Sparsity-Assisted Intelligent Condition Monitoring Method for Aero-engine Main Shaft Bearing

DING Baoqing[1], *WU Jingyao*[1], *SUN Chuang*[1], *WANG Shibin*[1], *CHEN Xuefeng*[1*], *LI Yinghong*[1,2]

1. State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, P.R. China;
2. Science and Technology on Plasma Dynamics Laboratory, Air Force Engineering University, Xi'an 710038, P.R. China

(Received 15 July 2020; revised 10 August 2020; accepted 15 August 2020)

Abstract: Weak feature extraction is of great importance for condition monitoring and intelligent diagnosis of aero-engine. Aimed at achieving intelligent diagnosis of aero-engine main shaft bearing, an enhanced sparsity-assisted intelligent condition monitoring method is proposed in this paper. Through analyzing the weakness of convex sparse model, i.e. the tradeoff between noise reduction and feature reconstruction, this paper proposes an enhanced-sparsity nonconvex regularized convex model based on Moreau envelope to achieve weak feature extraction. Accordingly, a sparsity-assisted deep convolutional variational autoencoders network is proposed, which achieves the intelligent identification of fault state through training denoised normal data. Finally, the effectiveness of the proposed method is verified through aero-engine bearing run-to-failure experiment. The comparison results show that the proposed method is good at abnormal pattern recognition, showing a good potential for weak fault intelligent diagnosis of aero-engine main shaft bearings.

Key words: aero-engine main shaft bearing; intelligent condition monitoring; feature extraction; sparse model; variational autoencoders; deep learning

CLC number: TH17 **Document code:** A **Article ID:** 1005-1120(2020)04-0508-09

0 Introduction

Bearing condition monitoring is one of the most important parts in engine health management (EHM), which is a typical symbol of advanced aircraft engines. However, the fault diagnosis of aero-engine main shaft bearing is very difficult because of the complex engine structure and serious interference. The classical bearing condition monitoring methods mainly use specific indicators to identify the fault state, whose threshold value is set according to the manual experience. While the intelligent diagnosis method learns an effective intelligent monitoring model through incorporating intelligent algorithms and signal priors into the procedure of training samples[1-3]. Since aero-engines have high requirements for service safety, the failure state is a small probability event compared with the health state, and it is very difficult to capture effective failure data. Therefore, the key of intelligent diagnosis for aero-engine main shaft bearing is to train a model based on health state data and realize the intelligent identification of fault state.

High quality and low noise training samples are very important to improve the abnormal monitoring performance of intelligent methods. Filtering and sparsity-assisted fault diagnosis (SAFD) methods are two classical and advanced techniques for signal noise reduction and feature extraction. Filtering methods, such as wavelets[4], empirical mode decomposition (EMD)[5] and spectral kurtosis[6], usually convert original signal from time domain to a

*Corresponding author, E-mail address: chenxf@xjtu.edu.cn.

How to cite this article: DING Baoqing, WU Jingyao, SUN Chuang, et al. Sparsity-assisted intelligent condition monitoring method for aero-engine main shaft bearing[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2020, 37(4): 508-516.

http://dx.doi.org/10.16356/j.1005-1120.2020.04.002

specific transform domain based on physical priors. Such methods use the different coefficient distribution in the transform domain of noise and fault feature to achieve noise reduction and fault feature extraction. The SAFD method[7-9] is to incorporate such priors into regularization terms to construct a sparse model, and obtain the optimum solution of constructed model based on convex optimization theory to achieve the goal of reducing noise and extracting features. Compared with the traditional method, the SAFD method takes fully use of the physical priors in the framework of convex optimization, thus showing a better denoising performance. The classical sparsity-assisted diagnosis model generally constructs a convex model based on a convex sparse penalty[10], such as L1 norm penalty, which suppresses the amplitude of the estimated signal[11]. Therefore, both the traditional filtering-like feature extraction method and the classical SAFD method will suppress the denoised signal's amplitude, i.e. without balancing the performance of noise reduction and feature reconstruction. Constructing a sparse model with less signal underestimation is a key to improve the performance of weak feature extraction and intelligent condition monitoring.

Due to the measured signal of aero-engine bearing is seriously corrupted by strong noise, an enhanced sparsity-assisted intelligent condition monitoring method is proposed in this paper to realize weak feature extraction and intelligent identification of bearing fault state. The remainder of this paper is organized as follows. The basic idea of SAFD method is briefly introduced in Section 1. Section 2 describes in detail the intelligent monitoring method based on sparsity-assisted deep convolution variational autoencoders network. Then, the proposed method is verified through aero-engine bearing run-to-failure experiment in Section 3. Finally, conclusions are drawn in Section 4.

