characteristics, this method can autonomously locate and screen out the optimal frequency band rich in fault impact information. The whole implementation

process requires no manual participation or predefined parameter configuration in advance. Such an adaptive design effectively eliminates the need for manual parameter presets, thereby significantly improving the adaptability of fault feature extraction under complex working conditions—including strong noise interference and variable operating parameters—and further enhancing the accuracy of fault feature identification. The processing results obtained by applying the Autogram algorithm to the simulated fault signal are illustrated in Fig. 13. Through the adaptive frequency band selection mechanism of this algorithm, the sensitive frequency band that contains fault-related information is accurately located at the 5th decomposition level. This sensitive frequency band has a center frequency of 3671.875 Hz and a bandwidth of 156.25 Hz. Such results reflect evident discrepancies in

the selection of center frequency for the optimal frequency band between the Autogram algorithm and the former comparison algorithm adopted in this work. Nevertheless, similar to the comparative method, the Autogram algorithm also fails to clearly identify effective fault feature information under strong noise conditions after conducting signal reconstruction and envelope spectrum computation, indicating its limitations in complex interference environments.

The EWTk-HSK method constitutes a self-adaptive fault diagnosis procedure that integrates a Gaussian scale-space (GSS)-based spectrum segmentation strategy with harmonic spectral kurtosis (HSK)-guided optimal frequency band selection. The implementation steps are as follows: first, GSS is adopted in place of the conventional local-minimum approach to perform adaptive boundary partitioning of the original frequency spectrum. Subsequently, the empirical wavelet transform (EWT) is constructed based on the Meyer wavelet, thereby decomposing the signal into a set of narrowband modal components. For each resulting mode, the HSK value with respect to the prescribed fault characteristic frequency (and its first to sixth harmonics) is computed, and the frequency band exhibiting the maximum HSK is automatically selected as the optimal resonance band. Finally, Hilbert envelope demodulation is performed on the chosen band, while two quantitative indices, namely the harmonic energy ratio (HER) and the false peak-to-noise ratio (FPNR), are introduced to evaluate the intensity of fault impacts and the spectral-line clarity of the envelope spectrum, respectively.

With the target fault frequency set at 100 Hz, the analytical results, as illustrated in Fig. 14,

indicate that the algorithm identifies the high-frequency resonance band from 3920 Hz to 5000 Hz as containing the predominant fault information, where the HSK metric attains its maximum value. Nevertheless, after envelope demodulation of this band, the resultant envelope spectrum, although capable of revealing the fault characteristic frequency, fails to sufficiently suppress the background noise. Moreover, the harmonic multiples of the fault characteristic frequency remain indistinct and cannot be clearly resolved. These observations suggest that the demodulation performance of the proposed method remains suboptimal, thereby limiting its reliability for practical bearing fault diagnosis.

Simulation experiments are carried out under controllable extreme noise conditions with $\text{SNR} = -10$ dB, where the signal is mixed with preset periodic fault impulses, harmonic interferences, random accidental shocks and Gaussian white noise. The simulation results verify the theoretical effectiveness of the CGFS model: benefiting from frequency-domain graph construction and hybrid correlation entropy screening strategy, the proposed method can accurately locate the optimal graph-order band and reconstruct periodic fault impulses. Complete fault characteristic frequency and its multiple harmonics can be observed in the envelope spectrum. In contrast, Fast Kurtogram, Autogram and the EWTk-HSK method fail to retrieve fault features under such heavy simulated noise. The simulation validates that the CGFS possesses promising anti-interference potential for weak fault feature extraction, and provides theoretical support and parameter reference for subsequent practical experimental verification.

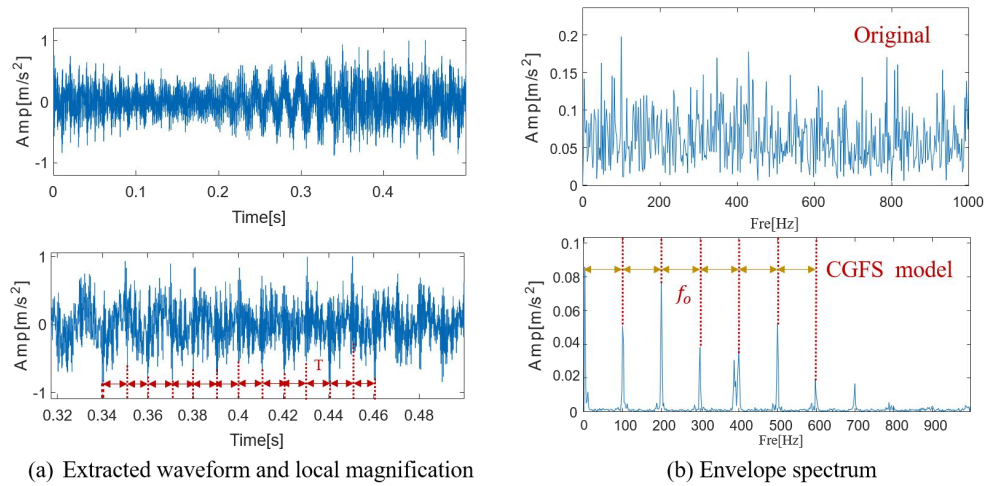


Fig. 11. Time-domain waveforms and envelope spectra obtained using the CGFS method

