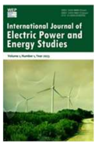
International Journal of Electric Power and Energy Studies
ISSN: 3006-2004 (Print), ISSN: 3006-0826 (Online) | Volume 5, Number 2, Year 2026
DOI: https://doi.org/10.62051/ijepes.v5n2.04
Journal homepage: https://wepub.org/index.php/IJEPES/index



Fault Evolution Trajectory Reconstruction and Manifold Interpolation Method based on Parallel Transport Mechanism

Yuanhong Liu, Qi Li, Jieqi Chen

College of Electrical Information Engineering, Northeast Petroleum University, Daqing 163318, Heilongjiang, China

ABSTRACT

In industrial intelligent operation and maintenance, the bearing fault signals, after dimensionality reduction through manifold learning, tend to exhibit discontinuous "fragmented" characteristics, making it difficult to depict the continuous evolution pattern of the fault. Moreover, the absence of intermediate state samples can easily lead to diagnostic errors. Therefore, this paper proposes a fault evolution trajectory reconstruction and manifold interpolation method based on parallel transport mechanism (TR-MIPT). This method utilizes local principal component analysis to construct the manifold topology, and generates pseudo-time degradation sequences based on geodesic metric; through tangent space estimation, orthogonal Procrustes alignment and discrete Riemannian connection, it realizes the distortion-free transmission of evolution gradients across coordinate systems; and combines local normal curvature and second-order geometric retract mapping to generate virtual samples and continuous evolution trajectories. Experimental results on the datasets from Case Western Reserve University and Northeast Petroleum University show that TR-MIPT can effectively reconstruct the degradation path, reduce the reconstruction error of intermediate states, and improve the diagnostic accuracy for small sample sizes and unseen intermediate-state faults.

KEYWORDS

Fault Evolution Trajectory; Manifold Learning; Parallel Transport; Local Principal Component Analysis; Discrete Riemannian Connection; Fault Diagnosis.

1. INTRODUCTION

In the intelligent operation and maintenance of industrial equipment, the vibration signals of rolling bearings usually exhibit high-dimensional and nonlinear characteristics, and due to the limitations of data acquisition strategies, they often exist in a discontinuous "fragmented" form. This makes traditional static feature analysis methods difficult to accurately quantify the true severity of high-dimensional feature spaces, and when intermediate state data is missing, it is prone to generate large diagnostic errors, unable to reveal the continuous dynamic laws of fault progressive evolution[1].

To address the issues of missing intermediate states and small sample sizes, researchers have proposed various interpolation and completion methods: Linear interpolation fills the missing points through simple weighting, with small computational cost but presenting a stepped trajectory, which is difficult to reflect the continuous degradation process[2,3]. Cubic spline interpolation can construct a smooth curve to ensure the continuity of the first and second derivatives, but it may still deviate from the true trajectory when the sample spacing is uneven or the curvature changes significantly[4,5]; Weighted K-neighborhood interpolation or Bayesian multiple interpolation methods enhance the robustness of small sample prediction[6]. Although these methods can complete discrete data, they still have difficulty guaranteeing continuity and geometric consistency in high-dimensional spaces. In terms of mathematical theory, industrial process signals are mapped to the tangent space, and by extracting main features through PCA or dimensionality reduction techniques, high dimensional data modeling can be achieved[7,8]; At the same time, extending PCA to manifold principal curves and combining with Riemannian metrics can efficiently extract and classify high-dimensional sensor or industrial data features[9,10].

In response to the above problems, this paper proposes a fault evolution trajectory reconstruction and manifold interpolation method based on the parallel movement mechanism (TR-MIPT). This method achieves smooth transmission of cross-coordinate system evolution gradients through local tangent vector estimation and orthogonal alignment, and combines local normal curvature and second-order geometric contraction mapping to generate high-fidelity virtual samples, thereby realizing continuous trajectory reconstruction under discrete small sample conditions. Experiments show that TR-MIPT outperforms traditional linear interpolation and cubic spline methods in fault evolution trajectory visualization, severity quantification, and small sample classification tasks, and can achieve high-precision identification of intermediate state faults.

2. IMPLEMENTATION OF THE TR-MIPT ALGORITHM

To address the problem of missing evolutionary information caused by data discontinuity, this paper proposes a fault evolution trajectory reconstruction and manifold interpolation method (TR-MIPT) based on the manifold parallel movement mechanism. As shown in Fig. 1, this algorithm mainly consists of three core steps: degradation sequence construction, evolution vector field estimation, and second-order geometric trajectory generation.

