Research Article

New Multifeature Information Health Index (MIHI) Based on a Quasi-Orthogonal Sparse Algorithm for Bearing Degradation Monitoring

Xiao Zhang [1], **Tengyi Peng** [2], **Shilong Sun** [2], **and Yu Zhou** [3]

[1]*College of Computer Science, South-Central University for Nationalities, Wuhan 430074, China*
[2]*School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen 518055, China*
[3]*College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China*

Correspondence should be addressed to Shilong Sun; shilosun-c@my.cityu.edu.hk

Data-driven intelligent prognostic health management (PHM) systems have been widely investigated in the area of defective bearing signals. These systems can provide precise information on condition monitoring and diagnosis. However, existing PHM systems cannot identify the accurate degradation trend and the current fault types simultaneously. Given that different fault types have various effects on the mechanical system, the corresponding maintenance strategies also vary. Then, choosing the appropriate maintenance strategy according to the future fault type can reduce the maintenance cost of the equipment operation. Therefore, a multifeature information health index (MIHI) must be developed to trace various bearing degradation trends with various types of faults simultaneously. This paper reports a new quasi-orthogonal sparse project algorithm that can mutually convert the degraded processing feature vector sets (such as spectrum) for each type of fault to orthogonal approximate spatial straight lines. The algorithm builds a MIHI through the spectrum of current state measured points. The MIHI is then transformed by a quasi-orthogonal sparse project algorithm to trace the various bearing degradation trends and recognize the fault type simultaneously. The case study of bearing degradation data demonstrates that this approach is effective in assessing the various degradation trends of different fault types.

1. Introduction

With the progress of science and technology, building an effective intelligent prognostic health management (PHM) system has become necessary for some critical components. Engineers deploy various types of sensors to detect the health conditions of a single component. However, how to deal with sensor information effectively and thus assist the intelligent health management of equipment are yet to be extensively researched. Remaining useful life (RUL) estimation is one of the key factors in asset condition-based maintenance, prognostics, and health management [1]. The basic principle of RUL estimation is how to construct the health index [2]. An appropriate health index can effectively improve the accuracy of RUL prediction. To detect the incipient faults and offer effective information to the RUL

estimation and prognostics model, the health index plays an important role as a bridge connecting the sensor measured signal with the assets' health condition prognostics. Some statistical indicators, such as root mean square (RMS) and the kurtosis of the measured vibration signal [3], are used to indicate the assets' conditions for RUL estimation and prognostics. Moreover, some studies utilize time-frequency features, such as waveform entropy [4]. Many dimensionality reduction methods have been favored by researchers, for example, kernel principal component analysis is used in [5] and isometric feature mapping is used in [6]. In recent studies, researchers have developed some effective statistical indicators, i.e., using multiscale fuzzy entropy [7], Kullback–Leibler divergence combined with Gaussian process regression [8], and a neural network to conduct a health index to monitor the equipment condition directly [9, 10].

The above health index building methods are used in prognostics, and they mostly map the information from various sensor signals or feature extractors into an indicator full of information. This process is called data fusion, which has three categories: feature-level fusion, decision-level fusion, and data-level fusion [11]. Feature-level fusion methods rely on prior knowledge of the degradation mechanism and physical models to analyse the input data. Goebel [12] utilized a feature-level fusion model that consisted of principal component analysis, filtering, smoothing, normalization, and log-transformation methods to feed information to the RUL estimation model and predict the breakage of paper webs in a paper-making machine at the wet end. Ma [13] reported a multiple-view feature fusion to predict lithium-ion battery RUL. Decision-level methods fuse high-level decisions made based on individual sensor data and do not rely on raw signal feature extraction. Niu et al. [14] combined the advantages of wavelet analysis on transient signals and decision-level fusion techniques to improve the accuracy of fault diagnosis. Wei [15] proposed a decision-level data fusion method to map the individual sensor signal into reliable data to improve the ability of the quality control system in additive manufacturing and RUL estimation of aircraft engines.

In contrast to feature- and decision-level fusion methods that focus on the physical meaning, data-level fusion methods pay attention to mining the embedding feature suitable to the task from the raw data. In RUL prediction and condition monitoring fields, data-level fusion methods always build the health index based on some properties that the researchers hope can improve the RUL estimation ability. Data-level data fusion methods are suitable for dealing with complex system situations because these systems hardly build an effective model to fuse various signals, but data-level can monitor the machinery system state according to the requirements of monitoring task, which has a stronger versatility. In summary, the data-level fusion method is concentrated on the property of required tasks. Some scholars have focused on investigations of properties on the health index. Two properties of the health index are proposed by Liu [16]. First, it must have good monotonicity; second, the variance of the failure threshold of multiple experiments must be minimized. Chehade [17] proposed another proposition: separability. They thought the greater the difference of the health index between two observations, the more validated and reliable the health index. In [2], the development of the health index focused on monotonicity and separability, and it converted the transform matrix solving the problem to a convex optimization process. Yan [18] focused on locating informative frequency bands, determining the optimal fault frequency band for health indicating. Some researchers combined the health index building with statistical analysis. For example, Kim [19] proposed a linear multisensor information fusion method to build a health index and derive the best linear unbiased estimator of the fusion coefficients. At the same time, some researchers studied the methods based on the Kalman filter to fuse the sensor signals (e.g., Markov model [20]). To deal with the sensor selection problem from the multiple sensors situation, Liu [21] studied the perspective of the quality of the signal provided by sensors and proposed a signal-to-noise ratiometric method to combine various sensor signals, develop a health index, and monitor asset degradation. However, from the current study of bearing health condition monitoring, the sparsity used in building a health index is rarely considered. The lack of sparse terms leads to overfitting and reduces the generalization ability of the health index. In data-driven model building, the sparse penalty term is also used to make models have additional reusability and avoid overfitting. Sparse models built via lasso (L1-normalization) can achieve interpretable feature selection [22], and lasso has been widely investigated in the fields of biology and medicine. For some features with adjacent relevance, the fused lasso is proposed to make feature weights sparse and smooth [23].

Given that the high-dimensional feature is difficult to calculate and observe, some scholars use the health index to monitor equipment's operating condition intuitively and conduct prognostic health management. The degradation trend of different types of faults varies. Meanwhile, information on the current faults' evolution direction can also further improve the accuracy of RUL estimation. Therefore, this study aims to provide a method for constructing a health index that can indicate various fault type degradation trends from the fusion of the features at the same time. The contributions of this study can be summarized as follows:

(1) This work proposes an orthogonal proposition of developing a health index. This proposition focuses on expanding the discrepancy of different fault type degradations, which is the basic idea of the multi-feature information health index (MIHI).

(2) The MIHI uses a quasi-orthogonal sparse projection algorithm to convert the spectral features of the current measurement point into a low-dimensional vector. This low-dimensional vector can simultaneously represent the bearing degradation trend of multiple types of faults.

(3) The weight sparse and the weight difference sparse are added to MIHI to build an objection function and improve the general applicability.

(4) The optimization problem expression of the proposed quasi-orthogonal sparse project algorithm is a nonconvex function with constraint. The fast-solving algorithm based on iteration is given.

2. Proposed Quasi-Orthogonal Sparse Project Algorithm

In this section, a quasi-orthogonal sparse project algorithm is developed to transfer the spectrum feature to a low-dimensions' vector health index for bearing degradation monitoring. To efficiently trace various fault type degradation trends, the project algorithm should satisfy the five properties:

Property 1. Monotonicity: once an initial fault occurs, the trend of the degradation signals should be monotonic [16].

Property 2. Sensitivity: the health index should be sensitive to bearing components that generate abnormal defects, that is, the health index can separate the normal and abnormal health conditions of bearings [24].

Property 3. Orthogonality: to indicate the various types of bearing fault effectively, the health index is built as a vector, and the size of the health index vector should be the same as the number of the fault types. To avoid confusion of the fault types in the health index, each dimension of the health index should be orthogonal.

Property 4. Weight sparsity: to prevent overfitting, the weight of the features with low correlation to Properties 1–3 should be small. We also hope that the weights of these features can be set as 0 to achieve feature selection.

Property 5. Weight difference sparsity: under rotating machinery operating conditions with load, working frequency has some small fluctuations (especially, bearing ball pass frequency). The weights' difference sparsity helps the projection matrix to become more flexible and deal with the fluctuating frequencies. The principle of MIHI is extracting the spectrum characteristics, which can track the degradation trend and distinguish different fault types from degradation datasets of various faults. Given that the bearing fault impulse signal is a pseudocyclization signal (the period of impulse random fluctuates around a mean value), the characteristic frequency of the fault fluctuates in a small interval. The fused lasso sparse term makes HIMI's weight matrix focus on the frequency band around the mean characteristic frequency rather than the signal characteristic frequency. Hence, this term can prevent overfitting.