1 Sparsity-Assisted Fault Diagnosis Model

In order to obtain high quality and low noise training samples to support intelligent identification of fault abnormal state, the SAFD strategy for noise reduction and feature enhancement is firstly studied. The core idea of the SAFD method is to represent the fault feature signal with specific transformation or redundant dictionary based on signal sparse prior, and reconstruct the characteristic signal to realize the extraction and recognition of fault features. The classical sparsity-assisted fault diagnosis convex model, i.e. basis pursuit denoising problem, is shown as

$$F(\boldsymbol{x})=\min_{x}\frac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}\right\|_2^2+\lambda\left\|\boldsymbol{x}\right\|_1 \tag{1}$$

where $\boldsymbol{y}$ is the measured noisy signal, $\boldsymbol{D}$ the sparse representation dictionary, $\boldsymbol{x}$ the sparse coefficient, λ the regularization parameter, and $\|\cdot\|_1$ the convex penalty constructed with L1 norm function.

Eq.(1) is the synthesis model of the basis pursuit de-noising (BPDN) problem, while its analysis model[12-13] is shown as

$$F(\boldsymbol{x})=\min_{x}\frac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{x}\right\|_2^2+\lambda\left\|\boldsymbol{A}\boldsymbol{x}\right\|_1 \tag{2}$$

where $\boldsymbol{A}$ denotes the sparse transform.

The traditional SAFD method cannot take into account the two important properties of signal processing, namely noise reduction and signal reconstruction. Taking the BPDN problem (shown in Eq.(1) and Eq.(2)) as an example, parameter λ balances the data fidelity and sparse regularization term. The larger λ is, the less the sparse coefficient is and the less the noise component is, which means the better noise reduction performance. But at the same time, the more suppressed the amplitude of $\boldsymbol{x}$ is, the smaller the energy of reconstructed feature signal is, which means worse reconstruction performance.

In order to overcome this problem, an enhanced sparsity-assisted fault diagnosis model is proposed, namely nonconvex regularized convex sparse model. This model makes fully use of the good amplitude-preserving performance of nonconvex penalties and the advantage of convex model, i.e. easy to be solved, thus combining the performance of noise reduction and feature reconstruction.

The relationship among L0 norm, L1 norm and minimax concave (MC) penalty is shown in Fig.1. L0 norm is the definition of sparsity and means the number of nonzero elements, which will make the sparse model constructed with L0 norm be nonconvex. L1 norm is a convex function closest to the L0 norm, which can make the constructed sparse model be convex, but will underestimate the amplitude of feature signal. MC penalty[14] is a nonconvex function, which is between L0 and L1 norms. By setting appropriate parameters, the sparse model constructed with MC penalty can be convex, thus making it convenient to solve the model with the convex optimization method. This kind of nonconvex penalty provides a possible way to realize signal noise reduction and feature reconstruction simultaneously.

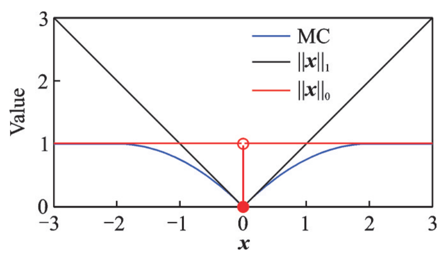

Fig.1 Relationship among MC, $\|x\|_0$ and $\|x\|_1$

2 Sparsity-Assisted Intelligent Condition Monitoring Method

Motivated by the construction idea of MC penalty, this section proposes a kind of nonconvex penalty construction method, and builds a convex sparse model based on such nonconvex penalty. Accordingly, an intelligent condition monitoring method based on sparsity-assisted deep convolution variational autoencoders network is proposed in this seciton.

2.1 Nonconvex penalty construction based on Moreau envelope

Through analyzing the definition of MC penalty, the relationship among MC function, Huber function and Moreau envelope is firstly studied in this section. Then a nonconvex penalty construction method based on generalized Moreau envelope is proposed.

Supposing that ϕ: $\mathbf{R}\rightarrow\mathbf{R}$, s: $\mathbf{R}\rightarrow\mathbf{R}$ and f: $\mathbf{R}^N\rightarrow\mathbf{R}$, we show the definition of three functions in Eqs.(3)—(6).

The MC penalty in scalar form can be defined as

$$\phi(x)=\begin{cases}|x|-\dfrac{1}{2}x^2 & |x|\leqslant 1\\ \dfrac{1}{2} & |x|>1\end{cases} \tag{3}$$

The definition of Huber function in scalar form is described as

$$s(x)=\begin{cases}\dfrac{1}{2}x^2 & |x|\leqslant 1\\ |x|-\dfrac{1}{2} & |x|>1\end{cases} \tag{4}$$

It can be also written as another form

$$s(x)=\min_{v\in\mathbf{R}}\left\{|v|+\frac{1}{2}(x-v)^2\right\} \tag{5}$$

The Moreau envelope is defined as

$$f^{\mathrm{M}}(x)=\inf_{v\in\mathbf{R}^N}\left\{f(v)+\frac{1}{2}\|x-v\|_2^2\right\} \tag{6}$$

Based on the above definition formula, it can be found that the three functions have the following relationships.