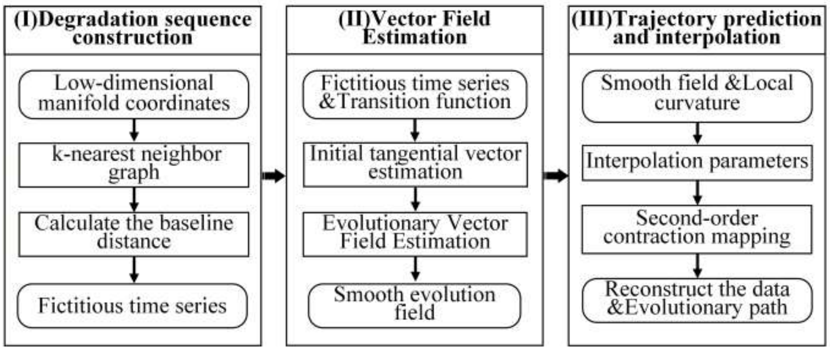


Fig. 1 TR-MIPT Algorithm Framework

2.1. Construction of Degraded Sequences based on Geodesic Distance

When dealing with fault data that exhibits highly nonlinear manifold structures, the traditional Euclidean distance metric is equivalent to seeking shortcuts outside the manifold, which cannot truly reflect the intrinsic dynamic process of equipment performance degradation. Therefore, this paper uses geodesics on the manifold surface to replace the Euclidean distance to define the pseudo-time severity of faults.

Given a sample set of discrete states in a high-dimensional feature space $Q=\{q_0,q_1,\ldots,q_N\}$, q_0 represents the health reference center.Firstly, using the Euclidean distance, find the k nearest neighbors for each sample, construct a local connectivity graph and define the weight matrix:

$$W_{ij}=\begin{cases}\| q_i-q_j\|_2, & j\in KNN(i)\\ \infty, & j\notin KNN(i)\end{cases} \tag{1}$$

In the formula: W_{ij} represents the edge weight between nodes; $\| q_i - q_j \|_2$ represents the Euclidean distance between two points in the original feature space; $KNN(i)$ represents the set of k nearest neighbors of sample point q_i;

Subsequently, based on the Dijkstra algorithm, the shortest path length from the health center q_0 to any sample point q_i in the graph is approximated for calculation, and this is used as the geodesic severity index $S(q_i)$.

$$S(q_i) \approx \min_{\text{path}} \sum_{a,b \in \text{path}} W_{ab} \tag{2}$$

In the formula: path represents the set of valid paths in a connected graph from the reference point q_0 to the target point q_i.

After the calculation is completed, all discrete samples will be re-ordered according to the principle of monotonic increase of $S(q_i)$, thereby obtaining a pseudo-time degradation sequence $Q = \{q_0, q_1, \dots, q_N\}$ with a monotonic severity order.

2.2. Local Tangent Space Estimation and Parallel Translation Alignment

In discrete sequences, the directly calculated difference vectors are constrained by the sampling interval and noise interference, often deviating from the manifold surface. Therefore, they need to be projected onto the local tangent space and achieve geometric alignment across time and space through parallel movement. As shown in Fig. 2.

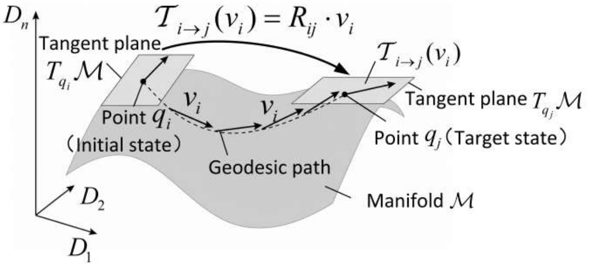


Fig. 2 Illustration of Evolution Vector Field Based on Parallel Transport

Extract samples within the local neighborhood of the state point q_i and construct a covariance matrix from them, then perform singular value decomposition on it:

$$C_i = U_i \Lambda_i U_i^T \tag{3}$$

In the formula: C_i represents the covariance matrix; $U_i \in \mathbb{R}^{D \times d}$ is the orthogonal basis matrix composed of the first d principal components; Λ_i is the corresponding singular value diagonal matrix; D represents the dimension of the original high-dimensional feature space; d is the intrinsic dimension of the low-dimensional manifold.