In summary, the quasi-orthogonal sparse project algorithm for building a sparse multi-information feature health index is expected to detect bearing fault types and evaluate the bearing degradation process monotonically and sensitively at the same time. The proposed quasi-orthogonal project algorithm is based on the traditional linear fitting method, and the equation is defined as follows:

$$\mathbf{HI}\mathrm{MI}_i = \mathbf{W} \times \mathbf{f}_i, \tag{1}$$

where $\mathbf{W} \in \mathbb{R}^{k \times m}$ is the projection matrix, m denotes the size of the spectrum feature $\mathbf{f}_i$ of the ith observation, k is the fault type number, and $\mathbf{HI}\mathrm{MI}_i \in \mathbb{R}^{k \times 1}$. Figure 1 shows the example of equation (1).

To deal with various fault type monitoring tasks, the proposed algorithm needs different types of fault degradation datasets as prior knowledge. We denote the jth fault type degradation data as $\mathbf{F}_j \in \mathbb{R}^{m \times n_j}$, where n_j denotes the number of observation epochs of the jth fault type degradation. We also denote the ith observation epoch feature vector of the jth fault type degradation data as $\mathbf{f}_{j,i}$.

In addition, the projection matrix can be denoted as $\mathbf{W} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \ldots, \mathbf{w}_k^T]^T$, and the weights in this projection matrix $\mathbf{W}$ need to be calculated row by row. The solving process of the jth row of the projection matrix $\mathbf{w}_j$ can be described as an optimization problem, and the formula of this optimization is provided as follows:

$$\underset{\mathbf{w}_j}{\text{maximize}} \left\{ D^2_{\text{sensitive}}(\mathbf{w}_j) - D^2_{\text{orthogonality}}(\mathbf{w}_j) - \alpha D^2_{\text{monotonousness}}(\mathbf{w}_j) - \beta \sum_{q=1}^{m} |w_{j,q}| - \gamma \sum_{q=2}^{m} |w_{j,q} - w_{j,q-1}| \right\}, \tag{2}$$

where α is defined to balance sensitivity and monotonicity (corresponding to Property 1 and 2), β is the sparsity penalty parameter (corresponding to Property 4), and γ is the weight difference sparsity penalty parameter (corresponding to Property 5). μ_j denotes the mean vector of the normal state observation matrix $\mathbf{N}_j = [\mathbf{f}_{j,1}, \mathbf{f}_{j,2}, \ldots, \mathbf{f}_{j,n_{\text{nor}}}]$, and n_{nor} is the number of the first few observation epochs (generally, the bearing signals can be assumed as normal signals in the first few observation epochs), and n_{nor} corresponds to Property 3.

Then, $D^2_{\text{sensitive}}(\mathbf{w}_j)$ can be calculated by

$$D^2_{\text{sensitive}}(\mathbf{w}_j) = \sum_{i=1}^{n_j} \left\| \mathbf{w}_j \times (\mathbf{f}_{j,i} - \mu_j) \right\|_2^2. \tag{3}$$

To ensure that the health index can sensitively monitor the abnormal condition, $D^2_{\text{sensitive}}(\mathbf{w}_j)$ needs to be as large as possible.

$D^2_{\text{monotonousness}}(\mathbf{w}_j)$ corresponds to Property 1. To reduce computational complexity, health index monotonicity relies on $D^2_{\text{sensitive}}(\mathbf{w}_j)$ and $D^2_{\text{monotonousness}}(\mathbf{w}_j)$. $D^2_{\text{monotonousness}}(\mathbf{w}_j)$ is used to evaluate the stability of the different health indexes of two adjacent observation epochs, and it can be calculated by

$$D^2_{\text{monotonousness}}(\mathbf{w}_j) = \sum_{i=2}^{n_j} \left\| \mathbf{w}_j \times (\mathbf{f}_{j,i} - \mathbf{f}_{j,i-1}) \right\|_2^2, \tag{4}$$

where $D^2_{\text{monotonousness}}(\mathbf{w}_j)$ can only assess the stability but not the monotonicity of the health index. However, enlarging $D^2_{\text{sensitive}}(\mathbf{w}_j)$ and shrinking $D^2_{\text{monotonousness}}(\mathbf{w}_j)$ simultaneously can effectively approximate the realization of monotonicity. To achieve orthogonality, we need to reduce the health index difference between the other fault type data and the current fault type health state. $D^2_{\text{orthogonality}}(\mathbf{w}_j)$ is calculated by
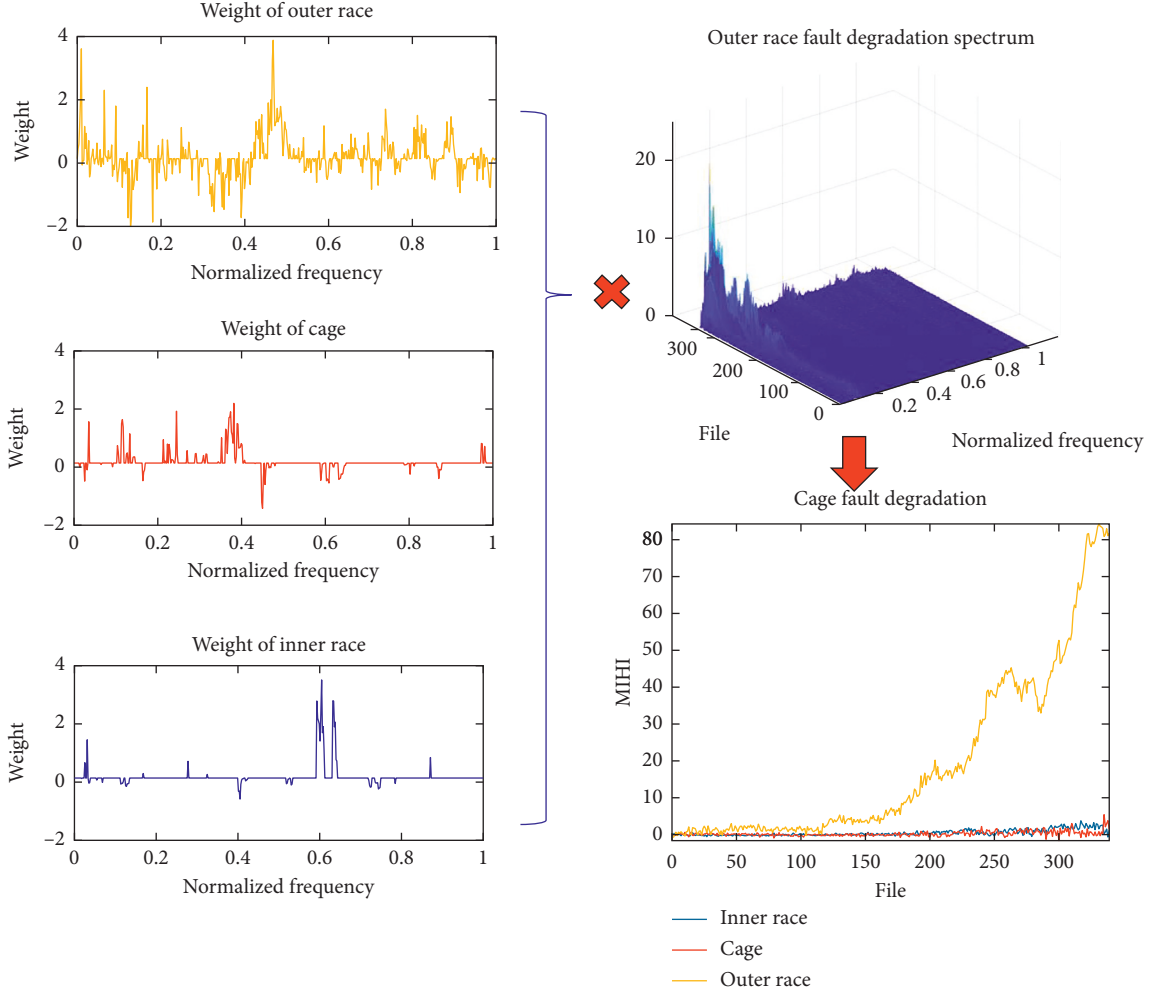
FIGURE 1: Illustrative example for building MIHI from a spectrum.

$$D^2_{\text{orthogonality}}\left(\mathbf{w}_j\right) = \sum_{v=1}^{j-1}\sum_{i=1}^{n_j}\left\|\mathbf{w}_j\times\left(\mathbf{f}_{v,i}-\boldsymbol{\mu}_j\right)\right\|_2^2 + \sum_{v=j+1}^{k}\sum_{i=1}^{n_j}\left\|\mathbf{w}_j\times\left(\mathbf{f}_{v,i}-\boldsymbol{\mu}_j\right)\right\|_2^2 + \sum_{i=1}^{n_{\text{nor}}}\left\|\mathbf{w}_j\times\left(\mathbf{f}_{j,i}-\boldsymbol{\mu}_j\right)\right\|_2^2, \tag{5}$$

where $D^2_{\text{orthogonality}}\left(\mathbf{w}_j\right)$ assesses the health index value difference of other fault types of degradation data with the jth fault type normal condition data. To realize orthogonality, $D^2_{\text{orthogonality}}\left(\mathbf{w}_j\right)$ needs to be as small as possible.