(1) Huber function is a special case of the Moreau function when $f(v)=\|\cdot\|_1$.

(2) MC penalty can be constructed with the Huber function

$$\phi(x)=|x|-s(x) \tag{7}$$

Inspired by this relationship, the scalar MC can be extended to multivariate form, thus forming a nonconvex penalty construction strategy based on generalized Moreau envelope.

Suppose $\Gamma_0(\mathbf{R}^N)$ denotes the set of proper lower semi-continuous convex functions $\mathbf{R}^N\rightarrow\mathbf{R}\cup\{+\infty\}$, f_B^{M}: $\mathbf{R}^N\rightarrow\mathbf{R}$ and $B\in\mathbf{R}^{M\times N}$. The generalized Moreau envelope is defined as

$$f_B^{\mathrm{M}}(x)=\inf_{v\in\mathbf{R}^N}\left\{f(v)+\frac{1}{2}\|B(x-v)\|_2^2\right\} \tag{8}$$

If $f(v)=\|\cdot\|_1$, we have the generalized Huber function, shown as

$$S_B(x)=\min_{v\in\mathbf{R}^N}\left\{\|v\|_1+\frac{1}{2}\|B(x-v)\|_2^2\right\} \tag{9}$$

Accordingly, the generalized MC (GMC) penalty can be deduced as

$$\psi_B(x)=\|x\|_1-S_B(x) \tag{10}$$

2.2 Enhanced sparsity-assisted fault diagnosis (ESAFD) convex model

Assuming that feature signal is sparse in the transform domain $\boldsymbol{A}$, an enhanced sparsity-assisted fault diagnosis model can be constructed by combining proposed GMC penalty and data fidelity term, shown as

$$F_B(\boldsymbol{x}) = \min_{\boldsymbol{x}\in\mathbf{R}^N} \frac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\right\|_2^2 + \lambda\psi_B(\boldsymbol{x}) \tag{11}$$

If

$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \preceq \frac{1}{\lambda}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} \tag{12}$$

then $F_B(\boldsymbol{x})$ is a convex function.

Proof: Write $F_B(\boldsymbol{x})$ as

$$\begin{aligned} F_B(\boldsymbol{x}) = &\frac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\right\|_2^2 + \lambda\left\|\boldsymbol{x}\right\|_1 - \\ &\lambda \min_{\boldsymbol{v}\in\mathbf{R}^N}\left\{\left\|\boldsymbol{v}\right\|_1 + \frac{1}{2}\left\|\boldsymbol{B}(\boldsymbol{x}-\boldsymbol{v})\right\|_2^2\right\} = \\ &\max_{\boldsymbol{v}\in\mathbf{R}^N}\left\{\frac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\right\|_2^2 + \lambda\left\|\boldsymbol{x}\right\|_1 - \right. \\ &\left.\lambda\left\|\boldsymbol{v}\right\|_1 - \frac{\lambda}{2}\left\|\boldsymbol{B}(\boldsymbol{x}-\boldsymbol{v})\right\|_2^2\right\} = \\ &\max_{\boldsymbol{v}\in\mathbf{R}^N}\left\{\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}-\lambda\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})\boldsymbol{x} + \right. \\ &\left.\lambda\left\|\boldsymbol{x}\right\|_1 + g(\boldsymbol{x},\boldsymbol{v})\right\} = \\ &\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}-\lambda\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})\boldsymbol{x} + \\ &\lambda\left\|\boldsymbol{x}\right\|_1 + \max_{\boldsymbol{v}\in\mathbf{R}^N} g(\boldsymbol{x},\boldsymbol{v}) \end{aligned} \tag{13}$$

where $g(\boldsymbol{x},\boldsymbol{v})$ is affine in $\boldsymbol{x}$, which is convex. Since L1 norm is convex, the objective function $F_B(\boldsymbol{x})$ is convex if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}-\lambda\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$ is positive semidefinite, which is equivalent to Eq.(12).

The design of parameter $\boldsymbol{B}$ determines not only the convexity of the model, but also the way of GMC implementing regularizations. In order to design a better $\boldsymbol{B}$, we can incorporate the physical prior of the processed signal into the design procedure. Since the objective function of the proposed model is convex, which is a saddle-point problem, according to the convex optimization theory, the forward-backward splitting algorithm can be used to obtain the optimum solution.

2.3 Sparsity-assisted deep convolutional variational autoencoders network monitoring model

It is a semi-supervised learning task to learn the discriminant model based on the health state data and to detect the fault state. In terms of the task requirement, variational autoencoders show a good application potential for its strong compression and signal reconstruction property.