The intrinsic dimension is determined by the cumulative contribution rate of the singular values of the local neighborhood covariance matrix. Considering that the singular value inflection points may not be obvious under noisy conditions, a maximum likelihood estimation method is further employed to conduct cross-validation on d. When the results of both methods are consistent, the one obtained is taken as the final intrinsic dimension; in this paper's experiments, d is set to 3.

The initial discrete tangent coordinate $\tilde{v}_i \in \mathbb{R}^d$ for the evolution from the current state to the next state can be approximately obtained by projecting onto the local tangent plane:

$$\tilde{v}_i = U_i^T (q_{i+1} - q_i) \tag{4}$$

Due to the influence of the manifold curvature, the tangent spaces at different points are not parallel to each other. To transfer the evolution trend at point q_i without distortion to the tangent space at the target point q_j, this paper uses orthogonal Pohl alignment to solve the optimal discrete Riemann connection matrix $R_{ij} \in \mathbb{R}^{d \times d}$. Firstly, the SVD decomposition is performed on the overlapping matrix of the spatial basis:

$$U_j^T U_i = A \Sigma B^T \tag{5}$$

Obtained the optimal parallel translation and rotation matrix:

$$R_{ij} = AB^T \tag{6}$$

Thus, the parallel transport operator $\mathcal{T}_{i \to j}$ is defined on the Riemannian manifold, which maps the source tangent vector to the target tangent space without distortion.

$$\mathcal{T}_{i \to j}(\tilde{v}_i) = U_j R_{ij} \tilde{v}_i \tag{7}$$

To suppress high-frequency vibration noise while retaining the macroscopic degradation trend, a local sliding window is constructed to perform Gaussian-weighted smoothing on the tangential vectors after parallel movement:

$$\bar{v}_j = \sum_{i \in \Omega_j} w_{ij} \mathcal{T}_{i \to j}(\tilde{v}_i) \tag{8}$$

Among them, the smoothness weight w_{ij} is determined by the pseudo time span:

$$w_{ij} = \frac{1}{Z} \exp\left(-\frac{(S(q_i) - S(q_j))^2}{2\sigma^2} \right) \tag{9}$$

In the formula: Ω_j represents the set of sliding window indices centered on the target interpolation point q_j ; w_{ij} is the Gaussian kernel weight based on the pseudo-time distance; Z is the normalization factor to ensure that the sum of the weights is 1; σ is the Gaussian kernel width parameter that controls the rate of time decay.

2.3. The Manifold Interpolation Method based on Second-order Geometric Correction

After obtaining a smooth tangential evolution field, if only first-order linear interpolation is performed on the discrete points, the generated virtual samples will still deviate from the curved manifold surface. This paper employs a second-order retract mapping based on Taylor expansion to approximate the exponential mapping on the Riemannian manifold, and reconstructs the trajectory within the missing data domain.

According to the theory of differential geometry, the normal acceleration of a curve on a manifold is entirely determined by the degree of curvature of the manifold surface. Based on the aforementioned local tangent space, the projection operator P_{N_j} of the normal space at the state point q_j can be obtained by removing the tangential components through the identity matrix:

$$P_{N_j} = I - U_j U_j^T \tag{10}$$

In the formula: P_{N_j} is the linear operator for projecting onto the local normal space; $I \in \mathbb{R}^{D\times D}$ is the high-dimensional identity matrix; $U_j U_j^T$ is the projection operator of the local tangent space. Using the discrete central difference to approximate the local principal curvature vector κ_j at this point:

$$\kappa_j \approx P_{N_j}\left(q_{j-1} - 2q_j + q_{j+1}\right) \tag{11}$$

Ultimately, by combining the first-order smooth tangent vector field $\overline{v}_j$ with the second-order local normal curvature κ_j, a high-precision manifold trajectory interpolation formula with step size $\delta \in [0,1]$ between adjacent discrete states is derived:

$$\hat{q}(\delta) = q_j + \delta \cdot \overline{v}_j + \frac{1}{2}\delta^2 \kappa_j \tag{12}$$

3. EXPERIMENTAL SETUP AND RESULT ANALYSIS

3.1. Dataset Introduction

The experiment utilized the bearing benchmark dataset from Case Western Reserve University (CWRU) in the United States. Data were selected for three fault modes: inner ring, outer ring, and rolling element, with damage diameters of 0.178mm (early stage), 0.356mm (mid-stage), and 0.533mm (late stage), respectively. Combined with the healthy benchmark state, a multi-level damage evolution sequence was constructed. The data set used in this paper: the load was 0 HP, the rotational speed was 1797 r/min, and the sampling frequency was 12 kHz. To consider the algorithm's generalization ability across different data sets, this paper also adopted the bearing data set from Northeast Petroleum University (NEPU). The working conditions were set as follows: the sampling frequency was 10 kHz, the motor rotational speed was 1400 r/min, the load was 1 HP, and the damage diameter was 0.178mm.