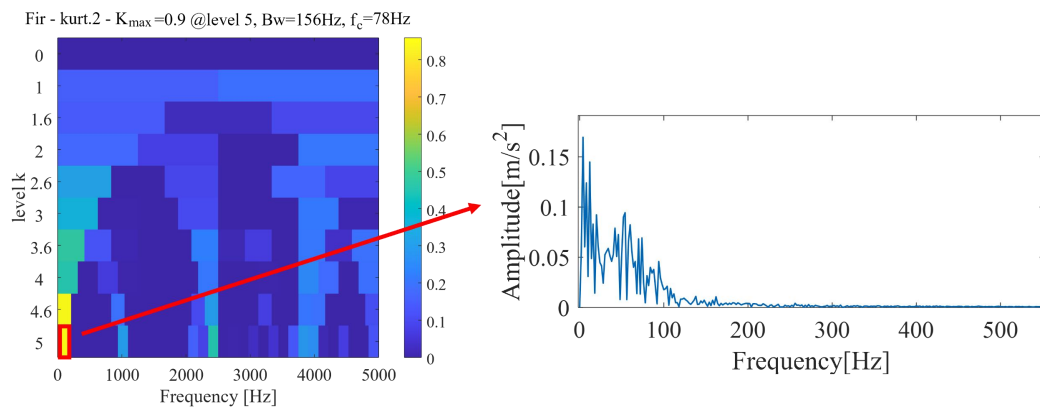


Fig. 12. Envelope spectrum obtained using the Fast Kurtogram method

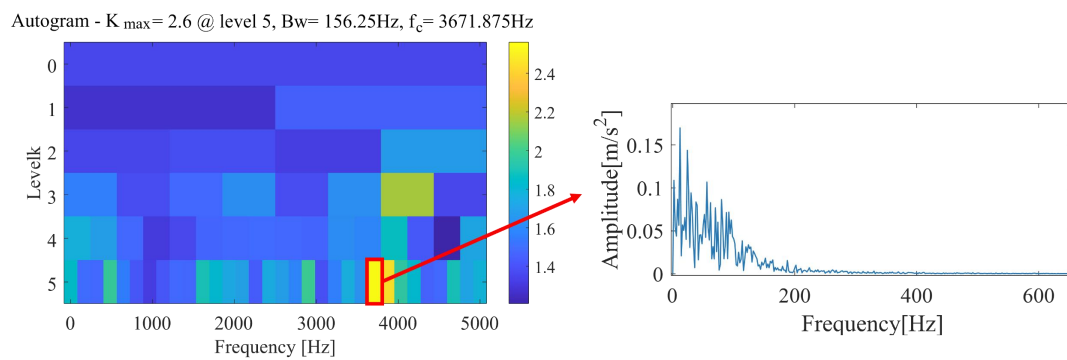


Fig. 13. Envelope spectrum obtained using the Autogram method

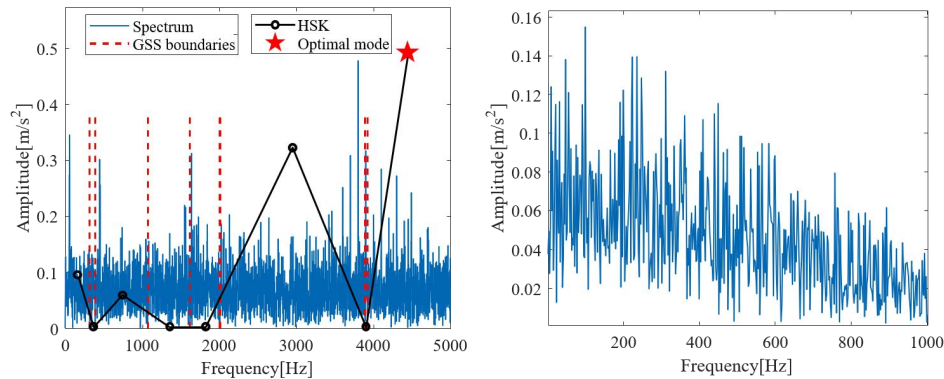


Fig. 14. Envelope spectrum obtained using the EWTk-HSK method

5. Experimental verifications

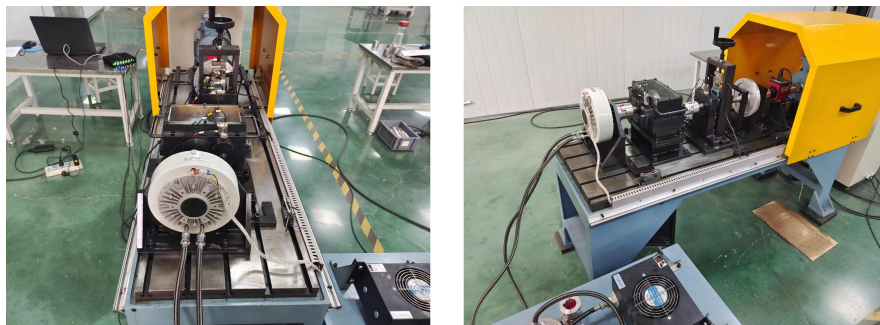


Fig. 15. Test bench

Bearings are indispensable in wind turbines. As the core transmission component of rotating machinery, their operating condition directly determines the overall reliability of the equipment. Failures can easily trigger a chain reaction of shutdowns or even safety accidents. Therefore, bearing fault diagnosis is of significant engineering importance. To

demonstrate the practicality of the proposed method, this section processes real-time data collected from an experimental platform. The signal sampling frequency was set to 16,384 Hz, and all bearing experimental data were acquired from the platform shown in Fig. 15.