A deep convolutional variational autoencoders network is constructed in this paper by combining the structure of deep convolutional neural network and variational autoencoders. The model contains two sub-modules, named encoder $f_\theta(\mu,\sigma|\boldsymbol{x})$ and decoder $g_\phi(\boldsymbol{x}|\boldsymbol{z})$, where $\boldsymbol{x}$ denotes input data and $\boldsymbol{z}$ denotes compressed code. It is supposed that all variables z_i are independent distributed as normal distribution with the mean of μ_i and the variance of σ_i. In the updating procedure, the model learns to establish an effective mapping $g_\phi(\bar{\boldsymbol{x}}|f_\theta(\mu,\sigma|\boldsymbol{x}))$ between the original signal $\boldsymbol{x}$ and the reconstructed signal $\bar{\boldsymbol{x}}$. The mean square error loss function $\left\|\boldsymbol{x}-\bar{\boldsymbol{x}}\right\|_2^2$ is used to guide the reconstruction process. It is also assumed that each variable z_i has a standard normal prior distribution $p(z_i|\boldsymbol{x})\sim\boldsymbol{N}(0,1)$. The Kullback-Leibler (KL) divergence between the conditional posterior distribution and the prior distribution is used as a loss function to constrain their similarities. The loss function of whole model can be expressed as

$$L = \left\|\boldsymbol{x}-\bar{\boldsymbol{x}}\right\|_2^2 + \omega\sum_{j=1}^{d}(\mu_j^2+\sigma_j^2-\log\sigma_j^2-1) \tag{14}$$

where d represents the dimension of compressed code, and ω is introduced to adjust the weight between reconstruction error and distribution constraint.

The encoders contain five layers of convolutional blocks with channel size of 32, 64, 128, 256, and 512, respectively. Each block represents a combination of 2D-convolutional layer, batch normalization layer and leaky ReLU layer. After obtaining the latent codes of each μ_i and σ_i from the encoders, the resampling process is implemented by sampling each z_i from a normal distribution with the parameters of μ_i and σ_i. The decoders have the similar structure of encoders with the convolutional layer replaced by transposed convolutional layer and arranged in reverse order.

Combining the denoised feature signal of ESAFD and deep convolution variational autoencoders network, a sparsity-assisted intelligent monitoring method is proposed in this paper. Its algorithm diagram is shown in Fig. 2. Firstly, the denoised feature signal is extracted from original signal by the ESAFD method, and then the denoised signal is used as the input to train the deep convolution variational autoencoders network. The reconstruction error of variational autoencoders is utilized as the bearing condition monitoring index. If the reconstruction error is above the threshold value, the tested signal is different from the trained model corresponding to health status, meaning the abnormal status of the tested bearing occurs. Then the early warning is activated timely, thus achieving intelligent identification of fault state.

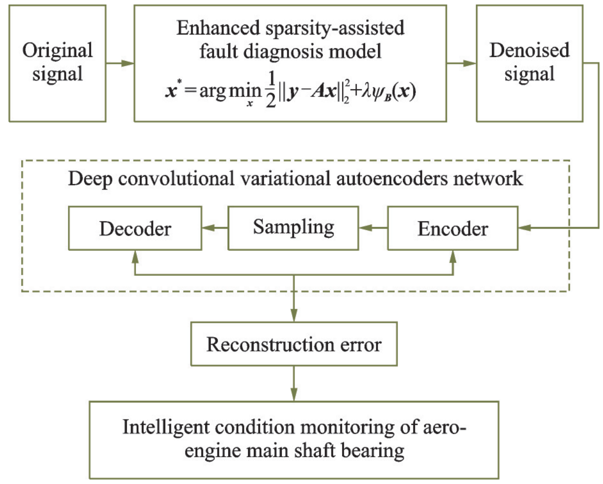

Fig.2 Sparsity-assisted intelligent monitoring model for aero-engine main shaft bearing

3 Experiment Verification

In this section, we verify the proposed method using an aero-engine bearing run-to-failure experiment. The test rig consists of several parts, such as mechanical system, drive and lubrication system, loading system, and electrical control system, as shown in Fig.3. The load system applies 11 kN radial load and 2 kN axial load to the test bearing, respectively. Vibration signals are collected by two accelerometers mounted in the horizontal and vertical directions of the bearing seat, with a sampling frequency of 20 kHz.