To reduce computation complexity, $D^2_{\text{sensitive}}\left(\mathbf{w}_j\right)$, $D^2_{\text{monotonousness}}\left(\mathbf{w}_j\right)$, and $D^2_{\text{orthogonality}}\left(\mathbf{w}_j\right)$ are simplified as follows:

$$D^2_{\text{sensitive}}\left(\mathbf{w}_j\right) = \mathbf{w}_j\times\left\{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\left(\mathbf{f}_{i,j}-\boldsymbol{\mu}_j\right)\times\left(\mathbf{f}_{i,j}-\boldsymbol{\mu}_j\right)^T\right]\right\}\times\mathbf{w}_j^T,$$
$$= \mathbf{w}_j\times\mathbf{S}^j_{\text{sensitive}}\times\mathbf{w}_j^T, \tag{6}$$

$$D^2_{\text{monotonousness}}\left(\mathbf{w}_j\right) = \mathbf{w}_j\times\left\{\sum_{j=2}^{n_j}\left[\left(\mathbf{f}_{j,i}-\mathbf{f}_{j,i-1}\right)\times\left(\mathbf{f}_{j,i}-\mathbf{f}_{j,i-1}\right)^T\right]\right\}\times\mathbf{w}_j^T,$$
$$= \mathbf{w}_j\times\mathbf{S}^j_{\text{monotonousness}}\times\mathbf{w}_j^T, \tag{7}$$

$$D_{\text{orthogonality}}^2(\mathbf{w}_j) = \mathbf{w}_j \times \left\{ \sum_{v=1}^{j-1}\sum_{i=1}^{n_j}\left[\left(\mathbf{f}_{v,i}-\boldsymbol{\mu}_j\right)\times\left(\mathbf{f}_{v,i}-\boldsymbol{\mu}_j\right)^T\right] + \sum_{v=j+1}^{k}\sum_{i=1}^{n_j}\left[\left(\mathbf{f}_{v,i}-\boldsymbol{\mu}_j\right)\times\left(\mathbf{f}_{v,i}-\boldsymbol{\mu}_j\right)^T\right] \right.$$
$$\left. + \sum_{i=1}^{n_{\text{nor}}}\left[\left(\mathbf{f}_{j,i}-\boldsymbol{\mu}_j\right)\times\left(\mathbf{f}_{j,i}-\boldsymbol{\mu}_j\right)^T\right] \right\} \times \mathbf{w}_j^T,$$
$$= \mathbf{w}_j \times \mathbf{S}_{\text{orthogonality}}^j \times \mathbf{w}_j^T. \tag{8}$$

Substituting equations (6)–(8) into equation (2), referring to the calculation from Fisher's discriminant ratio, and then rewriting equation (2), we obtain

$$\underset{\mathbf{w}_j}{\text{maximize}} \left\{ \mathbf{w}_j \mathbf{S}_{\text{sensitive}}^j \mathbf{w}_j^T - \beta_j \sum_{q=1}^{m}\left|\widehat{\sigma_{j,q}} w_{j,q}\right| - \gamma_j \sum_{q=2}^{m}\left|\widehat{\sigma_{j,q}} w_{j,q} - \widehat{\sigma_{j,q-1}} w_{j,q-1}\right| \right\},$$
$$\text{subject to } \mathbf{w}_j \mathbf{S}_{\text{subject}}^j \mathbf{w}_j^T \le 1, \tag{9}$$

where $\mathbf{S}_{\text{subject}}^j = \mathbf{S}_{\text{orthogonality}}^j + \alpha^2 \mathbf{S}_{\text{monotonousness}}^j$. $\widehat{\sigma_{j,q}}$ gives more penalty to the features that cause MIHI monotonicity and orthogonality showing fluctuations. $\widehat{\sigma}_j = (\widehat{\sigma_{j,1}}, \widehat{\sigma_{j,2}}, \ldots, \widehat{\sigma_{j,m}})$ is calculated by the following formula:

$$\widehat{\sigma_{j,q}} = \text{std}\left(\mathbf{O}_j(:,q)\right) + \alpha^2 \times \text{std}\left(\mathbf{P}_j(:,q)\right), \tag{10}$$

where $\mathbf{O}_j = [\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_{j-1}, \mathbf{N}_j, \mathbf{F}_{j+1}, \ldots, \mathbf{F}_k]$ and $\mathbf{P}_j = [\mathbf{f}_{j,2} - \mathbf{f}_{j,1}, \mathbf{f}_{j,2} - \mathbf{f}_{j,1}, \ldots, \mathbf{f}_{j,n_j} - \mathbf{f}_{j,n_j-1}]$. Moreover, in equation (9), $\beta_j = \beta\|\mathbf{S}_{\text{subject}}^{-1/2}\mathbf{S}_{\text{sensitive}}\mathbf{S}_{\text{subject}}^{-1/2}\|$ and $\gamma_j = \gamma\|\mathbf{S}_{\text{subject}}^{-1/2}\mathbf{S}_{\text{sensitive}}\mathbf{S}_{\text{subject}}^{-1/2}\|$, where $\|.\|$ indicates the largest eigenvalue.

Lastly, according to the health index of jth fault type value changing trend is positive or negative, multiplying 1 or -1 with $\mathbf{w}_j$ to ensure the health index has an increasing trend.

3. Solving Process

3.1. Weight Matrix Solving.
Figure 2 shows the solving flowchart of the proposed quasi-orthogonal sparse project algorithm. First, the historical data of various types of faults are reprocessed, transferring all observation signals to the spectrum feature $[\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_k]$. Second, the parameters α, β, γ, and n_{nor} are set, and each row of the weight matrix $\mathbf{W}$ is solved in turns.

3.2. Solution Detail of the Weight Vector.
In this section, the solution of equation (10) is provided. In general, equation (9) cannot be solved using tools from convex optimization. According to [25], we need to use a minimization-maximization algorithm to rewrite it. The first step is to construct an iterative from equation (9):

$$\underset{\mathbf{d}}{\text{minimize}} \left\{ \mathbf{d}\mathbf{S}_{\text{subject}}^j\mathbf{d}^T - 2\mathbf{d}\mathbf{S}_{\text{sensitive}}^j\left(\mathbf{w}_j^{(h)}\right)^T + \beta_j \sum_{q=1}^{m}\left|\widehat{\sigma_{j,q}} d_{j,q}\right| + \gamma_j \sum_{q=2}^{m}\left|\widehat{\sigma_{j,q}} d_{j,q} - \widehat{\sigma_{j,q-1}} d_{j,q-1}\right| \right\}. \tag{11}$$

From the result $\widehat{d}$ of equation (11), we can obtain $\mathbf{w}_j^{(h+1)} = \widehat{d}/\sqrt{\widehat{d}\mathbf{S}_{\text{subject}}^j\widehat{d}^T}$ (if $\widehat{d} = 0$, $\mathbf{w}^{(h)} = 0$).

The second step is to build a transfer matrix $\mathbf{R} \in \mathbb{R}^{(m-1)\times m}$:

$$\mathbf{R} = \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \widehat{\sigma_{j,1}} & 0 & \ldots & 0 \\ 0 & \widehat{\sigma_{j,2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \widehat{\sigma_{j,m}} \end{pmatrix}. \tag{12}$$

Substituting equation (12) into equation (11), we obtain

$$\underset{\mathbf{d}}{\text{minimize}} \left\{ \mathbf{d}\mathbf{S}_{\text{subject}}^j\mathbf{d}^T - 2\mathbf{d}\mathbf{S}_{\text{sensitive}}^j\left(\mathbf{w}^{(h)}\right)^T + \beta_j\left\|\widehat{\boldsymbol{\sigma}}_j \circ \mathbf{d}\right\|_1 + \gamma_j\|\mathbf{b}\|_1 \right\},$$
$$\text{subject to } \mathbf{R}\mathbf{d}^T = \mathbf{b}, \tag{13}$$

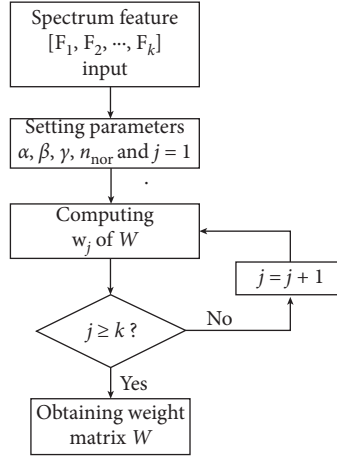where $\circ$ denotes Hadamard product. Then, the augmented Lagrange function of equation (13) is built as follows:

FIGURE 2: Solving weight matrix of the quasi-orthogonal project algorithm.