5.1 Bearing outer ring test signal

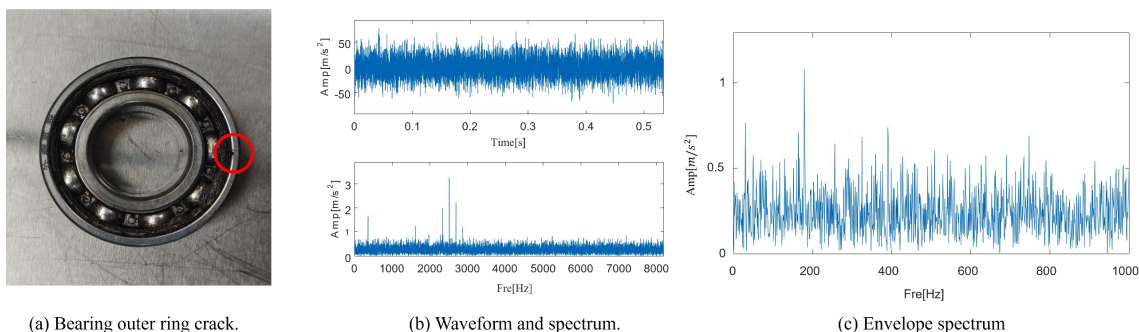


Fig. 16. Outer ring faults and outer ring test signals

The following section presents a detailed analysis of the experimental signal processing process for a bearing outer ring fault, aiming to systematically evaluate the fault feature extraction effect under actual experimental conditions. The specific fault type studied in this experiment is an outer ring crack, which is a common and typical fault form in rolling bearings, and its detailed fault morphology (including crack size, shape, and distribution) is clearly illustrated in Fig. 16(a). To simulate the actual operating state of the bearing, the motor speed was stably set to 3000 r/min during the experiment, ensuring the bearing operates under a constant rotational speed condition. The tested bearing adopted in the experiment is a 6205 deep groove ball bearing, a widely used type in industrial rotating machinery due to its simple structure and reliable performance. The specific structural dimensions of the tested 6205 deep groove ball bearing are strictly measured as follows: the bearing width is 15 mm, the inner diameter is 25 mm, and the outer diameter is 52 mm. With reference to these accurately acquired structural parameters and the corresponding computational formula for fault characteristic frequency, the theoretical fault characteristic frequency of the bearing outer race is calculated as 179.25 Hz. The raw time-domain waveform of the bearing fault signal displayed in Fig. 16 (b) fails to present obvious periodic impulse components related to outer race crack fault. This phenomenon primarily arises from background noise interference and the weak amplitude characteristics of fault signals under actual experimental conditions. Similarly, in the Hilbert envelope spectrum displayed in Fig. 16(c), which is commonly used for fault feature extraction, the theoretical fault characteristic frequency of 179.25 Hz is not effectively identified, indicating that the fault

features are severely masked in the original experimental signal.

Subsequently, the key frequency-domain spectral bands distributed within the last 100 orders of the experimental bearing fault signal are carefully selected for further analysis, as clearly illustrated in Fig. 17 (a). A fixed sliding window with a length of 10 and a step size of 10 is imposed on the screened spectral bands. Systematic calculation of the corresponding HCE indicator values is performed to assess the impulsiveness of signal components distributed across different frequency intervals, and the obtained computational outcomes are illustrated in Fig. 17 (b). The distribution characteristics of HCE indicator values enable clear determination of the optimal order band corresponding to the maximum indicator value at [8191–8200], which is regarded as the frequency range carrying the most prominent fault-related information. The time-domain waveform reconstructed by the proposed CGFS method is shown in Fig. 18(a), where prominent periodic impulse components closely correlated with bearing outer ring fault can be clearly observed. Such periodic pulses can precisely reflect the inherent impact characteristics induced by the contact interaction between bearing rolling elements and outer ring crack during operational processes. Furthermore, as shown in Fig. 18 (b), the bearing fault characteristic frequency in the Hilbert envelope spectrum forms a clear and prominent amplitude peak, which is free from obvious noise interference and can be easily distinguished. Meanwhile, the background noise in the spectrum is effectively suppressed to a very low level, which enables the fault feature information to be prominently displayed and ensures the accurate identification of the bearing outer ring fault.

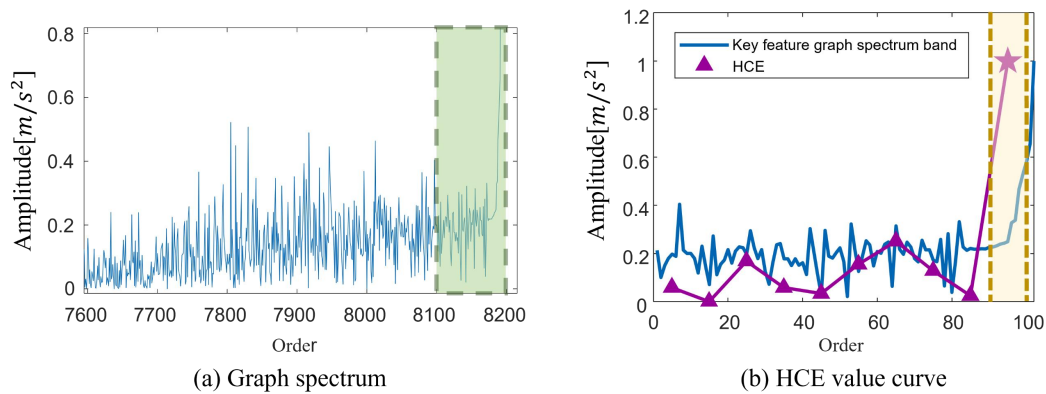


Fig. 17. Frequency domain spectra and HCE value curves obtained using the CGFS Model