29-dimensional statistical features are extracted from each vibration signal. Among them, 17 time-domain indicators include mean, standard deviation, root mean square, root mean square of mean, skewness, kurtosis, peak factor, margin factor, waveform factor, pulse factor, maximum value, minimum value, peak-to-peak value, absolute mean, absolute peak factor, second-order center distance, and energy; 12 frequency-domain indicators include mean of spectral amplitude, standard deviation of spectral amplitude, skewness of spectral amplitude, kurtosis of spectral amplitude, spectral frequency center of gravity, standard deviation of spectral frequency, root mean square of spectral frequency, 4/2 moment ratio of spectral frequency, spectral frequency energy ratio, spectral coefficient of variation, skewness of spectral frequency, and kurtosis of spectral frequency. For each type of fault in CWRU, each damage size contains 100 samples; while for each type of fault in NEPU, there is only one damage size and it contains 100 samples.

3.2. Fault Severity Quantitative Evaluation Experiment

This experiment aims to evaluate the temporal reliability of pseudo-time degraded sequences. Based on the geometric center of the healthy state data, the Euclidean distance and the manifold geodesic distance of the sample points in the high-dimensional space are calculated for different evolution stages of the three fault modes, and the fault severity index $S(y)$ corresponding to these distances is determined. The monotonicity and stage discrimination ability of the two metrics as the damage increases are qualitatively analyzed.

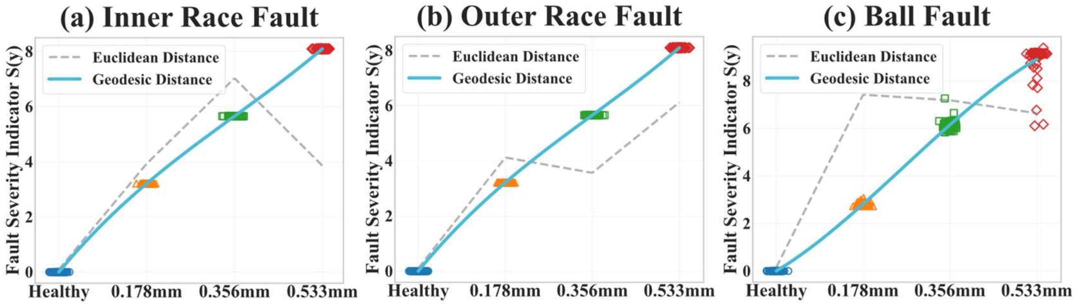


Fig. 3 Boxplots of Different Fault Severity Indicators

As shown in Fig. 3, the traditional Euclidean distance (gray dotted line) is insufficiently stable in the nonlinear manifold: as the degree of damage increases, some severely faulty samples are actually closer to the healthy center, indicating that the Euclidean distance is difficult to accurately represent the degradation distance along the manifold. In contrast, the severity index based on the geodesic line of the manifold (blue solid line) generally shows a monotonically increasing trend, and the box plot distribution is more compact, which can provide physically consistent temporal coordinates for trajectory interpolation. There is a certain overlap in the severity of samples in the late stage of rolling element faults, which may be related to rotation, revolution, and signal attenuation in the non-bearing area, rather than measurement failure.