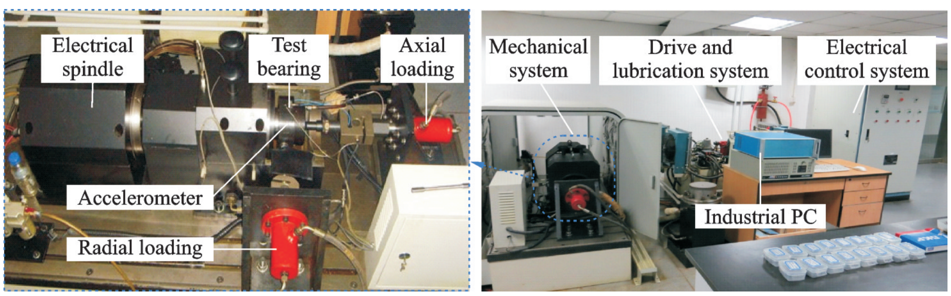

Fig.3 Bearing test rig

The test bearing is an H7018C angular contact ball bearing, with a pitch diameter of 117 mm, ball number of 27, ball diameter of 11.12 mm and a contact angle of 15°. The rotating speed is about 6 000 r/min. After about 156.3 working hours, the vibration signal of the bearing increases sharply and obvious periodic impact appears. Then we stop the test rig, and find a spall on the inner race with an area of 3 mm^2, shown in Fig.4. According to parameters of bearing and rotating speed, the ball passing frequency through inner raceway is about 1 475 Hz.

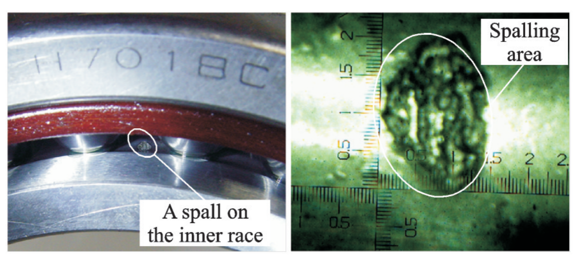

Fig.4 Spall on the bearing inner race

3.1 Verificatin 1: ESAFD for weak feature extraction

We verify the performance by comparing proposed ESAFD method with L1-based SAFD method. Both methods are applied to analyze the bearing vibration signal in the weak fault stage of the experiment, shown in Fig. 5. In this experiment, we set the TQWT[15] parameters to $Q=1, r=5, J=10$. The parameter λ of L1-based SAFD method is chosen to preserve 2% of the sparse coefficients to extract a significant fault impact characteristic. In order to better compare the performance, we set λ of ESAFD as the same value. For display convenience, the 100 Hz rotating speed is recorded as f_r, and the 1 475 Hz characteristic frequency is recorded as f.

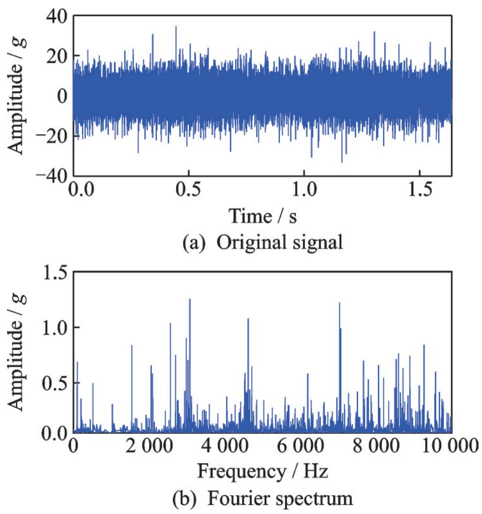

Fig.5 Original signal and its spectrum

The results of ESAFD and L1-based SAFD are shown in Fig. 6 and Fig. 7, respectively. The comparison results of denoised signal show that the amplitude of the fault feature signal extracted by L1-based SAFD method is small, while the ESAFD method extracts the bearing fault signal with high amplitude when reducing noise. The zoomed-in comparison results show that the ESAFD method can effectively extract the group sparse structure of transients, due to the periodic change of the load applying to bearing inner race. The comparison results of square envelope spectrum show that the ESAFD method is superior to the L1-based SAFD method, which can significantly extract the rotating frequency, characteristic frequency of bearing fault