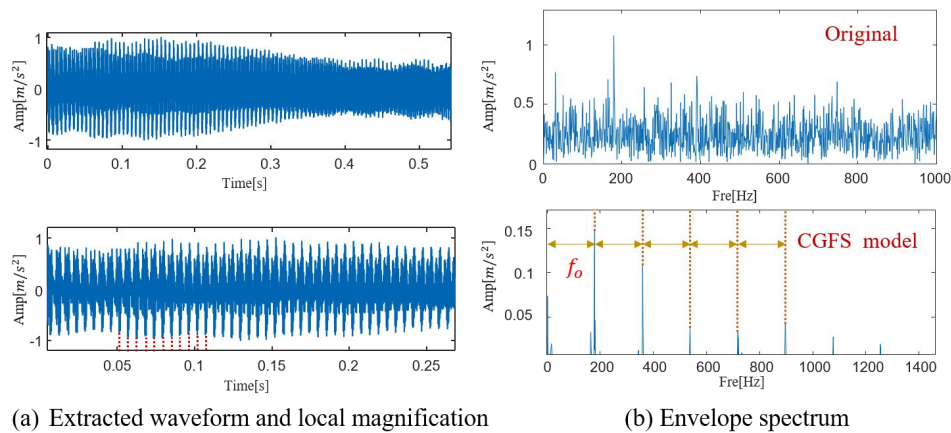


Fig.18. Time-domain waveforms and envelope spectra obtained using the CGFS method

For performance comparison, three algorithms commonly used in the field of fault diagnosis—namely, the Fast kurtogram, the Autogram, and the proposed EWTk-HSK method—are employed and tested against each other. To ensure fairness in the comparative study, all algorithms are applied to the same measured vibration signal of an outer-race fault.

Results from Fast Kurtogram are shown in Figure 19(a). The algorithm pinpoints the optimal sensitive frequency band at the 5-th decomposition level, with centre frequency 1408 Hz and bandwidth 256 Hz. Signal reconstruction and subsequent envelope spectrum calculation are performed on the

extracted band. Nevertheless, no spectral peak matching the outer-race fault characteristic frequency appears in the output spectrum, and fault discrimination cannot be completed. Figure 19(b) illustrates results of the Autogram algorithm. Its adaptively selected optimal band is also situated at the 5-th decomposition level with centre frequency 896 Hz and bandwidth 256 Hz. After signal reconstruction and envelope demodulation, valid outer-race fault features still cannot be extracted. The target fault frequency is set to 179 Hz, corresponding to the outer-race fault characteristic frequency, and the same bearing outer-race fault signal is analysed using the EWTk-HSK method, with the results presented in Fig.20. The

analytical results indicate that the algorithm identifies the high-frequency resonance band from 4063 Hz to 5000 Hz as containing the predominant fault information, where the HSK metric attains its maximum value. Nevertheless, after reconstruction and envelope demodulation of the signal in this band, no valid spectral peak corresponding to 179 Hz can be discerned in the resulting envelope spectrum; moreover, the background noise is not effectively suppressed, and the spectral lines exhibit a disordered structure. This finding suggests that, although the HSK metric can indicate the frequency band with concentrated energy, the selected band does not necessarily contain modulation information associated with fault impacts. In the present experiment, the EWTK-HSK method fails to achieve effective fault feature extraction, demonstrating insufficient reliability for bearing outer-race fault diagnosis.

Unlike ideal simulated signals, measured signals are contaminated by multiple interferences including transmission-chain coupled vibration, assembly gaps and sensor noise. Under such complex noise conditions, the Fast kurtogram, the Autogram, and the EWTK-HSK method are all unable to effectively extract the weak outer-race fault impact components from the mixed signal. In contrast, the CGFS model constructs the Laplacian matrix in the frequency domain, which avoids adverse effects of time-domain noise outliers on graph topology. Combined with adaptive selection of optimal graph-spectral orders via HCE, prominent fault-related spectral peaks and low noise floor are achieved in the final envelope spectrum. This set of outer-race fault experiments demonstrates that the proposed CGFS model can transfer fault-feature extraction capability from simulation environments to practical measured scenarios with favourable engineering practicability.