3.3. Visualization Experiment of Fault Evolution Trajectory

To evaluate the ability of TR-MIPT in filling data gaps and restoring the evolution of consecutive failures, this experiment reconstructs the trajectory in the reduced-dimensional three-dimensional embedding space. Starting from the early centroid and ending at the late centroid, a path is generated and compared with linear and cubic spline interpolation. The mid-term samples that were not involved in the generation serve as references. The algorithm for evaluating the retention ability of the manifold geometry and the evolution path through the trajectory traversing the intermediate state and the smoothness of the tangential vector is used

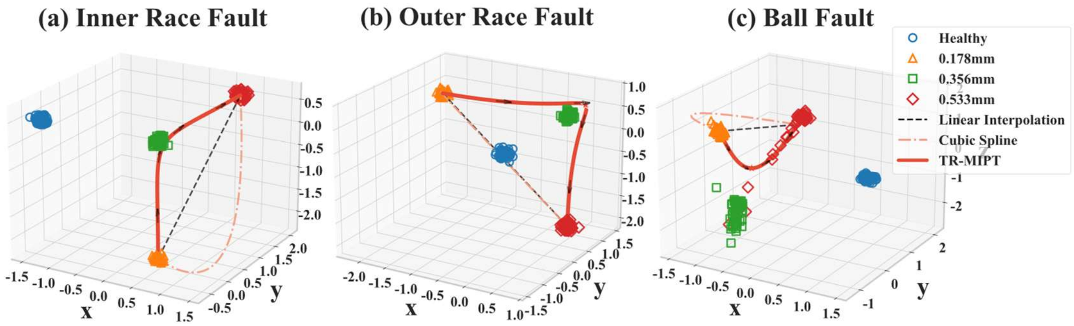


Fig. 4 Visualization of Trajectory Reconstruction Using Different Interpolation Methods

Fig. 4 shows that linear interpolation fails to take into account geometric constraints and thus the trajectory crosses the manifold gap, making it difficult to effectively cover the intermediate state; while cubic spline interpolation is smooth, it deviates more severely in areas with significant curvature changes. TR-MIPT achieves local coordinate smooth transition through the parallel movement mechanism, and reconstructs the path to be more in line with the actual degradation trend.

To quantitatively evaluate the reconstruction accuracy, in this experiment, the mid-term operating conditions were used as the test set. The predicted intermediate state points generated by the three algorithms were compared with the true centroids in each dimension, and an embedding space reconstruction error index was introduced. Let the centroid position vector of the true sample cluster in the embedding space be P_{true}, and the position vector of the corresponding predicted point generated by the algorithm be P_{pred}. The reconstruction error is defined as:

$$E_{recon} = \left\| P_{pred} - P_{true} \right\|_2 \tag{13}$$

Table 1 presents a detailed comparison of the reconstruction errors between the intermediate state coordinates generated by linear interpolation, cubic spline interpolation and TR-MIPT, and the true values under three fault modes.

Table 1. Comparison of Intermediate-State Reconstruction Errors for Different Algorithms

Algorithm\Reconstruction Error	Inner ring failure	Outer ring failure	Rolling element failure
Linear interpolation	1.7257	1.7302	2.4767
Tricubic spline interpolation	2.6479	2.6293	3.0123
TR-MIPT	0.2029	0.3676	1.1913

As shown in Table 1, different fault modes exhibit significant nonlinear geometric structures in the low-dimensional manifold space, resulting in large deviations from linear interpolation based on the Euclidean assumption. In contrast, TR-MIPT effectively captures the intrinsic geometric laws of the manifold through the mechanism of parallel movement of tangent vectors, reconstructs with lower errors, and achieves effective reconstruction of the fault evolution trajectory. Additionally, the rolling element fault is affected by self-rotation, revolution, and load zone switching, making the dynamics more complex; after mapping to the low-dimensional manifold, local curvature fluctuations and changes in adjacent tangent space directions are more obvious, increasing the approximate error of parallel movement, and thus the intermediate state reconstruction error is higher than that of inner and outer ring faults.

To further verify the cross-dataset generalization ability of TR-MIPT, this paper conducts trajectory reconstruction experiments on the NEPU bearing dataset. Considering that this dataset only contains healthy state and fixed damage state, and lacks real intermediate state samples, this paper selects three state pairs, namely "healthy - inner ring failure", "healthy - rolling element failure", and "healthy - outer ring failure", for trajectory reconstruction, and uses linear interpolation and cubic spline interpolation as comparisons. Since it is impossible to directly calculate the reconstruction error of the intermediate state, further the angle error between the tangential vector at the trajectory endpoints and the main direction of the real sample cluster is adopted to evaluate the rationality of the boundary of the reconstructed trajectory.