$$\underset{\mathbf{d}}{\text{minimize}}\left\{\left\|\mathbf{A}\mathbf{d}^T - \mathbf{c}\right\|_2^2 + \beta_j\left\|\widehat{\sigma}_j \circ \mathbf{d}\right\|_1 + \gamma_j\|\mathbf{b}\|_1 - \mathbf{e}\left(\mathbf{R}\mathbf{d}^T - \mathbf{b}\right) + \frac{\eta}{2}\left\|\mathbf{R}\mathbf{d}^T - \mathbf{b}\right\|_2^2\right\}, \tag{14}$$

where $\mathbf{A} = \mathbf{Q} \times \sqrt{\mathbf{G}}$, $\mathbf{Q}$ and $\mathbf{G}$ satisfy $\mathbf{S}_{\text{subject}}^j = \mathbf{Q}^T\mathbf{G}\mathbf{Q}$, $\mathbf{G}$ is a diagonal matrix, and $\mathbf{c} = \mathbf{w}_j^{(h-1)} \times \mathbf{S}_{\text{sensitive}}^j \times \text{inv}(\mathbf{A})$. By using the linearized alternating direction method [26], iterative augmented Lagrange function equation (14) can be rewritten as three subequations:

$$\begin{cases} \mathbf{d}^{(z+1)} = \underset{\mathbf{d}}{\text{minimize}}\left\{\beta_j\left\|\widehat{\sigma}_j \circ \mathbf{d}\right\|_1 + \frac{l}{2}\left\|\mathbf{d} - \mathbf{d}^{(z)} + \frac{\widehat{\mathbf{X}}^T\left(\widehat{\mathbf{X}}\left(\mathbf{d}^{(z)}\right)^T - \widehat{\mathbf{y}}^{(z)}\right)}{l}\right\|_2^2\right\}, \\[4mm] \mathbf{b}^{(z+1)} = \underset{\mathbf{b}}{\text{minimize}}\left\{\gamma_j\|\mathbf{b}\|_1 + \frac{\eta}{2}\left\|\mathbf{b} - \mathbf{R}\left(\mathbf{d}^{(z+1)}\right)^T + \frac{\mathbf{e}^{(z)}}{\eta}\right\|_2^2\right\}, \\[4mm] \mathbf{e}^{(z+1)} = \mathbf{e}^{(z)} - \eta\left(\mathbf{R}\left(\mathbf{d}^{(z+1)}\right)^T - \mathbf{b}^{(z+1)}\right), \end{cases} \tag{15}$$

where $\widehat{\mathbf{X}} = \left(\left(\sqrt{2}\,\mathbf{A}\right)^T, \left(\sqrt{\eta}\,\mathbf{R}\right)^T\right)^T$, $\widehat{\mathbf{y}}^{(w-1)} = \left(\left(\sqrt{2}\,\mathbf{c}\right)^T, \left(\sqrt{\eta}\,\left(\mathbf{b}^{(w-1)} + \mathbf{e}^{(w-1)}/\eta\right)\right)^T\right)^T$, and l is the approximation parameter. To ensure the convergence of equation (15), $l > \rho\left(\mathbf{A}^T\mathbf{A} + \eta\mathbf{R}^T\mathbf{R}\right)$. Using a soft-threshold algorithm can obtain the closed-form solution of equation (15):

$$\begin{cases} \mathbf{d}^{(z+1)} = \text{shrinkage}\left(\left(\mathbf{d}^{(z)} - \frac{\widehat{\mathbf{X}}^T\left(\widehat{\mathbf{X}}\left(\mathbf{d}^{(z)}\right)^T - \widehat{\mathbf{y}}^{(z)}\right)}{l}\right), \left(\left(\frac{\beta_j}{l}\right)\widehat{\sigma}_j\right)\right), \\[4mm] \mathbf{b}^{(z+1)} = \text{shrinkage}\left(\left(\mathbf{R}\left(\mathbf{d}^{(z+1)}\right)^T - \frac{\mathbf{e}^{(z)}}{\eta}\right), \left(\frac{\gamma_j}{\eta}\right)\right), \\[4mm] \mathbf{e}^{(z+1)} = \mathbf{e}^{(z)} - \eta\left(\mathbf{R}\left(\mathbf{d}^{(z+1)}\right)^T - \mathbf{b}^{(z+1)}\right), \end{cases} \tag{16}$$

where shrinkage $(\mathbf{v}_1, \mathbf{v}_2) \triangleq \text{sign}(\mathbf{v}_1) \cdot \max\{0, |\mathbf{v}_1| - \mathbf{v}_2\}$.

Following the above deduction, the solving process of equation (9) can be summarized as follows (Algorithm 1):

4. Case Study

4.1. Data of Illustrative Example. In this illustrative example, the bearing run-to-failure data are studied. Bearing fault datasets have three types: inner race fault, cage, and outer race fault. The run-to-failure data from XJTU are measured under the condition of 11 kN load, 2250 rpm speed, and 25.6 kHz sampling frequency. The data files that come from bearing 2_1, bearing 2_3, and bearing 2_5 in reference. Table 1 is used to calculate the weight matrix.

In each observation epoch, the length of the signal that the sensor collected is more than 20000; thus, the size of the spectrum generated via FFT transform can be more than 10000. To reduce computational complexity, in this case, FFT only generates 512 dimensions' spectrums as the quasi-orthogonal sparse project algorithm input data.

4.2. Results and Analysis. Figure 3 shows the performance degradation assessment of MIHI for the different bearing fault monitoring. Here, the balance parameter α is set to 1, n_{nor} is set to 10, β is $2e^{-4}$, and $\gamma = 2e^{-5}$. The blue line denotes the condition monitoring health index of the inner race fault, the red line represents the cage fault monitoring, and the green color is the degradation trend of the outer race fault. To facilitate the observation, the MIHI value of each fault type monitor curve will minus the average of the first 50 files' MIHI value.

Figure 4 shows time-domain features that can indicate an occurrence of an incipient-bearing fault. To monitor the bearing health state, some time-domain features are used to quantify the bearing run-to-failure data, such as standard deviation and kurtosis. In Figure 4, we use the dataset bearing 2_5 to illustrate the superiority of the MIHI for bearing health monitoring. However, the time-domain features do not have monotonic trending, which is not beneficial to the assessment of bearing degradation performance and prognostics. Meanwhile, the MIHI monitoring curve not only has a strong monotonic trending line but also can indicate an incipient-bearing fault.

To illustrate further the advantages of the proposed method, the natural variability of the proposed MIHI and how it is used for fault detection and incipient fault diagnosis are also provided. First, the MIHI at observation epochs 1–50 in a normal stage is used as a historical normal dataset. Second, whether the normal state dataset obeys the Gaussian distribution is checked. At a significance level of 5%, the MIHI normal state datasets from bearings 2_1, 2_3, and 2_5 all satisfy the normal distribution conditions. Therefore, the Gaussian distribution assumption of the normal stage is accepted. Lastly, the three-sigma rule is used to detect a bearing abnormality, and the statistical threshold can be used as an early warning baseline for fault detection and beginning of degradation assessment.

The proposed MIHI monitoring curve family and its corresponding incipient fault threshold are plotted in Figure 5. Combined with the statistical threshold, the proposed MIHI can realize bearing incipient fault diagnosis and continuous detection of the bearing degradation process.

4.3. Hyperparameter Analysis. The main hyperparameters of the proposed algorithm α, β, γ, and n_{nor} measure monotonicity, orthogonality, weight sparsity, and weight difference sparsity. These four hyperparameters are empirically chosen. In this section, the hyperparameter selection suggestion and the hyperparameters' effect on the final HIMI are studied.

First, a function is built to evaluate the MIHI condition monitoring curve's monotonicity:

$$\text{Monotonicity} = \sum_{j=1}^{k} \frac{\sum_{i=2}^{n_j} \left| \text{sign}\left(\text{HIMI}_{i,j}^{j} - \text{HIMI}_{i,j}^{j} \right) \right|}{n_j - 1}, \quad (17)$$

where $\text{HIMI}_{i,j}^{j}$ means the jth value of ith observation $\mathbf{HIMI}_i^{j} \in \mathbb{R}^{k \times 1}$, which is calculated by $\mathbf{HIMI}_i^{j} = \mathbf{W} \times \mathbf{f}_{j,i}$. The larger the value of monotonicity, the better monotonicity the MIHI curve has.

Second, to indicate the orthogonality of the MIHI, a formula is utilized to calculate orthogonality:

$$\text{Orthogonality} = \sum_{j=1}^{k} \frac{\sum_{u=1}^{k} \text{std}\left(\mathbf{HIMI}_{u,:}^{j} \right) - \text{std}\left(\mathbf{HIMI}_{j,:}^{j} \right)}{\left| \max\left(\mathbf{HIMI}_{j,:}^{j} \right) - \min\left(\mathbf{HIMI}_{j,:}^{j} \right) \right|}, \quad (18)$$

where $\mathbf{HIMI}^{j} \in \mathbb{R}^{k \times n_j}$, which is calculated by $\mathbf{HIMI}^{j} = \mathbf{W} \times [\mathbf{f}_{j,1}, \mathbf{f}_{j,2}, \ldots, \mathbf{f}_{j,n_j}]$, and $\mathbf{HIMI}_{j,:}^{j}$ denotes the jth row of $\mathbf{HIMI}^{j}$. The smaller the value of orthogonality, the better orthogonality the MIHI curve family has.