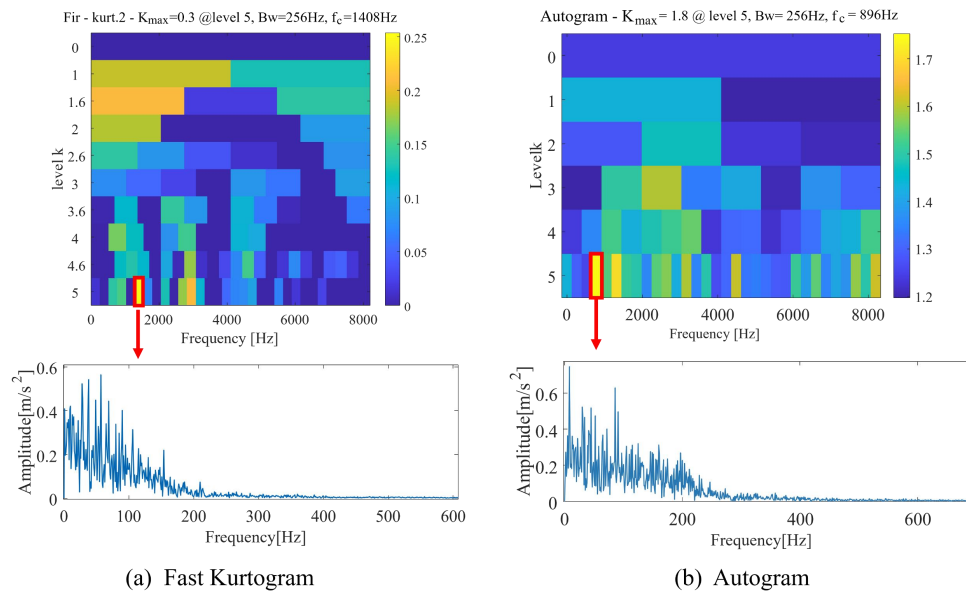


Fig.19. Envelope spectra obtained using the Fast Kurtogram and Autogram methods

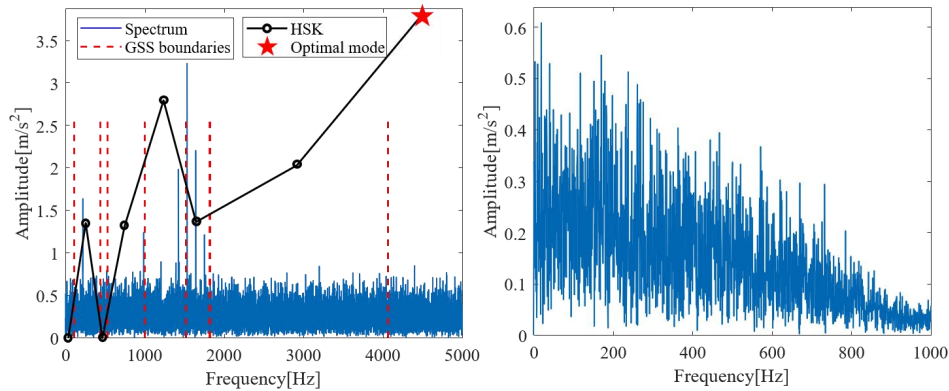


Fig. 20. Envelope spectrum obtained using the EWTk-HSK method

5.2 Bearing inner ring test signal

To further verify the generalisation performance of the CGFS model across diverse fault modes, experimental validation is performed for bearing inner-race crack faults. As a typical destructive fault in rolling bearings, inner-race cracks propagate continuously during operation and may lead

to sudden bearing failure. Fault morphology is shown in Fig.21(a). The 6205 deep-groove ball bearing is still utilised while experimental conditions such as rotational speed remain unchanged. According to bearing geometric dimensions, the theoretical inner-race fault characteristic frequency is calculated as 270.749 Hz, which serves as the reference benchmark for fault identification.

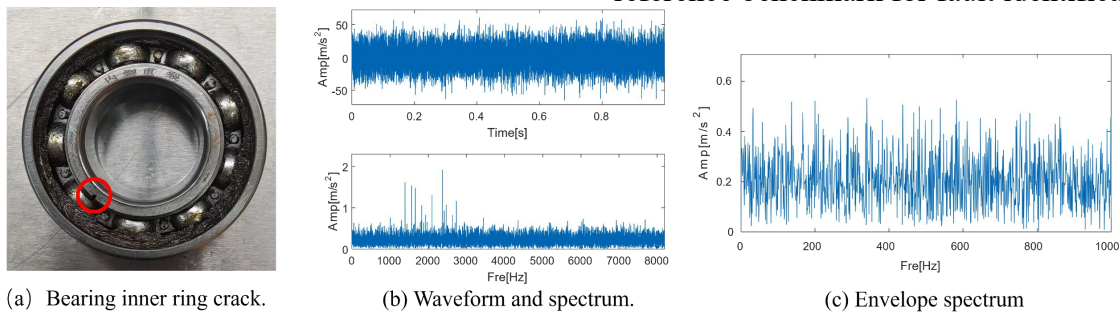


Fig.21. Inner ring fault and inner ring test signal

The raw time-domain vibration waveform of inner-race fault is presented in Fig.21(b). Since impulses induced by inner-race damage are modulated by shaft rotation, fault pulses are inherently weaker and further corrupted by background noise. Periodic impulses cannot be directly observed in time-domain representation. Hilbert envelope demodulation is applied to the raw signal and the result is shown in Fig.21(c). Neither the characteristic frequency of 270.749 Hz nor its sidebands can be

identified in the envelope spectrum. Fault information is masked by heavy noise and conventional demodulation approaches fail.

Similarly, the last 100 orders of high-order graph-spectral segments are selected as candidate feature intervals, as shown in Fig.22(a). A sliding window (length = 10, step = 10) traverses the candidate orders and HCE values are calculated window-by-window to obtain the indicator curve in Fig.22(b). The optimal

reconstruction order interval [8191, 8200] is determined according to the HCE maximum.

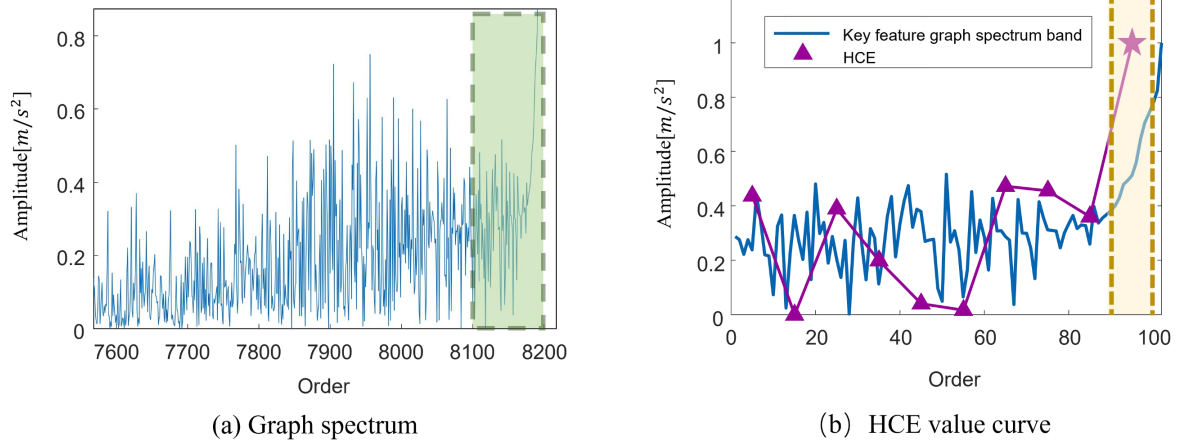


Fig.22. Frequency-domain spectra and HCE value curves obtained using the CGFS method