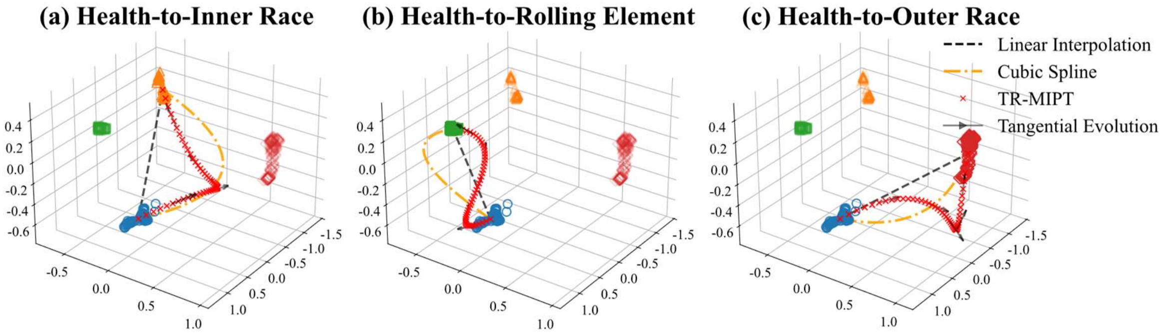


Fig. 5 Visualization of Trajectory Reconstruction

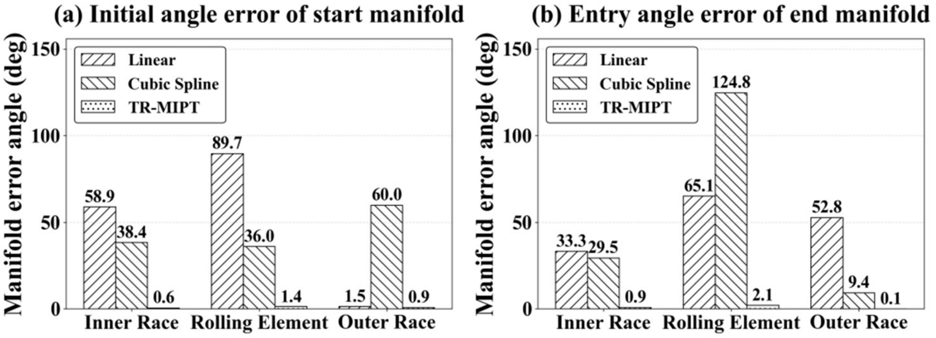


Fig. 6 Comparison of Manifold Error Angles

As shown in Figs 5 and 6, the trajectory direction deviations of linear interpolation and cubic spline interpolation are relatively large; in contrast, the direction errors of TR-MIPT at both the starting and ending points are controlled within 2.1°. This indicates that this method can still maintain the consistency of the fault evolution direction even when there are no real intermediate state observations.

3.4. Small Sample Classification Accuracy Experiment

This experiment aims to quantitatively evaluate the classification gain of the virtual samples generated by TR-MIPT on the CWRU dataset in the scenario of intermediate state absence. Using the KNN classifier, the training set consists of early and late real faults, and the test set is composed of 25 mid-term real samples of each type. Through the stepwise enhancement strategy, 15 virtual mid-term samples of each type are generated along the manifold trajectory, and different mid-term virtual samples of different fault modes are successively added to observe the impact of the expansion of the training set on the classification accuracy and confusion matrix of the test set. Among them, A, B, and C respectively represent the inner race, the outer race and the rolling element.