In this illustrative example, the final fault type number is 3 (inner race, cage, and outer race); thus, $k = 3$. Assume that the first fault degradation dataset is bearing 2_1 in Table 1 and consists of 491 files; thus, $n_1 = 491$. The spectrum from the first file's signal in bearings 2_1, 2_3, and 2_5 is $\mathbf{f}_{1,1}$, $\mathbf{f}_{2,1}$, and $\mathbf{f}_{3,1}$, respectively. Our experiments show that even though the hyperparameter n_{nor} ($n_{\text{nor}} = 10$) is set as a small value, the quasi-orthogonal sparse project algorithm can still help the MIHI curve family to obtain good orthogonality. For the rest of this section, the hyperparameter n_{nor} selection is not studied.

The resulting heatmap of monotonicity is shown in Figure 6, and that of orthogonality is shown in Figure 7. We study four α values, i.e., 0, 0.4, 1, and 2. In each heatmap, we study eight β and γ values, which make up 64 combinations. Figures 6 and 7 not only indicate the hyperparameters' (α, β, and γ) effect on the monotonicity and orthogonality of the MIHI monitoring curve family but also offer the selection reference of hyperparameters α, β, and γ.

The hyperparameter α decides the MIHI monitoring curve's monotonicity. As shown in Figure 6, the bigger the parameter α, the more monotonic the MIHI monitoring curve is. However, Figure 7 indicates that the big α causes the

Input $\mathbf{S}_{sensitive}^{j}$, $\mathbf{S}_{subject}^{j}$, $\hat{\sigma}_j$, β_j, γ_j, and let $\mathbf{w}_j^{(1)}$ be the maximum eigenvector of $\mathrm{inv}(\mathbf{S}_{subject}^{j}) \times \mathbf{S}_{sensitive}^{j}$.
For $h = 1, 2, \ldots$ until convergence $(\|\mathbf{w}_j^{(h)} - \mathbf{w}_j^{(h-1)}\|_2/\max(1, \|\mathbf{w}_j^{(h)}\|_2) < 0.0001)$
   Set $\eta > 0$, $l > \rho(\mathbf{A}^T\mathbf{A} + \eta\mathbf{R}^T\mathbf{R})$, $\mathbf{d}^{(1)} = \mathbf{0}$, $\mathbf{e}^{(1)} = \mathbf{0}$, $\mathbf{d}^{(1)} = \mathbf{w}_j^{(h)}$
    For $z = 1, 2, \ldots$ until convergence $(\|\mathbf{d}^{(z)} - \mathbf{d}^{(z-1)}\|_2/\max(0, \|\mathbf{d}^{(z)}\|_2) < 0.0001)$:
    Let $\mathbf{d}^{(z+1)}$, $\mathbf{b}^{(z+1)}$, $\mathbf{e}^{(z+1)}$ be the solution of equation (16).
     End. Let $\mathbf{w}_j^{(h+1)} = \hat{d}/\sqrt{\hat{d}\mathbf{S}_{subject}^{j}\hat{d}^T}$.
End. Let $\mathbf{w}_j = \mathbf{w}_j^{(h)}$.

ALGORITHM 1: Algorithm 1.

TABLE 1: Detail of the illustrative example.

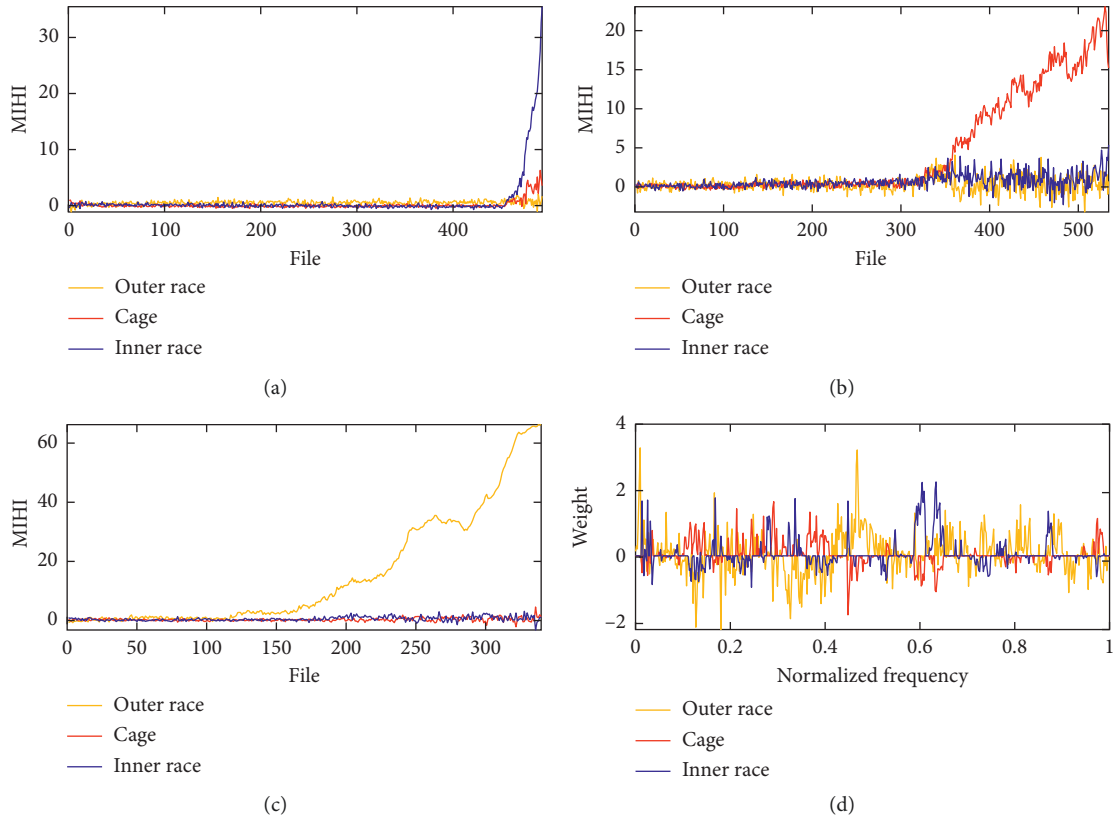| Condition	Bearing no.	Time duration	Number of measurement files	Final fault
Speed: 37.5 Hz	Bearing 2_1	8 h 11 min	491	Inner race
Load: 11 kN	Bearing 2_3	8 h 53 min	533	Cage
	Bearing 2_5	5 h 39 min	339	Outer rage

(a)

(b)

(c)

(d)

FIGURE 3: Results of degradation monitoring and weight of the component. (a) Inner race fault. (b) Cage fault. (c) Outer race fault. (d) Weight matrix (blue inner race, red cage, and yellow outer race).

MIHI monitoring curve family's orthogonality to decrease. Moreover, from all the heatmaps in Figures 6 and 7, $\alpha = 1$.

When prerequirement Properties 1–3 of the HIMI curve family are met, the sparser the weight matrix $\mathbf{W}$ is, the better overfitting is avoided. The heatmaps in Figures 6 and 7 show that β should be smaller than 0.0002 and γ/β should be smaller than 0.1; otherwise, the weight matrix $\mathbf{W}$ is too sparse to map the spectrum features into a MIHI monitoring

curve family, which has good monotonicity and orthogonality.