Inverse graph Fourier transform is implemented based on this optimal-order interval to obtain the reconstructed denoised time-domain signal. The waveform and local magnification are plotted in Fig.23(a), where periodic fault impulses are clearly observed, corresponding to impact responses generated when rolling elements sweep across

inner-race cracks. The Hilbert envelope spectrum of reconstructed data is given in Fig.23(b). The inner-race fault characteristic frequency of 270.749 Hz, together with its modulation sidebands and harmonics, can be clearly distinguished. Background noise is effectively suppressed and accurate identification of inner-race faults is realised.

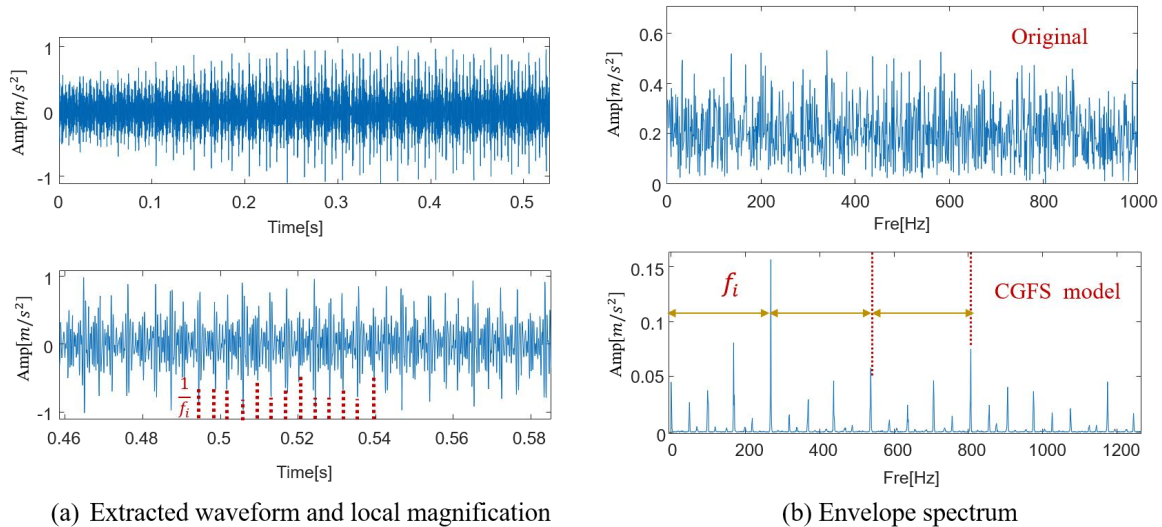


Fig.23. Time-domain waveforms and envelope spectra obtained using the CGFS method

To compare performance between the proposed algorithm and traditional methods for inner-race faults, Fast Kurtogram and

Autogram are reused to process the identical inner-race fault dataset. Experimental

parameters are kept consistent with previous outer-race comparative tests.

Processing results of Fast Kurtogram are shown in Fig.24(a). The algorithm selects the frequency band at the 5-th decomposition level with centre frequency 128 Hz and bandwidth 256 Hz. After signal reconstruction and envelope spectrum computation, no obvious spectral peaks related to inner-race faults emerge and fault-feature extraction fails. For the Autogram algorithm in Fig.24(b), the adaptively selected optimal band lies at the 5-th decomposition level with centre frequency 6528 Hz and bandwidth 256 Hz. Even after signal reconstruction and envelope demodulation, the characteristic frequency and sideband information of inner-race faults still cannot be extracted. To evaluate the diagnostic performance of the EWTk-HSK method for inner-race faults, the target frequency is set to 270 Hz, and the results are presented in Fig. 25. The method identifies the frequency band from 1454 Hz to 2698 Hz as containing the predominant fault information, with the HSK metric reaching its maximum within this interval. Nevertheless, after reconstruction and envelope demodulation of this band, no valid spectral peak corresponding to 270 Hz can be discerned in the resulting envelope spectrum; sideband structures are absent, and the background noise is not effectively suppressed. These findings suggest that, in the present experiment, the algorithm is

incapable of achieving reliable diagnosis of inner-race faults.

The underlying reason lies in the fact that impulses of inner-race faults are modulated by shaft rotation, producing sophisticated sideband structures in vibration signals and increasing the difficulty of fault-feature extraction. The Fast Kurtogram and the Autogram employ spectral kurtosis as the frequency-band selection criterion; however, under strong noise and modulation coupling, they are susceptible to interference, which readily leads to biased band selection. In contrast, the EWTk-HSK method adopts harmonic spectral kurtosis, emphasising the harmonic energy at the target frequency, yet it exhibits weak adaptability to modulation sideband structures. Consequently, all three methods possess inherent limitations under such complex operating conditions, making it difficult to ensure diagnostic reliability. Benefiting from high-order graph-spectral decomposition and the Hybrid Correlation Entropy, the CGFS model jointly measures time-domain periodic impulse properties and frequency-domain energy distribution. It adaptively locks graph-spectral orders containing fault information and successfully recovers inner-race fault characteristic frequencies as well as their sidebands buried in noise. Combining experimental results of both outer-race and inner-race cases, it is validated that the proposed algorithm possesses outstanding fault-feature extraction capability and noise-resistance for diverse rolling-bearing damage modes.