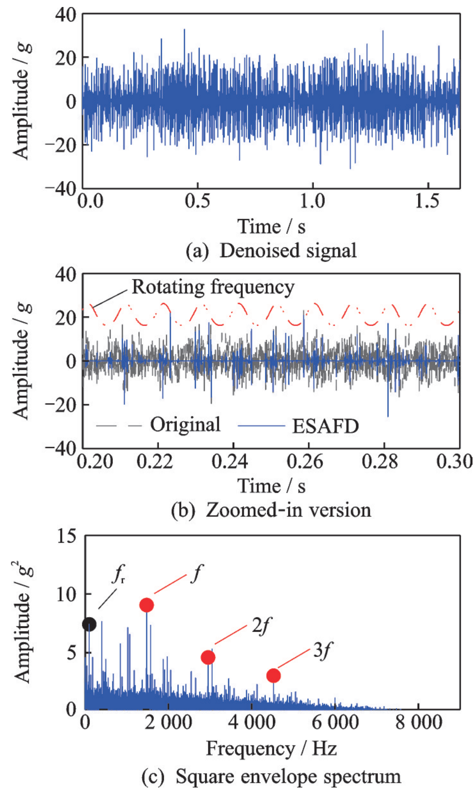

Fig.6 Denoised result by ESAFD method

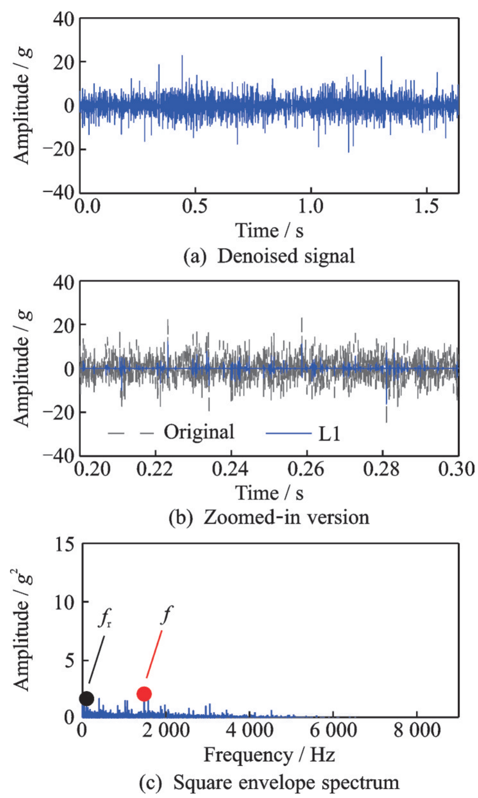

Fig.7 Denoised result by L1-based SAFD

(1 475 Hz) and its double and triple frequency component, thus realizing the effective identification of bearing inner-race fault. Therefore, compared with

L1-based SAFD method, the proposed ESAFD method can give better consideration to both noise reduction and feature reconstruction, showing a stronger ability of extracting weak features.

3.2 Verificatin 2: Sparsity-assisted intelligent monitoring method for fault status identification

The performance verification of the intelligent monitoring model is carried out using the aero-engine bearing run-to-failure experiment. The curve of root-mean-square (RMS) value in the run-to-failure experiment is shown in Fig.8. It is found that the RMS index jumps slightly at about 139.8 h, corresponding to the bearing initial fault. At about 147.6 h, the RMS index jumps with a larger value, meaning the extension stage of bearing fault. At about 156.3 h, the test ends. In order to verify the performance of intelligent identification for bearing status, the data of the former 72 h are selected as the training set, and the data after 72 h are selected as the test set. We use 139.8 h as threshold value to divide the test samples into two groups, i.e. the sample collected before 139.8 h belongs to bearing healthy state, while the others correspond to bearing failure state.

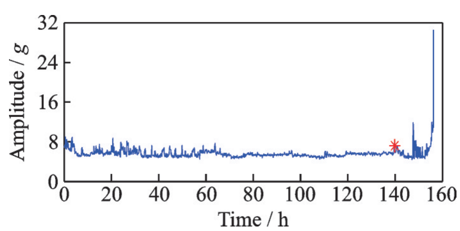

Fig.8 RMS of the vibration in the run-to-failure experiment

During the implementation of the algorithm, the data are firstly processed by the ESAFD model to obtain the denoised signal. In order to better compare results, the spectrum of original signal and denoised signal are taken as the input of the autoencoders respectively. After 200 training iterations, the reconstruction error of the deep convolutional variational autoencoders is used as the monitoring index. The probability density function (PDF) curves of the reconstruction error based on the original signal and denoised signal are shown in Fig.9. It can be seen that the intelligent monitoring method based on denoised signal behaves better, i.e. the distribution of health test set matchs the normal training set, while the PDF of fault test set is different from the traning set, meaning the monitoring accuracy of the denoised signal outperforms the original signal. The denoised signals of the ESAFD method can significantly strengthen the fault feature information, and distinguish the distribution of health state data from abonormal state data under the reconstruction error index, which helps achieve the precise identification of bearing fault status.

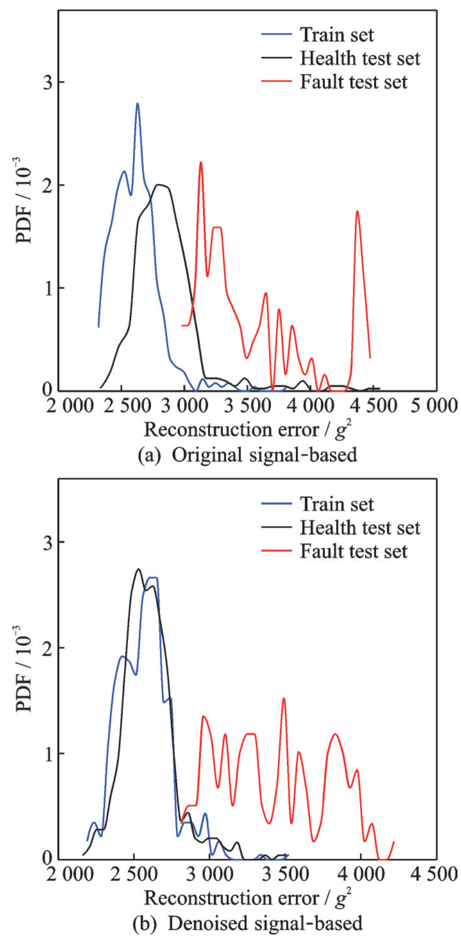

Fig.9 PDF for reconstruction signal of deep convolutional variational autoencoders