When the quasi-orthogonal sparse project algorithm only focuses on the first three properties: monotonicity, sensitivity, and orthogonality (β and $\gamma = 0$), the MIHI monitoring curve family has the best monotonicity and orthogonality. However, the weight matrix is severely overfitted. The degradation process of bearing 2_3 is used to
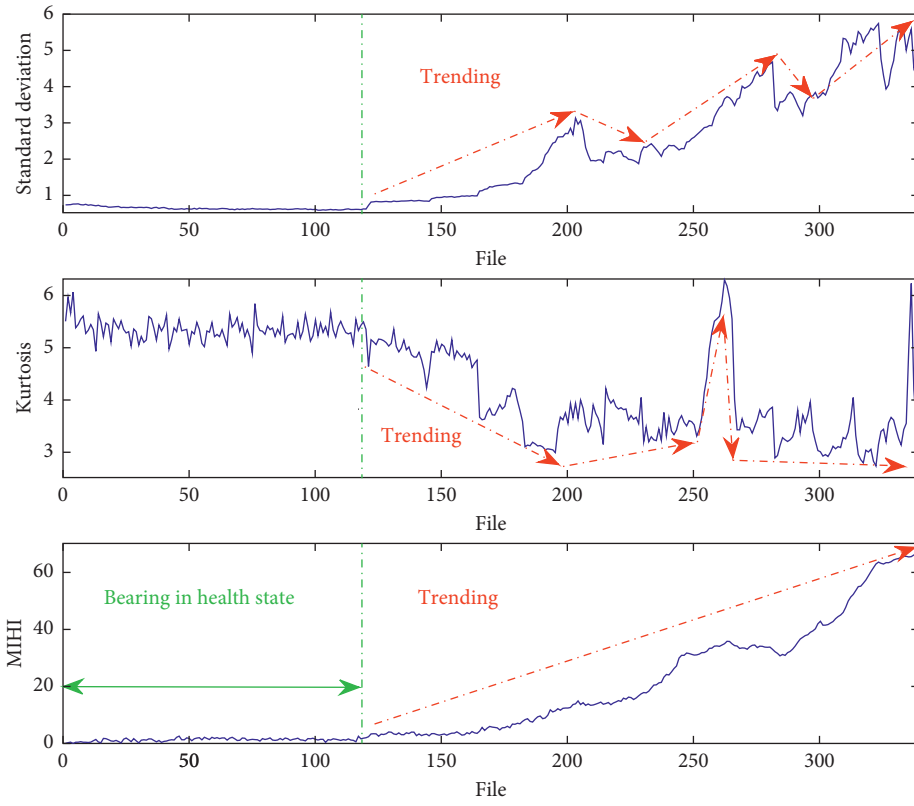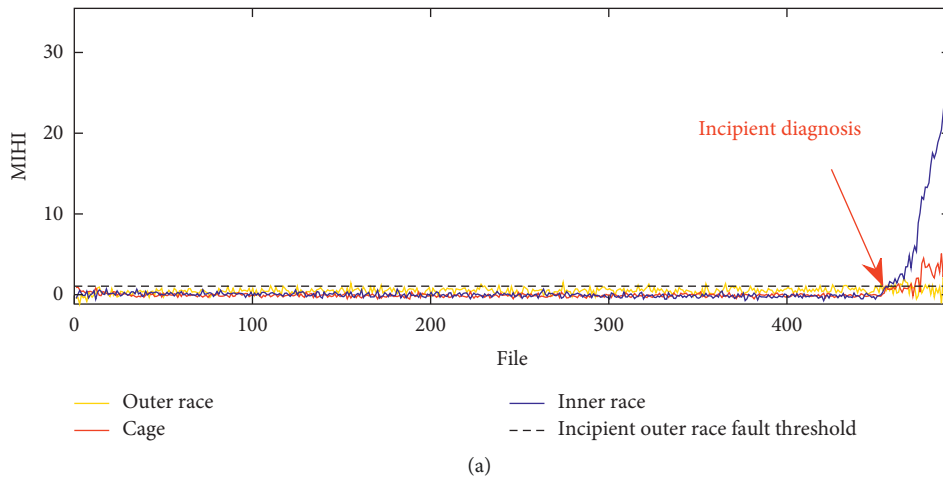
FIGURE 4: Bearing health state monitoring with different health indexes. (a) Standard deviation. (b) Kurtosis. (c) MIHI.
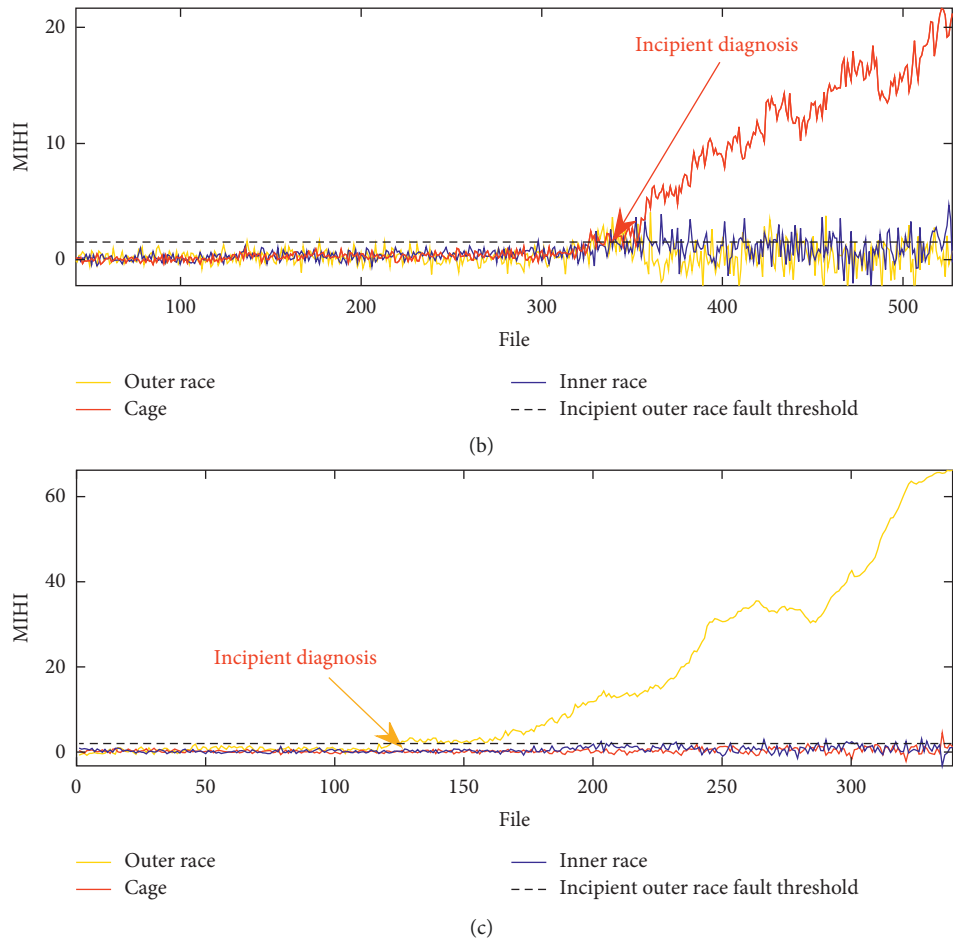

(a)

FIGURE 5: Continued.

(b)

(c)

FIGURE 5: Bearing health state monitoring with proposed MIHI. (a) Inner race fault degradation. (b) Cage fault degradation. (c) Outer race fault degradation.
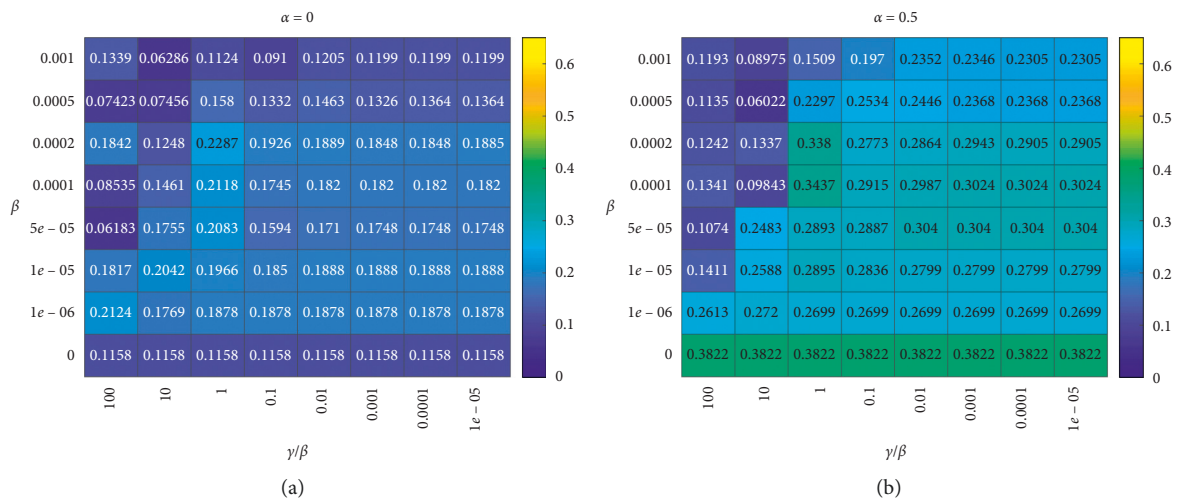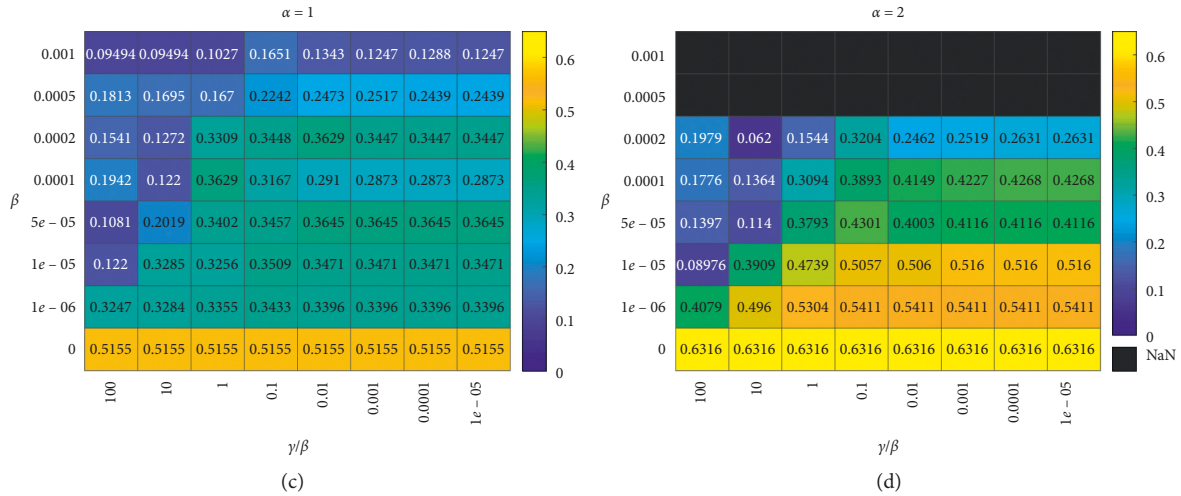

(a)

(b)

FIGURE 6: Continued.