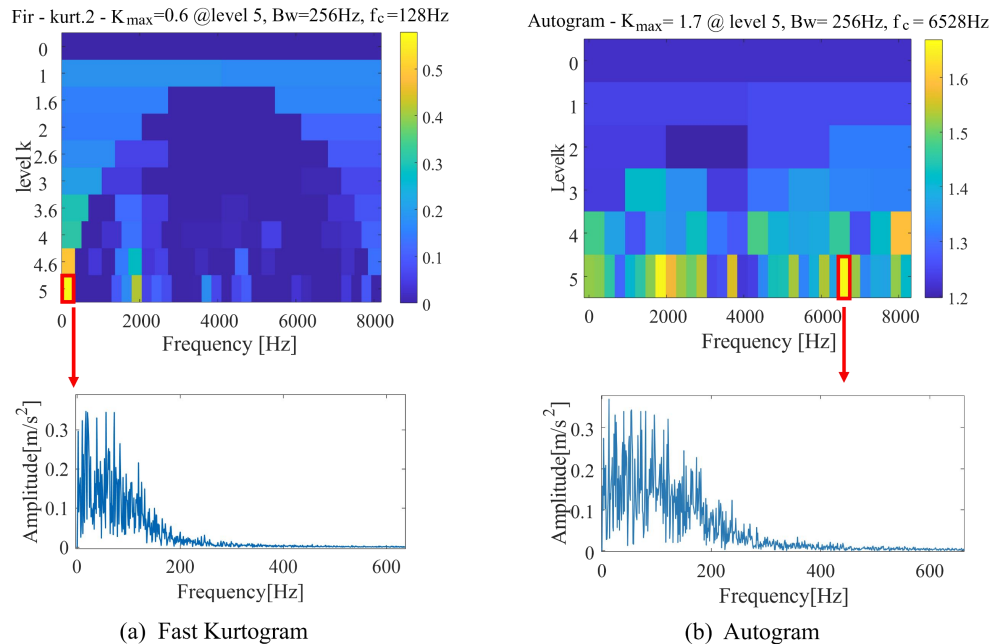


Fig.24. Envelope spectra obtained using the Fast Kurtogram and Autogram methods

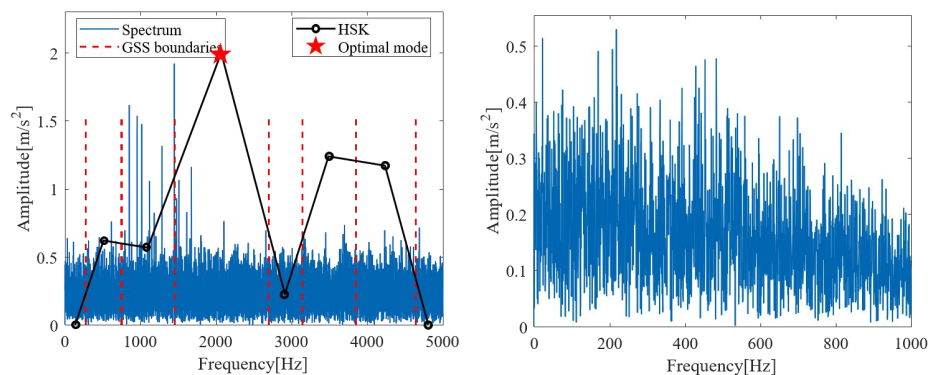


Fig. 25. Envelope spectrum obtained using the EWTk-HSK method

6. Conclusions

This paper proposes the Correlation Graph Fourier Spectrum Model (CGFS Model) specifically designed to address the challenging feature separation problem that arises when processing bearing fault signals under strong noise conditions. This problem often leads to the obscuration of weak fault features by background noise, making it difficult for traditional methods to achieve accurate fault diagnosis. Based on the Laplacian matrix derived from the frequency-

domain spectrum, the CGFS model implements a transformation of the original fault signal from the time domain to the graph domain by virtue of the graph Fourier transform (GFT). This transformation process effectively maps the time-domain characteristics of the fault signal to the graph domain, enabling the model to better isolate and extract weak fault-related information that may be masked by background noise, thus laying a solid foundation for subsequent fault feature extraction and identification.

This transformation effectively avoids the adverse influence of noise extrema existing in the time-domain waveform on the construction of the graph framework, thereby simultaneously achieving two key objectives: enhancement of fault feature dimensionality and effective noise reduction. The enhanced feature dimensionality improves the distinguishability of fault-related components, while noise reduction minimizes the interference of irrelevant noise, laying a solid foundation for subsequent fault feature extraction. Subsequently, to overcome the inherent limitations of traditional fault diagnosis methods in establishing effective connections between the time domain and the graph domain—limitations that often result in incomplete feature extraction and inaccurate fault localization—a linear graph order separation scheme is specially designed and integrated into the CGFS model. This scheme enables efficient and accurate separation of graph orders corresponding to different signal components, ensuring the effective transmission of fault feature information between domains. Additionally, the proposed hybrid correlation entropy indicator is employed to quantitatively evaluate and characterize the fault features present in different graph orders, which facilitates the accurate localization of fault features in both the inner and outer races of wind turbine bearings. Comprehensive simulation results, obtained under various strong noise conditions, demonstrate that the proposed CGFS method can more effectively extract the fault features of rolling bearings compared with traditional algorithms. The method exhibits better robustness against complex background noise and higher diagnostic accuracy in identifying both inner-race and outer-race faults, fully verifying its practical applicability in wind turbine bearing fault diagnosis.