In order to compare the performance of anomaly monitoring, the 98%, 99% and 100% quantiles of reconstruction error based on the training sample are selected as the anomaly threshold. Samples with reconstruction error higher than the threshold will be judged as abnormal state. Model accuracy (Acc), health detection rate (HDR), fault detection rate (FDR) and false alarm rate (FAR) are selected to quantitatively evaluate the performance. In this paper, true positive N_{TP} can be defined as the number of health samples identified as healthy state. Similarly, the definitions of true negative N_{TN}, false positive N_{FP} and false negative N_{FN} can be obtained. The calculation formulas of several indexes are as follows

$$\mathrm{Acc}=\frac{N_{TP}+N_{TN}}{N_{TP}+N_{TN}+N_{FP}+N_{FN}} \tag{15}$$

$$\mathrm{HDR}=\frac{N_{TP}}{N_{TP}+N_{FN}} \tag{16}$$

$$\mathrm{FDR}=\frac{N_{TN}}{N_{TP}+N_{FP}} \tag{17}$$

$$\mathrm{FAR}=\frac{N_{FN}}{N_{TN}+N_{FN}} \tag{18}$$

Comparison results in Table 1 are obatined by 10 randomized repeat experiments on deep convolutional variational autoencoders. It can be seen that compared with original signal-based method, the sparsity-assisted intelligent monitoring model makes significant improvement in Acc, HDR, FDR and FAR. The average improvements of these four indicators are 5.6%, 2.3%, 20.3% and 7.3%, respectively. It is proved that the proposed intelligent monitoring method can identify anomaly state sensitively and stably, which effectively support the intelligent identification of aero-engine bearing weak fault state.

Table 1 Comparison results between original signal-based and denoised signal-based intelligent monitoring %

Monitoring method	Threshold	Acc↑	HDR↑	FDR↑	FAR↓
Original signal-based	98	87.3	85.9	93.5	38.7
	99	90.0	92.4	79.7	28.2
	100	86.7	99.0	32.5	11.0
Denoised signal-based	98	93.0	92.0	97.5	26.0
	99	93.8	93.5	95.2	22.6
	100	94.0	98.6	73.9	7.4

4 Conclusions

Aimed at achieving the intelligent monitoring of aero-engine main shaft bearing, this paper incorporates sparse denoising into the training procedure of deep learning to propose a new intelligent condtion monitoring method. The conclusions are drawn as follows:

(1) This paper establishes an enhanced sparsity-assisted fault diangois convex model based on constructed nonconvex penalty and convexity condition. The main advantage of ESAFD is that it can preserve the amplitude of bearing signal while reducing noise, showing a good ability of extracting weak features from strong noisy signal.

(2) An intelligent condition monitoring method based on sparsity-assisted deep convolutional variational autoencoders network is proposed in this paper. The main advantage of the proposed method is that it makes fully use of sparse denoised feature signal and obtains a sensitive and stable training model, thus realizing accurate identification of abnormal state.

(3) The effectiveness and feasibility of the proposed method is verified through aero-engine bearing run-to-failure experiment. It is proved that the proposed method is good at abnormal pattern recognition, showing a good potential for weak fault intelligent diagnosis of aero-engine main shaft bearings.

References

[1] HARRIS B W, MILO M W, ROAN M J. A general anomaly detection approach applied to rolling element bearings via reduced-dimensionality transition matrix analysis[J]. Proceedings of the Institution of Mechanical Engineers Part C: Journal of Mechanical Engineering Science, 2016, 230(13): 2169-2180.

[2] WU J, ZHAO Z, SUN C, et al. Fault-attention generative probabilistic adversarial autoencoder for machine anomaly detection[J]. IEEE Transactions on Industrial Informatics, 2020. DOI: 10.1109/TII.2020.2976752.

[3] SUN C, MA M, ZHAO Z, et al. Sparse deep stacking network for fault diagnosis of motor[J]. IEEE Transactions on Industrial Informatics, 2018, 14(7): 3261-3270.

[4] YAN R, GAO R X, CHEN X. Wavelets for fault diagnosis of rotary machines: A review with applications[J]. Signal Processing, 2014, 96: 1-15.