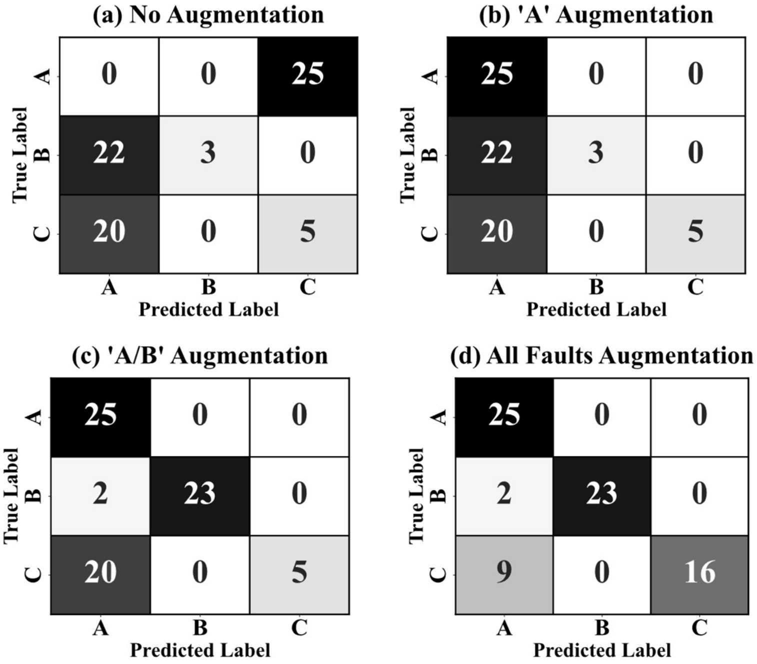


Fig. 7 Matrix for Small-Sample Fault Classification

As shown in Fig. 7, without the enhanced baseline, the classifier misclassifies the intermediate state samples, and its generalization ability is insufficient in the absence of intermediate state information. After introducing the virtual intermediate state samples generated by TR-MIPT, the classifier gradually learns the intermediate state features and corrects the initial misjudgment; even in complex areas such as rolling element faults, it still effectively guides the decision boundary. The results show that the virtual data generated by TR-MIPT can effectively improve the model's ability to recognize samples with missing intermediate states.

4. CONCLUSION

This paper addresses the issues of discontinuous, fragmented distribution of industrial bearing monitoring data and the absence of intermediate states, which lead to an increase in diagnostic errors. To solve these problems, the TR-MIPT algorithm is proposed. This method estimates the manifold tangent space structure based on local PCA, constructs a pseudo-time degradation sequence using geodesic metrics, and realizes the geometric consistent transmission of local evolution gradients through orthogonal Procrustes alignment and parallel moving operators. On this basis, local normal curvature and second-order retract mapping are combined to generate intermediate state virtual samples and continuous evolution trajectories. Experimental results show that this method can effectively restore the fault degradation trajectory, reduce the intermediate state reconstruction error, and improve the fault diagnosis accuracy under small sample conditions. It should be noted that the performance of TR-MIPT still depends on the identifiability of the local manifold structure; in low signal-to-noise ratio conditions, impact features may be masked by background noise, resulting in decreased stability of tangent space estimation, parallel moving alignment, and curvature estimation. Further generalization evaluations under cross operating conditions and low signal-to-noise ratio conditions will be conducted by combining more public datasets and complex industrial field data in the future.

REFERENCES

[1] Chen S, Zheng X. A bearing fault diagnosis method with improved symplectic geometry mode decomposition and feature selection[J]. Measurement Science and Technology, 2024, 35(4): 046111.

[2] Lv Y, Han Q, Xue S. Data anomaly repair method based on fuzzy voting and multi-segment interpolation[J]. Scientific Reports, 2025, 15: 10234.

[3] Chen, X.-L.; Wang, R.-X.; Wang, J.; Zhou, J.-L. Industrial process monitoring and fault diagnosis based on hybrid discriminant analysis. Acta Automatica Sinica, 2020, 46(8): 1600–1613.

[4] Kim SW, Kim YI. A data imputation approach for missing power consumption measurements in water-cooled centrifugal chillers[J]. Energies, 2025, 18(11): 2779.

[5] Li, N.; Ding, H.; Sun, X.-C.; Liu, Z.-P.; Pu, G.-S. Intelligent fault diagnosis of shearer based on simplified interval kernel global-local feature fusion. Journal of China Coal Society, 2024, 49(2): 452–463.

[6] Wu, B.-L.; Qi, X.-L.; Wang, Z.-Y.; Ye, X.-D.; Zheng, J.-D. Rolling bearing fault diagnosis based on improved semi-supervised LTSA and BA-SVM. Bearing, 2020, 38(5): 23–31.

[7] Smith A, Laubach B, Castillo I, Zavala VM. Data analysis using Riemannian geometry and applications to chemical engineering[J]. Computers & Chemical Engineering, 2022.

[8] Gao W, Ma Z, Gan W, Liu S. Dimensionality reduction of SPD data based on Riemannian manifold tangent spaces and isometry[J]. Entropy, 2021, 23(9): 1117.

[9] Xu J, Grosse-Wentrup M. Tangent space spatial filters for interpretable and efficient Riemannian classification[J]. Journal of Neural Engineering, 2020, 17(4): 046017.

[10] Li J, He D, Wei Z, et al. CEEMDAN and adaptive distance embedding for fault diagnosis of train bogie bearing[J]. Measurement Science and Technology, 2025, 36(3): 036127.