Acknowledgments

The authors would like to gratefully acknowledge the CGN (Beijing) New Energy Technology Co., Ltd. research project (S-Y2025DBN).

References

- [1] P. He, R. Hong, H. Wang, and C. Lu, "Fatigue life analysis of slewing bearings in wind turbines," *Int. J. Fatigue*, vol. 111, pp. 233–242, Jun. 2018.
- [2] J. Chen, Y. Xu, T. Liu, M. Yang, K. Zhang, and H. Zhang, "A convex optimization difference analysis model for intelligent fault detection and diagnosis of gearboxes," *Mech. Syst. Signal Process.*, vol. 251, p. 114207, 2026.
- [3] K. Zhang, M. Li, F. Jin, T. Liu, J. Guo, and Y. Xu, "Variational filtering differential strategy with amplitude modulation for fault feature extraction," *Meas. Sci. Technol.*, vol. 37, p. 106105, 2026.
- [4] C. Zhang, S. Wei, G. Dong, Y. Zeng, G. Zhu, X. Zhou, and F. Liu, "Time-domain sparsity-based bearing fault diagnosis methods using pulse signal-to-noise ratio," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–4, Apr. 2024.
- [5] G. Yu, H. Dong, W. Wang, A. Wang, M. Sun, and F. Li, "Wavelet-enhanced time-frequency analysis method for bearing fault detection of rotating machinery," *IEEE Access*, vol. 12, pp. 120948–120961, 2024.
- [6] X. Zou, H. Zhang, Z. Jiang, K. Zhang, and Y. Xu, "Toward accurate extraction of bearing fault modulation characteristics with novel time-frequency modulation bispectrum and modulation Gini index analysis," *Mech. Syst. Signal Process.*, vol. 219, p. 111629, Oct. 2024.

- [7] W. Bao, S. Liu, Z. Liu, et al., "Generalized synchroextracting transform: algorithm and applications," *Mech. Syst. Signal Process.*, vol. 224, p. 112116, 2025.
- [8] H. Jin, X. Li, J. Yu, T. Wang, and Q. Yun, "Kolmogorov-Arnold convolutional squeeze-and-excitation network for bearing fault diagnosis," *Appl. Acoust.*, vol. 240, p. 110903, Jan. 2025.
- [9] L. Meng, Y. Wu, Z. Xue, D. Zhang, K. Zhang, and F. Li, "DRN-PISSM: a pitch bearing fault diagnosis network integrating dilated residual convolution and a Physics-Informed State-Space model," *Mech. Syst. Signal Process.*, vol. 259, p. 114787, 2026.
- [10] K. Zhang, Y. Liu, L. Zhang, C. Ma, and Y. Xu, "Frequency slice graph spectrum model and its application in bearing fault feature extraction," *Mech. Syst. Signal Process.*, vol. 226, p. 112383, Mar. 2025.
- [11] Z. Wang, X. Yu, Y. Guo, et al., "Research on high-order Vese Osher model based on partial differential equation and its application in planetary gearbox fault diagnosis," *IEEE Trans. Instrum. Meas.*, 2026.
- [12] H. Li, J. Lin, J. Jiao, Z. Liu, W. Li, and Y. Hu, "In-band noise-suppressed bidirectional reweighted cyclostationary analysis for acoustic emission-monitored fault diagnosis," *Mech. Syst. Signal Process.*, vol. 257, p. 114619, 2026.
- [13] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, Mar. 2011.
- [14] D. A. Spielman, "Spectral Graph Theory and its Applications," in *Proc. 48th Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, Providence, RI, USA, 2007, pp. 29–38.
- [15] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
- [16] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Appl. Comput. Harmon. Anal.*, vol. 40, no. 2, pp. 260–291, Mar. 2016.
- [17] A. Sandryhaila and J. M. F. Moura, "Discrete Signal Processing on Graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.
- [18] L. Ou, D. Yu, and H. Yang, "A new rolling bearing fault diagnosis method based on GFT impulse component extraction," *Mech. Syst. Signal Process.*, vol. 81, pp. 162–182, Dec. 2016.
- [19] Y. Gao, D. Yu, and H. Wang, "Fault diagnosis of rolling bearings using weighted horizontal visibility graph and graph Fourier transform," *Measurement*, vol. 149, p. 107036, Jan. 2020.
- [20] L. Ou and D. Yu, "Fault diagnosis of roller bearings based on Laplacian energy feature extraction of path graphs," *Measurement*, vol. 91, pp. 168–176, Sep. 2016.
- [21] J. Tang, "Mechanical system reliability analysis using a combination of graph theory and Boolean function," *Reliab. Eng. Syst. Saf.*, vol. 72, no. 1, pp. 21–30, Apr. 2001.
- [22] H. Ding, Z. Wei, and Y. Ye, "Large-Scale Spectral Graph Neural Networks via Laplacian Sparsification," in *Proc. 31st ACM SIGKDD Conf. Knowl. Discov. Data Mining*

<https://doi.org/10.37965/jdmd.2026.1722>

(KDD), Toronto, ON, Canada, Aug. 2025, pp. 1–11.

[23] M. Ma, K. Wu, X. Gao, X. Teng, and Y. Wang, “AutoSGNN: Automatic Propagation Mechanism Discovery for Spectral Graph Neural Networks,” in Proc. 39th AAAI Conf. Artif. Intell. (AAAI), Philadelphia, PA, USA, 2025, pp. 1–9.