[5] LEI Y, LIN J, HE Z, et al. A review on empirical mode decomposition in fault diagnosis of rotating machinery[J]. Mechanical Systems and Signal Processing, 2013, 35(1/2): 108-126.

[6] ANTONI J. The infogram: Entropic evidence of the signature of repetitive transients[J]. Mechanical Systems and Signal Processing, 2016, 74: 73-94.

[7] ZHANG H, CHEN X, DU Z, et al. Nonlocal sparse model with adaptive structural clustering for feature extraction of aero-engine bearings[J]. Journal of Sound and Vibration, 2016, 368: 223-248.

[8] HE G, DING K, LIN H. Fault feature extraction of rolling element bearings using sparse representation[J]. Journal of Sound and Vibration, 2016, 366: 514-527.

[9] CAI G, CHEN X, HE Z. Sparsity-enabled signal decomposition using tunable Q-factor wavelet transform for fault feature extraction of gearbox[J]. Mechanical Systems and Signal Processing, 2013, 41 (1/2) : 34-53.

[10] CHEN S, DONOHO D L, SAUNDERS M A. Atomic decomposition by basis pursuit[J]. SIAM Review, 2001, 43(1): 129-159.

[11] SELESNICK I. Sparse regularization via convex analysis[J]. IEEE Transactions on Signal Processing, 2017, 65(17): 4481-4494.

[12] WANG S, SELESNICK I W, CAI G, et al. Synthesis versus analysis priors via generalized minimax-concave penalty for sparsity-assisted machinery fault diagnosis[J]. Mechanical Systems and Signal Processing, 2019, 127: 202-233.

[13] ELAD M, MILANFAR P, RUBINSTEIN R. Analysis versus synthesis in signal priors[J]. Inverse Problems, 2007, 23(3): 947-968.

[14] BAYRAM I. On the convergence of the iterative shrinkage/thresholding algorithm with a weakly convex penalty[J]. IEEE Transactions on Signal Processing, 2016, 64(6): 1597-1608.

[15] SELESNICK I. Wavelet Transform with tunable Q-factor[J]. IEEE Transactions on Signal Processing, 2011, 59(8): 3560-3575.

Acknowledgements This work was supported by the National Natural Science Foundations of China (Nos.91860125, 51705398), the National Key Basic Research Program of China (No. 2015CB057400) and the Shaanxi Province 2020 Natural Science Basic Research Plan (No. 2020JQ-042).

Authors Dr. **DING Baoqing** received the B.S. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2013, where he is now pursuing the Ph.D. degree. His current research interests include sparse signal representation, time-frequency analysis and intelligent fault diagnosis for rotating machine.

Prof. **CHEN Xuefeng** received the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, China, in 2004. He is currently a full professor and the dean of the School of Mechanical Engineering, Xi'an Jiaotong University. His current research interests include finite element method, mechanical system and signal processing, diagnosis and prognosis for complicated industrial systems, smart structures, aeroengine fault diagnosis, and wind turbine system monitoring.

Author contributions Dr. **DING Baoqing** designed the study, complied the models and wrote the manuscript. Mr. **WU Jingyao** contributed to data analysis, result interpretation and manuscript revision. Dr. **SUN Chuang** contributed to the discussion and revision of the study. Dr. **WANG Shibin** contributed to the experiment analysis. Prof. **CHEN Xuefeng** contributed to the design and discussion of the study. Prof. **LI Yinghong** contributed to the discussion and background of the study. All authors commented on the manuscript draft and approved the submission.

Competing interests The authors declare no competing interests.

(Production Editor: ZHANG Huangqun)

稀疏驱动的航空发动机主轴承智能监测研究

丁宝庆[1]，武靖耀[1]，孙　闯[1]，王诗彬[1]，陈雪峰[1]，李应红[1,2]

（1.西安交通大学机械制造系统工程国家重点实验室，西安 710049，中国；
2.空军工程大学等离子体动力学重点实验室，西安 710038，中国）

摘要：微弱特征提取是航空发动机健康监测与智能诊断的关键技术之一。本文针对航空发动机主轴承微弱故障智能监测难题，基于信号先验提出增强稀疏驱动的智能监测方法。通过分析经典凸稀疏诊断模型难以兼顾信号降噪与特征重构性能的缺陷，构建基于莫罗包络理论的非凸正则凸优化增强稀疏模型，以实现微弱特征提取；进而提出稀疏驱动的深度卷积变分自编码网络智能监测方法，通过对健康状态稀疏降噪样本的训练实现对故障异常状态的智能识别。通过航空发动机主轴承疲劳寿命试验的工程案例对提出方法进行性能验证，结果表明：增强稀疏驱动的智能监测方法具有良好的异常状态智能识别能力，能够有效支撑航空发动主轴承微弱故障的智能监测与诊断。

关键词：航空发动机主轴承；智能监测；特征提取；稀疏模型；变分自编码；深度学习