FIGURE 6: Monotonicity heatmap of hyperparameters α, β, and γ. (a) $\alpha = 0$, (b) $\alpha = 0.5$, (c) $\alpha = 1$, and (d) $\alpha = 2$.
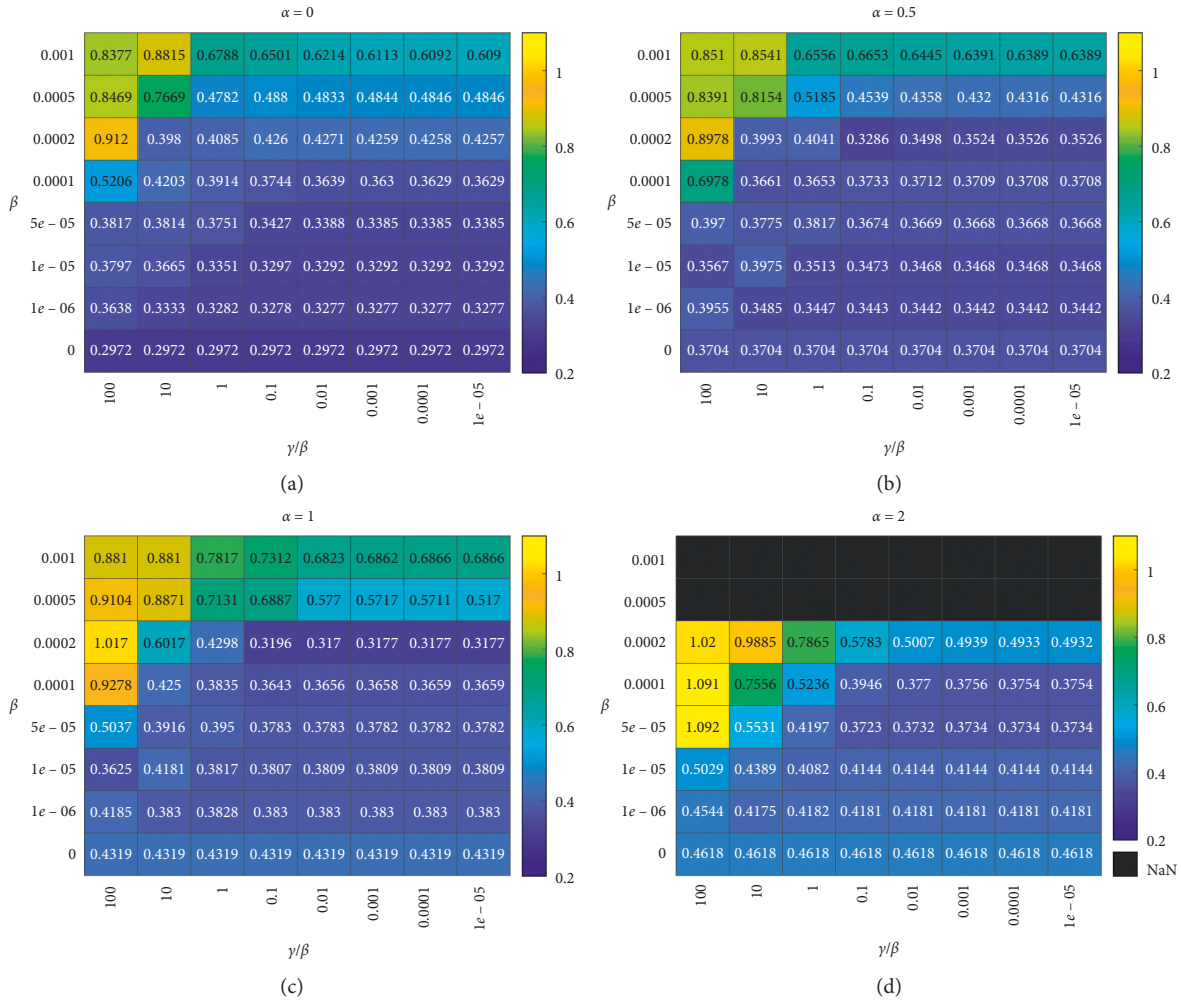
FIGURE 7: Orthogonality heatmap of hyperparameters α, β, and γ. (a) $\alpha = 0$, (b) $\alpha = 0.5$, (c) $\alpha = 1$, and (d) $\alpha = 2$.
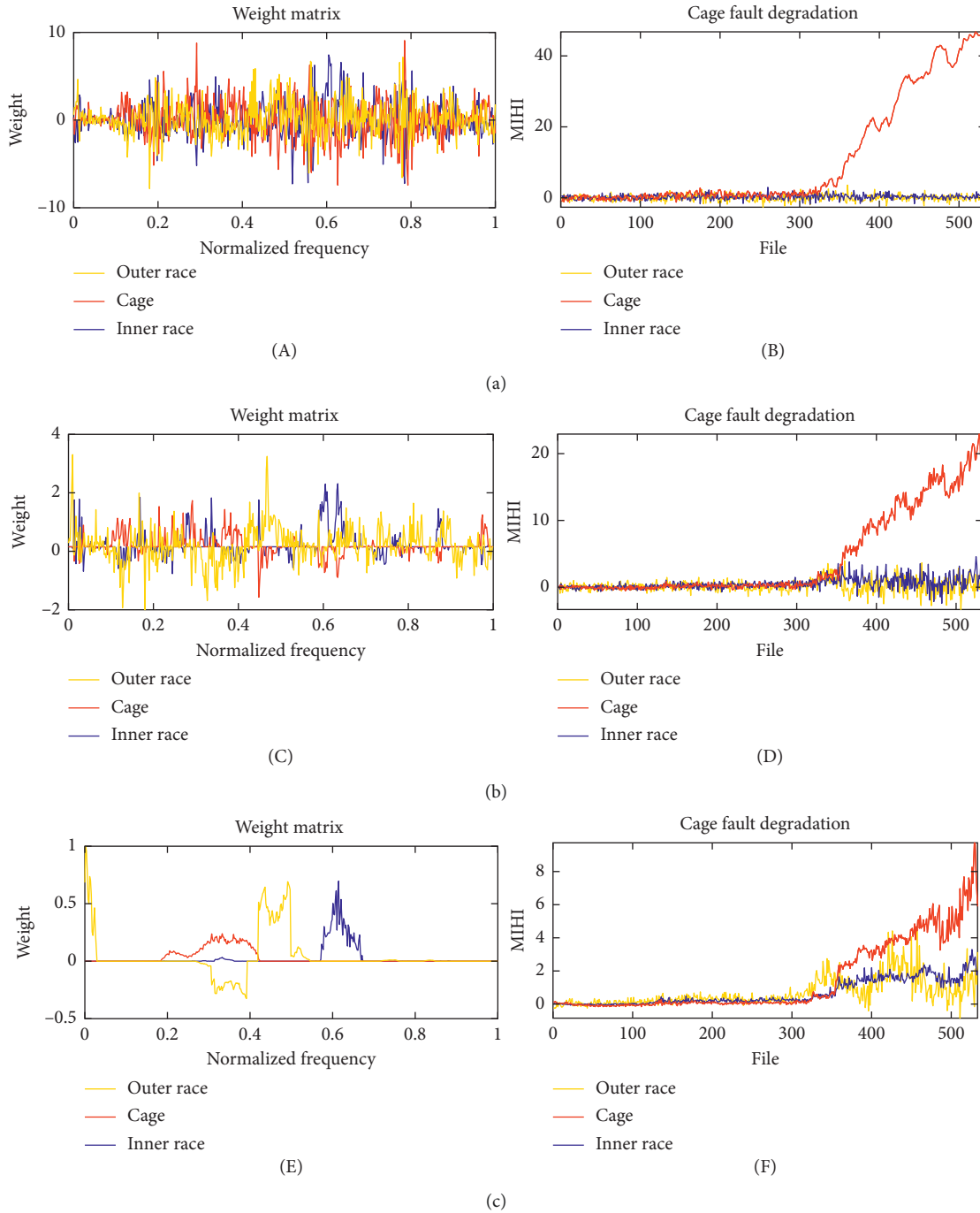
FIGURE 8: MIHI example of hyperparameters' change; the hypermeters $\alpha = 1$ and $n_{nor} = 10$ are the same, but β and γ are different; (a) $\beta = 0$ and $\gamma = 0$. (b) $\beta = 2e^{-4}$ and $\gamma = 2e^{-5}$. (c) $\beta = 5e^{-4}$ and $\gamma = 5e^{-3}$.

illustrate the effect of hyperparameters β and γ in Figure 8. Compared with the MIHI curve family when $\beta = 2e^{-4}$ and $\gamma = 2e^{-5}$, as shown in Figure 8(b), the monotonicity and orthogonality of the MIHI curve family when β and $\gamma = 0$ are stronger (Figure 8(a)), and the weight matrix fits all spectrum components. However, bearing fault characteristic information is not distributed in all frequency bands, and the weight matrix should only focus on the

spectrum components related to the fault. The weight matrix calculated when β and $\gamma = 0$ also fits with substantial noise, thus reducing the generality of MIHI. As shown in Figure 8(c), if the weight matrix is too sparse, then the MIHI monitoring curve family cannot satisfy Properties 1, 2, and 3. Thus, we recommend using hyperparameter heatmaps to select suitable hyperparameters when applying the proposed method.

5. Conclusions

This study proposed a method of building a MIHI to trace the various bearing degradation trends with various types of faults. The proposed method is a linear transform algorithm that maps high-dimensional observation features into low-dimension MIHI, indicating the bearing's various fault type degradation trends at the same time. Meanwhile, inspired by the orthogonal vector, we proposed utilizing orthogonality to develop the health index of bearing degradation trend monitoring. This algorithm also introduces weight sparsity and weight difference sparsity to avoid overfitting. The proposed algorithm has explicit and simple mathematical expressions, and the process of calculation does not rely on the complex optimization algorithm. Therefore, the method is suitable to deal with situations that have high-dimensional observation features.

Nomenclature

k:	Number of fault types
m:	The dimension of the spectrum feature
α:	MIHI balance parameter
β:	MIHI sparsity penalty parameter
γ:	MIHI weight difference sparsity penalty parameter
n_{nor}:	Number of the epoch of bearing normal state
η:	Lagrangian penalty parameter when solving
l:	Approximation parameter when solving
$w_{j,q}$:	j − row and q − column value of $\mathbf{W}$
$\mathbf{W} \in \mathbb{R}^{k \times m}$:	Project matrix of MIHI
$\mathbf{w}_j \in \mathbb{R}^{1 \times m}$:	j row of $\mathbf{W}$
$\mathbf{F}_j \in \mathbb{R}^{m \times n_j}$:	Spectrum feature matrix of jth fault type degradation process data
$\mathbf{f}_{j,i} \in \mathbb{R}^{m \times 1}$:	Spectrum feature of jth fault type ith observation epoch
$\mathbf{N}_j \in \mathbb{R}^{m \times n_{\text{nor}}}$:	Normal state observation matrix of jth fault type
$\boldsymbol{\mu}_j \in \mathbb{R}^{m \times 1}$:	Mean vector of $\mathbf{N}_j$
$\widehat{\sigma}_j \in \mathbb{R}^{1 \times m}$:	Project matrix j − row's feature penalty vector
$\mathbf{A} \in \mathbb{R}^{m \times m}$, $\mathbf{b} \in \mathbb{R}^{1 \times (m-1)}$, $\mathbf{c} \in \mathbb{R}^{m \times 1}$, $\mathbf{d} \in \mathbb{R}^{1 \times m}$, $\mathbf{e} \in \mathbb{R}^{1 \times (m-1)}$, $\mathbf{Q} \in \mathbb{R}^{m \times m}$, $\mathbf{G} \in \mathbb{R}^{m \times m}$:	Intermediate variables in solving process

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation - a review on the statistical data driven approaches," *European Journal of Operational Research*, vol. 213, no. 1, pp. 1–14, 2011.

[2] T. Yan, D. Wang, T. Xia, and L. Xi, "A generic framework for degradation modeling based on fusion of spectrum amplitudes," *IEEE Transactions on Automation Science and Engineering*, pp. 1–12, 2020.

[3] Y. Shao and K. Nezu, "Prognosis of remaining bearing life using neural networks," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 214, no. 3, pp. 217–230, 2005.

[4] B. Zhang, S. Zhang, and W. Li, "Bearing performance degradation assessment using long short-term memory recurrent network," *Computers in Industry*, vol. 106, pp. 14–29, 2019.

[5] M. Zhao, B. Tang, and Q. Tan, "Bearing remaining useful life estimation based on time-frequency representation and supervised dimensionality reduction," *Measurement*, vol. 86, pp. 41–55, 2016.

[6] Y. Qin, S. Xiang, Y. Chai, and H. Chen, "Macroscopic-microscopic attention in LSTM networks based on fusion features for gear remaining life prediction," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 12, pp. 10865–10875, 2020.

[7] A. Rai and J.-M. Kim, "A novel health indicator based on information theory features for assessing rotating machinery performance degradation," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6982–6994, 2020.

[8] P. Shankar Kumar, L. A. Kumaraswamidhas, and S. K. Laha, "Bearing degradation assessment and remaining useful life estimation based on Kullback-Leibler divergence and Gaussian processes regression," *Measurement*, vol. 174, 2021.

[9] Y. Chen, G. Peng, Z. Zhu, and S. Li, "A novel deep learning method based on attention mechanism for bearing remaining useful life prediction," *Applied Soft Computing*, vol. 86, 2020.

[10] N. Ghosh, Y. B. Ravi, A. Patra et al., "Estimation of tool wear during CNC milling using neural network-based sensor fusion," *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 466–479, 2007.

[11] J. Llinas and D. L. Hall, "An introduction to multi-sensor data fusion," in *Proceedings of the 1998 IEEE International Symposium on Circuits and Systems ISCAS'98 (Cat. No.98CH36187)*, Monterey, CA, USA, May-June 1998.

[12] K. Goebel and P. Bonissone, "Prognostic information fusion for constant load systems, presented at the 2005 7th International Conference on Information Fusion," 2005.

[13] C. Ma, X. Zhai, Z. Wang et al., "State of health prediction for lithium-ion batteries using multiple-view feature fusion and support vector regression ensemble," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 9, pp. 2269–2282, 2018.

[14] G. Niu, A. Widodo, J.-D. Son, B.-S. Yang, D.-H. Hwang, and D.-S. Kang, "Decision-level fusion based on wavelet decomposition for induction motor fault diagnosis using transient current signal," *Expert Systems with Applications*, vol. 35, no. 3, pp. 918–928, 2008.

[15] Y. Wei, D. Wu, and J. Terpenny, "Decision-level data fusion in quality control and predictive maintenance," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 1, pp. 184–194, 2021.

[16] K. Liu, N. Z. Gebraeel, and J. Shi, "A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis," *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 3, pp. 652–664, 2013.

[17] A. Chehade and Z. Shi, "Sensor fusion via statistical hypothesis testing for prognosis and degradation analysis," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 4, pp. 1774–1787, 2019.

[18] T. Yan, D. Wang, M. Zheng, T. Xia, E. Pan, and L. Xi, "Fisher's discriminant ratio based health indicator for locating informative frequency bands for machine performance degradation assessment," *Mechanical Systems and Signal Processing*, vol. 162, 2022.

[19] M. Kim, C. Song, and K. Liu, "A generic health index approach for multisensor degradation modeling and sensor selection," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 3, pp. 1426–1437, 2019.

[20] M. Dong and D. He, "Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis," *European Journal of Operational Research*, vol. 178, no. 3, pp. 858–878, 2007.

[21] K. Liu, A. Chehade, and C. Song, "Optimize the signal quality of the composite health index via data fusion for degradation modeling and prognostic analysis," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 3, pp. 1504–1514, 2017.

[22] R. Tibshirani, "Regression shrinkage and selection via the lasso: a retrospective," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 3, pp. 273–282, 2011.

[23] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 1, pp. 91–108, 2005.

[24] C. Song, K. Liu, and X. Zhang, "A generic framework for multisensor degradation modeling based on supervised classification and failure surface," *IISE Transactions*, vol. 51, no. 11, pp. 1288–1302, 2019.

[25] D. M. Witten and R. Tibshirani, "Penalized classification using Fisher's linear discriminant," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 5, pp. 753–772, 2011.

[26] X. Wang and X. Yuan, "The linearized alternating direction method of multipliers for dantzig selector," *SIAM Journal on Scientific Computing*, vol. 34, no. no. 5, pp. A2792–A2811, 2012.